
Section 1

Principles of Sound and Hearing

Sound would be of little interest if we could not hear. It is through the production and perception of sounds that it is possible to communicate and monitor events in our surroundings. Some sounds are functional, others are created for aesthetic pleasure, and still others yield only annoyance. Obviously a comprehensive examination of sound must embrace not only the physical properties of the phenomenon but also the consequences of interaction with listeners.

This section deals with sound in its various forms, beginning with a description of what it is and how it is generated, how it propagates in various environments, and, finally, what happens when sound impinges on the ears and is transformed into a perception. Part of this examination is a discussion of the factors that influence the opinions about sound and spatial qualities that so readily form when listening to music, whether live or reproduced.

Audio engineering, in virtually all its facets, benefits from an understanding of these basic principles. A foundation of technical knowledge is a useful instrument, and, fortunately, most of the important ideas can be understood without recourse to complex mathematics. It is the intuitive interpretation of the principles that is stressed in this section; more detailed information can be found in the reference material.

In This Section:

Chapter 1.1: The Physical Nature of Sound	1-7
Introduction	1-7
Sound Waves	1-7
Complex Sounds	1-11
Phase	1-11
Spectra	1-11
Dimensions of Sound	1-16
References	1-19
Chapter 1.2: Sound Propagation	1-21
Introduction	1-21
Inverse-Square and Other Laws	1-21
Sound Reflection and Absorption	1-22

1-2 Section One

Interference: The Sum of Multiple Sound Sources	1-24
Diffraction	1-28
Refraction	1-30
References	1-31
Chapter 1.3: Resonance	1-33
Introduction	1-33
Fundamental Properties	1-33
Resonance in Pipes	1-36
Resonance in Rooms and Large Enclosures	1-39
Resonance in Small Enclosures: Helmholtz Resonators	1-40
Horns	1-41
References	1-41
Chapter 1.4: The Physical Nature of Hearing	1-43
Introduction	1-43
Anatomy of the Ear	1-43
Psychoacoustics and the Dimensions of Hearing	1-45
Loudness	1-45
Loudness as a Function of Frequency and Amplitude	1-45
Loudness as a Function of Bandwidth	1-47
Loudness as a Function of Duration	1-47
Measuring the Loudness of Complex Sounds	1-47
Masking	1-49
Simultaneous Masking	1-49
Temporal Masking	1-50
Acoustic Reflex	1-51
Pitch	1-51
Timbre, Sound Quality, and Perceptual Dimensions	1-52
Audibility of Variations in Amplitude and Phase	1-56
Perception of Direction and Space	1-57
Monaural Transfer Functions of the Ear	1-58
Interaural Differences	1-60
Localization Blur	1-61
Lateralization versus Localization	1-61
Spatial Impression	1-63
Distance Hearing	1-63
Stereophonic Imaging	1-64
Summing Localization with Interchannel Time/Amplitude Differences	1-66
Effect of Listener Position	1-66
Stereo Image Quality and Spaciousness	1-70
Special Role of the Loudspeakers	1-70
Sound in Rooms: The General Case	1-71
Precedence Effect and the Law of the First Wavefront	1-71
Binaural Discrimination	1-72
References	1-72

Reference Documents for this Section: _____

- Backus, John: *The Acoustical Foundations of Music*, Norton, New York, N.Y., 1969.
- Batteau, D. W.: "The Role of the Pinna in Human Localization," *Proc. R. Soc. London*, B168, pp. 158–180, 1967.
- Benade, A. H.: *Fundamentals of Musical Acoustics*, Oxford University Press, New York, N.Y., 1976.
- Beranek, Leo L: *Acoustics*, McGraw-Hill, New York, N.Y., 1954.
- Blauert, J., and W. Lindemann: "Auditory Spaciousness: Some Further Psychoacoustic Studies," *J. Acoust. Soc. Am.*, vol. 80, 533–542, 1986.
- Blauert, J: *Spatial Hearing*, translation by J. S. Allen, M.I.T., Cambridge. Mass., 1983.
- Bloom, P. J.: "Creating Source Elevation Illusions by Spectral Manipulations," *J. Audio Eng. Soc.*, vol. 25, pp. 560–565, 1977.
- Bose, A. G.: "On the Design, Measurement and Evaluation of Loudspeakers," presented at the 35th convention of the Audio Engineering Society, preprint 622, 1962.
- Buchlein, R.: "The Audibility of Frequency Response Irregularities" (1962), reprinted in English translation in *J. Audio Eng. Soc.*, vol. 29, pp. 126–131, 1981.
- Denes, Peter B., and E. N. Pinson: *The Speech Chain*, Bell Telephone Laboratories, Waverly, 1963.
- Durlach, N. I., and H. S. Colburn: "Binaural Phenomena," in *Handbook of Perception*, E. C. Carterette and M. P. Friedman (eds.), vol. 4, Academic, New York, N.Y., 1978.
- Ehara, Shiro: "Instantaneous Pressure Distributions of Orchestra Sounds," *J. Acoust. Soc. Japan*, vol. 22, pp. 276–289, 1966.
- Fletcher, H., and W. A. Munson: "Loudness, Its Definition, Measurement and Calculation," *J. Acoust. Soc. Am.*, vol. 5, pp. 82–108, 1933.
- Fryer, P.: "Loudspeaker Distortions—Can We Rear Them?," *Hi-Fi News Record Rev.*, vol. 22, pp. 51–56, 1977.
- Gabrielsson, A., and B. Lindstrom: "Perceived Sound Quality of High-Fidelity Loudspeakers." *J. Audio Eng. Soc.*, vol. 33, pp. 33–53, 1985.
- Gabrielsson, A., and H. Siogren: "Perceived Sound Quality of Sound-Reproducing Systems," *J. Acoust. Soc. Am.*, vol. 65, pp. 1019–1033, 1979.
- Haas, H.: "The Influence of a Single Echo on the Audibility of Speech," *Acustica*, vol. I, pp. 49–58, 1951; English translation reprinted in *J. Audio Eng. Soc.*, vol. 20, pp. 146–159, 1972.
- Hall, Donald: *Musical Acoustics—An Introduction*, Wadsworth, Belmont, Calif., 1980.
- International Electrotechnical Commission: *Sound System Equipment*, part 10, *Programme Level Meters*, Publication 268-1 0A, 1978.
- International Organization for Standardization: *Normal Equal-Loudness Contours for Pure Tones and Normal Threshold for Hearing under Free Field Listening Conditions*, Recommendation R226, December 1961.

1-4 Section One

- Jones, B. L., and E. L. Torick: "A New Loudness Indicator for Use in Broadcasting," *J. SMPTE*, Society of Motion Picture and Television Engineers, White Plains, N.Y., vol. 90, pp. 772–777, 1981.
- Kuhl, W., and R. Plantz: "The Significance of the Diffuse Sound Radiated from Loudspeakers for the Subjective Hearing Event," *Acustica*, vol. 40, pp. 182–190, 1978.
- Kuhn, G. F.: "Model for the Interaural Time Differences in the Azimuthal Plane," *J. Acoust. Soc. Am.*, vol. 62, pp. 157–167, 1977.
- Kurozumi, K., and K. Ohgushi: "The Relationship between the Cross-Correlation Coefficient of Two-Channel Acoustic Signals and Sound Image Quality," *J. Acoust. Soc. Am.*, vol. 74, pp. 1726–1733, 1983.
- Main, Ian G.: *Vibrations and Waves in Physics*, Cambridge, London, 1978.
- Mankovsky, V. S.: *Acoustics of Studios and Auditoria*, Focal Press, London, 1971.
- Meyer, J.: *Acoustics and the Performance of Music*, Verlag das Musikinstrument, Frankfurt am Main, 1987.
- Morse, Philip M.: *Vibrations and Sound*, 1964, reprinted by the Acoustical Society of America, New York, N.Y., 1976.
- Olson, Harry F.: *Acoustical Engineering*, Van Nostrand, New York, N.Y., 1957.
- Pickett, J. M.: *The Sounds of Speech Communications*, University Park Press, Baltimore, MD, 1980.
- Pierce, John R.: *The Science of Musical Sound*, Scientific American Library, New York, N.Y., 1983.
- Piercy, J. E., and T. F. W. Embleton: "Sound Propagation in the Open Air," in *Handbook of Noise Control*, 2d ed., C. M. Harris (ed.), McGraw-Hill, New York, N.Y., 1979.
- Plomp, R.: *Aspects of Tone Sensation—A Psychophysical Study*, Academic, New York, N.Y., 1976.
- Rakerd, B., and W. M. Hartmann: "Localization of Sound in Rooms, II—The Effects of a Single Reflecting Surface," *J. Acoust. Soc. Am.*, vol. 78, pp. 524–533, 1985.
- Rasch, R. A., and R. Plomp: "The Listener and the Acoustic Environment," in D. Deutsch (ed.), *The Psychology of Music*, Academic, New York, N.Y., 1982.
- Robinson, D. W., and R. S. Dadson: "A Redetermination of the Equal-Loudness Relations for Pure Tones," *Br. J. Appl. Physics*, vol. 7, pp. 166–181, 1956.
- Scharf, B.: "Loudness," in E. C. Carterette and M. P. Friedman (eds.), *Handbook of Perception*, vol. 4, *Hearing*, chapter 6, Academic, New York, N.Y., 1978.
- Shaw, E. A. G., and M. M. Vaillancourt: "Transformation of Sound-Pressure Level from the Free Field to the Eardrum Presented in Numerical Form," *J. Acoust. Soc. Am.*, vol. 78, pp. 1120–1123, 1985.
- Shaw, E. A. G., and R. Teranishi: "Sound Pressure Generated in an External-Ear Replica and Real Human Ears by a Nearby Sound Source," *J. Acoust. Soc. Am.*, vol. 44, pp. 240–249, 1968.

Principles of Sound and Hearing 1-5

- Shaw, E. A. G.: "Aural Reception," in A. Lara Saenz and R. W. B. Stevens (eds.), *Noise Pollution*, Wiley, New York, N.Y., 1986.
- Shaw, E. A. G.: "External Ear Response and Sound Localization," in R. W. Gatehouse (ed.), *Localization of Sound: Theory and Applications*, Amphora Press, Groton, Conn., 1982.
- Shaw, E. A. G.: "Noise Pollution—What Can be Done?" *Phys. Today*, vol. 28, no. 1, pp. 46–58, 1975.
- Shaw, E. A. G.: "The Acoustics of the External Ear," in W. D. Keidel and W. D. Neff (eds.), *Handbook of Sensory Physiology*, vol. V/I, *Auditory System*, Springer-Verlag, Berlin, 1974.
- Shaw, E. A. G.: "Transformation of Sound Pressure Level from the Free Field to the Eardrum in the Horizontal Plane," *J. Acoust. Soc. Am.*, vol. 56, pp. 1848–1861, 1974.
- Stephens, R. W. B., and A. E. Bate: *Acoustics and Vibrational Physics*, 2nd ed., E. Arnold (ed.), London, 1966.
- Stevens, W. R.: "Loudspeakers—Cabinet Effects," *Hi-Fi News Record Rev.*, vol. 21, pp. 87–93, 1976.
- Sundberg, Johan: "The Acoustics of the Singing Voice," in *The Physics of Music*, introduction by C. M. Hutchins, Scientific American/Freeman, San Francisco, Calif., 1978.
- Tonic, F. E.: "Loudness—Applications and Implications to Audio," *dB*, Part 1, vol. 7, no. 5, pp. 27–30; Part 2, vol. 7, no. 6, pp. 25–28, 1973.
- Toole, F. E., and B. McA. Sayers: "Lateralization Judgments and the Nature of Binaural Acoustic Images," *J. Acoust. Soc. Am.*, vol. 37, pp. 319–324, 1965.
- Toole, F. E.: "Loudspeaker Measurements and Their Relationship to Listener Preferences," *J. Audio Eng. Soc.*, vol. 34, part 1, pp. 227–235, part 2, pp. 323–348, 1986.
- Toole, F. E.: "Subjective Measurements of Loudspeaker Sound Quality and Listener Performance," *J. Audio Eng. Soc.*, vol. 33, pp. 2–32, 1985.
- Voelker, E. J.: "Control Rooms for Music Monitoring," *J. Audio Eng. Soc.*, vol. 33, pp. 452–462, 1985.
- Ward, W. D.: "Subjective Musical Pitch," *J. Acoust. Soc. Am.*, vol. 26, pp. 369–380, 1954.
- Waterhouse, R. V., and C. M. Harris: "Sound in Enclosed Spaces," in *Handbook of Noise Control*, 2d ed., C. M. Harris (ed.), McGraw-Hill, New York, N.Y., 1979.
- Wong, G. S. K.: "Speed of Sound in Standard Air," *J. Acoust. Soc. Am.*, vol. 79, pp. 1359–1366, 1986.
- Zurek, P. M.: "Measurements of Binaural Echo Suppression," *J. Acoust. Soc. Am.*, vol. 66, pp. 1750–1757, 1979.
- Zwislocki, J. J.: "Masking—Experimental and Theoretical Aspects of Simultaneous, Forward, Backward and Central Masking," in E. C. Carterette and M. P. Friedman (eds.), *Handbook of Perception*, vol. 4, *Hearing*, chapter 8, Academic, New York, N.Y., 1978.

Chapter**1.1****The Physical Nature of Sound****Floyd E. Toole****E. A. G. Shaw, G. A. Daigle, M. R. Stinson****1.1.1 Introduction**

Sound is a physical disturbance in the medium through which it is propagated. Although the most common medium is air, sound can travel in any solid, liquid, or gas. In air, sound consists of localized variations in pressure above and below normal atmospheric pressure (*compressions and rarefactions*).

Air pressure rises and falls routinely, as environmental weather systems come and go, or with changes in altitude. These fluctuation cycles are very slow, and no perceptible sound results, although it is sometimes evident that the ears are responding in a different way to these *infra-sonic* events. At fluctuation frequencies in the range from about 20 cycles per second up to about 20,000 cycles per second the physical phenomenon of sound can be perceived as having pitch or tonal character. This generally is regarded as the *audible* or *audio-frequency range*, and it is the frequencies in this range that are the concern of this chapter. Frequencies above 20,000 cycles per second are classified as *ultrasonic*.

1.1.2 Sound Waves

The essence of sound waves is illustrated in Figure 1.1.1, which shows a tube with a piston in one end. Initially, the air within and outside the tube is all at the prevailing atmospheric pressure. When the piston moves quickly inward, it compresses the air in contact with its surface. This energetic compression is rapidly passed on to the adjoining layer of air, and so on, repeatedly. As it delivers its energy to its neighbor, each layer of air returns to its original uncompressed state. A longitudinal sound pulse is moving outward through the air in the tube, causing only a passing disturbance on the way. It is a pulse because there is only an isolated action, and it is longitudinal because the air movement occurs along the axis of sound propagation. The rate at which the pulse propagates is the speed of sound. The pressure rise in the compressed air is proportional to the velocity with which the piston moves, and the perceived loudness of the resulting sound pulse

1-8 Principles of Sound and Hearing

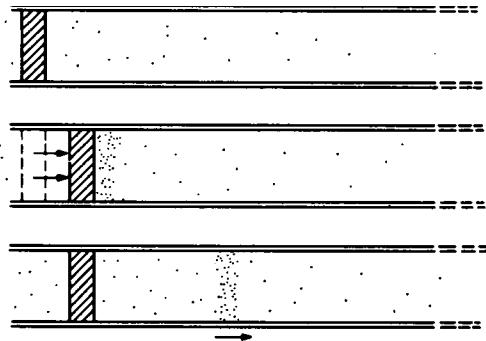


Figure 1.1.1 Generation of a longitudinal sound wave by the rapid movement of a piston in the end of a tube, showing the propagation of the wave pulse at the speed of sound down the length of the tube.

is related to the incremental amplitude of the pressure wave above the ambient atmospheric pressure.

Percussive or impulsive sounds such as these are common, but most sounds do not cease after a single impulsive event. Sound waves that are repetitive at a regular rate are called *periodic*. Many musical sounds are periodic, and they embrace a very wide range of repetitive patterns. The simplest of periodic sounds is a pure tone, similar to the sound of a tuning fork or a whistle. An example is presented when the end of the tube is driven by a loudspeaker reproducing a recording of such a sound (Figure 1.1.2). The pattern of displacement versus time for the loudspeaker diaphragm, shown in Figure 1.1.2b, is called a *sine wave* or *sinusoid*.

If the first diaphragm movement is inward, the first event in the tube is a pressure compression, as seen previously. When the diaphragm changes direction, the adjacent layer of air undergoes a *pressure rarefaction*. These cyclic compressions and rarefactions are repeated, so that the sound wave propagating down the tube has a regularly repeated, periodic form. If the air pressure at all points along the tube were measured at a specific instant, the result would be the graph of air pressure versus distance shown in Figure 1.1.2c. This reveals a smoothly sinusoidal waveform with a repetition distance along the tube symbolized by λ (lambda), the *wavelength* of the periodic sound wave.

If a pressure-measuring device were placed at some point in the tube to record the instantaneous changes in pressure at that point as a function of time, the result would be as shown in Figure 1.1.2d. Clearly, the curve has the same shape as the previous one except that the horizontal axis is time instead of distance. The periodic nature of the waveform is here defined by the time period T , known simply as the *period* of the sound wave. The inverse of the period, $1/T$, is the *frequency* of the sound wave, describing the number of repetition cycles per second passing a fixed point in space. An ear placed in the path of a sound wave corresponding to the musical tone middle C would be exposed to a frequency of 261.6 cycles per second or, using standard scientific terminology, a frequency of 261.6 hertz (Hz). The perceived loudness of the tone would depend on the magnitude of the pressure deviations above and below the ambient air pressure.

The parameters discussed so far are all related by the *speed of sound*. Given the speed of sound and the duration of one period, the wavelength can be calculated as follows:

$$\lambda = cT \quad (1.1.1)$$

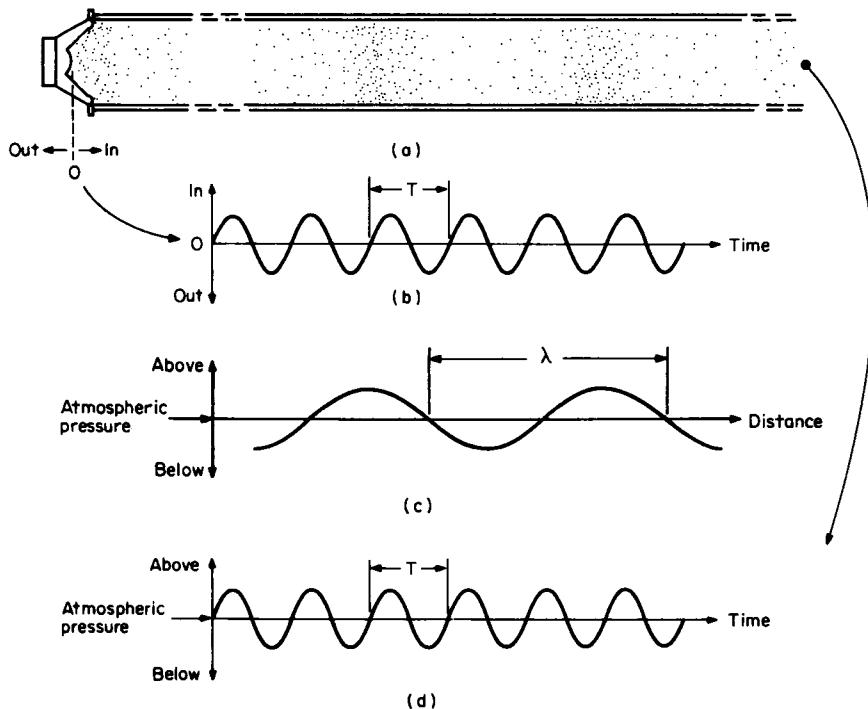


Figure 1.1.2 Characteristics of sound waves: (a) A periodic sound wave, a sinusoid in this example, is generated by a loudspeaker placed at the end of a tube. (b) Waveform showing the movement of the loudspeaker diaphragm as a function of time: displacement versus time. (c) Waveform showing the instantaneous distribution of pressure along a section of the tube: pressure versus distance. (d) Waveform showing the pressure variation as a function of time at some point along the tube: pressure versus time.

where:

λ = wavelength

c = speed of sound

T = period

By knowing that the frequency $f = 1/T$, the following useful equation and its variations can be derived:

$$\lambda = \frac{c}{f} \quad f = \frac{c}{\lambda} \quad c = f\lambda \quad (1.1.2)$$

The speed of sound in air at a room temperature of 22°C (72°F) is 345 m/s (1131 ft/s). At any other ambient temperature, the speed of sound in air is given by the following approximate relationships [1, 2]:

1-10 Principles of Sound and Hearing

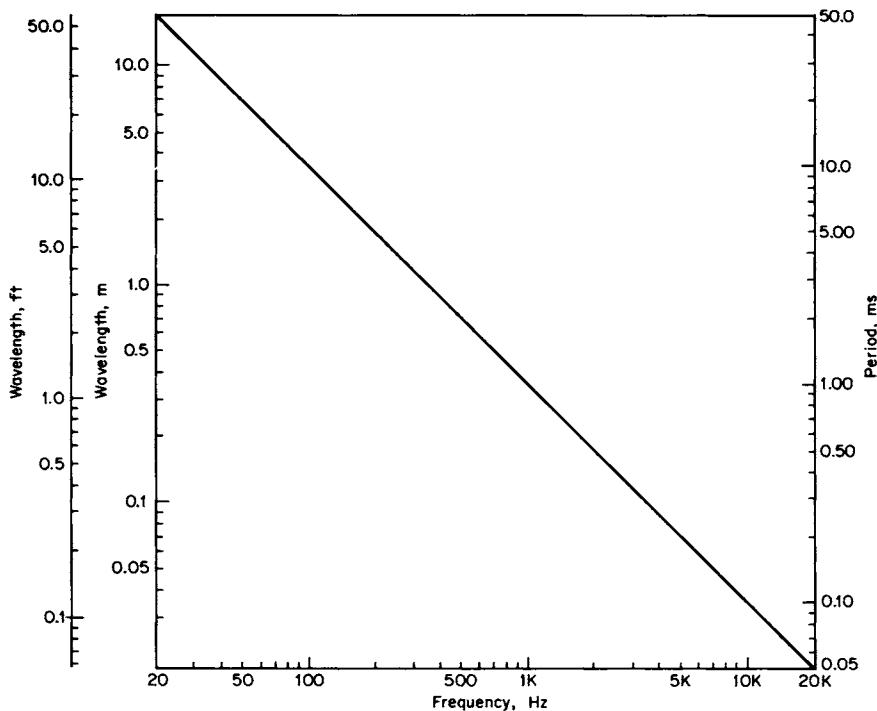


Figure 1.1.3 Relationships between wavelength, period, and frequency for sound waves in air.

$$c(m/s) = 331.29 + 0.607t(\text{ }^{\circ}\text{C}) \quad (1.1.3)$$

or

$$c(m/s) = 1051.5 + 1.106t(\text{ }^{\circ}\text{F}) \quad (1.1.4)$$

where t = ambient temperature.

The relationships between the frequency of a sound wave and its wavelength are essential to understanding many of the fundamental properties of sound and hearing. The graph of Figure 1.1.3 is a useful quick reference illustrating the large ranges of distance and time embraced by audible sounds. For example, the tone middle C with a frequency of 261.6 Hz has a wavelength of 1.3 m (4.3 ft) in air at 20°C. In contrast, an organ pedal note at C1, 32.7 Hz, has a wavelength of 10.5 m (34.5 ft), and the third-harmonic overtone of C8, at 12,558 Hz, has a wavelength of 27.5 mm (1.1 in). The corresponding periods are, respectively, 3.8 ms, 30.6 ms, and 0.08 ms. The contrasts in these dimensions are remarkable, and they result in some interesting and troublesome effects in the realms of perception and audio engineering. For the discussions that follow it is often more helpful to think in terms of wavelengths rather than in frequencies.

1.1.2a Complex Sounds

The simple sine waves used for illustration reveal their periodicity very clearly. Normal sounds, however, are much more complex, being combinations of several such pure tones of different frequencies and perhaps additional transient sound components that punctuate the more sustained elements. For example, speech is a mixture of approximately periodic vowel sounds and staccato consonant sounds. Complex sounds can also be periodic; the repeated wave pattern is just more intricate, as is shown in Figure 1.1.4a. The period identified as T_1 applies to the *fundamental frequency* of the sound wave, the component that normally is related to the characteristic pitch of the sound. Higher-frequency components of the complex wave are also periodic, but because they are typically lower in amplitude, that aspect tends to be disguised in the summation of several such components of different frequency. If, however, the sound wave were analyzed, or broken down into its constituent parts, a different picture emerges: Figure 1.1.4b, c, and d. In this example, the analysis shows that the components are all *harmonics*, or whole-number multiples, of the fundamental frequency; the higher-frequency components all have multiples of entire cycles within the period of the fundamental.

To generalize, it can be stated that all *complex periodic waveforms* are combinations of several harmonically related sine waves. The shape of a complex waveform depends upon the relative amplitudes of the various harmonics and the position in time of each individual component with respect to the others. If one of the harmonic components in Figure 1.1.4 is shifted slightly in time, the shape of the waveform is changed, although the frequency composition remains the same (Figure 1.1.5). Obviously a record of the time locations of the various harmonic components is required to completely describe the complex waveform. This information is noted as the *phase* of the individual components.

1.1.2b Phase

Phase is a notation in which the time of one period of a sine wave is divided into 360° . It is a relative quantity, and although it can be defined with respect to any reference point in a cycle, it is convenient to start (0°) with the upward, or positive-going, zero crossing and to end (360°) at precisely the same point at the beginning of the next cycle (Figure 1.1.6). *Phase shift* expresses in degrees the fraction of a period or wavelength by which a single-frequency component is shifted in the time domain. For example, a phase shift of 90° corresponds to a shift of one-fourth period. For different frequencies this translates into different time shifts. Looking at it from the other point of view, if a complex waveform is time-delayed, the various harmonic components will experience different phase shifts, depending on their frequencies.

A special case of phase shift is a *polarity reversal*, an inversion of the waveform, where all frequency components undergo a 180° phase shift. This occurs when, for example, the connections to a loudspeaker are reversed.

1.1.2c Spectra

Translating time-domain information into the frequency domain yields an *amplitude-frequency spectrum* or, as it is commonly called, simply a *spectrum*. Figure 1.1.7a shows the spectrum of the waveform in Figures 1.1.4 and 1.1.5, in which the height of each line represents the amplitude of that particular component and the position of the line along the frequency axis identifies its frequency. This kind of display is a *line spectrum* because there are sound components at only

1-12 Principles of Sound and Hearing

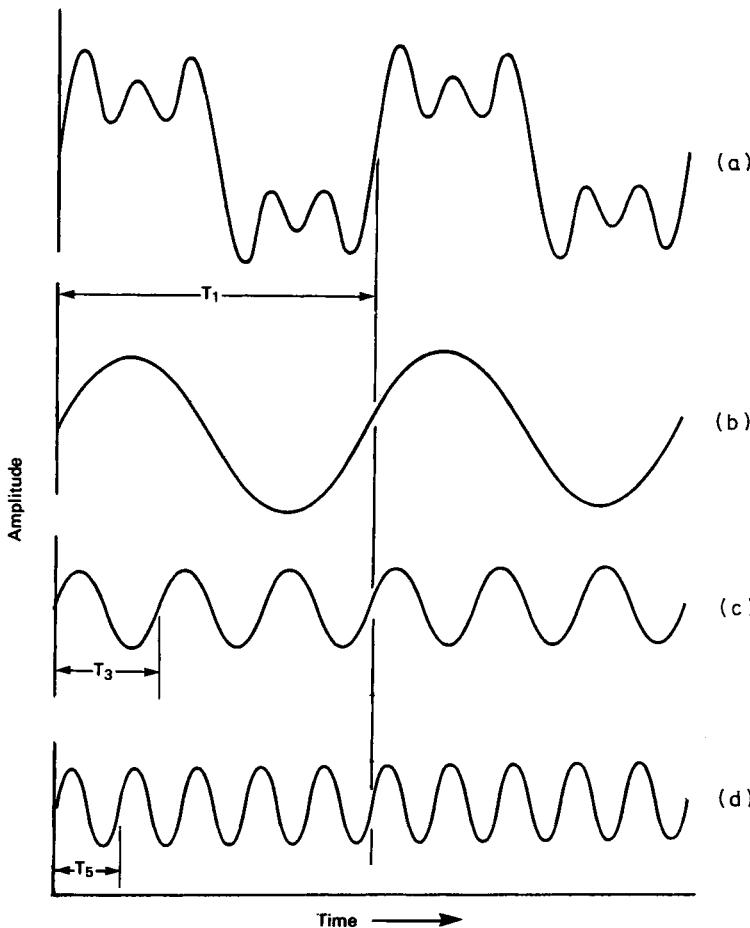


Figure 1.1.4 A complex waveform constructed from the sum of three harmonically related sinusoidal components, all of which start at the origin of the time scale with a positive-going zero crossing. Extending the series of odd-harmonic components to include those above the fifth would result in the complex waveform progressively assuming the form of a square wave. (a) Complex waveform, the sum of b, c, and d. (b) Fundamental frequency. (c) Third harmonic. (d) Fifth harmonic.

certain specific frequencies. The phase information is shown in Figure 1.1.7b, where the difference between the two waveforms is revealed in the different *phase-frequency spectra*.

The equivalence of the information presented in the two domains—the waveform in the time domain and the amplitude- and phase-frequency spectra in the frequency domain—is a matter of considerable importance. The proofs have been thoroughly worked out by the French mathematician Fourier, and the well-known relationships bear his name. The breaking down of waveforms into their constituent sinusoidal parts is known as *Fourier analysis*. The construction of complex

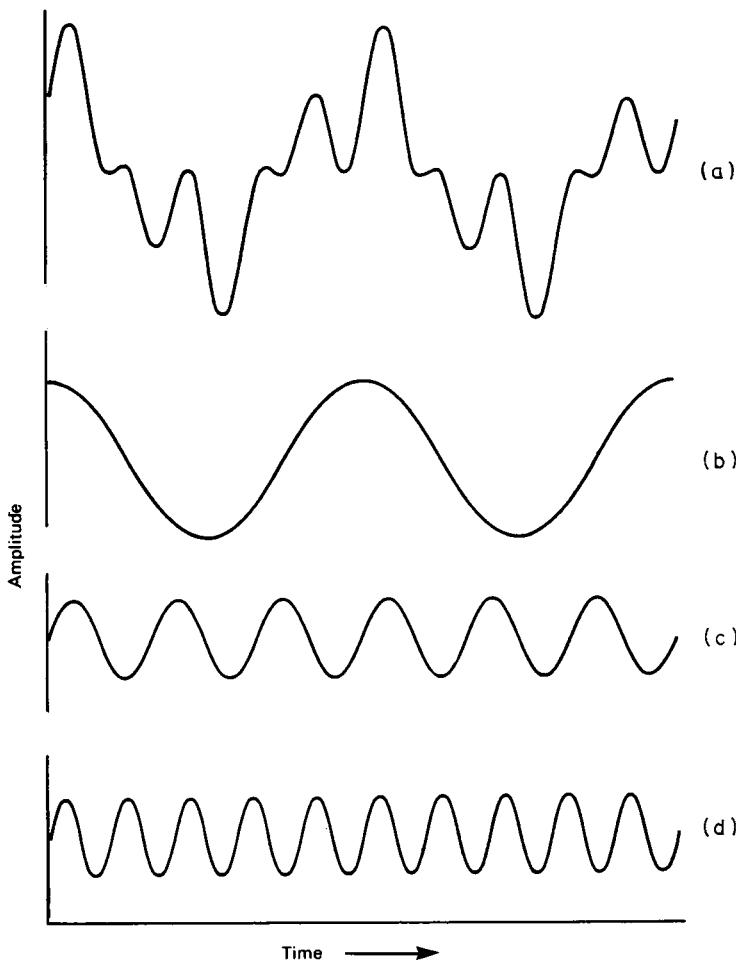


Figure 1.1.5 A complex waveform with the same harmonic-component amplitudes as in Figure 1.1.4, but with the starting time of the fundamental advanced by one-fourth period: a phase shift of 90° .

waveshapes from summations of sine waves is called *Fourier synthesis*. *Fourier transformations* permit the conversion of time-domain information into frequency-domain information, and vice versa. These interchangeable descriptions of waveforms form the basis for powerful methods of measurement and, at the present stage, provide a convenient means of understanding audio phenomena. In the examples that follow, the relationships between time-domain and frequency-domain descriptions of waveforms will be noted.

Figure 1.1.8 illustrates the sound waveform that emerges from the larynx, the buzzing sound that is the basis for vocalized speech sounds. This sound is modified in various ways in its passage down the vocal tract before it emerges from the mouth as speech. The waveform is a series

1-14 Principles of Sound and Hearing

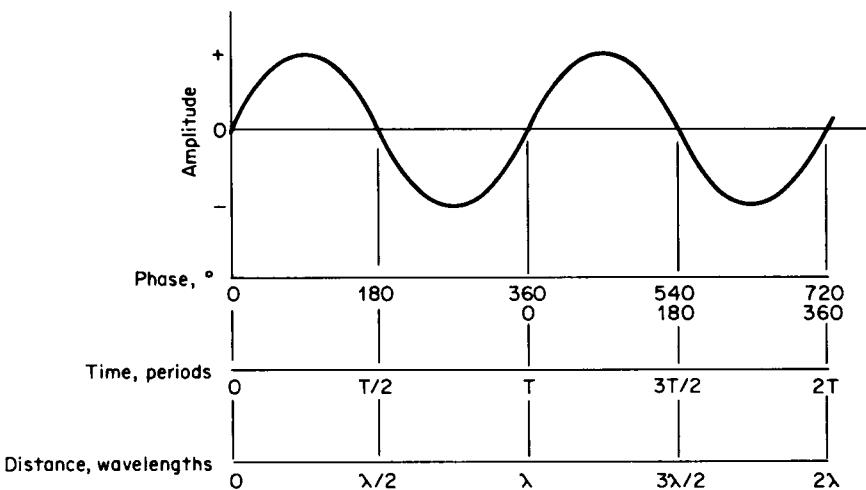


Figure 1.1.6 The relationship between the period T and wavelength λ of a sinusoidal waveform and the phase expressed in degrees. Although it is normal to consider each repetitive cycle as an independent 360° , it is sometimes necessary to sum successive cycles starting from a reference point in one of them.

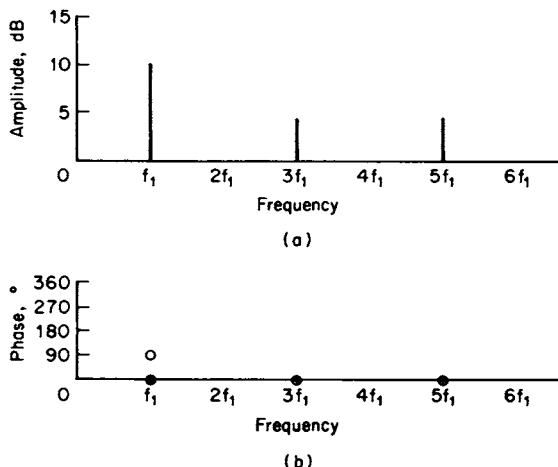


Figure 1.1.7 The amplitude-frequency spectra (a) and the phase-frequency spectra (b) of the complex waveforms shown in Figures 1.1.4 and 1.1.5. The amplitude spectra are identical for both waveforms, but the phase-frequency spectra show the 90° phase shift of the fundamental component in the waveform of Figure 1.1.5. Note that frequency is expressed as a multiple of the fundamental frequency f_1 . The numerals are the harmonic numbers. Only the fundamental f_1 and the third and fifth harmonics (f_3 and f_5) are present.

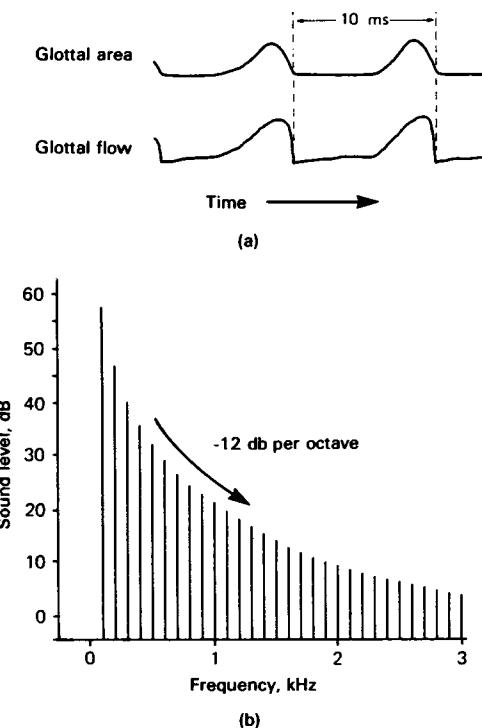


Figure 1.1.8 Characteristics of speech. (a) Waveforms showing the varying area between vibrating vocal cords and the corresponding airflow during vocalized speech as a function of time. (b) The corresponding amplitude-frequency spectrum, showing the 100-Hz fundamental frequency for this male speaker. (From [3]. Used with permission.)

of periodic pulses, corresponding to the pulses of air that are expelled, under lung pressure, from the vibrating vocal cords. The spectrum of this waveform consists of a harmonic series of components, with a fundamental frequency, for this male talker, of 100 Hz. The gently rounded contours of the waveform suggest the absence of strong high-frequency components, and the amplitude-frequency spectrum confirms it. The *spectrum envelope*, the overall shape delineating the amplitudes of the components of the line spectrum, shows a progressive decline in amplitude as a function of frequency. The amplitudes are described in *decibels*, abbreviated dB. This is the common unit for describing sound-level differences. The rate of this decline is about –12 dB per octave (an *octave* is a 2:1 ratio of frequencies).

Increasing the pitch of the voice brings the pulses closer together in time and raises the fundamental frequency. The harmonic-spectrum lines displayed in the frequency domain are then spaced farther apart but still within the overall form of the spectrum envelope, which is defined by the shape of the pulse itself. Reducing the pitch of the voice has the opposite effect, increasing the spacing between pulses and reducing the spacing between the spectral lines under the envelope. Continuing this process to the limiting condition, if it were possible to emit just a single pulse, would be equivalent to an infinitely long period, and the spacing between the spectral lines would vanish. The discontinuous, or *aperiodic*, pulse waveform therefore yields a *continuous* spectrum having the form of the spectrum envelope.

Isolated pulses of sound occur in speech as any of the variations of consonant sounds and in music as percussive sounds and as transient events punctuating more continuous melodic lines. All these aperiodic sounds exhibit continuous spectra with shapes that are dictated by the wave-

1-16 Principles of Sound and Hearing

forms. The leisurely undulations of a bass drum waveform contain predominantly low-frequency energy, just as the more rapid pressure changes in a snare drum waveform require the presence of higher frequencies with their more rapid rates of change. A technical waveform of considerable use in measurements consists of a very brief impulse which has the important feature of containing equal amplitudes of all frequencies within the audio-frequency bandwidth. This is moving toward a limiting condition in which an infinitely short event in the time domain is associated with an infinitely wide amplitude-frequency spectrum.

1.1.3 Dimensions of Sound

The descriptions of sound in the preceding section involved only pressure variation, and while this is the dimension that is most commonly referred to, it is not the only one. Accompanying the pressure changes are temporary movements of the air “particles” as the sound wave passes (in this context a particle is a volume of air that is large enough to contain many molecules while its dimensions are small compared with the wavelength). Other measures of the magnitude of the sound event are the displacement amplitude of the air particles away from their rest positions and the velocity amplitude of the particles during the movement cycle. In the physics of sound, the *particle displacement* and the *particle velocity* are useful concepts, but the difficulty of their measurement limits their practical application. They can, however, help in understanding other concepts.

In a normally propagating sound wave, energy is required to move the air particles; they must be pushed or pulled against the elasticity of the air, causing the incremental rises and falls in pressure. Doubling the displacement doubles the pressure change, and this requires double the force. Because the work done is the product of force times distance and both are doubled, the energy in a sound wave is therefore proportional to the square of the particle displacement amplitude or, in more practical terms, to the square of the sound pressure amplitude.

Sound energy spreads outward from the source in the three dimensions of space, in addition to those of amplitude and time. The energy of such a sound field is usually described in terms of the energy flow through an imaginary surface. The sound energy transmitted per unit of time is called *sound power*. The sound power passing through a unit area of a surface perpendicular to a specified direction is called the *sound intensity*. Because intensity is a measure of energy flow, it also is proportional to the square of the sound pressure amplitude.

The ear responds to a very wide range of sound pressure amplitudes. From the smallest sound that is audible to sounds large enough to cause discomfort there is a ratio of approximately 1 million in sound pressure amplitude, or 1 trillion (10^{12}) in sound intensity or power. Dealing routinely with such large numbers is impractical, so a logarithmic scale is used. This is based on the *bel*, which represents a ratio of 10:1 in sound intensity or sound power (the power can be acoustical or electrical). More commonly the decibel, one-tenth of a bel, is used. A difference of 10 dB therefore corresponds to a factor-of-10 difference in sound intensity or sound power. Mathematically this can be generalized as

$$\text{Level difference} = \log \frac{P_1}{P_2} \text{ bels} \quad (1.1.5)$$

or

Table 1.1.1 Various Power and Amplitude Ratios and their Decibel Equivalents*

Sound or electrical power ratio	Decibels	Sound pressure, voltage, or current ratio	Decibels
1	0	1	0
2	3.0	2	6.0
3	4.8	3	9.5
4	6.0	4	12.0
5	7.0	5	14.0
6	7.8	6	15.6
7	8.5	7	16.9
8	9.0	8	18.1
9	9.5	9	19.1
10	10.0	10	20.0
100	20.0	100	40.0
1,000	30.0	1,000	60.0
10,000	40.0	10,000	80.0
100,000	50.0	100,000	100.0
1,000,000	60.0	1,000,000	120.0

* Other values can be calculated precisely by using Eqs. (1.6) and (1.7) or estimated by using this table and the following rules:

Power ratios that are multiples of 10 are converted into their decibel equivalents by multiplying the appropriate exponent by 10. For example, a power ratio of 1000 is 10^3 , and this translates into $3 \times 10 = 30$ dB. Since power is proportional to the square of amplitude, the exponent of 10 must be doubled to arrive at the decibel equivalent of an amplitude ratio.

Intermediate values can be estimated by combining values in this table by means of the rule that the multiplication of power or amplitude ratios is equivalent to adding level differences in decibels. For example, increasing a sound level by 27 dB requires increasing the power by a ratio of 500 (20 dB is a ratio of 100, and 7 dB is a ratio of 5; the product of the ratios is 500). The corresponding increase in sound pressure or electrical signal amplitude is a factor of just over 20 (20 dB is a ratio of 10, and 7 dB falls between 6.0 and 9.5 and is therefore a ratio of something in excess of 2); the calculated value is 22.4. Reversing the process, if the output from a power amplifier is increased from 40 to 800 W, a ratio of 20, the sound pressure level would be expected to increase by 13 dB (a power ratio of 10 is 10 dB, a ratio of 2 is 3 dB, and the sum is 13 dB). The corresponding voltage increase measured at the output of the amplifier would be a factor of between 4 and 5 (by calculation, 4.5).

$$\text{Level difference} = 10 \log \frac{P_1}{P_2} \text{ decibels} \quad (1.1.6)$$

where P_1 and P_2 are two levels of power.

For ratios of sound pressures (analogous to voltage or current ratios in electrical systems) the squared relationship with power is accommodated by multiplying the logarithm of the ratio of pressures by 2, as follows:

$$\text{Level difference} = 10 \log \frac{P_1^2}{P_2^2} = 20 \log \frac{p_1}{p_2} \text{ dB} \quad (1.1.7)$$

where p_1 and p_2 are sound pressures.

1-18 Principles of Sound and Hearing

Table 1.1.2 Typical Sound Pressure Levels and Intensities for Various Sound Sources*

Sound source	Sound pressure level, dB	Intensity, W/m^2	Listener reaction
Jet engine at 10 m	160	10^3	Immediate damage
	150		Painful feeling
	140		
	130		
SST takeoff at 500 m	120	1	Discomfort
Amplified rock music	110		
Chain saw at 1 m	100		fff
Power mower at 1.5 m	90	10^{-3}	
75-piece orchestra at 7 m	80		f
City traffic at 15 m	70		
Normal speech at 1 m	60	10^{-6}	p
Suburban residence	50		
Library	40		ppp
Empty auditorium	30	10^{-9}	
Recording studio	20		
Breathing	10		
	0†	10^{-12}	Inaudible

* The relationships illustrated in this table are necessarily approximate because the conditions of measurement are not defined. Typical levels should, however, be within about 10 dB of the stated values.

† 0-dB sound pressure level (SPL) represents a reference sound pressure of 0.0002 μbar , or 0.00002 N/m^2 .

The relationship between decibels and a selection of power and pressure ratios is given in Table 1.1.1. The footnote to the table describes a simple process for interpolating between these values, an exercise that helps to develop a feel for the meaning of the quantities.

The representation of the relative magnitudes of sound pressures and powers in decibels is important, but there is no indication of the absolute magnitude of either quantity being compared. This limitation is easily overcome by the use of a universally accepted reference level with which others are compared. For convenience the standard reference level is close to the smallest sound that is audible to a person with normal hearing. This defines a scale of *sound pressure level* (SPL), in which 0 dB represents a sound level close to the hearing-threshold level for middle and high frequencies (the most sensitive range). The SPL of a sound therefore describes, in decibels, the relationship between the level of that sound and the reference level. Table 1.1.2 gives examples of SPLs of some common sounds with the corresponding intensities and an indication of listener reactions. From this table it is clear that the musically useful range of SPLs extend from the level of background noises in quiet surroundings to levels at which listeners begin to experience auditory discomfort and nonauditory sensations of feeling or pain in the ears themselves.

While some sound sources, such as chain saws and power mowers, produce a relatively constant sound output, others, like a 75-piece orchestra, are variable. The sound from such an orchestra might have a *peak factor* of 20 to 30 dB; the momentary, or peak, levels can be this amount higher than the long-term average SPL indicated [4].

The sound power produced by sources gives another perspective on the quantities being described. In spite of some impressively large sounds, a full symphony orchestra produces only

about 1 acoustic watt when working through a typical musical passage. On crescendos with percussion, though, the levels can be of the order of 100 W. A bass drum alone can produce about 25 W of acoustic power of peaks. All these levels are dependent on the instruments and how they are played. Maximum sound output from cymbals might be 10 W; from a trombone, 6 W; and from a piano, 0.4 W [5]. By comparison, average speech generates about 25 μ W, and a present-day jet liner at takeoff between 50 and 100 kW. Small gasoline engines produce from 0.001 to 1.0 acoustic watt, and electric home appliances less than 0.01 W [6].

1.1.4 References

1. Beranek, Leo L: *Acoustics*, McGraw-Hill, New York, N.Y., 1954.
2. Wong, G. S. K.: "Speed of Sound in Standard Air," *J. Acoust. Soc. Am.*, vol. 79, pp. 1359–1366, 1986.
3. Pickett, J. M.: *The Sounds of Speech Communications*, University Park Press, Baltimore, MD, 1980.
4. Ehara, Shiro: "Instantaneous Pressure Distributions of Orchestra Sounds," *J. Acoust. Soc. Japan*, vol. 22, pp. 276–289, 1966.
5. Stephens, R. W. B., and A. E. Bate: *Acoustics and Vibrational Physics*, 2nd ed., E. Arnold (ed.), London, 1966.
6. Shaw, E. A. G.: "Noise Pollution—What Can be Done?" *Phys. Today*, vol. 28, no. 1, pp. 46–58, 1975.

Chapter
1.2
Sound Propagation

Floyd E. Toole

E. A. G. Shaw, G. A. Daigle, M. R. Stinson

1.2.1 Introduction

Sound propagating away from a source diminishes in strength at a rate determined by a variety of circumstances. It also encounters situations that can cause changes in amplitude and direction. Simple reflection is the most obvious process for directional change, but with sound there are also some less obvious mechanisms.

1.2.2 Inverse-Square and Other Laws

At increasing distances from a source of sound the level is expected to decrease. The rate at which it decreases is dictated by the directional properties of the source and the environment into which it radiates. In the case of a source of sound that is small compared with the wavelength of the sound being radiated, a condition that includes many common situations, the sound spreads outward as a sphere of ever-increasing radius. The sound energy from the source is distributed uniformly over the surface of the sphere, meaning that the intensity is the sound power output divided by the surface area at any radial distance from the source. Because the area of a sphere is $4\pi r^2$, the relationship between the sound intensities at two different distances is

$$\frac{I_1}{I_2} = \frac{r_2^2}{r_1^2} \quad (1.2.1)$$

where I_1 = intensity at radius r_1 , I_2 = intensity at radius r_2 , and

$$\text{Level difference} = 10 \log \frac{r_2^2}{r_1^2} = 20 \log \frac{r_2}{r_1} \text{ dB} \quad (1.2.2)$$

1-22 Principles of Sound and Hearing

This translates into a change in sound level of 6 dB for each doubling or halving of distance, a convenient mnemonic.

In practice, however, this relationship must be used with caution because of the constraints of real environments. For example, over long distances outdoors the absorption of sound by the ground and the air can modify the predictions of simple theory [1]. Indoors, reflected sounds can sustain sound levels to greater distances than predicted, although the estimate is correct over moderate distances for the *direct sound* (the part of the sound that travels directly from source to receiver without reflection). Large sound sources present special problems because the sound waves need a certain distance to form into an orderly wave-front combining the inputs from various parts of the source. In this case measurements in what is called the *near field* may not be representative of the integrated output from the source, and extrapolations to greater distances will contain errors. In fact the *far field* of a source is sometimes defined as being distances at which the inverse-square law holds true. In general, the far field is where the distance from the source is at least 2 to 3 times the distance between the most widely separated parts of the sound source that are radiating energy at the same frequency.

If the sound source is not small compared with the wavelength of the radiated sound, the sound will not expand outward with a spherical wavefront and the rate at which the sound level reduces with distance will not obey the inverse-square law. For example, a sound source in the form of a line, such as a long column of loudspeakers or a long line of traffic on a highway, generates sound waves that expand outward with a cylindrical wavefront. In the idealized case, such sounds attenuate at the rate of 3 dB for each doubling of distance.

1.2.3 Sound Reflection and Absorption

A sound source suspended in midair radiates into a *free field* because there is no impediment to the progress of the sound waves as they radiate in any direction. The closest indoor equivalent of this is an *anechoic room*, in which all the room boundaries are acoustically treated to be highly absorbing, thus preventing sounds from being reflected back into the room. It is common to speak of such situations as sound propagation in *full space*, or 4π steradians (sr; the units by which solid angles are measured).

In normal environments sound waves run into obstacles, such as walls, and the direction of their propagation is changed. Figure 1.2.1 shows the *reflection* of sound from various surfaces. In this diagram the pressure crests of the sound waves are represented by the curved lines, spaced one wavelength apart. The radial lines show the direction of sound propagation and are known as *sound rays*. For reflecting surfaces that are large compared with the sound wavelength, the normal *law of reflection* applies: the angle that the incident sound ray makes with the reflecting surface equals the angle made by the reflected sound ray.

This law also holds if the reflecting surface has irregularities that are small compared with the wavelength, as shown in Figure 1.2.1c, where it is seen that the irregularities have negligible effect. If, however, the surface features have dimensions similar to the wavelength of the incident sound, the reflections are *scattered* in all directions. At wavelengths that are small compared with the dimensions of the surface irregularities, the sound is also sent off in many directions but, in this case, as determined by the rule of reflections applied to the geometry of the irregularities themselves.

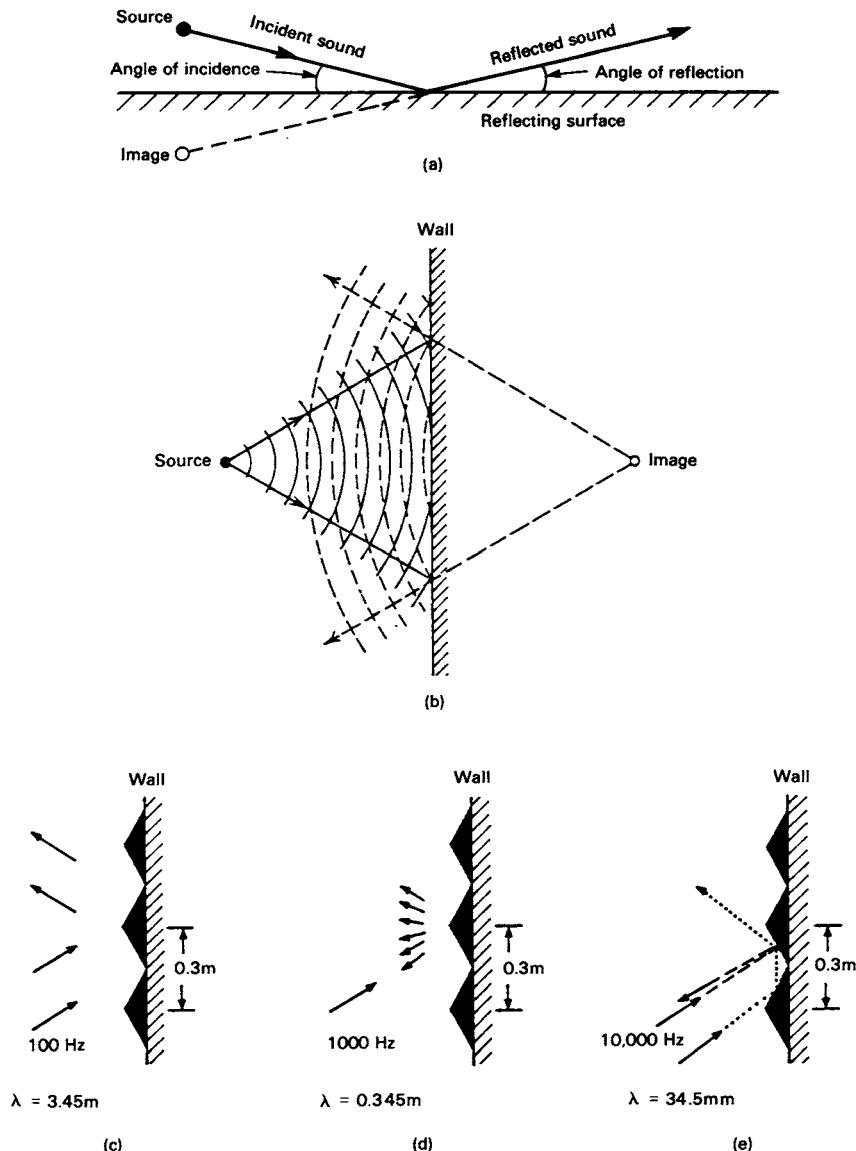


Figure 1.2.1 (a) The relationship between the incident sound, the reflected sound, and a flat reflecting surface, illustrating the law of reflection. (b) A more elaborate version of (a), showing the progression of wavefronts (the curved lines) in addition to the sound rays (arrowed lines). (c) The reflection of sound having a frequency of 100 Hz (wavelength 3.45 m) from a surface with irregularities that are small compared with the wavelength. (d) When the wavelength of the sound is similar to the dimensions of the irregularities, the sound is scattered in all directions. (e) When the wavelength of the sound is small compared with the dimensions of the irregularities, the law of reflection applies to the detailed interactions with the surface features.

1-24 Principles of Sound and Hearing

If there is perfect reflection of the sound, the reflected sound can be visualized as having originated at an image of the real source located behind the reflector and emitting the same sound power. In practice, however, some of the incident sound energy is *absorbed* by the reflecting surface; this fraction is called the *sound absorption coefficient* of the surface material. A coefficient of 0.0 indicates a perfect reflector, and a coefficient of 1.0 a perfect absorber; intermediate values indicate the portion of the incident sound energy that is dissipated in the surface and is not reflected. In general, the sound absorption coefficient for a material is dependent on the frequency and the angle of incidence of the sound. For simplicity, published values are normally given for octave bands of frequencies and for random angles of incidence.

1.2.3a Interference: The Sum of Multiple Sound Sources

The principle of *superposition* states that multiple sound waves (or electrical signals) appearing at the same point will add linearly. Consider two sound waves of identical frequency and amplitude arriving at a point in space from different directions. If the waveforms are exactly in step with each other, i.e., there is no phase difference, they will add perfectly and the result will be an identical waveform with double the amplitude of each incoming sound (6-dB-higher SPL). Such *in-phase* signals produce *constructive interference*. If the waveforms are shifted by one-half wavelength (180° phase difference) with respect to each other, they are *out of phase*; the pressure fluctuations are precisely equal and opposite, *destructive interference* occurs, and perfect cancellation results.

In practice, interference occurs routinely as a consequence of direct and reflected sounds adding at a microphone or a listener's ear. The amplitude of the reflected sound is reduced because of energy lost to absorption at the reflecting surface and because of inverse-square-law reduction related to the additional distance traveled. This means that constructive interference yields sound levels that are increased by less than 6 dB and that destructive interference results in imperfect cancellations that leave a residual sound level. Whether the interference is constructive or destructive depends on the relationship between the extra distance traveled by the reflection and the wavelength of the sound.

Figure 1.2.2 shows the direct and reflected sound paths for an omnidirectional source and receivers interacting with a reflecting plane. Note that there is an acoustically mirrored source, just as there would be a visually mirrored one if the plane were optically reflecting. If the distance traveled by the direct sound and that traveled by the reflected sound are different by an amount that is small and is also small compared with a wavelength of the sound under consideration (receiver R_1), the interference at the receiver will be constructive. If the plane is perfectly reflecting, the sound at the receiver will be the sum of two essentially identical sounds and the SPL will be about 6 dB higher than the direct sound alone. Constructive interference will also occur when the difference between the distances is an even multiple of half wavelengths. Destructive interference will occur for odd multiples of half wavelengths.

As the path length difference increases, or if there is absorption at the reflective surface, the difference in the sound levels of the direct and reflected sounds increases. For receivers R_2 and R_3 in Figure 1.2.2, the situation will differ from that just described only in that, because of the additional attenuation of the reflected signal, the constructive peaks will be significantly less than 6 dB and the destructive dips will be less than perfect cancellations.

For a fixed geometrical arrangement of source, reflector, and receiver, this means that at sufficiently low frequencies the direct and reflected sounds add. As the wavelength is reduced (fre-

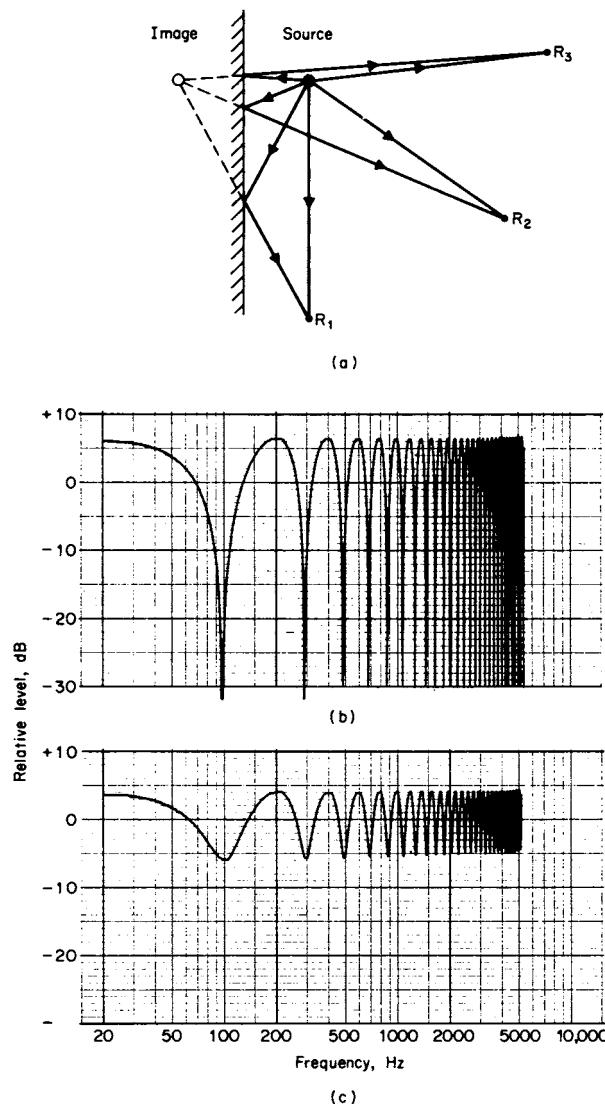


Figure 1.2.2 (a) Differing direct and reflected path lengths as a function of receiver location. (b) The interference pattern resulting when two sounds, each at the same sound level (0 dB) are summed with a time delay of just over 5 ms (a path length difference of approximately 1.7 m). (c) The reflection signal has been attenuated by 6 dB (it is now at a relative level of -6 dB, while the direct sound remains at 0 dB); the maximum sound level is reduced, and perfect nulls are no longer possible. The familiar comb-filtering pattern remains.

1-26 Principles of Sound and Hearing

quency rising), the sound level at the receiver will decline from the maximum level in the approach to the first destructive interference at $\lambda/2 = r_2 - r_1$, where the level drops to a null. Continuing upward in frequency, the sound level at the receiver rises to the original level when $\lambda = r_2 - r_1$, falls to another null at $3\lambda/2 = r_2 - r_1$, rises again at $2\lambda = r_2 - r_1$, and so on, alternating between maxima and minima at regular intervals in the frequency domain. The plot of the frequency response of such a transmission path is called an *interference pattern*. It has the visual appearance of a comb, and the phenomenon has also come to be called *comb filtering* (see Figure 1.2.2b).

Taking a more general view and considering the effects averaged over a range of frequencies, it is possible to generalize as follows for the influence of a single reflecting surface on the sound level due to the direct sound alone [2].

- When $r_2 - r_1$ is much less than a wavelength, the sound level at the receiver will be elevated by 6 dB or less, depending on the surface absorption and distances involved.
- When $r_2 - r_1$ is approximately equal to a wavelength, the sound level at the receiver will be elevated between 3 and 6 dB, depending on the specific circumstances.
- When $r_2 - r_1$ is much greater than a wavelength, the sound level at the receiver will be elevated by between 0 and 3 dB, depending on the surface absorption and distances involved.

A special case occurs when the sound source, such as a loudspeaker, is mounted in the reflecting plane itself. There is no path length difference, and the source radiates into a hemisphere of free space, more commonly called a *half space*, or 2π sr. The sound level at the receiver is then elevated by 6 dB at frequencies where the sound source is truly omnidirectional, which—in practice—is only at low frequencies.

Other reflecting surfaces contribute additively to the elevation of the sound level at the receiver in amounts that can be arrived at by independent analysis of each. Consider the situation in which a simple point monopole (omnidirectional) source of sound is progressively constrained by reflecting planes intersecting at right angles. In practice this could be the boundaries of a room that are immediately adjacent to a loudspeaker which, at very low frequencies, is effectively an omnidirectional source of sound. Figure 1.2.3 summarizes the relationships between four common circumstances, where the sound output from the source radiates into solid angles that reduce in stages by a factor of 2. These correspond to a loudspeaker radiating into free space (4π sr), placed against a large reflecting surface (2π sr), placed at the intersection of two reflecting surfaces (π sr), and placed at the intersection of three reflecting surfaces ($\pi/2$ sr). In all cases the dimensions of the source and its distance from any of the reflecting surfaces are assumed to be a small fraction of a wavelength. The source is also assumed to produce a constant volume velocity of sound output; i.e., the volumetric rate of air movement is constant throughout.

By using the principles outlined here and combining the outputs from the appropriate number of image sources that are acoustically mirrored in the reflective surfaces, it is found that the sound pressure at a given radius increases in inverse proportion to the reduction in solid angle; sound pressure increases by a factor of 2, or 6 dB, for each halving of the solid angle.

The corresponding sound intensity (the sound power passing through a unit surface area of a sphere of the given radius) is proportional to pressure squared. Sound intensity therefore increases by a factor of 4 for each halving of the solid angle. This also is 6 dB for each reduction in angle because the quantity is power rather than pressure.

Finally, multiplying the sound intensity by the surface area at the given radius yields the total sound power radiated into the solid angle. Because the surface area at each transition is reduced

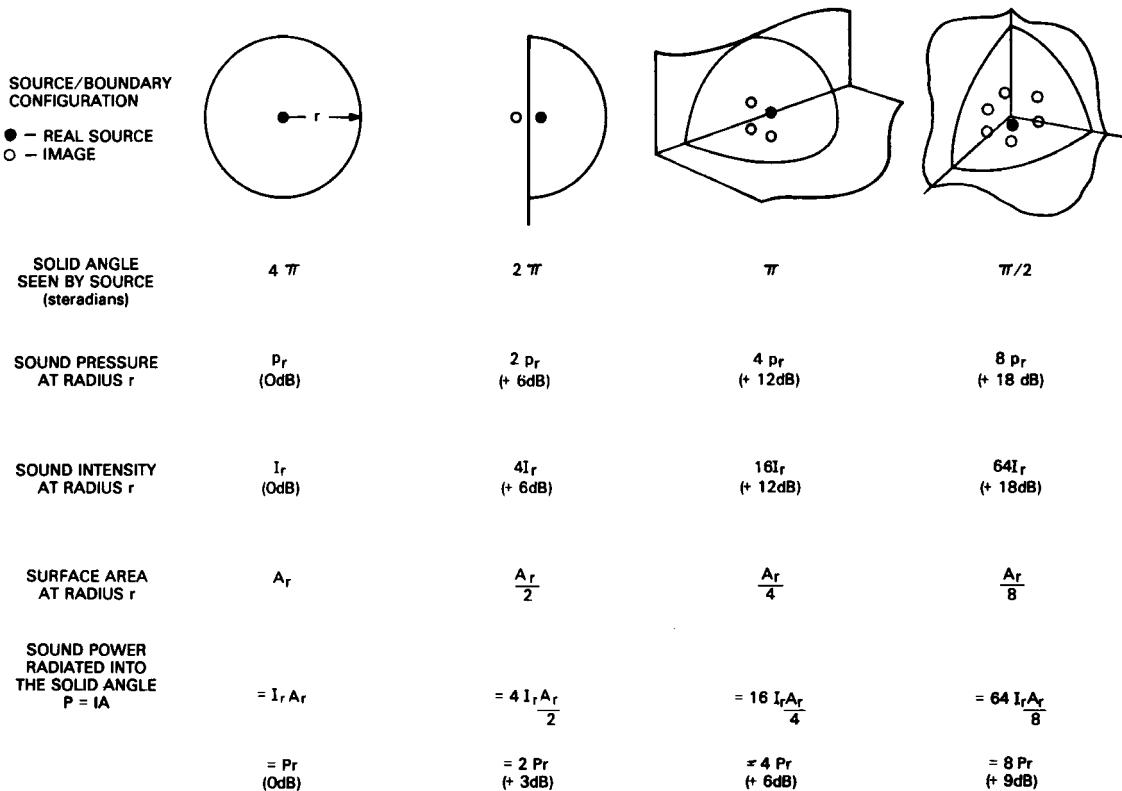


Figure 1.2.3 Behavior of a point monopole sound source in full space (4π) and in close proximity to reflecting surfaces that constrain the sound radiation to progressively smaller solid angles. (After [3].)

by a factor of 2, the total sound power radiated into the solid angle increases by a factor of 2, or 3 dB, for each halving of the solid angle.

By applying the reverse logic, reducing the solid angle by half increases the rate of energy flow into the solid angle by a factor of 2. At a given radius, this energy flows through half of the surface area that it previously did, so that the sound intensity is increased by a factor of 4; i.e., pressure squared is increased by a factor of 4. This means that sound pressure at that same radius is increased by a factor of 2.

The simplicity of this argument applies when the surfaces shown in Figure 1.2.3 are the only ones present; this can only happen outdoors. In rooms there are the other boundaries to consider, and the predictions discussed here will be modified by the reflections, absorption, and standing-wave patterns therein.

1-28 Principles of Sound and Hearing

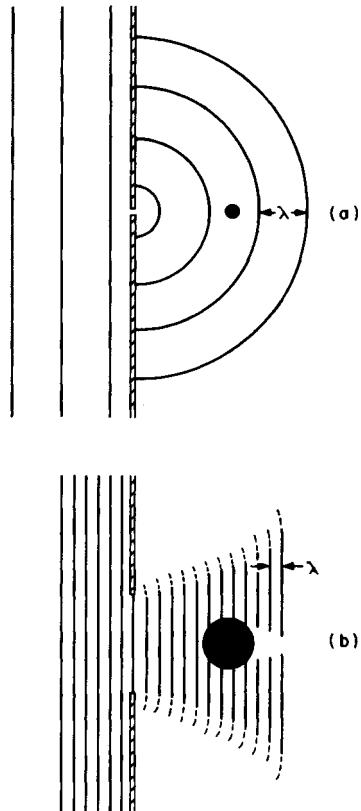


Figure 1.2.4 Stylized illustration of the diffraction of sound waves passing through openings and around obstacles. (a) The case where the wavelength is large compared with the size of the opening and the obstacle. (b) The case where the wavelength is small compared with the size of the opening and the obstacle.

1.2.3b Diffraction

The leakage of sound energy around the edges of an opening or around the corners of an obstacle results in a bending of the sound rays and a distortion of the wave-front. The effect is called *diffraction*. Because of diffraction it is possible to hear sounds around corners and behind walls—anywhere there might have been an “acoustical shadow.” In fact, acoustical shadows exist, but to an extent that is dependent on the relationship between the wavelength and the dimensions of the objects in the path of the sound waves.

When the openings or obstructions are small compared with the wavelength of the sound, the waves tend to spread in all directions and the shadowing effect is small. At higher frequencies, when the openings or obstructions are large compared with the wavelengths, the sound waves tend to continue in their original direction of travel and there is significant shadowing. Figure 1.2.4 illustrates the effect.

The principle is maintained if the openings are considered to be the diaphragms of loudspeakers. If one wishes to maintain wide dispersion at all frequencies, the radiating areas of the driver units must progressively reduce at higher frequencies. Conversely, large radiating areas can be used to restrict the dispersion, though the dimensions required may become impractically large at

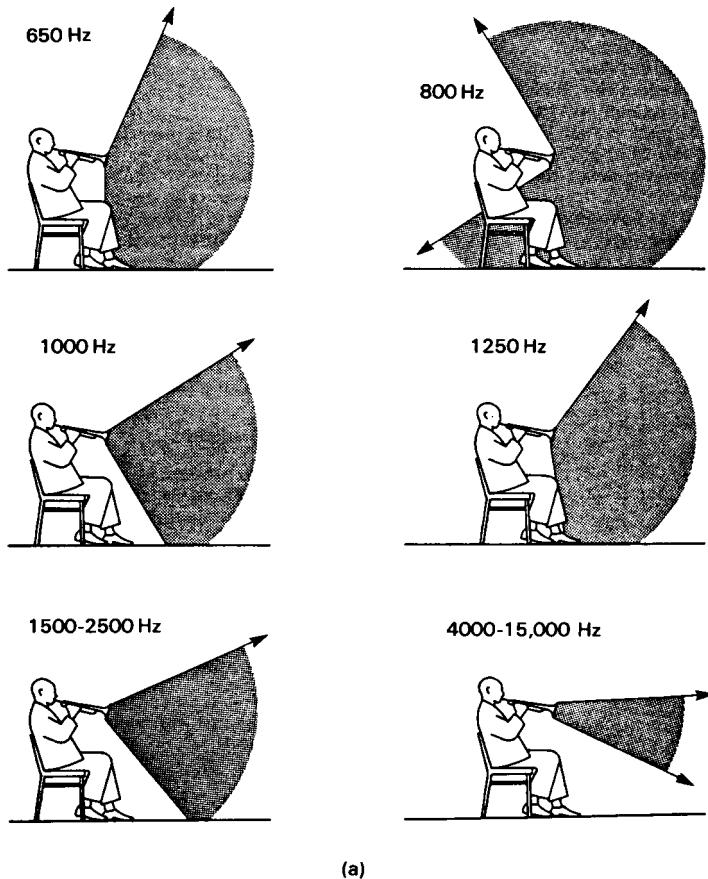


Figure 1.2.5 A simplified display of the main sound radiation directions at selected frequencies for: (a) a trumpet, (b, next page) a cello. (From [4]. Used with permission.)

low frequencies. As a consequence, most loudspeakers are approximately omnidirectional at low frequencies.

Sounds radiated by musical instruments obey the same laws. Low-frequency sounds from most instruments and the human voice radiate in all directions. Higher-frequency components can exhibit quite strong directional biases that are dependent on the size and orientation of the major sound-radiating elements. Figure 1.2.5a shows the frequency-dependent directivities of a trumpet, a relatively simple source. Compare this with the complexity of the directional characteristics of a cello (Figure 1.2.5b). It is clear that no single direction is representative of the total sound output from complex sound sources—a particular difficulty when it comes to choosing microphone locations for sound recordings. Listeners at a live performance hear a combination of all the directional components as spatially integrated by the stage enclosure and the hall itself.

1-30 Principles of Sound and Hearing

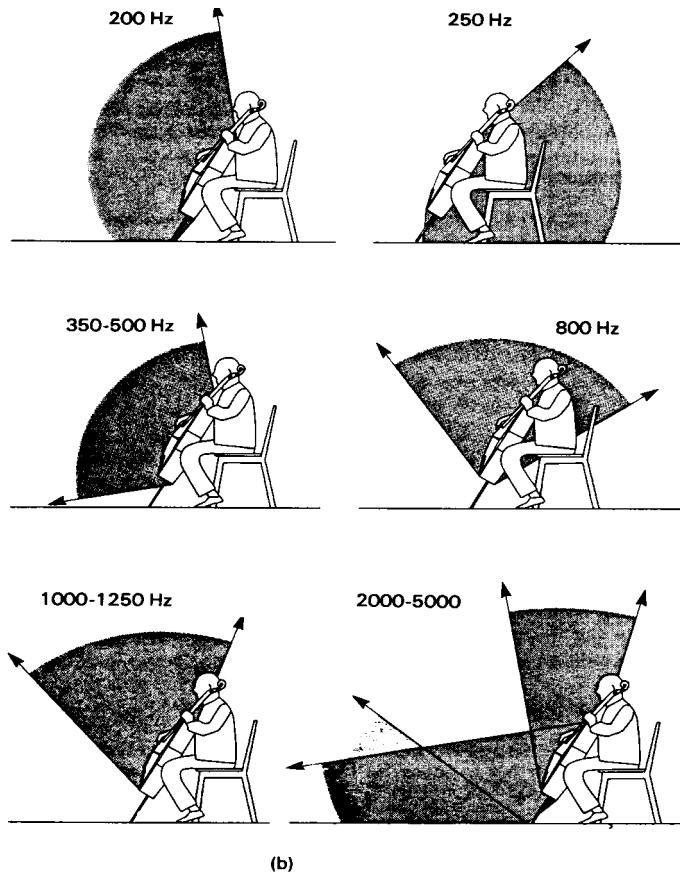


Figure 1.2.5b

1.2.3c Refraction

Sound travels faster in warm air than in cold and faster downwind than upwind. These factors can cause sound rays to be bent, or *refracted*, when propagating over long distances in vertical gradients of wind or temperature. Figure 1.2.6 shows the downward refraction of sound when the propagation is downwind or in a *temperature inversion*, as occurs at night when the temperature near the ground is cooler than the air higher up. Upward refraction occurs when the propagation is upwind or in a *temperature lapse*, a typical daytime condition when the air temperature falls with increasing altitude. Thus, the ability to hear sounds over long distances is a function of local climatic conditions; the success of outdoor sound events can be significantly affected by the time of day and the direction of prevailing winds.

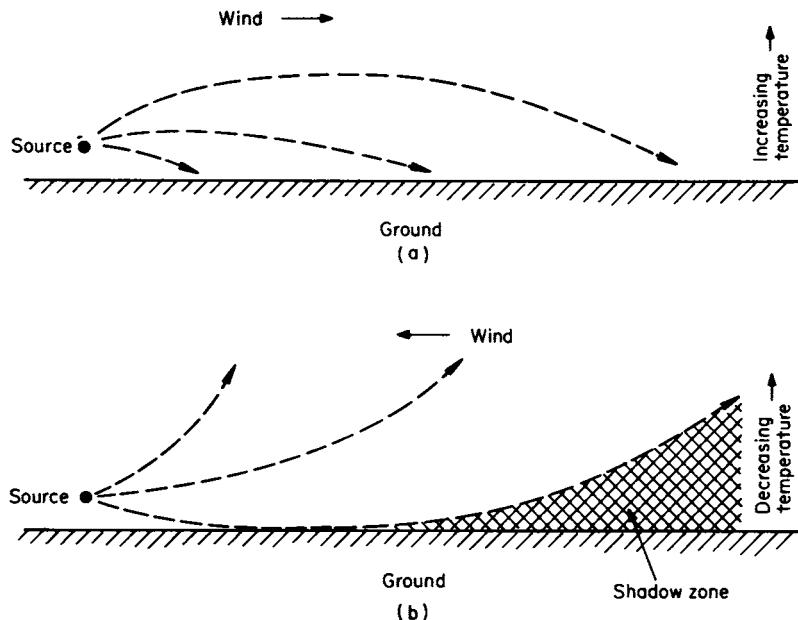


Figure 1.2.6 The refraction of sound by wind and by temperature gradients: (a) downwind or in a temperature inversion, (b) upwind or in a temperature lapse. (From [1]. Used with permission.)

1.2.4 References

1. Piercy, J. E., and T. F. W. Embleton: "Sound Propagation in the Open Air," in *Handbook of Noise Control*, 2d ed., C. M. Harris (ed.), McGraw-Hill, New York, N.Y., 1979.
2. Waterhouse, R. V., and C. M. Harris: "Sound in Enclosed Spaces," in *Handbook of Noise Control*, 2d ed., C. M. Harris (ed.), McGraw-Hill, New York, N.Y., 1979.
3. Olson, Harry F.: *Acoustical Engineering*, Van Nostrand, New York, N.Y., 1957.
4. Meyer, J.: *Acoustics and the Performance of Music*, Verlag das Musikinstrument, Frankfurt am Main, 1987.

Chapter
1.3
Resonance

Floyd E. Toole

E. A. G. Shaw, G. A. Daigle, M. R. Stinson

1.3.1 Introduction

A vibrating system of any kind that is driven by and is completely under the control of an external source of energy is in a state of *forced vibration*. The activity within such a system after the external force has been removed is known as *free vibration*. In this condition most systems exhibit a tendency to move at a natural or *resonant* frequency, declining with time at a rate determined by the amount of energy dissipation, or *damping*, in the system. The resonances in some musical instruments have little damping, as the devices are intended to resonate and produce sound at specific frequencies in response to inputs, such as impacts or turbulent airflow, that do not have any specific frequency characteristics. Most instruments provide the musician with some control over the damping so that the duration of the notes can be varied.

1.3.2 Fundamental Properties

If the frequency of the driving force is matched to the natural frequency of the resonant system, the magnitude of the vibration and the efficiency of the energy transfer are maximized.

These and other points are illustrated in Figure 1.3.1, which shows three versions of a resonant system having different amounts of damping. The term commonly used to describe this characteristic of resonant systems is the *quality factor*, Q , a measure of the lightness of damping in a system. The system in Figure 1.3.1a has a Q of 1; it is well damped. The system in Figure 1.3.1b is less well damped and has a Q of 10, while that in Figure 1.3.1c has little damping and is described as having a Q of 50. As a practical example, the resonance of a loudspeaker in an enclosure would typically have a Q of 1 or less. Panel resonances in enclosures might have Q s in the region of 10 or so. Resonances with a Q of 50 or more would be rare in sound reproducers but common in musical instruments.

On the left in Figure 1.3.1 can be seen the behavior of these systems when they are forced into oscillation by a pure tone tuned to the resonance frequency of the systems, 1000 Hz. When the

1-34 Principles of Sound and Hearing

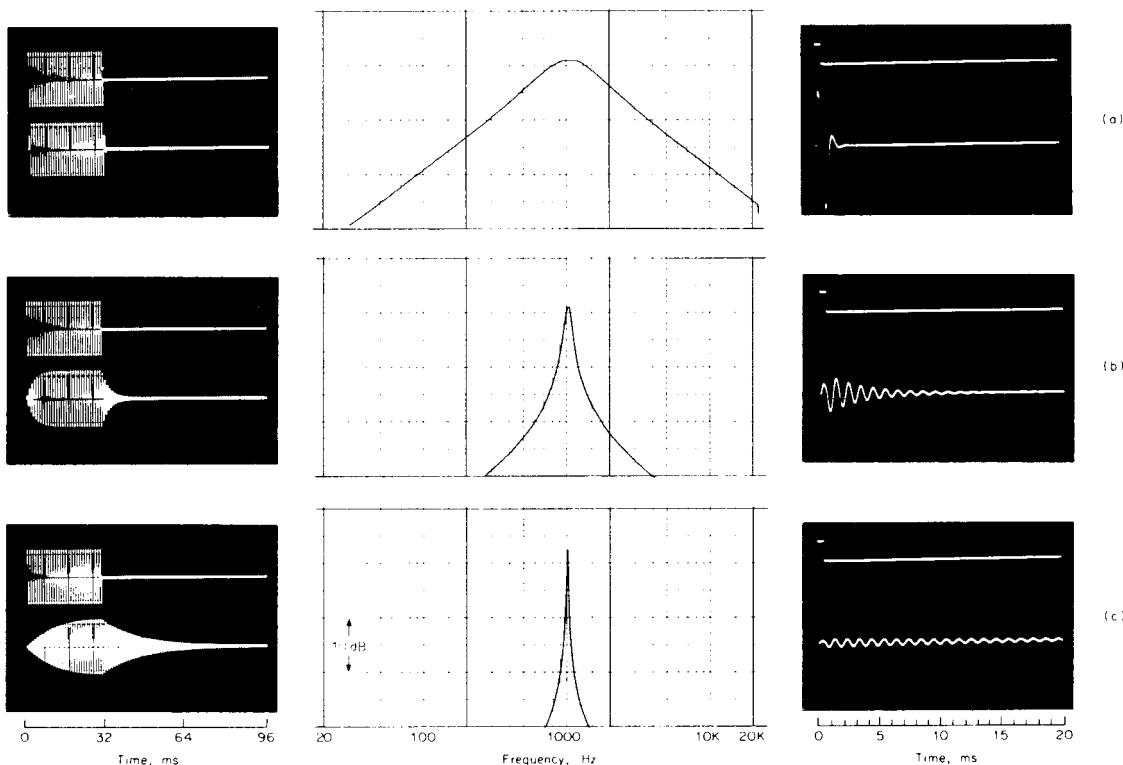


Figure 1.3.1 The frequency responses of three resonant systems and their behavior in conditions of forced and free vibration. The system shown in (a) has the least damping ($Q = 1$), system (b) has moderate damping ($Q = 10$), and the system shown in (c) has the least damping ($Q = 50$).

tone is turned on and off, the systems respond with a speed that is in inverse proportion to the Q . The low- Q resonance responds quickly to the onset of the tone and terminates its activity with equal brevity. The medium- Q system responds at a more leisurely rate and lets the activity decay at a similar rate after the cessation of the driving signal. The high- Q system is slow to respond to the driving signal and sustains the activity for some time after the interval of forced oscillation.

In the preceding example the forcing signal was optimized in frequency, in that it matched the resonance frequency of the system, and it was sustained long enough for the system to reach its level of maximum response. On the right of Figure 1.3.1 are shown the responses of these systems to an impulse signal brief in the time domain but having energy over a wide range of frequencies including that of the resonant system. In Figure 1.3.1a the low- Q system is shown responding energetically to this signal but demonstrating little sustained activity. In Figure 1.3.1b and 1.3.1c the higher- Q systems respond with progressively reduced amplitude but with progressively sustained ringing after the pulse has ended. Note that the ringing is recognizably at the resonance frequency, 1 cycle/ms.

In the center of Figure 1.3.1 are shown the amplitude-frequency responses or, more commonly, the *frequency responses* of the systems. These curves show the amplitude of system response when the frequency of a constant driving signal is varied from well below the resonance frequency to well above it. The low-*Q* system Figure 1.3.1a is seen to respond to signals over a wide frequency range, but the higher-*Q* systems become progressively more frequency-selective.

In this illustration, the maximum amplitudes of the system responses at resonance were adjusted to be approximately equal. Such is often the case in electronic resonators used in filters, frequency equalizers, synthesizers, and similar devices. In simple resonant systems in which everything else is held equal and only the damping is varied, the maximum amplitude response would be highest in the system with the least dissipation: the high-*Q* system, Figure 1.3.1c. Adding damping to the system would reduce the maximum amplitude, so that the system with the lowest *Q*, having the highest damping or losses, would respond to the widest range of frequencies, but with reduced amplitude [1].

Figure 1.3.2 shows the frequency responses of two systems with multiple resonances. In 1.3.2a the resonances are such that they respond independently to driving forces at single frequencies. In 1.3.2b an input at any single frequency would cause some activity in all the resonators but at different amplitudes in each one. The series of high-*Q* resonators in Figure 1.3.2a is characteristic of musical instruments, where the purpose is the efficient production of sound at highly specific frequencies. The overlapping set of low-*Q* resonators in Figure 1.3.2b are the filters of a parametric equalizer in which the frequency, *Q*, and amplitude of the filters are individually adjustable to provide a variable overall frequency response for a sound-recording or sound-reproducing system.

A special case of Figure 1.3.2b would be a multiway loudspeaker system intended for the reproduction of sounds of all kinds. In this case, the selection of loudspeaker units and their associated filters (crossovers) would be such that, in combination, they resulted in an overall amplitude response that is flat (the same at all frequencies) over the required frequency range. Such a system would be capable of accurately recreating any signal spectrum. For the loudspeaker or any system of multiple filters or resonant elements to accurately pass or reproduce a complex waveform, there must be no phase shift at the important frequencies. In technical terms this would be assessed by the *phase-frequency response*, or *phase response*, of the system showing the amount of phase shift at frequencies within the system bandwidth.

Resonant systems can take any of several forms of electrical, mechanical, or acoustical elements or combinations thereof. In electronics, resonators are the basis for frequency-selective or tuned circuits of all kinds, from radios to equalizers and music synthesizers. Mechanical resonances are the essential pitch determinants of tuning forks, bells, xylophones, and glockenspiels. Acoustical resonances are the essential tuning devices of organs and other wind instruments. Stringed instruments involve combinations of mechanical and acoustical resonances in the generation and processing of their sounds, as do reed instruments and the human voice.

The voice is a good example of a complex resonant system. The sound originates as a train of pulses emanating from the voice box. This excites a set of resonances in the vocal tract so that the sound output from the mouth is emphasized at certain frequencies. In spectral terms, the envelope of the line spectrum is modified by the frequency response of the resonators in the vocal tract. These resonances are called *formants*, and their frequencies contribute to the individual character of voices. The relative amplitudes of the resonances are altered by changing the physical form of the vocal tract so as to create different vowel sounds, as illustrated in Figure 1.3.3 [2–4].

1-36 Principles of Sound and Hearing

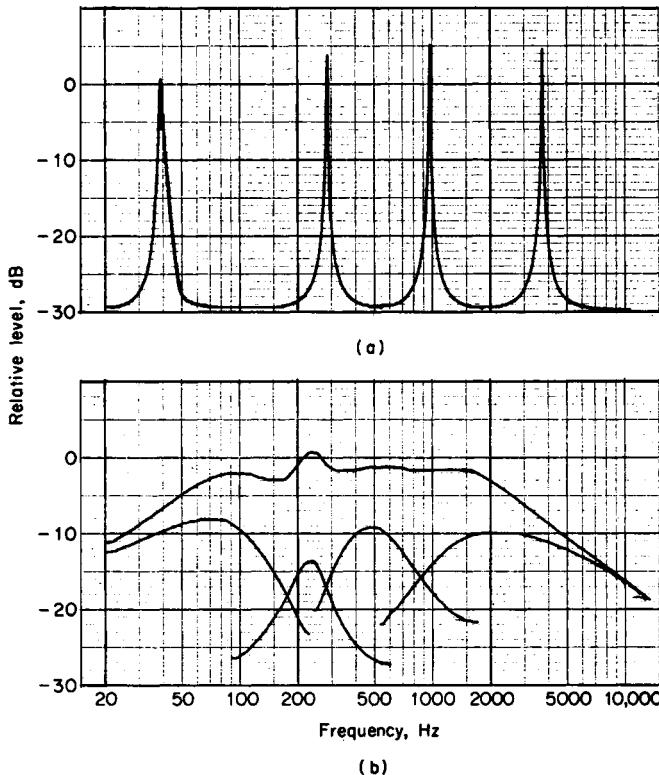


Figure 1.3.2 Two systems with multiple resonances: (a) well-separated high- Q resonances that can respond nearly independently of each other, as in the notes of a musical instrument; (b) the four filters of a parametric equalizer designed to produce overlapping low- Q resonance curves (bottom traces) which are combined to produce a total response (top trace) that may bear little resemblance to the individual contributions.

1.3.2a Resonance in Pipes

When the diameter of a pipe is small compared with the wavelength, sound will travel as plane waves perpendicular to the length of the pipe. At a closed end the sound is reflected back down the pipe in the reverse direction. At an open end, some of the sound radiates outward and the remainder is reflected backward, but with a pressure reversal (180° phase shift). The pressure distribution along the pipe is therefore the sum of several sound waves traveling backward and forward. At most frequencies the summation of these several waves results in varying degrees of destructive interference, but at some specific frequencies the interference is only constructive and a pattern stabilizes in the form of *standing waves*. At these frequencies, the wavelengths of the sounds are such that specific end conditions of the tube are simultaneously met by the waves traveling in both directions, the sounds reinforce each other, and a resonant condition exists.

Figures 1.3.4 and 1.3.5 show the first three *resonant modes* for pipes open at both ends and for those with one end closed. The open ends prevent the pressures from building up, but the par-

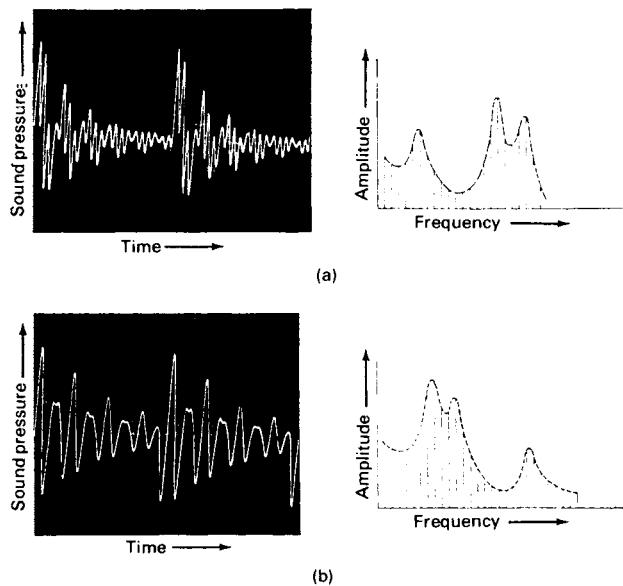


Figure 1.3.3 The waveforms and corresponding amplitude-frequency spectra of the vowel sounds “uh” (a) and “ah” (b). (From [3]. Used with permission.)

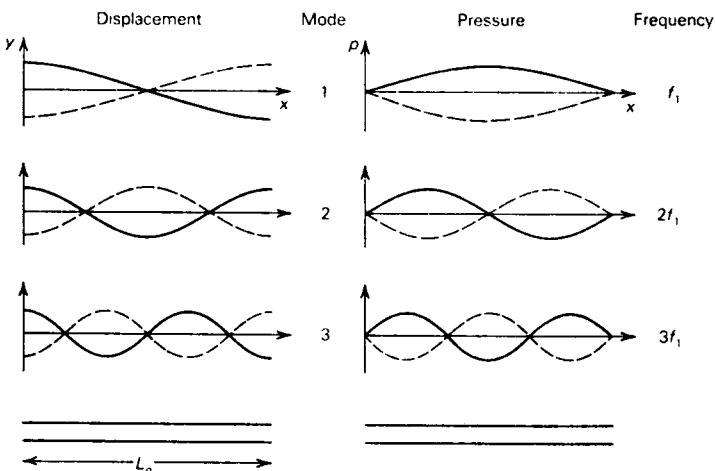


Figure 1.3.4 The first three resonant modes of air in a tube open at both ends. On the left are the patterns of particle displacement along the tube, showing the antinodes at the ends of the tube. At the right are the corresponding patterns of pressure, with the required nodes at the ends. The fundamental frequency is $c/2L_o$ (From [7]. Used with permission.)

1-38 Principles of Sound and Hearing

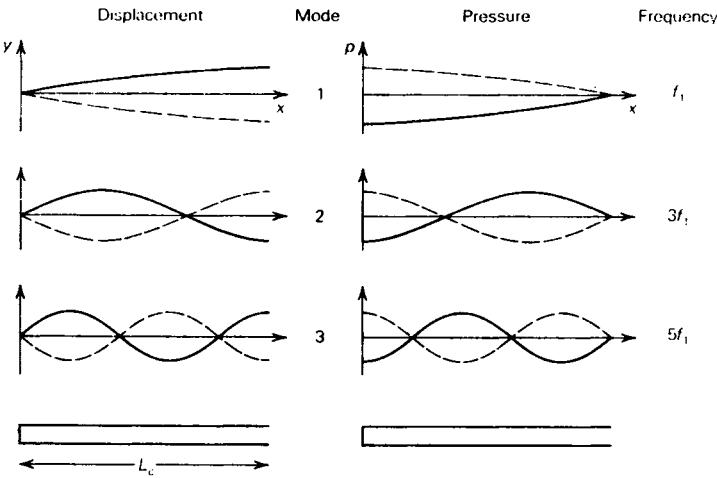


Figure 1.3.5 The first three resonant modes of air in a tube closed at one end. On the left are the patterns of particle displacement along the tube, and on the right are the pressure distributions. The fundamental frequency is $c/4L_o$ (From [7]. Used with permission.)

ticle displacements are unimpeded; the end condition for resonance is therefore a displacement maximum (*antinode*) and a pressure minimum (*node*) in the standing-wave pattern. A closed end does the reverse, forcing displacements to zero but permitting pressure to build up; the end condition for resonance is therefore a displacement node and a pressure antinode.

For a pipe open at both ends, the fundamental frequency has a wavelength that is double the length of the pipe; conversely, the pipe is one-half wavelength long. The fundamental frequency is therefore $f = c/2L_o$ where L is the length of the pipe in meters and c is the speed of sound: 345 m/s. Other resonances occur at all harmonically related frequencies: $2f_1$, $3f_1$, and so on.

A pipe closed at one end is one-quarter wavelength long at the fundamental resonance frequency; thus $f = c/4L_c$. In this case, however, the other resonances occur at odd harmonics only: $3f_1$, $5f_1$, and so on. A very simplistic view of the vocal tract considers it as a pipe, closed at the vocal cords, open at the mouth, and 175 mm long [4]. This yields a fundamental frequency of about 500 Hz and harmonics at 1500, 2500, and 3500 Hz. These are close to the formant frequencies appearing as resonance humps in the spectra of Figure 1.3.3.

Organ pipes are of both forms, although the pipes open at both ends produce the musically richer sound. To save space, pipes closed at one end are sometimes used for the lowest notes; these need be only one-fourth wavelength long, but they produce only odd harmonics.

In practice this simple theory must be modified slightly to account for what is called the *end correction*. This can be interpreted as the distance beyond the open end of the pipe over which the plane waves traveling down the pipe make the transition to spherical wavefronts as they diverge after passing beyond the constraints of the pipe walls. The pipe behaves as if it is longer than its physical length by an amount equal to 0.62 times its radius. If the pipe has a flange or opens onto a flat surface, the end correction is 0.82 times the radius.

1.3.2b Resonance in Rooms and Large Enclosures

Sounds propagating in rectangular rooms and large enclosures are subject to standing waves between the reflecting boundaries. In taking a one-dimensional view for illustration, sounds reflecting back and forth between two parallel surfaces form standing waves at frequencies satisfying the boundary conditions requiring pressure antinodes and particle displacement nodes at the reflecting surfaces. The fundamental resonance frequency is that at which the separation is one-half wavelength. Other resonances occur at harmonics of this frequency. This same phenomenon exists between all opposing pairs of parallel surfaces, establishing three sets of resonances, dependent on the length, width, and height, known as the *axial modes* of the enclosure. Other resonances are associated with sounds reflected from four surfaces and propagating in a plane parallel to the remaining two. For example, sound can be reflected from the four walls and travel parallel to the floor and ceiling. The three sets of these resonances are called *tangential modes*. Finally, there are resonances involving sounds reflected from all surfaces in the enclosure, called *oblique modes*. All these resonant modes, or *eigenmodes*, can be calculated from the following equation

$$f_n = \frac{c}{2} \sqrt{\left(\frac{n_x}{l_x}\right)^2 + \left(\frac{n_y}{l_y}\right)^2 + \left(\frac{n_z}{l_z}\right)^2} \quad (1.3.1)$$

where:

f_n = frequency of the n th mode

n_x, n_y, n_z = integers with independently chosen values between 0 and ∞

l_x, l_y, l_z = dimensions of enclosure, m (ft)

c = speed of sound, m/s (ft/s)

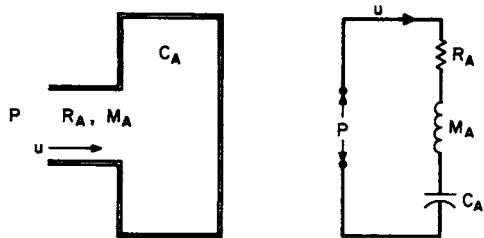
It is customary to identify the individual modes by a combination of n_x , n_y , and n_z , as in (2, 0, 0), which identifies the mode as being the second-harmonic resonance along the x dimension of the enclosure. All axial modes are described by a single integer and two zeros. Tangential modes are identified by two integers and one zero, and oblique modes by three integers. The calculation of all modes for an enclosure would require the calculation of Equation (1.3.1) for all possible combinations of integers for n_x , n_y , and n_z .

The sound field inside an enclosure is therefore a complex combination of many modes, and after the sound input has been terminated, they can decay at quite different rates depending on the amount and distribution of acoustical absorption on the room boundaries. Because some energy is lost at every reflection, the modes that interact most frequently with the room boundaries will decay first. The oblique modes have the shortest average distance between reflections and are the first to decay, followed by the tangential modes and later by the axial modes. This means that the sound field in a room is very complex immediately following the cessation of sound production, and it rapidly deteriorates to a few energetic tangential and axial modes [5, 6].

The ratio of length to width of an enclosure determines the distribution of the resonant modes in the frequency domain. The dimensions themselves determine the frequencies of the modes. The efficiency with which the sound source and receiver couple to the various modes determines the relative influence of the modes in the transmission of sound from the source to the receiver. These factors are important in the design of enclosures for specific purposes. In a listening or control room, for example, the locations of the loudspeakers and listeners are largely

1-40 Principles of Sound and Hearing

Figure 1.3.6 Physical representation of a Helmholtz resonator (left) and the corresponding symbolic representation as a series resonant acoustical circuit (right). Legend: P = sound pressure at mouth; u = volume velocity at the port = particle velocity \times port area; R_A = acoustical resistance; M_A = acoustical mass of the port; C_A = acoustical compliance of the volume.



determined by the geometrical requirements for good stereo listening and by restrictions imposed by the loudspeaker design. Accurate communication from the source to the receiver over a range of frequencies requires that the influential room modes be uniformly distributed in frequency. Clusters or gaps in the distribution of modes can cause sounds at some frequencies to be accentuated and others to be attenuated, altering the frequency response of the sound propagation path through the room. This causes the timbre of sounds propagated through the room to be changed.

Certain dimensional ratios have been promoted as having especially desirable mode distributions. Indeed, there are shapes like cubes and corridors that clearly present problems, but the selection of an ideal rectangular enclosure must accommodate the particular requirements of the application. Generalizations based on the simple application of Equation (1.3.1) assume that the boundaries of the enclosure are perfectly reflecting and flat, that all modes are equally energetic, and that the source and receiver are equally well coupled to them all. In practice it is highly improbable that these conditions will be met.

1.3.2c Resonance in Small Enclosures: Helmholtz Resonators

At frequencies where the wavelength is large compared with the interior dimensions of an enclosure, there is negligible wave motion because the sound pressure is nearly uniform throughout the volume. In these circumstances the *lumped-element* properties of the enclosed air dominate, and another form of resonance assumes control. Such *Helmholtz resonators* form an important class of acoustic resonators.

Figure 1.3.6 shows a simple cavity with a short ducted opening, like a bottle with a neck. Here the volume of air within the cavity acts as a spring for the mass of air in the neck, and the system behaves as the acoustical version of a mechanical spring-mass resonant system. It is also analogous to the electrical resonant circuit with elements as shown in the figure.

Acoustical compliance increases with the volume, meaning that the resonance frequency falls with increasing cavity volume. The acoustic mass (*inertance*) in the duct increases with the length of the duct and decreases with increasing duct area, leading to a resonance frequency that is proportional to the square root of the duct area and inversely proportional to the square root of the duct length.

Helmholtz resonators are the simplest form of resonating systems. They are found as the air resonance in the body of guitars, violins, and similar instruments, and they are the principal frequency-determining mechanism in whistles and ocarinas. They also describe the performance of loudspeaker-enclosure systems at low frequencies. The acoustical-mechanical-electrical analogs introduced here are the basis for the design of closed-box and reflex loudspeaker systems, result-

ing in closely predictable performance at low frequencies. At higher frequencies, standing waves form inside the box, and the tidy lumped-element concepts no longer apply.

1.3.2d Horns

If the open end of a tube has a diameter that is small compared with the wavelength of sound being propagated within it, most of the sound is reflected back into the tube, and if the wavelength is appropriate, standing waves result. At resonance, the acoustical activity is at its maximum, but the small tube opening is nevertheless a rather inefficient radiator of sound. If strong resonances are important and adequate input power is available, as in organ pipes, this is a desirable situation. Other devices, however, require the maintenance of strong standing waves, but with an improved radiation efficiency. With care this is achieved through the use of a flared development, or *horn*, at the end of the pipe. The shape and size of the horn determine, for every frequency, how much of the sound is reflected back into the tube and how much radiates outward.

The musical instruments of the brass family are all combinations of resonant pipes with a flaring bell at the output end. The shape of a trumpet bell, for example, is such that it has radiation efficiency that is low below about 1500 Hz and high above. By establishing strong resonances at the fundamental playing frequencies, the bell makes the instrument playable while imparting a bright sound character by efficiently radiating the higher harmonics of the basic pitch [7, 8].

On the other hand, a loudspeaker horn must have high radiation efficiency at all frequencies within its operating range; otherwise there will be resonances in a system that is intended to be free of such sources of tone color. The key to non-resonant behavior lies in the choice of flare shape and mouth size. The sound waves propagating outward must be allowed to expand at just the proper rate, maintaining close control over the directions of the particle velocities, so that the waves can emerge from the enlarged mouth with little energy reflected back to the loudspeaker. [5].

1.3.3 References

1. Main, Ian G.: *Vibrations and Waves in Physics*, Cambridge, London, 1978.
2. Pickett, J. M.: *The Sounds of Speech Communication*, University Park Press, Baltimore, MD, 1980.
3. Denes, Peter B., and E. N. Pinson: *The Speech Chain*, Bell Telephone Laboratories, Waverly, 1963.
4. Sundberg, Johan: "The Acoustics of the Singing Voice," in *The Physics of Music*, introduction by C. M. Hutchins, Scientific American/Freeman, San Francisco, Calif., 1978.
5. Morse, Philip M.: *Vibrations and Sound*, 1964, reprinted by the Acoustical Society of America, New York, N.Y., 1976.
6. Mankovsky, V. S.: *Acoustics of Studios and Auditoria*, Focal Press, London, 1971.
7. Hall, Donald E.: *Musical Acoustics: An Introduction*, Wadsworth, Belmont, Calif., 1980.

1-42 Principles of Sound and Hearing

8. Benade, A. H.: *Fundamentals of Musical Acoustics*, Oxford University Press, New York, N.Y., 1976.

Chapter**1.4****The Physical Nature of Hearing****Floyd E. Toole****E. A. G. Shaw, G. A. Daigle, M. R. Stinson****1.4.1 Introduction**

The process of hearing begins with acoustical modifications to the sound waves as they interact with the head and the external ear, the visible portion of the system. These acoustical changes are followed by others in the ear canal and by a conversion of the sound pressure fluctuations into mechanical displacements by the eardrum. Transmitted through the mechanical coupling system of the middle ear to the inner ear, the displacement patterns are partially analyzed and then encoded in the form of neural signals. The signals from the two ears are cross-compared at several stages on the way to the auditory centers of the brain, where finally there is a transformation of the streams of data into perceptions of sound and acoustical space.

By these elaborate means we are able to render intelligible acoustical signals that, in technical terms, can be almost beyond description. In addition to the basic information, the hearing process keeps us constantly aware of spatial dimensions, where sounds are coming from, and the general size, shape, and decor of the space around us—a remarkable process indeed.

1.4.2 Anatomy of the Ear

Figure 1.4.1a shows a cross section of the ear in a very simplified form in which the outer, middle, and inner ear are clearly identified. The head and the outer ear interact with the sound waves, providing acoustical amplification that is dependent on both direction and frequency, in much the same way as an antenna. At frequencies above about 2 kHz there are reflections and resonances in the complex folds of the *pinna* [1]. Consequently, sounds of some frequencies reach the *tympanic membrane* (eardrum) with greater amplitude than sounds of other frequencies. The amount of the sound pressure gain or loss depends on both the frequency and the angle of incidence of the incoming sound. Thus, the external ear is an important first step in the perceptual process, encoding sounds arriving from different directions with distinctive spectral characters. For example, the primary resonance of the external ear, at about 2.6 kHz, is most sensitive to

1-44 Principles of Sound and Hearing

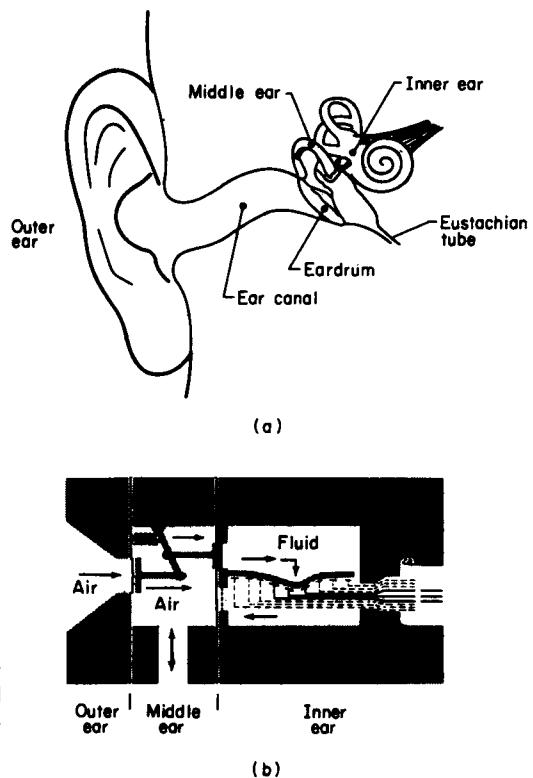


Figure 1.4.1 The human ear: (a) cross-sectional view showing the major anatomical elements, (b) a simplified functional representation.

sounds arriving from near 45° azimuth. This can be demonstrated by listening to a source of broadband sound while looking directly at it and then slowly rotating the head until one ear is pointing toward it. As the head is rotated through 45° , the sound should take on a “brighter” character as sounds in the upper midrange are accentuated. People with hearing problems use this feature of the ear to improve the intelligibility of speech when they unconsciously tilt the head, directing the ear toward the speaker. Continuing the rotation reveals a rapid dulling of the sound as the source moves behind the head. This is caused by acoustical shadowing due to diffraction by the pinna, a feature that helps to distinguish between front and back in sound localization.

At the eardrum the sound pressure fluctuations are transformed into movement that is coupled by means of the middle-ear bones (the *ossicular chain*) to the *oval window*, the input to the inner ear (*cochlea*). The middle ear increases the efficiency of sound energy transfer by providing a partial impedance match between sound in air, on the one hand, and wave motion in the liquid-filled inner ear, on the other. The inner ear performs the elaborate function of analyzing the sound into its constituent frequencies and converting the result into neural signals that pass up the auditory (eighth) nerve to the auditory cortex of the brain. From there sound is transformed into the many and varied perceptions that we take for granted. In the following discussions we shall be dealing with some of these functions in more detail.

1.4.3 Psychoacoustics and the Dimensions of Hearing

The physical dimensions of sound have parallels in the perceptual processes. The relationships are usually nonlinear, more complex than at first appearance, and somewhat variable among individuals as well as with time and experience. Nevertheless, they are the very essence of hearing.

The study of these relationships falls under the general umbrella of *psycho-acoustics*. A more specialized study, known as *psychophysics* or *psychometrics*, is concerned with quantification of the magnitudes of the sensation in relation to the magnitude of the corresponding physical stimulus.

1.4.3a Loudness

Loudness is the term used to describe the magnitude of an auditory sensation. It is primarily dependent upon the physical magnitude (sound pressure) of the sound producing the sensation, but many other factors are influential.

Sounds come in an infinite variety of frequencies, timbres, intensities, temporal patterns, and durations; each of these, as well as the characteristics of the individual listener and the context within which the sound is heard, has an influence on loudness. Consequently, it is impossible for a single graph or equation to accurately express the relationship between the physical quality and quantity of sound and the subjective impression of loudness. Our present knowledge of the phenomenon is incomplete, but there are some important experimentally determined relationships between loudness and certain measurable quantities of sound. Although it is common to present and discuss these relationships as matters of fact, it must always be remembered that they have been arrived at through the process of averaging the results of many experiments with many listeners. These are not precise engineering data; they are merely indicators of trends.

Loudness as a Function of Frequency and Amplitude

The relationship between loudness and the frequency and SPL of the simplest of sounds, the pure tone, was first established by Fletcher and Munson, in 1933 [2]. There have been several subsequent redeterminations of loudness relationships by experimenters incorporating various refinements in their techniques. The data of Robinson and Dadson [3], for example, provide the basis for the International Organization for Standardization (ISO) recommendation R226 [4]. The presentation of loudness data is usually in the form of *equal-loudness contours*, as shown in Figure 1.4.2. Each curve shows the SPLs at which tones of various frequencies are judged to sound equal in loudness to a 1-kHz reference tone; the SPL of the reference tone identifies the curve in units called *phons*. According to this method, the *loudness level* of a sound, in phons, is the SPL level of a 1-kHz pure tone that is judged to be equally loud.

The equal-loudness contours of Figure 1.4.2 show that the ears are less sensitive to low frequencies than to middle and high frequencies and that this effect increases as sound level is reduced. In other words, as the overall sound level of a broadband signal such as music is reduced, the bass frequencies will fade faster than middle or high frequencies. In the curves, this appears as a crowding together of the contours at low frequencies, indicating that, at the lower sound levels, a small change in SPL of low-frequency sounds produces the same change in loudness as a larger change in SPL at middle and high frequencies. This may be recognized as the basis for the loudness compensation controls built into many domestic hi-fi amplifiers, the pur-

1-46 Principles of Sound and Hearing

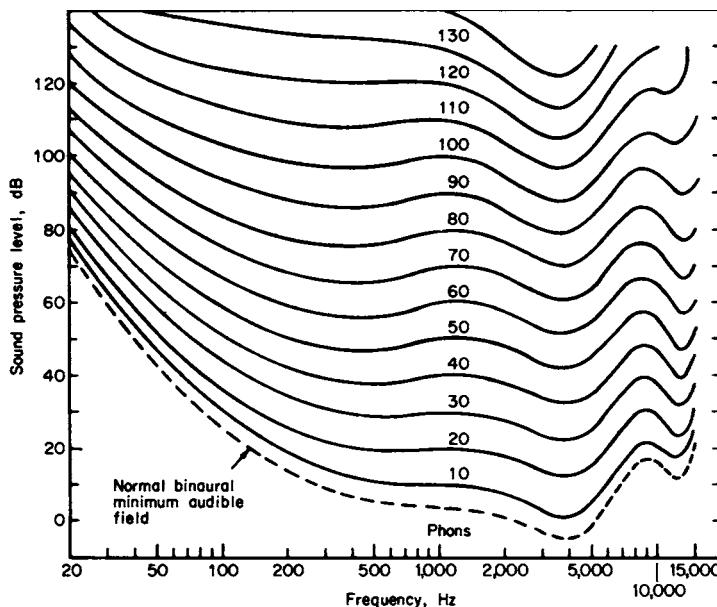


Figure 1.4.2 Contours of equal loudness showing the sound pressure level required for pure tones at different frequencies to sound as loud as a reference tone of 1000 Hz. (From ISO Recommendation R226.)

pose of which is to boost progressively the bass frequencies as the overall sound level is reduced. The design and use of such compensation have often been erroneous because of a confusion between the shape of the loudness contours themselves and the *differences* between curves at various phon levels [5]. Sounds reproduced at close to realistic levels should need no compensation, since the ears will respond to the sound just as they would to the “live” version of the program. By the same token, control-room monitoring at very high sound levels can result in program equalization that is not appropriate to reproduction at normal domestic sound levels (combined with this are the effects of temporary and permanent changes in hearing performance caused by exposure to loud sounds).

It is difficult to take the interpretations of equal-loudness contours much beyond generalizations since, as mentioned earlier, they are composites of data from many individuals. There is also the fact that they deal with pure tones and the measurements were done either through headphones (Fletcher and Munson [2]) or in an anechoic chamber (Robinson and Dadson [3]). The relationship between these laboratory tests and the common application for these data, the audition of music in normal rooms, is one that is only poorly established.

The lowest equal-loudness contour defines the lower limit of perception: the *hearing-threshold level*. It is significant that the ears have their maximum sensitivity at frequencies that are important to the intelligibility of speech. This optimization of the hearing process can be seen in various other aspects of auditory performance as well.

The rate of growth of loudness as a function of the SPL is a matter of separate interest. Units of *sones* are used to describe the magnitude of the subjective sensation. One sone is defined as

the loudness of a tone at the 40-phon loudness level. A sound of loudness 2 sones would be twice as loud, and a sound of 0.5 sone would be half as loud. The *loudness function* relating the subjective sensation to the physical magnitude has been studied extensively [6], and while there are consistencies in general behavior, there remain very large differences in the performance of individuals and in the effect of the temporal and spectral structure of the sound. A common approximation relates a change of 10 dB in SPL to a doubling or halving of loudness. Individual variations on this may be a factor of 2 or more, indicating that one is not dealing with precise data. For example, the growth of loudness at low frequencies, as shown in the curves of Figure 1.4.2, indicates a clear departure from the general rule. Nevertheless, it is worth noting that significant changes in loudness require large differences in SPL and sound power; a doubling of loudness that requires a 10-dB increase in sound level translates into a factor of 3.16 in sound pressure (or voltage) and a factor of 10 in power.

Loudness as a Function of Bandwidth

The studies of loudness that used pure tones leave doubts about how they relate to normal sounds that are complexes of several frequencies or continuous bands of sound extending over a range of frequencies. If the bandwidth of a sound is increased progressively while maintaining a constant overall measured sound level, it is found that loudness remains constant from narrow bandwidths up to a value called the *critical bandwidth*. At larger bandwidths, the loudness increases as a function of bandwidth because of a process known as *loudness summation*. For example, the broadband sound of an orchestra playing a chord will be louder than the simple sound of a flute playing a single note even when the sounds have been adjusted to the same SPL.

The critical bandwidth varies with the center frequency of the complex sound being judged. At frequencies below about 200 Hz it is fairly constant at about 90 Hz; at higher frequencies the critical bandwidth increases progressively to close to 4000 Hz at 15 kHz. The sound of the orchestra therefore occupies many critical bandwidths while the sound of the flute is predominantly within one band.

Loudness as a Function of Duration

Brief sounds can appear to be less loud than sounds with the same maximum sound level but longer duration. Experiments show that there is a progressive growth of loudness as signal duration is increased up to about 200 ms; above that, the relationship levels out. The implication is that the hearing system integrates sound energy over a time interval of about 200 ms. In reality, the integration is likely to be of neural energy rather than acoustical energy, which makes the process rather complicated, since it must embrace all the nonlinearities of the perceptual mechanism.

The practical consequence of this is that numerous temporal factors, such as duration, intermittency, repetition rate, and so on, all influence the loudness of sounds that are separate from SPL.

Measuring the Loudness of Complex Sounds

Given the numerous variables and uncertainties in ascertaining the loudness of simple sounds, it should come as no surprise that measuring the loudness of the wideband, complex, and ever-changing sounds of real life is a problem that has resisted simple interpretation. Motivated by the need to evaluate the annoyance value of sounds as well as the more neutral quantity of loudness,

1-48 Principles of Sound and Hearing

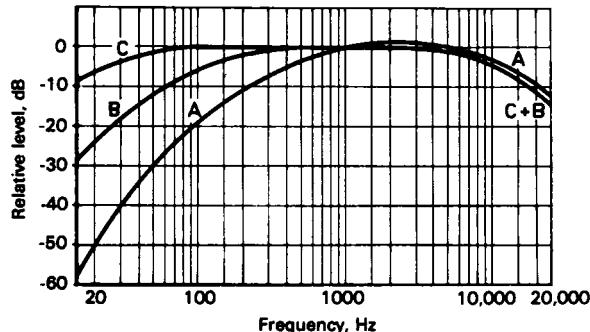


Figure 1.4.3 The standard frequency-weighting networks used in sound-level meters.

various methods have been developed for arriving at single-number ratings of complex sounds. Some methods make use of spectral analysis of the sound, adjusted by correction factors and weighting, to compute a single-number loudness rating. These tend to require expensive apparatus and are, at best, cumbersome to use; they also are most accurate for steady-state sounds.

Simplifying the loudness compensation permits the process to be accomplished with relatively straightforward electronics providing a direct-reading output in real time, a feature that makes the device practical for recording and broadcasting applications. Such devices are reported to give rather better indications of the loudness of typical music and speech program material than the very common and even simpler *volume-unit* (VU) meters or *sound-level meters* [7].

The VU meter responds to the full audio-frequency range, with a flat frequency response but with some control of its dynamic (time) response. A properly constructed VU meter should exhibit a response time of close to 300 ms, with an overswing of not more than 1.5 percent, and a return time similar to the response time. The dial calibrations and reference levels are also standardized. Such devices are therefore useful for measuring the magnitudes of steady-state signals and for giving a rough indication of the loudness of complex and time-varying signals, but they fail completely to take into account the frequency dependence of loudness.

The sound-level meters used for acoustical measurements are adjustable in both amplitude and time response. Various *frequency-weighting* curves, *A-weighting* being the most popular, acknowledge the frequency-dependent aspects of loudness, and “fast” and “slow” time responses deal differently with temporal considerations. Although these instruments are carefully standardized and find extensive use in acoustics, noise control, and hearing conservation, they are of limited use as program-level indicators. Figure 1.4.3 shows the common frequency-weighting options found in sound-level meters. *A-weighting* has become the almost universal choice for measurements associated with loudness, annoyance, and the assessment of hearing-damage risk.

Peak program meters (PPM) are also standardized [8], and they find extensive use in the recording and broadcast industries. However, they are used mainly as a means of avoiding overloading recorders and signal-processing equipment. Consequently, the PPM has a very rapid response (an integration time of about 10 ms in the normal mode), so that brief signal peaks are registered, and a slow return (around 3 s), so that the peak levels can be easily seen. These devices therefore are not useful indicators of loudness of fluctuating signals.

1.4.3b Masking

Listening to a sound in the presence of another sound, which for the sake of simplicity we shall call noise, results in the desired sound being, to some extent, less audible. This effect is called *masking*. If the noise is sufficiently loud, the signal can be completely masked, rendering it inaudible; at lower noise levels the signal will be partially masked, and only its apparent loudness may be reduced. If the desired sound is complex, it is possible for masking to affect only portions of the total sound. All this is dependent on the specific nature of both the signal and the masking sound.

In audio it is possible for the low-level sounds of music, for example, to be masked by background noise in a sound system. That same noise can mask distortion products, so the effects need not be entirely undesirable. In addition to the unwanted noises that have been implied so far, there can be masking of musical sounds by other musical sounds. Thus we encounter the interesting situation of the perceived sound of a single musical instrument modified by the sounds of other instruments when it is joined in an ensemble.

In addition to the partial and complete masking that occurs when two sounds occur simultaneously, there are instances of *temporal masking*, when the audibility of a sound is modified by a sound that precedes it in time (*forward masking*) or, strange as it may seem, by a sound that follows it (*backward masking*).

Simultaneous Masking

At the lowest level of audibility, the threshold, the presence of noise can cause a *threshold shift* wherein the amplitude of the signal must be increased to restore audibility. At higher sound levels the masked sound may remain audible but, owing to partial masking, its loudness can be reduced.

In *simultaneous masking* the signal and the masking sound coexist in the time domain. It is often assumed that they must also share the same frequency band. While this seems to be most effective, it is not absolutely necessary. The effect of a masking sound can extend to frequencies that are both higher and lower than those in the masking itself. At low sound levels a masking sound tends to influence signals with frequencies close to its own, but at higher sound levels the masking effect spreads to include frequencies well outside the spectrum of the masker. The dominant effect is an *upward spread* of masking that can extend several octaves above the frequency of the masking sound. There is also a *downward spread* of masking, but the effect is considerably less. In other words, a low-frequency masking sound can reduce the audibility of higher-frequency signals, but a high-frequency masking sound has relatively little effect on signals of lower frequency. Figure 1.4.4 shows that a simple masking sound elevates the hearing threshold over a wide frequency range but that the elevation is greater for frequencies above the masking sound.

In the context of audio, this means that we have built-in noise and distortion suppression. Background noises of all kinds are less audible while the music is playing but stand out clearly during the quiet intervals. Distortions generated in the recording and reproduction processes are present only during the musical sound and are therefore at least partially masked by the music itself. This is especially true for harmonic distortions, in which the objectionable distortion products are at frequencies higher than the masking sound—the sound that causes them to exist. Intermodulation-distortion products, on the other hand, are at frequencies both above and below

1-50 Principles of Sound and Hearing

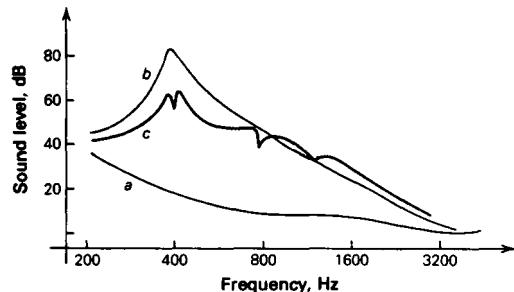


Figure 1.4.4 Detection threshold for pure tones of various frequencies: (a) in insulation, (b) in the presence of a narrow band (365 to 455 Hz) of masking noise centered on 400 Hz at a sound level of 80 dB, (c) in the presence of a masking tone of 400 Hz at 80 dB. (From [40]. Used with permission.)

the frequencies of the signals that produce the distortion. In this case, the upper distortion products will be subject to greater masking by the signal than the lower distortion products.

Studies of distortion have consistently noted that all forms of distortion are less audible with music than with simple signals such as single tones or combinations of tones; the more effective masking of the spectrally complex music signal is clearly a factor in this. Also noted is that intermodulation distortion is more objectionable than its harmonic equivalent. A simple explanation for this may be that not only are the difference-frequency components of intermodulation distortion unmusical, but they are not well masked by the signals that produce them.

Temporal Masking

The masking that occurs between signals not occurring simultaneously is known as *temporal masking*. It can operate both ways, from an earlier to a later sound (forward masking) or from a later to an earlier sound (backward masking). The apparent impossibility of backward masking (going backward in time) has a physiological explanation. It takes time for sounds to be processed in the peripheral auditory system and for the neural information to travel to the brain. If the later sound is substantially more intense than the earlier sound, information about it can take precedence over information about the earlier sound. The effect can extend up to 100 to 200 ms, but because such occurrences are rare in normal hearing, the most noteworthy auditory experiences are related to forward masking.

Forward masking results from effects of a sound that remain after the physical stimulus has been removed. The masking increases with the sound level of the masker and diminishes rapidly with time, although effects can sometimes be seen for up to 500 ms [9]. Threshold shifts of 10 to 20 dB appear to be typical for moderate sound levels, but at high levels these may reach 40 to 50 dB. Combined with these substantial effects is a broadening of the frequency range of the masking; at masker sound levels above about 80 dB maximum masking no longer occurs at the frequency of the masking sound but at higher frequencies.

There are complex interactions among the numerous variables in the masking process, and it is difficult to translate the experimental findings into factors specifically related to audio engineering. The effects are not subtle, however, and it is clear that in many ways they influence what we hear.

1.4.3c Acoustic Reflex

One of the less-known features of hearing is the *acoustic reflex*, an involuntary activation of the middle-ear muscles in response to sound and some bodily functions. These tiny muscles alter the transmission of sound energy through the middle ear, changing the quantity and quality of the sound that reaches the inner ear. As the muscles tighten, there may be a slight reduction in the overall sound level reaching the inner ear, but mainly there is a change in spectral balance as the low frequencies are rolled off. Below approximately 1 kHz the attenuation is typically 5 to 10 dB, but it can be as much as 30 dB.

The reflex is activated by sounds above 80- to 85-dB SPL, which led to the early notion that it was a protective mechanism; however, the most hazardous sounds are at frequencies that are little affected by the reflex, and, furthermore, the reflex is too slow to block the passage of loud transients. The reflex activates rather slowly, in 10 to 20 ms for loud sounds and up to 150 ms for sounds near the activation threshold; then, after an interval, it slowly relaxes. Obviously there have to be other reasons for its existence. Although there is still some speculation as to its purpose, the fact that it is automatically activated when we talk and when we chew suggests that part of the reason is simply to reduce the auditory effects of our own voice and eating sounds.

Some people can activate the reflex voluntarily, and they report a reduction in the loudness of low frequencies during the period of activation. The behavior of the reflex also appears to depend on the state of the listener's attention to the sound itself. This built-in tone control clearly is a complication in sound quality assessments since the spectral balance appears to be a function of sound level, the pattern of sound-level fluctuations in time, and the listener's attitude or attention to the sound.

1.4.3d Pitch

Pitch is the subjective attribute of frequency, and while the basic correspondence between the two domains is obvious—low pitch to low frequencies and high pitch to high frequencies—the detailed relationships are anything but simple.

Fortunately waveforms that are periodic, however complex they may be, tend to be judged as having the same pitch as sine waves of the same repetition frequency. In other words, when a satisfactory pitch match has been made, the fundamental frequency of a complex periodic sound and a comparison sinusoid will normally be found to have the same frequency.

The exceptions to this simple rule derive from those situations where there is no physical energy at the frequency corresponding to the perceived pitch. Examples of pitch being associated with a *missing fundamental* are easily demonstrated by using groups of equally spaced tones, such as 100, 150, and 200 Hz, and observing that the perceived pitch corresponds to the difference frequency, 50 Hz. Common experience with sound reproducers, such as small radios, that have limited low-frequency bandwidth, illustrates the strength of the phenomenon, as do experiences with musical instruments, such as some low-frequency organ sounds, that may have little energy at the perceived fundamental frequency.

Scientifically, pitch has been studied on a continuous scale, in units of *mels*. It has been found that there is a highly nonlinear relationship between subjectively judged ratios of pitch and the corresponding ratios of frequency, with the subjective pitch interval increasing in size with increasing frequency. All this, though, is of little interest to traditional musicians, who have organized the frequency domain into intervals having special tonal relationships. The *octave* is particularly notable because of the subjective similarity of sounds spaced an octave apart and the

1-52 Principles of Sound and Hearing

fact that these sounds commonly differ in frequency by factors of 2. The musical fifth is a similarly well-defined relationship, being a ratio of 3:2 in repetition frequencies. These and the other intervals used in musical terminology gain meaning as one moves away from sine waves, with their one-frequency purity, into the sounds of musical instruments with their rich collection of overtones, many of which are harmonically related. With either pure tones [10] or some instrumental sounds in which not all the overtones are exactly harmonically related, the subjective octave may differ slightly from the physical octave; in the piano this leads to what is called *stretched tuning* [11].

The incompatibility of the mel scale of pitch and the hierarchy of musical intervals remains a matter for discussion. These appear to be quite different views of the same phenomenon, with some of the difference being associated with the musical expertise of listeners. It has been suggested, for example, that the mel scale might be better interpreted as a scale of *brightness* rather than one of pitch [12]. With periodic sounds brightness and pitch are closely related, but there are sounds, such as bells, hisses, and clicks, that do not have all the properties of periodic sounds and yet convey enough of a sense of pitch to enable tunes to be played with them, even though they cannot be heard as combining into chords or harmony. In these cases, the impressions of brightness and pitch seem to be associated with a prominence of sound energy in a band of frequencies rather than with any of the spectral components (partials, overtones, or harmonics) that may be present in the sound. A separate confirmation of this concept of brightness is found in subjective assessments of reproduced sound quality, where there appears to be a perceptual dimension along a continuum of “darkness” to “brightness” in which brightness is associated with a frequency response that rises toward the high frequencies or in which there are peaks in the treble [13]. At this, we reach a point in the discussion where it is more relevant to move into a different but related domain.

1.4.3e Timbre, Sound Quality, and Perceptual Dimensions

Sounds may be judged to have the same subjective dimensions of loudness and pitch and yet sound very different from one another. This difference in *sound quality*, known as *timbre* in musical terminology, can relate to the tonal quality of sounds from specific musical instruments as they are played in live performance, to the character of tone imparted to all sounds processed through a system of recording and reproduction, and to the tonal modifications added by the architectural space within which the original performance or a reproduction takes place. Timbre is, therefore, a matter of fundamental importance in audio, since it can be affected by almost anything that occurs in the production, processing, storage, and reproduction of sounds.

Timbre has many dimensions, not all of which have been confidently identified and few of which have been related with any certainty to the corresponding physical attributes of sound. There is, for example, no doubt that the shape and composition of the frequency spectrum of the sound are major factors, as are the temporal behaviors of individual elements comprising the spectrum, but progress has been slow in identifying those measurable aspects of the signal that correlate with specific perceived dimensions, mainly because there are so many interactions between the dimensions themselves and between the physical and psychological factors underlying them.

The field of electronic sound synthesis has contributed much to the understanding of why certain musical instruments sound the way they do, and from this understanding have followed devices that permit continuous variations of many of the sound parameters. The result has been

progressively better imitations of acoustical instruments in electronic simulations, as well as an infinite array of new “instruments” exhibiting tonal colors, dynamics, and emotional connotations that are beyond the capability of traditional instruments. At the same time as this expansion of timbral variety is occurring on one front of technical progress, there is an effort on another front to faithfully preserve the timbre of real and synthesized instruments through the complex process of recording and reproduction. The original intentions of *high-fidelity reproduction* exist today in spite of the manifest abuses of the term in the consumer marketplace.

A fundamental problem in coming to grips with the relationship between the technical descriptions of sounds and the perception of timbre is in establishing some order in the choice and quantitative evaluation of words and phrases used by listeners to describe aspects of sound quality. Some of the descriptors are fairly general in their application and seem to fall naturally to quantification on a continuous scale from say, “dull” to “bright” or from “full” to “thin.” Others, though, are specific to particular instruments or lapse into poetic portrayals of the evoked emotions.

From carefully conducted assessments of reproduced sound quality involving forms of multivariate statistical analysis, it has become clear that the extensive list can be reduced to a few relatively independent dimensions. As might be expected, many of the descriptors are simply different ways of saying the same thing, or they are responses to different perceptual manifestations of the same physical phenomenon.

From such analyses can come useful clarifications of apparently anomalous results since these responses need not be unidirectional. For example, a relatively innocent rise in the high-frequency response of a sound reproducer might be perceived as causing violins to sound unpleasantly strident but cymbals to sound unusually clear and articulate. A nice sense of air and space might be somewhat offset by an accentuation of background hiss and vocal sibilants, and so on.

Inexperienced listeners tend to concentrate unduly on a few of the many descriptors that come to mind while listening, while slightly more sophisticated subjects may become confused by the numerous contradictory indications. Both groups, for different reasons, may fail to note that there is but a single underlying technical flaw. The task of critical listening is one that requires a broad perspective and an understanding of the meaning and relative importance of the many timbral clues that a varied musical program can reveal. Trained and experienced listeners tend to combine timbral clues in a quest for logical technical explanations for the perceived effects. However, with proper experimental controls and the necessary prompting through carefully prepared instructions and a questionnaire, listeners with little prior experience can arrive at similar evaluations of accuracy without understanding the technical explanations [14].

The following list of perceptual dimensions is derived from the work of Gabrielsson and various colleagues [13, 15], and is the basis for listening questionnaires used extensively by those workers and the author [14]. The descriptions are slightly modified from the original [13].

- *Clarity, or definition:* This dimension is characterized by adjectives such as clear, well defined, distinct, clean or pure, and rich in details or detailed, as opposed to adjectives such as diffuse, muddy or confused, unclear, blurred, noisy, rough, harsh, or sometimes rumbling, dull, and faint. High ratings in this dimension seem to require that the reproduction system perform well in several respects, exhibiting a wide frequency range, flat frequency response, and low nonlinear distortion. Systems with limited bandwidth, spectral irregularities due to resonances, or audible distortion receive lower ratings. Low-frequency spectral emphasis seems also to be detrimental to performance in this dimension, resulting in descriptions of

1-54 Principles of Sound and Hearing

rumbling, for the obvious reason, and dullness, probably due to the upward masking effects of the strong low frequencies. Increased sound levels result in increased clarity and definition.

- *Sharpness, or hardness, versus softness:* Adjectives such as sharp, hard, shrill, screaming, pointed, and clashing are associated with this dimension, contrasted with the opposite qualities of soft, mild, calm or quiet, dull, and subdued. A rising high-frequency response or prominent resonances in the high-frequency region can elicit high ratings in this dimension, as can certain forms of distortion. A higher or lower sound level also contributes to movement within this dimension, with reduced levels enhancing the aspect of softness.
- *Brightness versus darkness:* This dimension is characterized by the adjective bright, as opposed to dark, rumbling, dull, and emphasized bass. There appears to be a similar relationship between this dimension and the physical attributes of the sound system as exists with the preceding dimension, sharpness, or hardness, versus softness. In experiments, the two dimensions sometimes appear together and sometimes separately. The sense of pitch associated with brightness might be a factor in distinguishing between these two dimensions.
- *Fullness versus thinness:* This dimension also can appear in combination with brightness versus darkness, and there are again certain similarities in the relationship to measured spectrum balance and smoothness. There appears to be an association with the bandwidth of the system, especially at the low frequencies, and with sound level. It seems possible that this dimension is a representation of one encountered elsewhere as volume, which has been found to increase with increasing sound level but to decrease with increasing frequency.
- *Spaciousness:* Almost self-explanatory, this dimension elicits expressions of spacious, airy, wide, and open, as opposed to closed or shut up, narrow, and dry. The phenomenon appears to be related to poorly correlated sounds at the two ears of the listener. Other aspects of spaciousness are related to the spectrum of the reproduced sound. Gabrielsson points out that increased treble response enhances spaciousness, while reducing the bandwidth encourages a closed or shut-up impression. It is well known that the directional properties of the external ear (Figure 1.4.5) encode incoming sounds with spectral cues that can be significant influences in sound localization [16]. One such cue is a moving spectral notch and an increase in the sound level reaching the eardrum over the band from 5 to 10 kHz for progressively elevated sources (Figure 1.4.6). The appropriate manipulation of the sound spectrum in this frequency region can alone create impressions of height [17, 18] and, in this sense, alter the impression of spaciousness. It is worthy of note that the dimension of spaciousness is clearly observed in monophonic as well as stereophonic reproductions, indicating that it is a rather fundamental aspect of sound quality [13, 19].
- *Nearness:* Differences in the apparent proximity of sound sources are regularly observed in listening tests. It is clear that sound level affects perception of distance, especially for sounds such as the human voice that are familiar to listeners. Evidence from other studies indicates that impressions of distance are also influenced by the relationship between the direct, early-reflected, and reverberant sounds and the degree of coherence that exists in these sounds as they appear at the listener's ears [17].
- *Absence of extraneous sounds:* This dimension refers to nonmusical sounds that either exist in the original program material and are accentuated by aspects of the reproducer (such as tape hiss being aggravated by a treble boost) or are generated within the device itself (such as electronic amplifier clipping or mechanical noises from a loudspeaker).

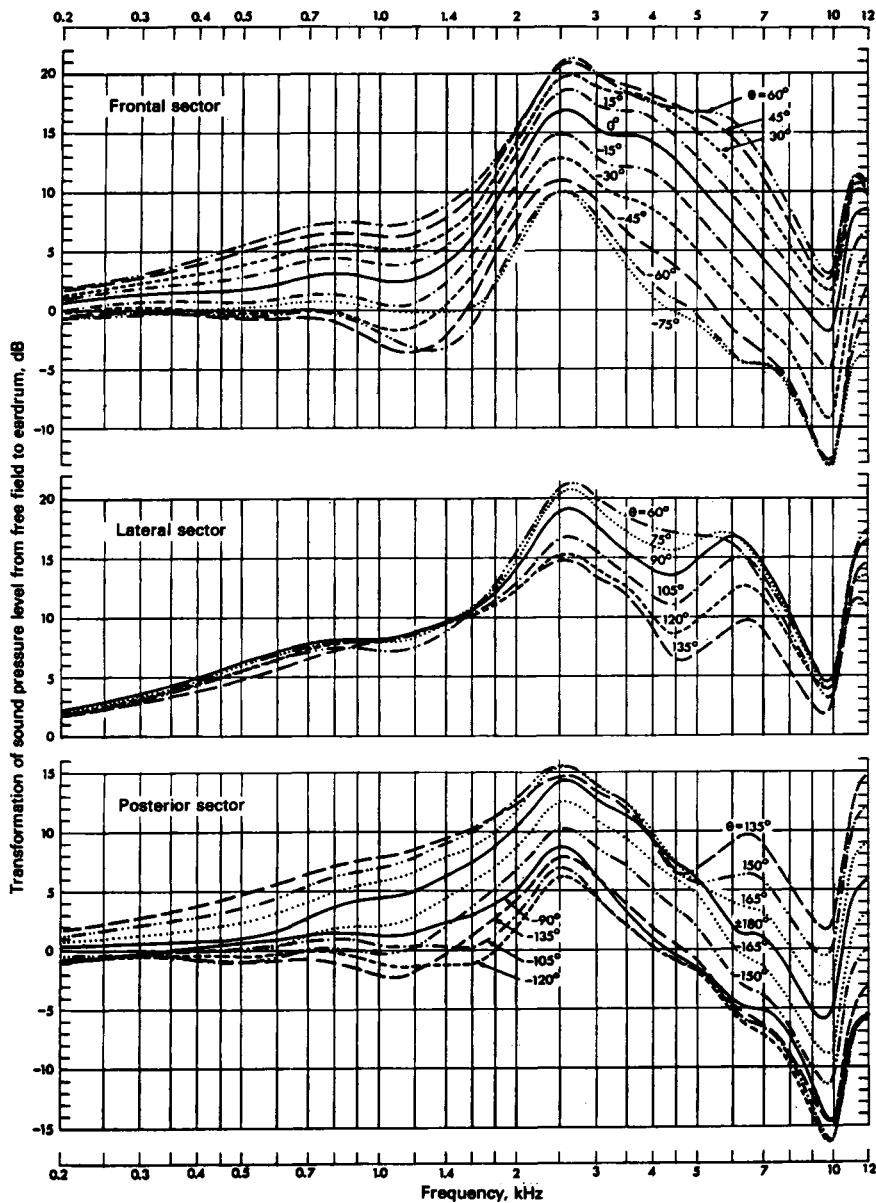


Figure 1.4.5 Family of curves showing the transformation of sound pressure level from the free field to the eardrum in the horizontal plane as a function of frequency, averaged over many listeners in several independent studies. The horizontal angles are referred to zero (the forward direction) and increase positively toward the ear in which the measurement is made and negatively away from it. (From [27]. Used with permission.)

1-56 Principles of Sound and Hearing

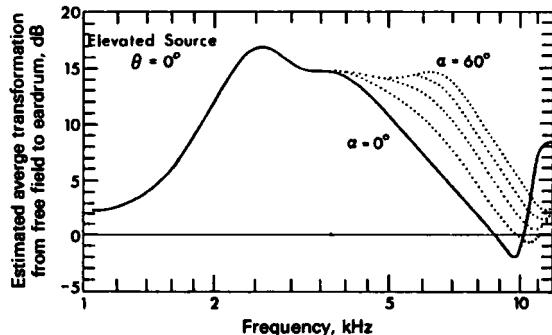


Figure 1.4.6 The estimated average transformation of sound pressure level from the free field to the eardrum as a function of frequency, showing the variations as a function of the angle of elevation for sounds arriving from the forward direction. (From [1]. Used with permission.)

- *Loudness:* This self-explanatory dimension is a useful check on the accuracy with which the sound levels of comparison sounds have been matched. It should, however, be noted that some listeners seem to regard the adjective loud as a synonym for sharp, hard, or painful.

The relative importance of these dimensions in describing overall sound quality changes slightly according to the specific nature of the devices under test, the form of the listener questionnaire, the program material, and, to some extent, the listeners themselves. In general, Gabrielson and colleagues [13, 15] have found that clarity, or definition, brightness versus darkness, and sharpness, or hardness, versus softness are major contributors to the overall impression of sound quality.

1.4.3f Audibility of Variations in Amplitude and Phase

Other things being equal, very small differences in sound level can be heard: down to a fraction of a decibel in direct A/B comparisons. Level differences that exist over only a small part of the spectrum tend to be less audible than differences that occupy a greater bandwidth. In other words, a small difference that extends over several octaves may be as significant as a much larger difference that is localized in a narrow band of frequencies. Spectral tilts of as little as 0.1 dB per octave are audible. For simple sounds the only audible difference may be loudness, but for complex sounds differences in timbre may be more easily detectable.

The audibility of phase shift is a very different matter. This hotly debated issue assumes major proportions because of the implication that if phase shifts are not audible, then the waveform of a complex sound, *per se*, is not important. Several independent investigations over many years have led to the conclusion that while there are some special signals and listening situations where phase effects can be heard, their importance when listening to music in conventional environments is small [19]. Psychophysical studies indicate that, in general, sensitivity to phase is small compared with sensitivity to the amplitude spectrum and that sensitivity to phase decreases as the fundamental frequency of the signal increases. At the same time, it appears to be phase shifts in the upper harmonics of a complex signal that contribute most to changes in timbre [20].

The notion that phase, and therefore waveform, information is relatively unimportant is consistent with some observations of normal hearing. Sounds from real sources (voices and musical instruments) generally arrive at our ears after traveling over many different paths, some of which may involve several reflections. The waveform at the ear therefore depends on various factors

other than the source itself. Even the argument that the direct sound is especially selected for audition and that later arrivals are perceptually suppressed does not substantially change the situation because sources themselves do not radiate waveforms that are invariably distinctive. With musical instruments radiating quite different components of their sound in different directions (consider the complexity of a grand piano or the cello, for example), the sum of these components—the waveform at issue—will itself be different at every different angle and distance; a recording microphone is in just such a situation.

The fact that the ear seems to be relatively insensitive to phase shifts would therefore appear to be simply a condition born of necessity. It would be incorrect to assume, however, that the phase performance of devices is *totally* unimportant. Spectrally localized phase anomalies are useful indicators of the presence of resonances in systems, and very large accumulations of phase shift over a range of frequencies can become audible as group delays.

While the presence of resonances can be inferred from phase fluctuations, their audibility may be better predicted from evidence in the amplitude domain [19]. It should be added that resonances of low Q in sound reproduction systems are more easily heard than those of higher Q [21–23]. This has the additional interesting ramification that evidence of sustained ringing in the time domain may be less significant than ringing that is rapidly damped; waveform features and other measured evidence that attract visual attention do not always correspond directly with the sound colorations that are audible in typical listening situations.

1.4.3g Perception of Direction and Space

Sounds are commonly perceived as arriving from specific directions, usually coinciding with the physical location of the sound source. This perception may also carry with it a strong impression of the acoustical setting of the sound event, which normally is related to the dimensions, locations, and sound-reflecting properties of the structures surrounding the listener and the sound source as well as objects in the intervening path.

Blauert, in his thorough review of the state of knowledge in this field [17], defines *spatial hearing* as embracing “the relationships between the locations of auditory events and other parameters—particularly those of sound events, but also others such as those that are related to the physiology of the brain.” This statement introduces terms and concepts that may require some explanation. The adjective *sound*, as in *sound event*, refers to a physical source of sound, while the adjective *auditory* identifies a perception. Thus, the perceived location of an auditory event usually coincides with the physical location of the source of sound. Under certain circumstances, however, the two locations may differ slightly or even substantially. The difference is then attributed to other parameters having nothing whatever to do with the physical direction of the sound waves impinging on the ears of the listener, such as subtle aspects of a complex sound event or the processing of the sound signals within the brain.

Thus have developed the parallel studies of *monaural*, or one-eared, hearing and *binaural*, or two-eared, hearing. Commercial sound reproduction has stimulated a corresponding interest in the auditory events associated with sounds emanating from a single source (*monophonic*) and from multiple sources that may be caused to differ in various ways (*stereophonic*). In common usage it is assumed that stereophonic reproduction involves only two loudspeakers, but there are many other possible configurations. In stereophonic reproduction the objective is to create many more auditory events than the number of real sound sources would seem to permit. This is accomplished by presenting to the listener combinations of sounds that take advantage of certain

1-58 Principles of Sound and Hearing

inbuilt perceptual processes in the brain to create auditory events in locations other than those of the sound events and in auditory spaces that may differ from the space within which the reproduction occurs.

Understanding the processes that create auditory events would ideally permit the construction of predictable auditory spatial illusions in domestic stereophonic reproduction, in cinemas, in concert halls, and in auditoria. Although this ideal is far from being completely realized, there are some important patterns of auditory behavior that can be used as guides for the processing of sound signals reproduced through loudspeakers as well as for certain aspects of listening room, concert hall, and auditorium design.

1.4.3h Monaural Transfer Functions of the Ear

Sounds arriving at the ears of the listener are subject to modification by sound reflection, diffraction, and resonances in the structures of the external ear, head, shoulders, and torso. The amount and form of the modification are dependent on the frequency of the sound and the direction and distance of the source from which the sound emanates. In addition to the effect that this has on the sensitivity of the hearing process, which affects signal detection, there are modifications that amount to a kind of directional encoding, wherein sounds arriving from specific directions are subject to changes characteristic of those directions.

Each ear is partially sheltered from sounds arriving from the other side of the head. The effect of diffraction is such that low-frequency sounds, with wavelengths that are large compared with the dimensions of the head, pass around the head with little or no attenuation, while higher frequencies are progressively more greatly affected by the directional effects of diffraction. There is, in addition, the acoustical interference that occurs among the components of sound that have traveled over paths of slightly different length around the front and back and over the top of the head.

Superimposed on these effects are those of the pinna, or external ear. The intriguingly complex shape of this structure has prompted a number of theories of its behavior, but only relatively recently have some of its important functions been properly put into perspective. According to one view, the folds of the pinna form reflecting surfaces, the effect of which is to create, at the entrance to the ear canal, a system of interferences between the direct and these locally reflected sounds that depends on the direction and distance of the incoming sound [24]. The small size of the structures involved compared with the wavelengths of audible sounds indicates that dispersive scattering, rather than simple reflection, is likely to be the dominant effect. Nevertheless, measurements have identified some acoustical interferences resembling those that such a view would predict, and these have been found to correlate with some aspects of localization [18, 25].

In the end, however, the utility of the theory must be judged on the basis of how effectively it explains the physical functions of the device and how well it predicts the perceptual consequences of the process. From this point of view, time-domain descriptions would appear to be at a disadvantage since the hearing process is demonstrably insensitive to the fine structure of signals at frequencies above about 1.5 kHz [17]. Partly for this reason most workers have favored descriptions in terms of spectral cues.

It is therefore convenient that the most nearly complete picture of external-ear function has resulted from examinations of the behavior of the external ear in the frequency domain. By carefully measuring the pressure distributions in the standing-wave patterns, the dominant reso-

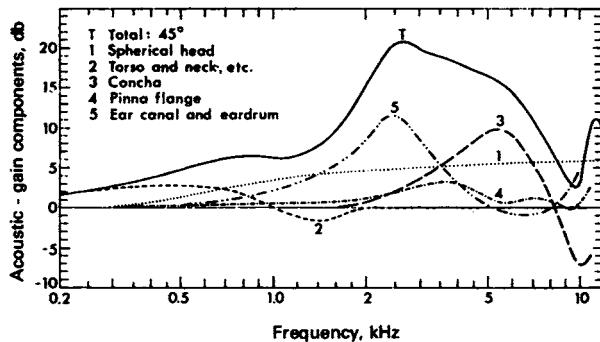


Figure 1.4.7 Contributions of various body parts to the total acoustic gain of the external hearing system for a sound source at a horizontal angle of 45°. Note that the interactions between these components prevent simple arithmetic addition of their individual contributions. (From [1]. Used with permission.)

nances in the external ear have been identified [26.] These have been related to the physical structures and to the measured acoustical performance of the external ear [1].

A particularly informative view of the factors involved in this discussion comes from an examination of curves showing the transformation of SPL from the free field to the eardrum [27]. These curves reveal, as a function of frequency, the amplitude modifications imposed on incident sounds by the external hearing apparatus. Figure 1.4.5 shows the family of curves representing this transformation for sounds arriving from different directions in the horizontal plane. Figure 1.4.6 shows the estimated transformations for sound sources at different elevations.

An interesting perspective on these data is shown in Figure 1.4.7, where it is possible to see the contributions of the various acoustical elements to the total acoustical gain of the ear. It should be emphasized that there is substantial acoustical interaction among these components, so that the sum of any combination of them is not a simple arithmetic addition. Nevertheless, this presentation is a useful means of acquiring a feel for the importance of the various components.

It is clear from these curves that there are substantial direction-dependent spectral changes, some rather narrowband in influence and others amounting to significant broadband tilts. Several studies in localization have found that, especially with pure tones and narrowband signals, listeners could attribute direction to auditory events resulting from sounds presented through only one ear (monaural localization) or presented identically in two ears, resulting in localization in the *median plane* (the plane bisecting the head vertically into symmetrical left-right halves). So strong are some of these effects that they can cause auditory events to appear in places different from the sound event, depending only on the spectral content of the sound. Fortunately such confusing effects are not common in the panorama of sounds we normally encounter, partly because of familiarity with the sounds themselves, but the process is almost certainly a part of the mechanism by which we are able to distinguish between front and back and between up and down, directions that otherwise would be ambiguous because of the symmetrical locations of the two ears.

1-60 Principles of Sound and Hearing

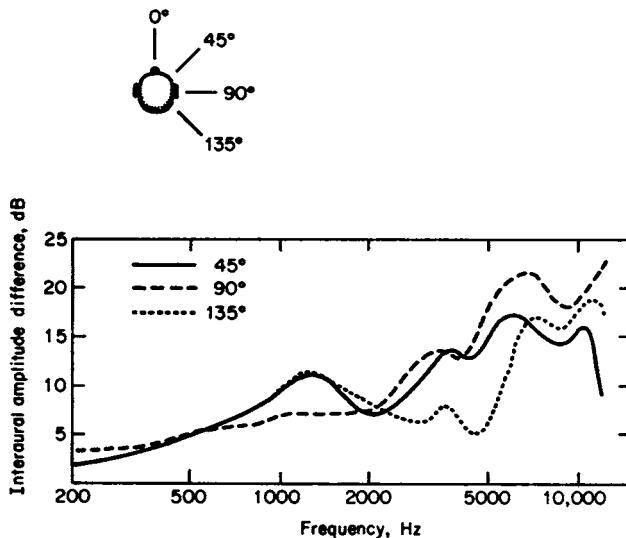


Figure 1.4.8 The interaural amplitude difference as a function of frequency for three angles of incidence. (After [28].)

Interaural Differences

As useful as the monaural cues are, it is sound localization in the horizontal plane that is dominant, and for this the major cues come from the comparison of the sounds at the two ears and the analysis of the differences between them. From the data shown in Figure 1.4.5 it is evident that there is a substantial frequency-dependent *interaural amplitude difference* (IAD) that characterizes sounds arriving from different horizontal angles. Because of the path length differences there will also be an associated *interaural time difference* (ITD) that is similarly dependent on horizontal angle.

Figure 1.4.8 shows IADs as a function of frequency for three angles of incidence in the horizontal plane. These have been derived from the numerical data in [28], from which many other such curves can be calculated.

The variations in IAD as a function of both frequency and horizontal angle are natural consequences of the complex acoustical processes in the external hearing apparatus. Less obvious is the fact that there is frequency dependency in the ITDs. Figure 1.4.9 shows the relationship between ITD and horizontal angle for various pure tones and for broadband clicks. Also shown are the predictive curves for low-frequency sounds, based on diffraction theory, and for high-frequency sounds, based on the assumption that the sound reaches the more remote ear by traveling as a creeping wave that follows the contour of the head. At intermediate frequencies (0.5 to 2 kHz) the system is dispersive, and the temporal differences become very much dependent on the specific nature of the signal [29, 30].

It is evident from these data that at different frequencies, especially the higher frequencies, there are different combinations of ITD and IAD associated with each horizontal angle of incidence. Attempts at artificially manipulating the localization of auditory events by means of fre-

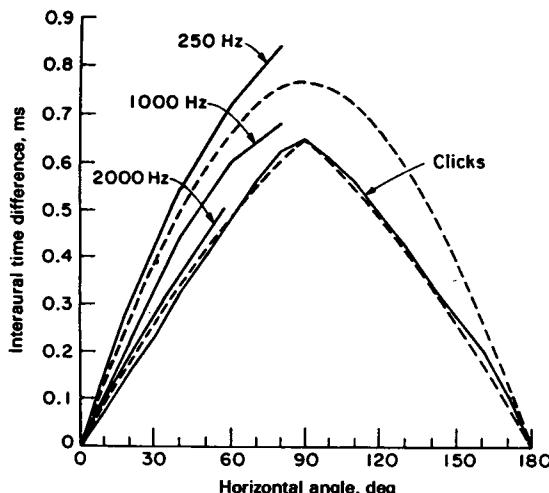


Figure 1.4.9 Interaural time difference as a function of horizontal angle. The curves show measured data for clicks and pure tones (solid lines) and predictive curves for low frequencies (top dashed curve), based on diffraction theory, and for high frequencies (bottom dashed curve), based on creeping-wave concepts. (From [41]. Used with permission.)

frequency-independent variations of these parameters are therefore unlikely to achieve the image size and positional precision associated with natural sound events.

Localization Blur

In normal hearing the precision with which we are able to identify the direction of sounds depends on a number of factors. The measure of this precision is called *localization blur*, the smallest displacement of the sound event that produces a just-noticeable difference in the corresponding auditory event. The concept of localization blur characterizes the fact that auditory space (the perception) is less precisely resolved than physical space and the measures we have of it.

The most precise localization is in the horizontal forward direction with broadband sounds preferably having some impulsive content. The lower limit of localization blur appears to be about 1° , with typical values ranging from 1 to 3° , though for some types of sound values of 10° or more are possible. Moving away from the forward axis, localization blur increases, with typical values for sources on either side of the head and to the rear being around 10 to 20° . Vertically, localization blur is generally rather large, ranging from about 5 to 20° in the forward direction to 30 to 40° behind and overhead [17].

Lateralization versus Localization

In exploring the various ways listeners react to interaural signal differences, it is natural that headphones be used, since the sounds presented to the two ears can then be independently con-

1-62 Principles of Sound and Hearing

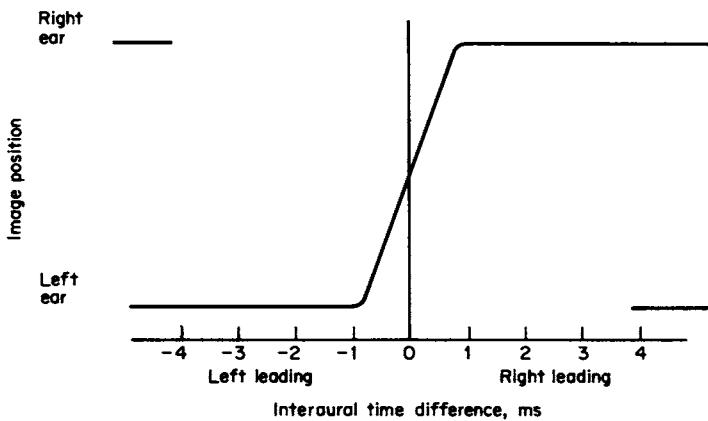


Figure 1.4.10 Perceived positions of the dominant auditory images resulting from impulsive signals (clicks) presented through headphones when the interaural time difference is varied.

trolled. The auditory events that result from this process are distinctive, however, in that the perceived images occur inside or very close to the head and image movement is predominantly lateral. Hence, this phenomenon has come to be known as *lateralization*, as opposed to *localization*, which refers to auditory events perceived to be external and at a distance. Overcoming the in-head localization characteristic of headphone listening has been a major difficulty, inhibiting the widespread use of these devices for critical listening.

In headphone listening it is possible to move the auditory event by independently varying the interaural time or amplitude difference. Manipulating interaural time alone yields auditory image trajectories of the kind shown in Figure 1.4.10, indicating that the ITD required to displace the auditory image from center completely to one side is about 0.6 ms, a value that coincides with the maximum ITD occurring in natural hearing (Figure 1.4.9). Although most listeners would normally be aware of a single dominant auditory image even when the ITD exceeds this normal maximum value, it is possible for there to be multiple auditory images of lesser magnitude, each with a distinctive tonal character and each occupying a different position in perceptual space. With complex periodic signals, experienced listeners indicate that some of these images follow trajectories appropriate to the individual harmonics for frequencies that are below about 1 kHz [31]. This spatial complexity would not be expected in normal listening to a simple sound source, except when there are delayed versions of the direct sounds caused by strong reflections or introduced electronically. The result, if there are several such delayed-sound components, is a confused and spatially dispersed array of images, coming and going with the changing spectral and temporal structure of the sound. It seems probable that this is the origin of the often highly desirable sense of spaciousness in live and reproduced musical performances.

The sensitivity of the auditory system to changes in ITD in the lateralization of auditory images, or *lateralization blur* is dependent on both the frequency and the amplitude of the signal. According to various experimenters, lateralization blur varies from around 2 μ s to about 60 μ s, increasing as a function of signal frequency and sound level, and is at a minimum point around ITD = 0.

Introducing an IAD displaces the auditory event toward the ear receiving the louder sound. An IAD of between 10 and 20 dB seems to be sufficient to cause the image to be moved completely to one side. The precise figure is difficult to ascertain because of the rapid increase in lateralization blur as a function of IAD; the auditory event becomes wider as it approaches the side of the head. Close to center, however, the lateralization blur is consistently in the vicinity of 1 to 2 dB.

Spatial Impression

Accompanying the auditory impression of images in any normal environment is a clear impression of the type and size of the listening environment itself. Two aspects appear to be distinguishable: *reverberance*, associated with the temporal stretching and blurring of auditory events caused by reverberation and late reflections; and *spaciousness*, often described as a spreading of auditory events so that they occupy more space than the physical ensemble of sound sources. Other descriptors such as ambience, width, or envelopment also apply. Spaciousness is a major determinant of listener preference in concert halls and as such has been the subject of considerable study.

In general, the impression of spaciousness is closely related to a lack of correlation between the input signals to the two ears. This appears to be most effectively generated by strong early lateral reflections (those arriving within about the first 80 ms after the direct sound). While all spectral components appear to add positively to the effect and to listener preference, they can contribute differently. Frequencies below about 3 kHz seem to contribute mainly to a sense of depth and envelopment, while high frequencies contribute to a broadening of the auditory event [32].

The acoustical interaction of several time-delayed and directionally displaced sounds at the ears results in a reduced interaural cross correlation; the sense of spaciousness is inversely proportional to this correlation. In other terms, there is a spectral and temporal incoherence in the sounds at the ears, leading to the fragmentation of auditory events as a function of both frequency and time. The fragments are dispersed throughout the perceptual space, contributing to the impression of a spatially extended auditory event.

1.4.3i Distance Hearing

To identify the distance of a sound source listeners appear to rely on a variety of cues, depending on the nature of the sound and the environment. In the absence of strong reflections, as a sound source is moved farther from a listener, the sound level diminishes. It is possible to make judgments of distance on this factor alone, but only for sounds that are familiar, where there is a memory of absolute sound levels to use as a reference. With any sound, however, this cue provides a good sense of relative distance.

In an enclosed space the listener has more information to work with, because as a sound source is moved away, there will be a change in the relationship between the direct sound and the reflected and reverberant sounds in the room. The hearing mechanism appears to take note of the relative strengths of the direct and indirect sounds in establishing the distance of the auditory event. When the sound source is close, the direct sound is dominant and the auditory image is very compact; at greater distances, the indirect sounds grow proportionately stronger until eventually they dominate. The size of the auditory event increases with distance, as does the localization blur.

1-64 Principles of Sound and Hearing

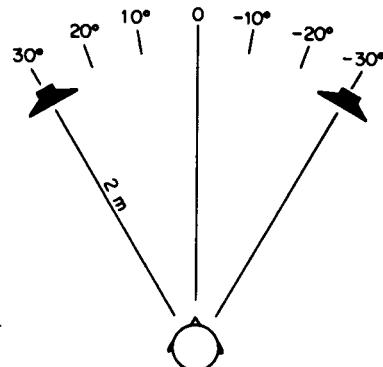


Figure 1.4.11 Standard stereophonic listening configuration.

1.4.3j Stereophonic Imaging

Consider the conventional stereophonic arrangement shown in Figure 1.4.11. If the two loudspeakers are radiating coherent sounds with identical levels and timing, the listener should perceive a single auditory event midway between the loudspeakers. This phantom, or virtual, sound source is the result of *summing localization*, the basis for the present system of two-channel stereophonic recording and reproduction.

Progressively increasing the time difference between the signals in the channels displaces the auditory event, or image, toward the side radiating the earlier sound until, at about 1 ms, the auditory image is coincident with the source of the earlier sound. At time differences greater than about 1 ms the perception may become spatially more dispersed, but the principal auditory event is generally perceived to remain at the position of the earlier sound event until, above some rather larger time difference, there will be two auditory events occurring separately in both time and space, the later of which is called an *echo*.

The region of time difference between that within which simple summing localization occurs and that above which echoes are perceived is one of considerable interest and complexity. In this region the position of the dominant auditory event is usually determined by the sound source that radiates the first sound to arrive at the listener's location. However, depending on the nature of the signal, simple summing localization can break down and there can be subsidiary auditory images at other locations as well. The later sound arrivals also influence loudness, timbre, and intelligibility in ways that are not always obvious.

The cause of this complexity can be seen in Figure 1.4.12, showing the sounds arriving at the two ears when the sound is symbolically represented by a brief impulse. It is immediately clear that the fundamental difference between the situation of summing localization and that of natural localization is the presence of four sound components at the ears instead of just two.

In all cases the listener responds to identical ear input signals by indicating a single auditory event in the forward direction. Note, however, that in both stereo situations the signals at the two ears are not the same as the signals in normal localization. Thus, even though the spatial aspects have been simulated in stereo, the sounds at the two ears are modified by the *acoustical crosstalk* from each speaker to the opposite ear, meaning that perfectly accurate timbral reproduction for these sounds is not possible. This aspect of stereo persists through all conditions for time-difference manipulation of the auditory image, but with amplitude-difference manipulation the effect

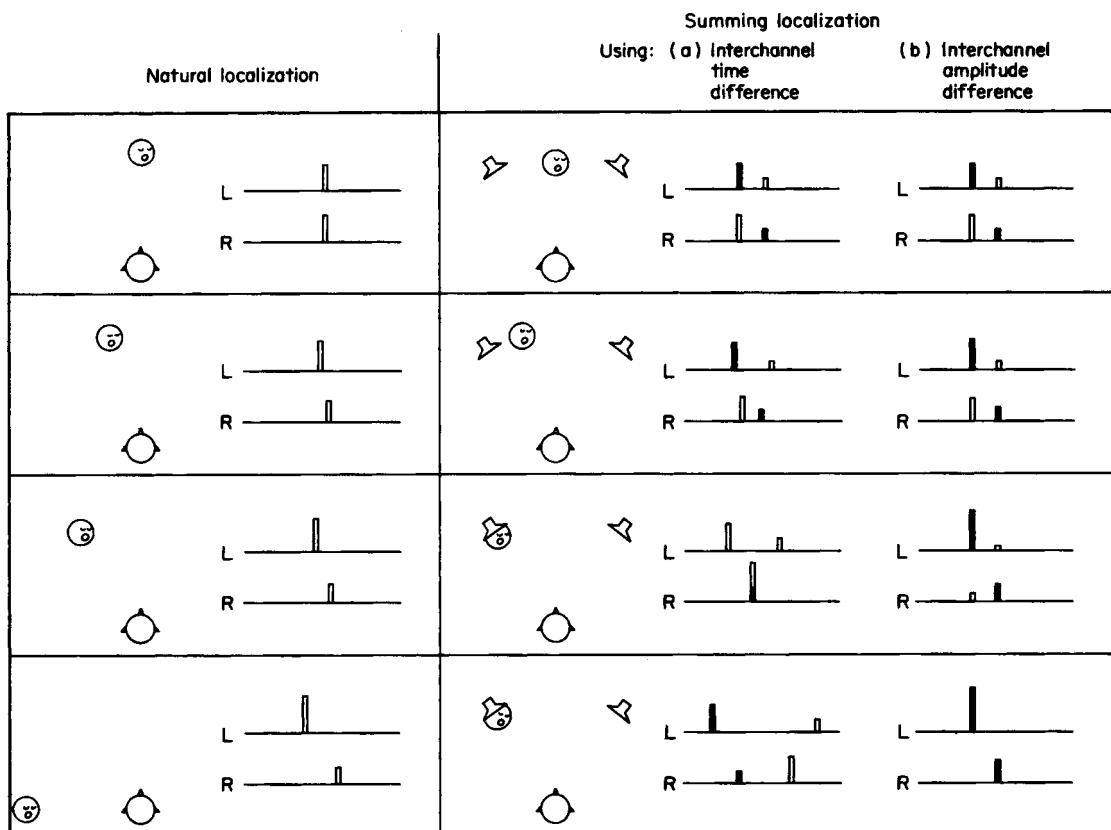


Figure 1.4.12 Comparison between sound localization in natural listening and localization in stereophonic listening within the range of simple summing. For the purposes of this simplified illustration, the sound waveform is an impulse. To the right of the pictorial diagrams showing a listener receiving sound from either a single source (natural localization) or a stereo pair of loudspeakers (summing localization) are shown the sounds received by the left and right ears of the listener. In the stereo illustrations, sounds from the left loudspeaker are indicated by dark bars and sounds from the right loudspeaker by light bars.

diminishes with increasing amplitude difference until, in the extreme, the listener hears only sound from a single speaker, a monophonic presentation.

Although impressions of image movement between the loudspeakers can be convincingly demonstrated by using either interchannel time or amplitude differences, there is an inherent limitation in the amount of movement: in both cases the lateral displacement of the principal auditory event is bounded by the loudspeakers themselves.

With time differences, temporal masking inhibits the contributions of the later arrivals, and the localization is dominated by the first sound to arrive at each ear. With small time differences the image can be moved between the loudspeakers, the first arrivals are from different loud-

1-66 Principles of Sound and Hearing

speakers, and it can be seen that an interchannel time difference is perceived as an ITD. At larger values of interchannel time difference the first arrivals are from the same loudspeaker, and the dominant auditory image remains at that location. This is because of the *law of the first wave-front*, also known as the precedence effect, according to which the dominant auditory event is perceived to be coincident with the loudspeaker radiating the earlier sound. The other sound components are still there nonetheless, and they can contribute to complexity in the spatial illusion as well as to changes in timbre.

With amplitude differences (also known as *intensity stereo*), the temporal pattern of events in the two ears is unchanged until the difference approaches infinity. At this point, the ears receive signals appropriate to a simple sound source with the attendant sound and localization accuracy. It is a real (monophonic) sound source generating a correspondingly real auditory event.

1.4.3k Summing Localization with Interchannel Time and Amplitude Differences

Figure 1.4.13 shows the position of the auditory image as a function of interchannel time difference for the conventional stereophonic listening situation shown in Figure 1.4.11. The curves shown are but a few of the many that are possible since, as is apparent, the trajectory of the auditory image is strongly influenced by signal type and spectral composition.

In contrast, the curves in Figure 1.4.14, showing the position of the auditory image as a function of interchannel amplitude difference, are somewhat more orderly. Even so, there are significant differences in the slopes of the curves for different signals.

With a signal like music that is complex in all respects, it is to be expected that, at a fixed time or amplitude difference, the auditory event will not always be spatially well defined or positionally stable. There are situations where experienced listeners can sometimes identify and independently localize several coexisting auditory images. Generally, however, listeners are inclined to respond with a single compromise localization, representing either the “center of gravity” of a spatially complex image display or the dominant component of the array. If the spatial display is ambiguous, there can be a strong flywheel effect in which occasional clear spatial indications from specific components of the sound engender the perception that all of that sound is continuously originating from a specific region of space. This is especially noticeable with the onset of transient or any small mechanical sounds that are easily localized compared with the sustained portion of the sounds.

The blur in stereo localization, as in natural localization, is least for an image localized in the forward direction, where, depending on the type of sound, the *stereo localization blur* is typically about 3 to 7°. With the image fully displaced by amplitude difference ($IAD = 30$ dB), the blur increases to typical values of 5 to 11°. With the image fully displaced by means of time difference ($ITD = 1$ ms), the blur increases to typical values of 10 to 16° [17].

Effect of Listener Position

Sitting away from the line of symmetry between the speakers causes the central auditory images to be displaced toward the nearer loudspeaker. Interaural time differences between the sound arrivals at the ears are introduced as the path lengths from the two speakers change. Within the first several inches of movement away from the axis of symmetry, the sound components from the left and right loudspeakers remain the first arrivals at the respective ears. In this narrow region it is possible to compensate for the effect of the ITD by adding the appropriate opposite bias of interchannel amplitude difference (see Figure 1.4.15). This process is known as *time-*

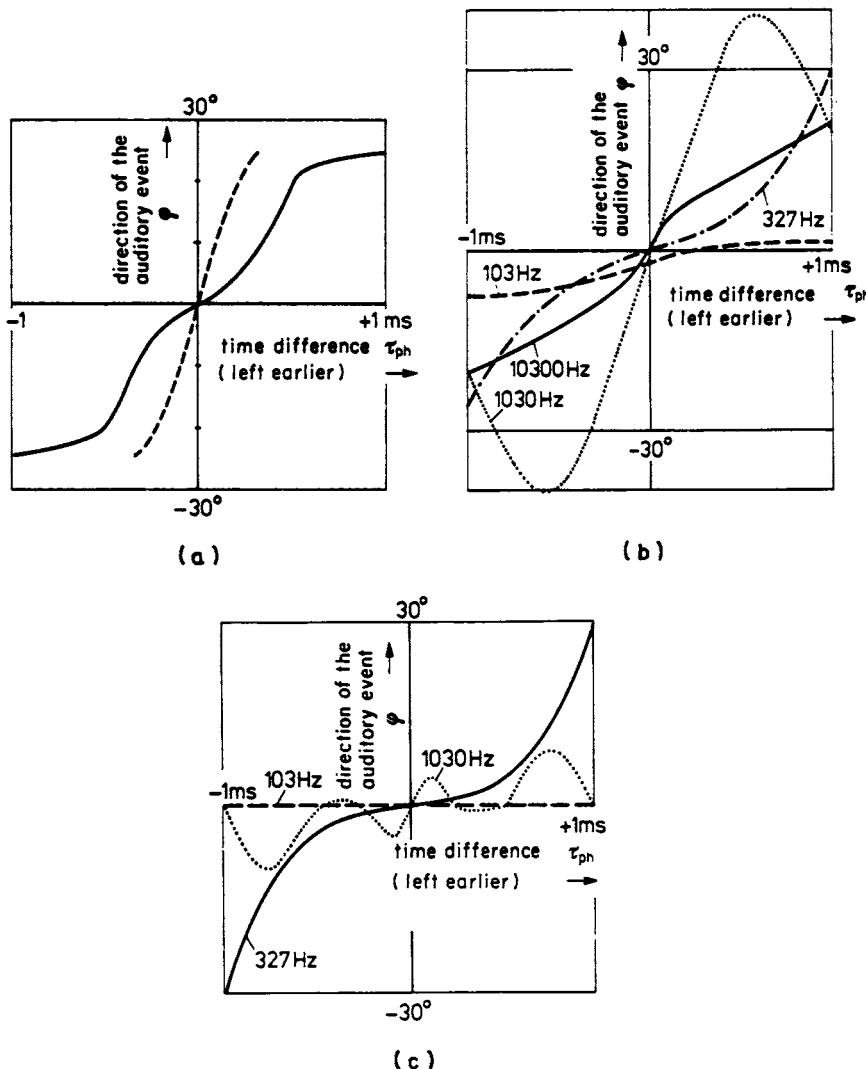


Figure 1.4.13 Direction of auditory images perceived by a listener in the situation of Figure 1.4.11 when the interchannel time difference is varied from 0 to +1 ms (left channel earlier) and -1 ms (right channel earlier). The curves show the results using different sounds: (a) dashed line = speech, solid line = impulses; (b) tone bursts; (c) continuous tones. (From [17]. Used with permission.)

intensity trading, and it is the justification for the balance control on home stereo systems, supposedly allowing the listener to sit off the axis of symmetry and to compensate for it by introducing an interchannel amplitude bias. There are some problems, however, the first one being that the trading ratio is different for different sounds, so that the centering compensations do not work

1-68 Principles of Sound and Hearing

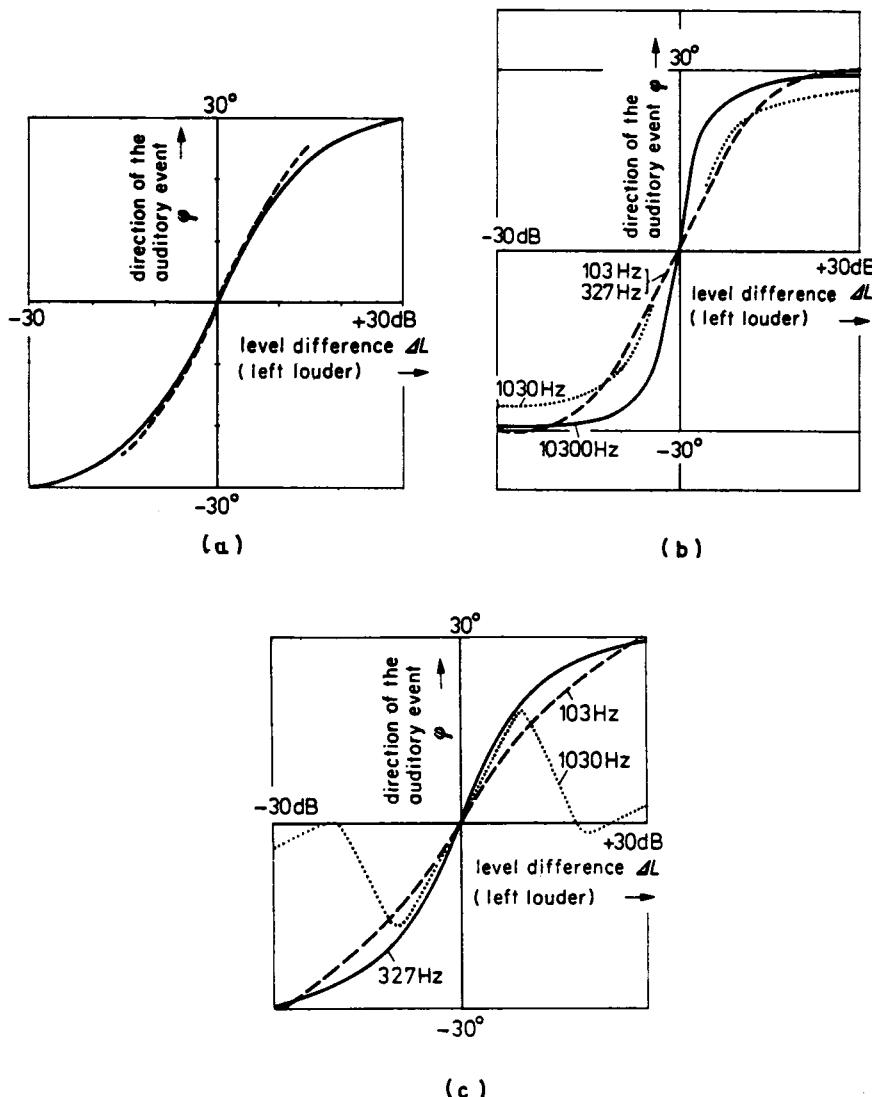


Figure 1.4.14 Direction of auditory images perceived by a listener in the situation of Figure 1.4.11 when the interchannel amplitude difference is varied from 0 to +30 dB (left louder) and -30 dB (right louder). The curves show the results with different sounds: (a) dashed line = speech, solid line = impulses; (b) tone bursts; (c) continuous tones. (From [17]. Used with permission.)

equally for all components of a complex signal; the image becomes blurred. The second problem arises when the listener moves beyond the limited range discussed previously, the simple form of summing localization breaks down, and the more complicated precedence effect comes into effect. In this region, it is to be expected that the auditory image will become rather muddled,

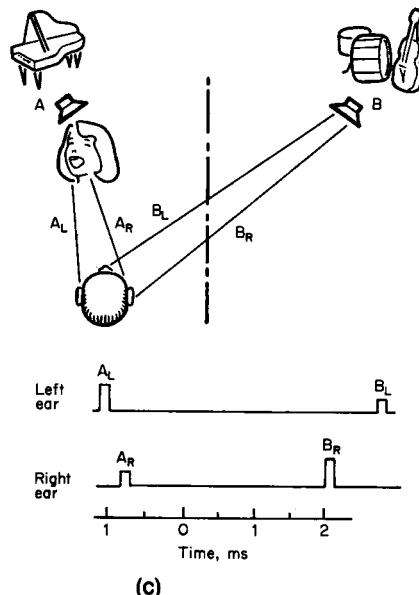
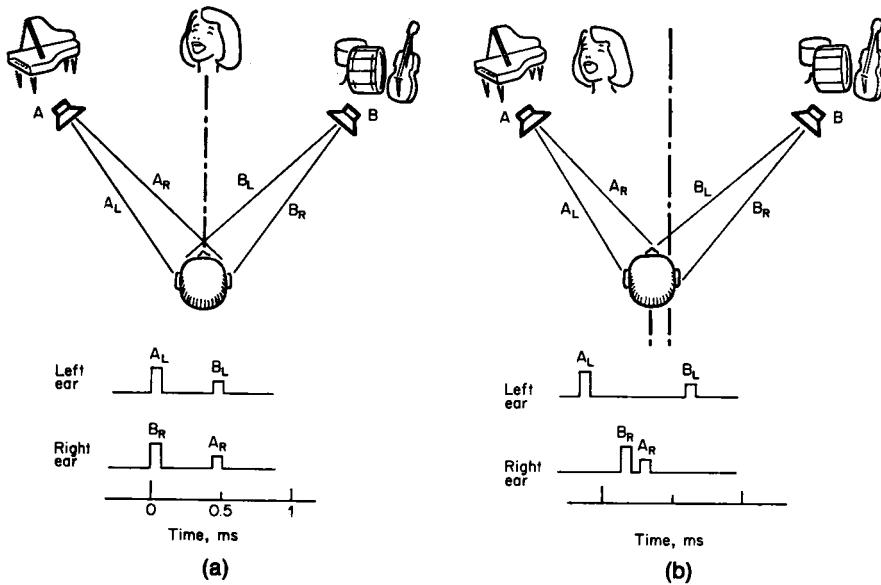


Figure 1.4.15 Sequence of events as a listener moves progressively away from the axis of symmetry in stereophonic listening.

1-70 Principles of Sound and Hearing

increasing in size and spaciousness. Localization will tend to be related to the center of gravity of a spatially diffuse auditory event rather than of a specific compact event. Nevertheless, in recordings of ensembles with natural ambience the trading may be judged to be satisfactory, since the initial effect is, by design, rather diffuse. As the listener moves about, there will also be progressive changes to the timbre due to the directional properties of the loudspeakers and wave interference between essentially similar sounds arriving at the ears with various time delays.

Stereo Image Quality and Spaciousness

The position of auditory events is but a part of the total spatial impression. In stereo reproduction as in live performances, listeners appreciate the aspect of spaciousness as long as it creates a realistic impression. The process by which an impression of spaciousness is generated in stereo is much the same as in normal hearing—a reduction in interaural cross correlation. The tradeoff is also similar, in that as the feeling of space increases, the width of the auditory images also increases [33]. The extent to which the interchannel cross-correlation coefficient is altered to manipulate these effects is, therefore, a matter of artistic judgment depending on the type of music involved.

Special Role of the Loudspeakers

In the production of stereophonic recordings the impressions of image position, size, and spaciousness are controlled by manipulating the two-channel signals. However, the impressions received by the listener are also affected by the loudspeakers used for reproduction and their interaction with the listening room.

The directionality of the loudspeakers and the location of reflecting room boundaries together determine the relative strengths of the direct, early-reflected, and reverberant sounds that impinge on the listener. To the extent that the reflected sounds can reduce the correlation between the sounds at the two ears, it is clear that loudspeakers with substantial off-axis sound radiation can enhance the sense of spaciousness. For this to be effective, however, the listening room boundaries must be sound-reflecting at least at the points of the first reflections, especially the wall (lateral) reflections.

There is evidence that listeners in domestic situations prefer a certain amount of locally generated spaciousness [19, 34, 35]. In part this may be due to the more natural spatial distribution of the reflected sounds in the listening room as opposed to the recorded ambient sounds which are reproduced only as direct sounds from the loudspeakers. Loudspeakers placed in a room where the early reflections have been absorbed or directional loudspeakers placed in any type of room would be expected to yield a reproduction lacking spaciousness. This, it seems, is preferred by some listeners at home and many audio professionals in the control room [35, 36] especially with popular music. The fact that opinions are influenced by the type of music, individual preferences, and whether the listening is done for production or for pleasure makes this a matter for careful consideration. Once selected, the loudspeaker and the room tend to remain as fixed elements in a listening situation.

1.4.4 Sound in Rooms: The General Case

Taking the broadest view of complex sound sources, we can consider the combination of real sources and their reflected images as multiple sources. In this way, it is possible to deal with situations other than the special case of stereophonic reproduction.

1.4.4a Precedence Effect and the Law of the First Wavefront

For well over 100 years it has been known that the first sound arrival dominates sound localization. The phenomenon is known as the *law of the first wavefront* or the *precedence effect*. With time delays between the first and second arrivals of less than about 1 ms we are in the realm of simple summing localization. At longer delays the location of the auditory event is dictated by the location of the source of the first sound, but the presence of the later arrival is indicated by a distinctive timbre and a change in the spatial extent of the auditory event; it may be smeared toward the source of the second sound. At still longer time delays the second event is perceived as a discrete echo.

These interactions are physically complex, with many parametric variations possible. The perceived effects are correspondingly complex, and—as a consequence—the literature on the subject is extensive and not entirely unambiguous.

One of the best-known studies of the interaction of two sound events is that by Haas [37], who was concerned with the perception and intelligibility of speech in rooms, especially where there is sound reinforcement. He formed a number of conclusions, the most prominent of which is that for delays in the range of 1 to 30 ms, the delayed sound can be up to 10 dB higher in level than the direct sound before it is perceived as an echo. Within this range, there is an increase in loudness of the speech accompanied by “a pleasant modification of the quality of the sound (and) an apparent enlargement of the sound source.” Over a wide range of delays the second sound was judged not to disturb the perception of speech, but this was found to depend on the syllabic rate. This has come to be known as the *Haas effect*, although the term has been extensively misused because of improper interpretation.

Examining the phenomenon more closely reveals a number of effects related to sound quality and to the localization dominance of the first-arrived sound. In general, the precedence effect is dependent on the presence of transient information in the sounds, but even this cannot prevent some interference from reflections in rooms. Several researchers have noted that high frequencies in delayed sounds were more disturbing than low frequencies, not only because of their relative audibility but because they were inclined to displace the localization. In fact, the situation in rooms is so complicated that it is to be expected that interaural difference cues will frequently be contradictory, depending on the frequency and temporal envelope of the sound. There are suggestions that the hearing process deals with the problem by means of a running plausibility analysis that pieces together evidence from the eyes and ears [38]. That this is true for normal listening where the sound sources are visible underlines the need in stereo reproduction to provide unambiguous directional cues for those auditory events that are intended to occupy specific locations.

1-72 Principles of Sound and Hearing

1.4.4b Binaural Discrimination

The *cocktail-party effect*, in which it is demonstrably easier to carry on a conversation in a crowded noisy room when listening with two ears than with one, is an example of *binaural discrimination*. The spatial concentration that is possible with two ears has several other ramifications in audio. Reverberation is much less obtrusive in two-eared listening, as are certain effects of isolated reflections that arrive from directions away from that of the direct sound. For example, the timbral modifications that normally accompany the addition of a signal to a time-delayed duplicate (comb filtering) are substantially reduced when the delayed component arrives at the listener from a different direction [39]. This helps to explain the finding that listeners frequently enjoy the spaciousness from lateral reflections without complaining about the coloration. In this connection it has been observed that the disturbing effects of delayed sounds are reduced in the presence of room reverberation [37] and that reverberation tends to reduce the ability of listeners to discriminate differences in the timbre of sustained sounds like organ stops and vowels [20].

1.4.5 References

1. Shaw, E. A. G.: "The Acoustics of the External Ear," in W. D. Keidel and W. D. Neff (eds.), *Handbook of Sensory Physiology*, vol. V/I, *Auditory System*, Springer-Verlag, Berlin, 1974.
2. Fletcher, H., and W. A. Munson: "Loudness, Its Definition, Measurement and Calculation," *J. Acoust. Soc. Am.*, vol. 5, pp. 82–108, 1933.
3. Robinson, D. W., and R. S. Dadson: "A Redetermination of the Equal-Loudness Relations for Pure Tones," *Br. J. Appl. Physics*, vol. 7, pp. 166–181, 1956.
4. International Organization for Standardization: *Normal Equal-Loudness Contours for Pure Tones and Normal Threshold for Hearing under Free Field Listening Conditions*, Recommendation R226, December 1961.
5. Tonic, F. E.: "Loudness—Applications and Implications to Audio," *dB*, Part 1, vol. 7, no. 5, pp. 27–30; Part 2, vol. 7, no. 6, pp. 25–28, 1973.
6. Scharf, B.: "Loudness," in E. C. Carterette and M. P. Friedman (eds.), *Handbook of Perception*, vol. 4, *Hearing*, chapter 6, Academic, New York, N.Y., 1978.
7. Jones, B. L., and E. L. Torick: "A New Loudness Indicator for Use in Broadcasting," *J. SMPTE*, Society of Motion Picture and Television Engineers, White Plains, N.Y., vol. 90, pp. 772–777, 1981.
8. International Electrotechnical Commission: *Sound System Equipment*, part 10, *Programme Level Meters*, Publication 268-1 0A, 1978.
9. Zwislocki, J. J.: "Masking—Experimental and Theoretical Aspects of Simultaneous, Forward, Backward and Central Masking," in E. C. Carterette and M. P. Friedman (eds.), *Handbook of Perception*, vol. 4, *Hearing*, chapter 8, Academic, New York, N.Y., 1978.
10. Ward, W. D.: "Subjective Musical Pitch," *J. Acoust. Soc. Am.*, vol. 26, pp. 369–380, 1954.
11. Backus, John: *The Acoustical Foundations of Music*, Norton, New York, N.Y., 1969.

12. Pierce, John R.: *The Science of Musical Sound*, Scientific American Library, New York, N.Y., 1983.
13. Gabrielsson, A., and H. Siogren: "Perceived Sound Quality of Sound-Reproducing Systems," *J. Acoust. Soc. Am.*, vol. 65, pp. 1019–1033, 1979.
14. Toole, F. E.: "Subjective Measurements of Loudspeaker Sound Quality and Listener Performance," *J. Audio Eng. Soc.*, vol. 33, pp. 2–32, 1985.
15. Gabrielsson, A., and B. Lindstrom: "Perceived Sound Quality of High-Fidelity Loudspeakers." *J. Audio Eng. Soc.*, vol. 33, pp. 33–53, 1985.
16. Shaw, E. A. G.: "External Ear Response and Sound Localization," in R. W. Gatehouse (ed.), *Localization of Sound: Theory and Applications*, Amphora Press, Groton, Conn., 1982.
17. Blauert, J: *Spatial Hearing*, translation by J. S. Allen, M.I.T., Cambridge. Mass., 1983.
18. Bloom, P. J.: "Creating Source Elevation Illusions by Spectral Manipulations," *J. Audio Eng. Soc.*, vol. 25, pp. 560–565, 1977.
19. Toole, F. E.: "Loudspeaker Measurements and Their Relationship to Listener Preferences," *J. Audio Eng. Soc.*, vol. 34, part 1, pp. 227–235, part 2, pp. 323–348, 1986.
20. Plomp, R.: *Aspects of Tone Sensation—A Psychophysical Study*, Academic, New York, N.Y., 1976.
21. Buchlein, R.: "The Audibility of Frequency Response Irregularities" (1962), reprinted in English translation in *J. Audio Eng. Soc.*, vol. 29, pp. 126–131, 1981.
22. Stevens, W. R.: "Loudspeakers—Cabinet Effects," *Hi-Fi News Record Rev.*, vol. 21, pp. 87–93, 1976.
23. Fryer, P.: "Loudspeaker Distortions—Can We Rear Them?," *Hi-Fi News Record Rev.*, vol. 22, pp. 51–56, 1977.
24. Batteau, D. W.: "The Role of the Pinna in Human Localization," *Proc. R. Soc. London*, B168, pp. 158–180, 1967.
25. Rasch, R. A., and R. Plomp: "The Listener and the Acoustic Environment," in D. Deutsch (ed.), *The Psychology of Music*, Academic, New York, N.Y., 1982.
26. Shaw, E. A. G., and R. Teranishi: "Sound Pressure Generated in an External-Ear Replica and Real Human Ears by a Nearby Sound Source," *J. Acoust. Soc. Am.*, vol. 44, pp. 240–249, 1968.
27. Shaw, E. A. G.: "Transformation of Sound Pressure Level from the Free Field to the Eardrum in the Horizontal Plane," *J. Acoust. Soc. Am.*, vol. 56, pp. 1848–1861, 1974.
28. Shaw, E. A. G., and M. M. Vaillancourt: "Transformation of Sound-Pressure Level from the Free Field to the Eardrum Presented in Numerical Form," *J. Acoust. Soc. Am.*, vol. 78, pp. 1120–1123, 1985.
29. Kuhn, G. F.: "Model for the Interaural Time Differences in the Azimuthal Plane," *J. Acoust. Soc. Am.*, vol. 62, pp. 157–167, 1977.

1-74 Principles of Sound and Hearing

30. Shaw, E. A. G.: "Aural Reception," in A. Lara Saenz and R. W. B. Stevens (eds.), *Noise Pollution*, Wiley, New York, N.Y., 1986.
31. Toole, F. E., and B. McA. Sayers: "Lateralization Judgments and the Nature of Binaural Acoustic Images," *J. Acoust. Soc. Am.*, vol. 37, pp. 319–324, 1965.
32. Blauert, J., and W. Lindemann: "Auditory Spaciousness: Some Further Psychoacoustic Studies," *J. Acoust. Soc. Am.*, vol. 80, 533–542, 1986.
33. Kurozumi, K., and K. Ohgushi: "The Relationship between the Cross-Correlation Coefficient of Two-Channel Acoustic Signals and Sound Image Quality," *J. Acoust. Soc. Am.*, vol. 74, pp. 1726–1733, 1983.
34. Bose, A. G.: "On the Design, Measurement and Evaluation of Loudspeakers," presented at the 35th convention of the Audio Engineering Society, preprint 622, 1962.
35. Kuhl, W., and R. Plantz: "The Significance of the Diffuse Sound Radiated from Loudspeakers for the Subjective Hearing Event," *Acustica*, vol. 40, pp. 182–190, 1978.
36. Voelker, E. J.: "Control Rooms for Music Monitoring," *J. Audio Eng. Soc.*, vol. 33, pp. 452–462, 1985.
37. Haas, H.: "The Influence of a Single Echo on the Audibility of Speech," *Acustica*, vol. I, pp. 49–58, 1951; English translation reprinted in *J. Audio Eng. Soc.*, vol. 20, pp. 146–159, 1972.
38. Rakerd, B., and W. M. Hartmann: "Localization of Sound in Rooms, II—The Effects of a Single Reflecting Surface," *J. Acoust. Soc. Am.*, vol. 78, pp. 524–533, 1985.
39. Zurek, P. M.: "Measurements of Binaural Echo Suppression," *J. Acoust. Soc. Am.*, vol. 66, pp. 1750–1757, 1979.
40. Hall, Donald: *Musical Acoustics—An Introduction*, Wadsworth, Belmont, Calif., 1980.
41. Durlach, N. I., and H. S. Colburn: "Binaural Phenomena," in *Handbook of Perception*, E. C. Carterette and M. P. Friedman (eds.), vol. 4, Academic, New York, N.Y., 1978.

Section
2

The Audio Spectrum

Intensity, duration, and repetition in time are perceptually important characteristics of audio signals. Consider, for example, speech, music, and natural and electronically generated sounds when heard by a listener. Most audio signals do have rather complicated waveforms in time, however, and these are difficult to analyze visually. The *spectrum* of the signal offers an alternative representation which displays the strengths of the signal's oscillating parts arranged in order of increasing oscillation. The spectrum also contains information about relative displacements or time shifts of these oscillating parts. In simple terms, the spectrum is a decomposition of the signal into several different oscillating components that later can be reassembled to re-create the original signal. All the information in the signal is contained in its spectrum, but the spectrum is a different way of representing the signal.

Frequency—the number of oscillations per second, or hertz—is a significant concept associated with the spectrum. Time is no longer explicitly used but is implicitly contained in the notion of frequency. A time interval, called a *period* and equal to the time taken for one full oscillation, is associated with every frequency, however. The period is simply the reciprocal of frequency (number of oscillations per second). A signal's overall repetitive nature as well as any hidden periodicities are revealed in its spectrum. The relative importance of the individual frequency components is also clear even though this may not be obvious from inspection of the signal itself. In the spectrum, frequency is the independent variable, or *domain*, rather than time.

These two different ways of viewing the signal, in time or in frequency, are called the *time domain* and the *frequency domain*, respectively. The two domains are interrelated by a mathematical operation known as a *transformation*, which either resolves the frequency components from a time-domain signal or reconstructs the signal from its frequency components. Insight into audio signal properties is gained by careful study of the signal in each domain. Furthermore, if the signal is passed through a system, the effects of that system on the signal also will be observed in both domains. The spectrum of the output signal can reveal important signal modifications such as, for example, which frequency components are reinforced or reduced in strength, which are delayed, or what is added, missing, or redistributed. Comparison of spectra can be used to identify and measure signal corruption or signal distortion. Thus, the spectrum plays a significant role in both signal analysis and signal processing.

With more advanced mathematical techniques it is possible to combine the two domains and form a *joint-domain* representation of the signal. This representation forms the basis for what is

2-2 Section Two

called *time-frequency analysis*. Its justification is that tones or pitch (which are frequency like) can exist and be perceived over a short time interval, after which they may change as indicated by notes in a musical score, for example. The *spectrogram* used in speech research is an early example of this approach. The objective of time-frequency analysis is to locate the signal energy in various frequency ranges during different time intervals.

Computers software is available to perform rapid transformations between time and frequency domains or to generate joint-domain representations of signals. Many computationally difficult or burdensome operations are carried out quickly and accurately. With the aid of a computer, virtually all the interesting audio spectrum and signal characteristics can be captured, displayed, and analyzed.

In computer-aided analysis of audio signals, *discrete-time* signals are used. These are formed by sampling the actual continuum of signal values at equally spaced instants in time. In principle, no information is lost through the sampling process if it is performed properly. Advanced digital signal analysis techniques play an important role both in objective technical assessment of audio equipment and in human auditory perception of sound quality.

In summary, analysis of signal and spectrum characteristics or, simply, *spectral analysis* is a quantitative means to assess audio signals and audio signal-processing systems as well as general audio quality. Additionally, certain features contained in or derived from the spectrum do correlate well with human perception of sound. Although the basis of spectral analysis is mathematical, considerable insight and understanding can be gained from a study of the several examples of time-domain and frequency-domain interrelationships provided in this section.

In This Section:

Chapter 2.1: Signals and Spectra	2-7
Introduction	2-7
Signal Energy and Power	2-7
Sinusoids and Phasor Representation	2-8
Line Spectrum	2-10
Fourier-Series Analysis	2-12
Discrete Fourier Series	2-14
Spectral Density and Fourier Transformation	2-16
Impulse Signal	2-18
Power Spectrum	2-19
Analytic Signal	2-20
Bibliography	2-22
Chapter 2.2: Spectral Changes and Linear Distortion	2-25
Introduction	2-25
Distortion Mechanisms	2-25
Linear Range	2-26
Spectra Comparison	2-26
Sinusoidal Steady-State Measurements	2-27
Some Effects of Frequency Response on Transient Signals	2-32
Phase Delay and Group Delay	2-37
Distortionless Processing of Signals	2-39
Linear Phase and Minimum Phase	2-42

Bandwidth and Rise Time	2-45
Echo Distortion	2-46
Classifications of Phase Distortion	2-48
Bibliography	2-50

Reference Documents for This Section:

- Bendat, J. S., and A. G. Riersol: *Engineering Applications of Correlation and Spectral Analysis*, Wiley, New York, 1980.
- Bendat, J. S., and A. G. Piersol: *Random Data: Analysis and Measurement Procedures*, Wiley-Interscience, New York, N.Y., 1971.
- Blinchikoff, H. J., and A. I. Zverev: *Filtering in the Time and Frequency Domains*, Wiley, New York, N.Y., 1976.
- Bloom, P. J., and D. Preis: "Perceptual Identification and Discrimination of Phase Distortions," *IEEE ICASSP Proc.*, pp. 1396–1399, April 1983.
- Bode, H. W.: *Network Analysis and Feedback Amplifier Design*, Van Nostrand, New York, N.Y., 1945.
- Bracewell, R.: *The Fourier Integral and Its Applications*, McGraw-Hill, New York, N.Y., 1965.
- Cheng, D. K.: *Analysis of Linear Systems*, Addison-Wesley, Reading, Mass., 1961.
- Childers, D. G.: *Modern Spectral Analysis*, IEEE, New York, N.Y., 1978.
- Connor, F. R.: *Signals*, Arnold, London, 1972.
- Deer, J. A., P. J. Bloom, and D. Preis: "Perception of Phase Distortion in All-Pass Filters," *J. Audio Eng. Soc.*, vol. 33, no. 10, pp. 782–786, October 1985.
- Di Toro, M. J.: "Phase and Amplitude Distortion in Linear Networks," *Proc. IRE*, vol. 36, pp. 24–36, January 1948.
- Guillemin, E. A.: *Communication Networks*, vol. 11, Wiley, New York, N.Y., 1935.
- Henderson, K. W., and W. H. Kautz: "Transient Response of Conventional Filters," *IRE Trans. Circuit Theory*, CT-5, pp. 333–347, December 1958.
- Hewlett-Packard: "Application Note 63—Section II, Appendix A, "Table of Important Transforms," Hewlett-Packard, Palo Alto, Calif, pp. 37, 38, 1954.
- Jenkins, G. M., and D. G. Watts: *Spectral Analysis and Its Applications*, Holden-Day, San Francisco, Calif., 1968.
- Kharkevich, A. A.: *Spectra and Analysis*, English translation, Consultants Bureau, New York, N.Y., 1960.
- Kupfmuller, K.: *Die Systemtheorie der elektrischen Nachrichtenübertragung*, S. Hirzel Verlag, Stuttgart, 1968.
- Lane, C. E.: "Phase Distortion in Telephone Apparatus," *Bell Syst. Tech. J.*, vol. 9, pp. 493–521, July 1930.
- Lathi, B. P.: *Signals, Systems and Communications*, Wiley, New York, N.Y., 1965.

2-4 Section Two

- Lynn, P. A.: *An Introduction to the Analysis and Processing of Signals*, 2nd ed. Macmillan, London, 1982.
- Mallinson, J. C.: "Tutorial Review of Magnetic Recording." *Proc. IEEE*, vol. 62, pp. 196–208, February 1976.
- Members of the Technical Staff of Bell Telephone Laboratories: *Transmission Systems for Communications*, 4th ed., Western Electric Company, Technical Publications, Winston-Salem, N.C., 1971.
- Oppenheim, A. V., and R. W. Schafer: *Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, N.J., 1975.
- Panter, P. F.: *Modulation, Noise and Spectral Analysis*, McGraw-Hill, New York, N.Y., 1965.
- Papoulis, A.: *Signal Analysis*, McGraw-Hill, New York, N.Y., 1977.
- Papoulis, A.: *The Fourier Integral and Its Applications*, McGraw-Hill, New York, N.Y., 1962.
- Peus, S.: "Microphones and Transients," *db Mag.*, translated from *Radio Mentor* by S. Temmer, vol. 11, pp. 35–38, May 1977.
- Preis, D.: "A Catalog of Frequency and Transient Responses," *J. Audio Eng. Soc.*, vol. 25, no. 12, pp. 990–1007, December 1977.
- Pries, D.: "Audio Signal Processing with Transversal Filters," *IEEE Conf. Proc.*, 1979 ICASSP, pp. 310–313, April 1979.
- Preis, D.: "Hilbert-Transformer Side-Chain Phase Equalizer for Analogue Magnetic Recording," *Electron. Lett.*, vol. 13, pp. 616–617, September 1977.
- Preis, D.: "Impulse Testing and Peak Clipping," *J. Audio Eng. Soc.*, vol. 25, no. 1, pp. 2–14, January 1977.
- Preis, D.: "Least-Squares Time-Domain Deconvolution for Transversal-Filter Equalizers," *Electron. Lett.*, vol. 13, no. 12, pp. 356–357, June 1977.
- Preis, D.: "Linear Distortion," *J. Audio Eng. Soc.*, vol. 24, no. 5, pp. 346–367, June 1976.
- Pries, D.: "Measures and Perception of Phase Distortion in Electroacoustical Systems," *IEEE Conf. Proc.*, 1980 ICASSP, pp. 490–493, 1980.
- Pries, D.: "Phase Equalization for Analogue Magnetic Recorders by Transversal Filtering," *Electron. Lett.*, vol. 13, pp. 127–128, March 1977.
- Pries, D.: "Phase Equalization for Magnetic Recording," *IEEE Conf. Proc.*, 1981 ICASSP, pp. 790–795, March 1981.
- Preis, D.: "Phase Distortion and Phase Equalization in Audio Signal Processing—A Tutorial Review," *J. Audio Eng. Soc.*, vol. 30, no. 11, pp. 774–794, November 1982.
- Pries, D., and C. Bunks: "Three Algorithms for the Design of Transversal-Filter Equalizers," *Proc. 1981 IEEE Int. Symp. Circuits Sys.*, pp. 536–539, 1981.
- Pries, D., and P. J. Bloom: "Perception of Phase Distortion in Anti-Alias Filters," *J. Audio Eng. Soc.*, vol. 32, no. 11, pp. 842–848, November 1984.

- Preis, D., F. Hlawatsch, P. J. Bloom, and J. A. Deer: "Wigner Distribution Analysis of Filters with Perceptible Phase Distortion," *J. Audio Eng. Soc.*, December 1987.
- Rabiner, L. R., and C. M. Rader (eds.): *Digital Signal Processing*, IEEE, New York, N.Y., 1972.
- Schwartz, M.: *Information Transmission, Modulation and Noise*, McGraw-Hill, New York, N.Y., 1970.
- Small, R. H.: "Closed-Box Loudspeaker Systems, Part 1: Analysis," *J. Audio Eng. Soc.*, vol. 20, pp. 798–808, December 1972.
- Totzek, U., and D. Press: "How to Measure and Interpret Coherence Loss in Magnetic Recording," *J. Audio Eng. Soc.*, December 1987.
- Totzek, U., D. Preis, and J. F. Boebme: "A Spectral Model for Time-Base Distortions and Magnetic Recording," *Archiv. fur Elektronik und Ubertragungstechnik*, vol. 41, no. 4, pp. 223–231, July-August 1987.
- Westman, H. P. (ed.): *ITT Reference Data for Radio Engineers*, Howard W. Sams, New York, N.Y., 1973.
- Wheeler, H. A.: "The Interpretation of Amplitude and Phase Distortion in Terms of Paired Echos," *Proc. IRE*, vol. 27, pp. 359–385, June 1939.
- Williams, A. B.: *Active Filter Design*, Artech House, Dedham, Mass., 1975.
- Zverev, A. I.: *Handbook of Filter Synthesis*, Wiley, New York, N.Y., 1967.

Chapter
2.1
Signals and Spectra

Douglas Preis

2.1.1 Introduction

Signals serve several purposes in audio engineering. Primarily, they carry information, for example, electrical analogs of music or speech or numerical data representing such information. Discrete-time signals, formed from sampled values of continuous signals, are now used extensively in digital recording, processing, storage, and reproduction of audio signals. Signals devised and used solely to elicit a response from an audio system are called test signals. Control signals modify the internal operation of signal-processing devices. Certain signals, such as electronic thermal noise, magnetic-tape hiss, or quantization noise in digital systems, may be present but unwanted.

Essential to a deeper understanding of all kinds of signals is the spectrum. The spectrum is defined in slightly different ways for different classes of signals, however. For example, deterministic signals have a mathematical functional relationship to time that can be described by an equation, whereas nondeterministic signals, such as noise generated by a random process, are not predictable but are described only by their statistical properties. Their spectra are defined in different ways. There are also two types of deterministic signals, classified by total energy content or average energy content; and, again, their spectra are defined differently. All spectral representations provide information about the underlying oscillatory content of the signal. This content can be concentrated at specific frequencies or distributed over a continuum of frequencies, or both.

2.1.2 Signal Energy and Power

A deterministic, real-valued signal $f(t)$ is called a *finite-energy* or *transient* signal if

$$0 < \int_{-\infty}^{\infty} f^2(t) dt < \infty \quad (2.1.1)$$

where t is time. The integrand $f^2(t)$ can be interpreted as the instantaneous power (energy/time) if $f(t)$ is assumed to be a time-varying voltage across a $1\text{-}\Omega$ resistor. The numerical value of the

2-8 The Audio Spectrum

integral in Equation (2.1.1) is the signal's total energy. A *finite power* deterministic signal satisfies

$$0 < \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} f^2(t) dt < \infty \quad (2.1.2)$$

That is, the average energy per time or average power is finite. For example, $f(t)$ could be a constant dc voltage existing for all time across a $1\text{-}\Omega$ resistor.

Fundamental to understanding spectral analysis is an elementary periodic signal, that is, one that oscillates with constant frequency and does not decay as time progresses. This simplest oscillating signal is called a *sinusoid*. It predicts, for example, the motion of a swinging pendulum (without friction) or the exchange of energy between inductor and capacitor in a lossless resonant circuit. The sinusoid is the solution to a differential equation that describes a wide variety of physical oscillatory and vibrational phenomena.

2.1.2a Sinusoids and Phasor Representation

The sinusoidal signal illustrated in Figure 2.1.1a and described mathematically by

$$f(t) = A \cos(\omega t + \theta) \quad (2.1.3)$$

is a finite-power signal characterized by its real amplitude A , radian frequency, ω rad/s, and θ , which is a constant phase angle (in radians). The quantity $\omega/(2\pi)$ is the cyclic frequency or number of oscillations per second, called hertz, and equals the reciprocal of the sinusoid's period T , the time taken for one full oscillation. These relationships are illustrated in Figure 2.1.1a.

The peak, or maximum, value of the signal is A . Its root-mean-square (rms) value found from taking the square root of the expression in Equation (2.1.2), is $A/\sqrt{2} = 0.7071A$, and the average value of $|f(t)|$ over one period is $2A/\pi = 0.6366A$. The average power, from Equation (2.1.2), equals $A^2/2$, or simply the rms value squared. By appropriately changing the phase angle θ , a pure cosine wave or sine wave is realized. For example, $\theta = 0, \pi/2, \pi, 3\pi/2, 2\pi$ corresponds to $A \cos(\omega t), -A \sin(\omega t), -A \cos(\omega t), A \sin(\omega t)$, respectively. If the phase of a sine wave is increased by $\pi/2$ (that is, 90° positive phase shift or phase lead), it becomes a cosine wave. A phase shift of π rad, or 180° , inverts the polarity of a sinusoid.

By using the Euler identity,

$$e^{jx} = \cos(x) + j \sin(x) \quad (2.1.4)$$

where $j = \sqrt{-1}$, $f(t)$ in Equation (2.1.3), can also be written as the real part of a time-varying complex number (*phasor*), namely

$$f(t) = \text{real} [A e^{j(\omega t + \theta)}] \quad (2.1.5a)$$

A conceptual picture of Equation (2.1.5a), called a phasor diagram, is given in Figure 2.1.1b. The tip of the arrow describes the locus of points in the complex plane of the expression in brackets

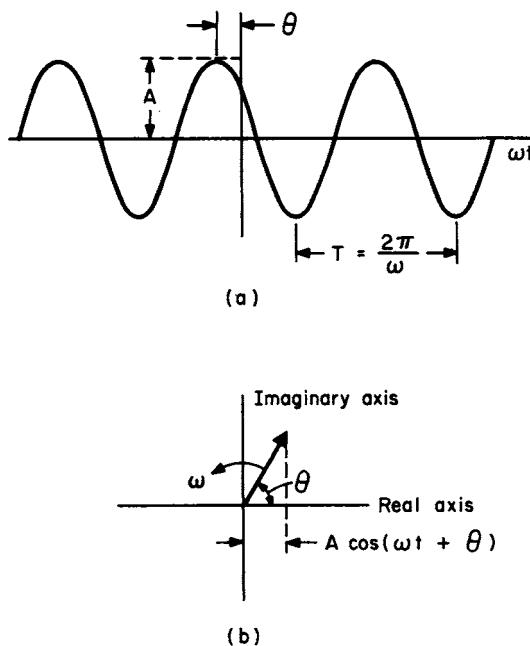


Figure 2.1.1 Sinusoid characteristics: (a) sinusoid $A \cos(\omega t + \theta)$, (b) phasor diagram representation of sinusoid.

in Equation (2.1.5a) as the arrow itself rotates counterclockwise with angular velocity ω rad/s. As time progresses, the projection (or shadow) of the arrow's length A onto the horizontal (real) axis has the values $A \cos(\omega t + \theta)$. By convention, the phasor diagram is drawn when $t = 0$ or, equivalently, showing the phase θ of the sinusoid relative to a cosine reference phasor, $\cos(\omega t)$, which would lie on the positive real axis. Note that multiplication of this phasor by $e^{j\psi}$ advances its phase by ψ rad, as can be shown by using Equation (2.1.5a). For example, multiplication of the phasor by $j = e^{j\pi/2}$ advances its phase by $\pi/2$, or 90° .

The phasor concept can be extended on the basis of the following two identities derived from Equation (2.1.4)

$$\cos(\omega t) = \frac{1}{2}[e^{j\omega t} + e^{-j\omega t}] \quad (2.1.5b)$$

and

$$\sin(\omega t) = \frac{1}{2}[e^{j\omega t} - e^{-j\omega t}] \quad (2.1.5c)$$

2-10 The Audio Spectrum

where the negative signs in the exponents are associated with ω implying negative angular velocity $-\omega$. Thus, $\cos(\omega t)$ and $\sin(\omega t)$ in Equation (2.1.4) each can be interpreted simply as a sum of two counter-rotating phasors, as shown in Figure 2.1.2. On the left side of this figure, the imaginary parts of the two phasors always cancel because they are equal and opposite, and the real parts add to form $\cos(\omega t)$. On the right side, imaginary parts cancel at all times, and the real parts add to yield $\sin(\omega t)$.

There are two ways to represent a sinusoid $A \cos(\omega t + \theta)$ in terms of counter-rotating phasors. The simpler method is to scale the lengths of the phasors for $\cos(\omega t)$ in Figure 2.1.2 by the amplitude A and then to change the phase of each phasor by rotating it in its direction of rotation θ rad. This operation corresponds to advancing or leading the phase of the $+\omega$ phasor and advancing or leading the phase of the $-\omega$ phasor by θ rad. The other way makes use of the trigonometric identity

$$A \cos(\omega t + \theta) = A \cos\theta \cos(\omega t) - A \sin\theta \sin(\omega t) \quad (2.1.6)$$

which states that a sinusoid with an arbitrary phase angle can always be represented as the sum of a cosine wave and a sine wave whose coefficients are $A \cos\theta$ and $-A \sin\theta$, respectively. Depending on the phase angle θ , these coefficients can be positive, negative, or zero. When either is negative, the polarity of the associated sine wave or cosine wave is just inverted; i.e., its phase is changed by π rad, or 180° , so that the negative sign is absorbed and the coefficients represent nonnegative amplitudes. So, the phasors for $\cos(\omega t)$ and $\sin(\omega t)$ in Figure 2.1.2 with their lengths multiplied by the coefficients in Equation (2.1.6) also represent $A \cos(\omega t + \theta)$.

2.1.2b Line Spectrum

An alternative to counter-rotating phasors is the *line spectrum* (discrete spectrum). Because each phasor shown in Figure 2.1.2 contains only three pieces of information, its length (amplitude), its angular velocity $+\omega$ or $-\omega$, and its phase (measured relative to the positive horizontal axis shown in Figure 2.1.2), the same information can be presented differently as shown in Figure 2.1.3. At points $-\omega$ and $+\omega$ on the horizontal frequency axis, a vertical line is drawn whose length equals that of its corresponding phasor to give the amplitude. The phase of each phasor is plotted above or below its frequency on a separate graph called the *phase spectrum*. The range of the phase spectrum is from $-\pi$, or -180° , to $+\pi$, or 180° . Consider, for example, $\sin(\omega t)$ as represented on the right side of Figure 2.1.2 and Figure 2.1.3. The phase of the counterclockwise phasor is $-\pi/2$, or -90° , and that of clockwise phasor is $\pi/2$, or $+90^\circ$.

The real advantage of the line-spectrum representation is evident when the signal is composed of several different frequency components $\pm\omega_1, \pm\omega_2, \pm\omega_3, \dots$, because a separate phasor diagram would be needed for each frequency pair $\pm\omega_n$, whereas all frequencies can be displayed in one line-spectrum plot. A slight and apparent disadvantage is that the concept of negative frequencies is used and signal amplitude is split evenly between the positive and negative frequencies. This representation is a matter of convention, however. If only the total amplitude at each specific frequency is of interest, then phases can be ignored and line spectra drawn at dc (zero frequency) and only at positive values of frequency. The amplitudes are simply doubled at positive frequencies from their values in the “two-sided” line-spectra representation. It is also possible, by a different convention, to define the spectrum by using only positive frequencies and simply giving the amplitude A and phase θ for each sinusoid in the form of Equation (2.1.3). In

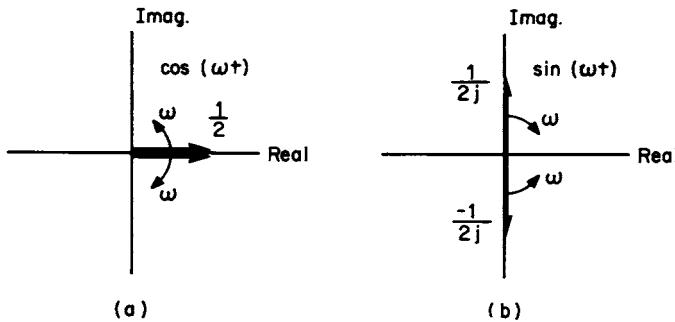


Figure 2.1.2 Waveform representations: (a) cosine wave, (b) sine wave represented by counter-rotating phasors.

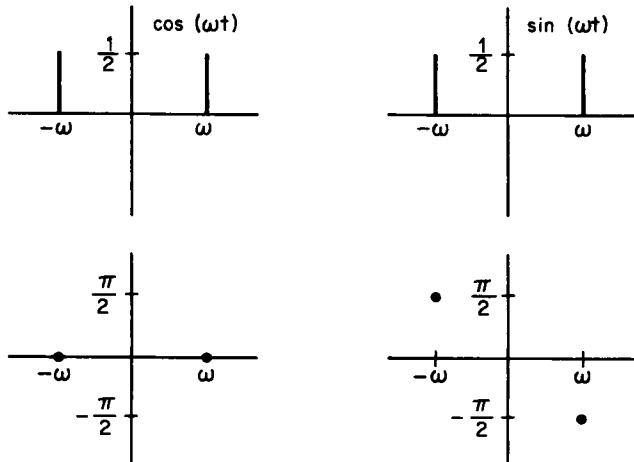


Figure 2.1.3 Line-spectrum representation of $\cos(\omega)$ at left and $\sin(\omega)$ at right.

some applications, it is more useful to give the power $A^2/2$, rather than the amplitude A , associated with each spectral line.

The complex exponentials not only have an interesting interpretation as counter-rotating phasors but also have the advantage of representing arbitrary sinusoids without explicitly using the sine and cosine functions themselves, or having to take the real part of a complex number as in Equation (2.1.5a). Because complex exponentials form the basis for all mathematical transformations between the time domain and frequency domain, their use has become the method of choice.

While a sinusoid mathematically represents a *pure tone* at a specific frequency $\omega_0/2\pi = 1/T$ Hz, most periodic musical sounds or periodic signal waveforms have *harmonic structure*, meaning that they also contain frequencies that are integer multiples of the lowest or *fundamental* frequency ω_0 which determines the period T . These higher frequencies do, in general, have different

2-12 The Audio Spectrum

amplitudes and phases. Within the period T , each higher frequency or higher harmonic will have an integer number, e.g., 2 or 3 or 4, etc., of full oscillations. The line spectrum is a meaningful way to represent such a signal in the frequency domain. Before the line spectrum can be drawn, the amplitudes and phases of the frequency components in the time waveform are required. These can be found from a specific mathematical analysis of the waveform called Fourier analysis. An understanding of Fourier analysis is essential to the general concept of the spectrum, frequency-domain representation of signals, transformations, and linear signal processing.

2.1.2c Fourier-Series Analysis

To determine whether a periodic $f(t)$ has a sinusoidal component, $\cos(\omega)$ for example, at frequency ω the product $f(t) \cos(\omega t)$ is integrated over one full period $T = 2\pi/\omega$. If, for example, $f(t) = A \cos(\omega t)$, then

$$\int_{-T/2}^{T/2} f(t) \cos(\omega t) dt = A \int_{-T/2}^{T/2} \cos^2(\omega t) dt = \frac{AT}{2} \quad (2.1.7)$$

because $\cos^2(\omega t) = 1/2 + 1/2 \cos(2\omega t)$ and the second term integrates to zero. Thus, the amplitude A of $\cos(\omega t)$ in $f(t)$ is simply $2/T$ times the integral, Equation (2.1.7). The left integral in Equation (2.1.7) would equal zero if $f(t) = A \cos(n\omega_0 t)$, $n \neq 1$, but $n = 0, 2, 3, \dots$. The same method applies by using $\sin(\omega t)$ if $f(t)$ contains the sinusoidal component $A \sin(\omega t)$. This forms the basis for *Fourier analysis*, that is, a systematic way to determine the harmonic content, in terms of $\sin(n\omega_0 t)$ and $\cos(n\omega_0 t)$, for $n = 0, 1, 2, \dots$, of a periodic $f(t)$ whose period $T = 2\pi/\omega_0$.

A *Fourier series* is a mathematical way to represent a real, periodic, finite power signal in terms of a sum of harmonically related sinusoids. If $f(t)$ is periodic with period T , it repeats itself every T so that

$$f(t+T) = f(t) \quad (2.1.8)$$

and its Fourier series is given by

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos(n\omega_0 t) + \sum_{n=1}^{\infty} b_n \sin(n\omega_0 t) \quad (2.1.9)$$

where $\omega_0 = 2\pi/T$ and the *Fourier coefficients* are real numbers given by

$$a_n = \frac{2}{T} \int_{-T/2}^{T/2} f(t) \cos(n\omega_0 t) dt \quad (2.1.10)$$

and

$$b_n = \frac{2}{T} \int_{-T/2}^{T/2} f(t) \sin(n\omega_0 t) dt \quad (2.1.11)$$

The frequency ω_0 is called the *fundamental frequency* (or first harmonic) and, for $n = 2, 3, 4, \dots, n_0$ are the second, third, fourth, etc., *harmonics*, respectively. The $a_0/2$ term is the average value or dc (zero-frequency) content of $f(t)$. The Fourier coefficients in Equations (2.1.10) and (2.1.11) can be interpreted as the *average content* of $\cos(n\omega_0 t)$ and $\sin(n\omega_0 t)$, respectively, in $f(t)$.

By using the Euler identity, Equations (2.1.4), (2.1.10), and (2.1.11) can be combined to yield

$$c_n = \frac{a_n - jb_n}{2} = \frac{1}{T} \int_{-T/2}^{T/2} f(t) e^{-j n \omega_0 t} dt \quad (2.1.12)$$

where each c_n is a *complex Fourier coefficient* whose magnitude and (phase) angle ϕ are, respectively

$$|c_n| = \sqrt{(a_n^2 + b_n^2)/2} \quad (2.1.13a)$$

$$\phi_n = \arctan\left(-\frac{b_n}{a_n}\right) \quad (2.1.13b)$$

By defining $c_0 = a_0/2$ and, for $n \neq 0$

$$c_{-n} = c_n^* = \frac{a_n + jb_n}{2} \quad (2.1.14)$$

the Fourier series Equation (2.1.9) can be written as

$$f(t) = c_0 + \sum_{n=1}^{\infty} c_n e^{jn\omega_0 t} + \int_{-\infty}^{\infty} c_n e^{jn\omega_0 t} \quad (n = -1) \quad (2.1.15a)$$

$$f(t) = \sum_{n=-\infty}^{\infty} c_n e^{jn\omega_0 t} \quad (2.1.15b)$$

which can be derived from Equation (2.1.9) by using Equations (2.1.12), (2.1.14), and (2.1.4). Equation (2.1.15b) is the *complex form* of the Fourier series and can be interpreted by using Equation (2.1.15a), which shows the sum of the dc term, positive-frequency terms, and negative-frequency terms (n has only negative values in the second sum), respectively.

2-14 The Audio Spectrum

When a periodic signal is represented in the form (2.1.15b), its corresponding line spectrum can be viewed, first, as pairs of pure sine waves and cosine waves with different real coefficients a_n and b_n , as prescribed by Equations (2.1.10) and (2.1.11) at positive frequencies $n\omega_0$ which sum to form sinusoids like Equation (2.1.6). Using identities (2.1.5), however, these sine waves and cosine waves also can be written as sums of complex exponentials $e^{jn\omega_0 t}$ and $e^{-jn\omega_0 t}$ with amplitudes and phases given by the complex coefficients c_n . Alternatively, the line spectrum can be viewed as sinusoids at positive frequencies $n\omega_0$ in the form (2.1.6) and each represented as the sum of a pair of phasors counter-rotating with angular velocities $+n\omega_0$ and $-n\omega_0$ whose lengths equal $|c_n|$, and whose initial phase angle is specified by the angle of c_n . In both interpretations negative frequencies are used, but this results in the very compact complex exponential notation in Equation (2.1.15b), where sine and cosine functions are no longer needed.

Figure 2.1.4 shows different periodic functions of time and their corresponding line spectra determined from the magnitude and phase (or angle) of the complex Fourier coefficients c_n .

2.1.2d Discrete Fourier Series

If a Fourier series has only a finite number of terms (i.e., its frequency band is limited) and the highest frequency equals $K\omega_0$ rad/s (the K th harmonic), then Equation (2.1.15b) simplifies to the finite sum

$$f(t) = \sum_{n=-K}^{K} c_n e^{jn\omega_0 t} \quad (2.1.16)$$

In certain applications the values of $f(t)$ are needed only at N instants in time (within the period T) that are spaced T/N apart. The values $f(t)$ at these N points in time, from Equation (2.1.16), are given by

$$f(mT/N) = \sum_{n=-K}^{K} c_n e^{jn\omega_0 mT/N} \quad (2.1.17)$$

where m is an integer. Choosing the number of points $N = 2K$, that is, the time interval between points equal to half of the period of the highest frequency ($T/N = T/2K$) and substituting $2\pi/T$ for ω_0 in Equation (2.1.17) gives

$$f(mT/N) = \sum_{n=-N/2}^{N/2} c_n w_N^{nm} \quad (2.1.18)$$

where

$$W_N = e^{j2\pi/N} \quad (2.1.19)$$

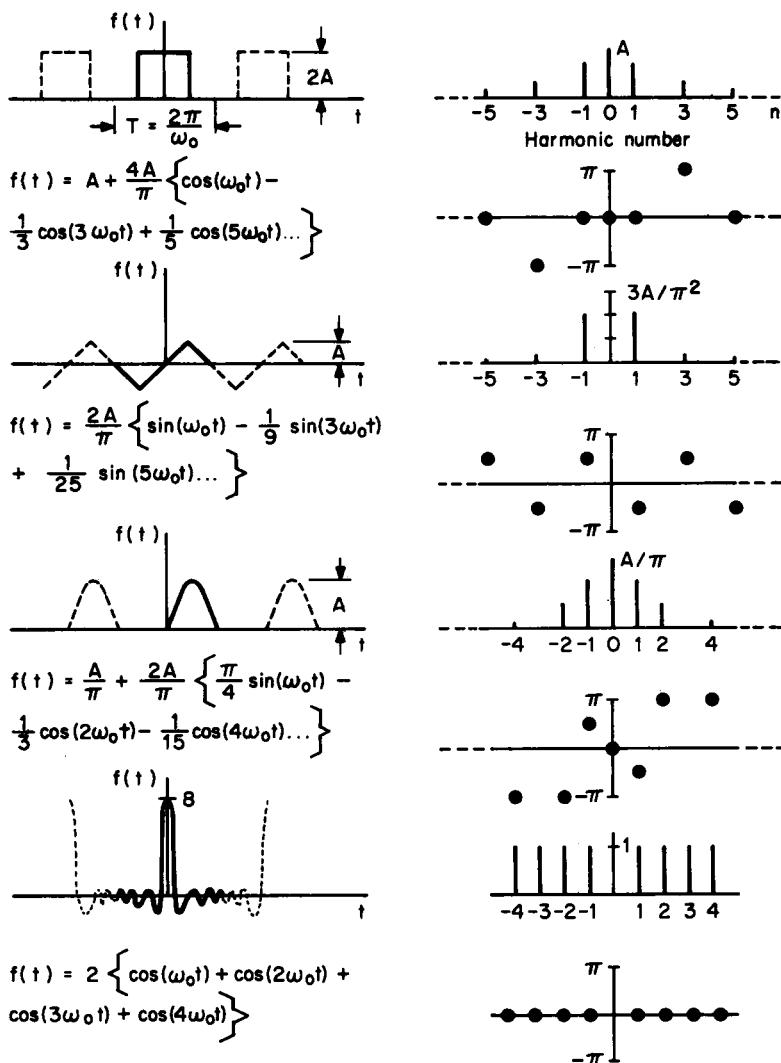


Figure 2.1.4 Periodic functions and their line spectra.

which is called the N th root of unity. The quantity w_N^m , for N values of m , mathematically represents specific points on the unit circle in the complex plane spaced $2\pi/N$, or $360^\circ/N$ apart (beginning from $e^{j0} = 1$). Each of these points is physically interpreted as the location of the tip of a phasor of unit length, which rotates through 2π rad, or 360° , during the period T at the instants in time mT/N . Because the phasor for the n th harmonic has an angular velocity n times faster than that of the fundamental ($n = 1$), the term w_N^{nm} appears in Equation (2.1.18). One way to interpret the term $2\pi/N$ in Equation (2.1.19) is that, for any one of the $N/2$ harmonics, only N discrete values of phase are possible after sampling, regardless of frequency. The coefficients c_n in Equation

2-16 The Audio Spectrum

(2.1.18) are complex numbers of the form $|c_n|e^{j\psi_n}$ that scale the unit amplitudes and shift the phases of the sampled harmonics represented by w_N^{mn} .

The Fourier coefficients are related to the sampled values of $f(t)$ by the similar equation

$$c_n = \frac{1}{N} \sum_{m=-M/2}^{M/2} f(mT/N) w_N^{-nm} \quad (2.1.20)$$

where the integer M is even and numerically equal to N . Equation (2.1.20) can be substituted into Equation (2.1.18), or vice versa, to yield an identity. These two equations define an N -point *discrete Fourier-series pair* which is often called a *discrete Fourier transform* (DFT) pair. They relate the N equally spaced sampled values of $f(t)$ in Equation (2.1.18) to its line spectrum, which contains $N/2$ negative-frequency components, a dc term, and $N/2$ positive-frequency components as well as their phases as prescribed by Equation (2.1.20). For an example, see the last line spectrum in Figure 2.1.4.

Equation (2.1.20) is a discrete-time version of Equation (2.1.12) and is valid only if the periodic function $f(t)$ is sampled throughout its period T at a rate that is twice the highest frequency it contains. If $f(t)$ contains harmonics higher than $N/2$, then the c_n coefficients in Equation (2.1.20) are not identical to those in Equation (2.1.12); they will become corrupted or *aliased* because higher frequencies (with a harmonic number greater than $N/2$) in $f(t)$ influence the value of c_n as computed from Equation (2.1.20), but these frequencies cannot be reconstructed by using the harmonics up to harmonic number $N/2$ as in Equation (2.1.18). Furthermore, the resulting line spectrum itself will be incorrect.

The DFT can be computed very efficiently from sampled values of $f(t)$, as contained in Equation (2.1.20), or $f(t)$ can be reconstructed at the sample points, as in Equation (2.1.18), using a numerical algorithm called the *fast Fourier transform* (FFT). The algorithm exploits, in computation, the fact that w_N^{mn} represents only N distinct values of phase.

2.1.2e Spectral Density and Fourier Transformation

Line spectra, where each line $|c_n|$ represents a sinusoid whose average power is $|c_n|^2 / 2$, cannot be used to represent the spectrum of a finite-energy signal $f(t)$ because the average power of $f(t)$, as given in Equation (2.1.2), is zero by definition. Furthermore, while a finite-energy signal can exist for all time, subject to the constraint given in Equation (2.1.1), it has no finite period T associated with itself. For finite-energy signals, an *amplitude-density spectrum* is defined in the following way

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt \quad (2.1.21)$$

The dimensions of $F(\omega)$ are amplitude multiplied by time or amplitude/(1/time) and are interpreted as amplitude/frequency. The quantity $F(\omega)$ is called *amplitude spectral density* or, simply, *spectral density* of $f(t)$. By comparing Equation (2.1.21) with Equation (2.1.12), $F(\omega)$ is seen to be defined for all rather than discrete frequencies, and the limits of integration include all time rather than one period. Similar to Fourier analysis in concept, the integral (2.1.21) extracts from

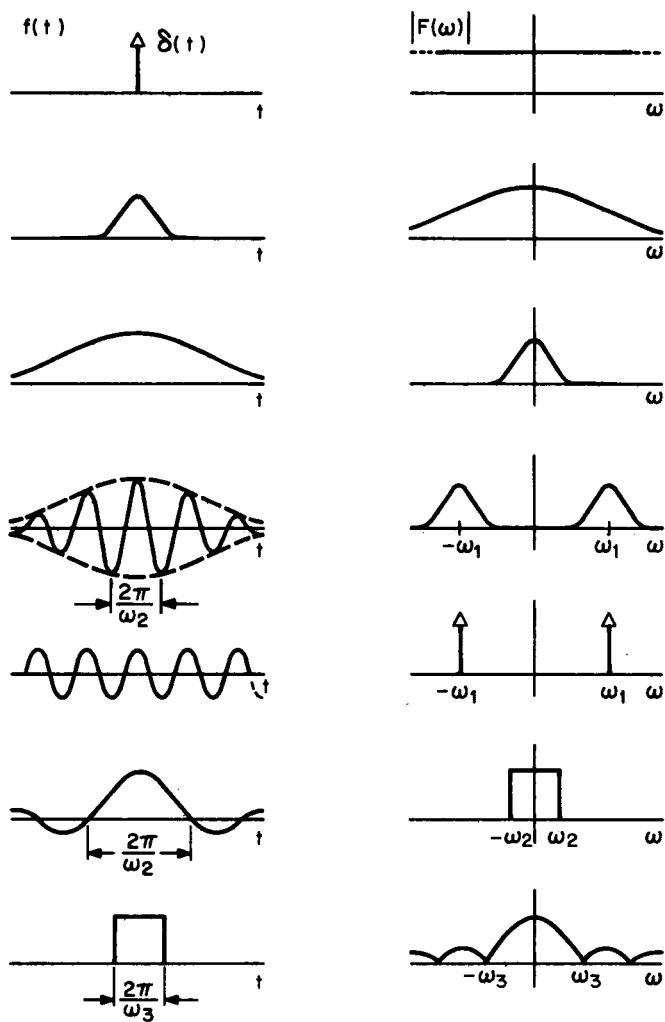


Figure 2.1.5 Amplitude spectral density $|F(\omega)|$ corresponding to time-domain signals $f(t)$.

$f(t)$ its average spectral density $F(\omega)$ and associated phase at each frequency ω throughout a continuum of frequencies. The original signal $f(t)$, again analogously to Fourier analysis, can be reconstructed from its spectral density by using the relation

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{j\omega t} d\omega \quad (2.1.22)$$

2-18 The Audio Spectrum

where all frequencies, instead of discrete frequencies, are used. Equations (2.1.21) and (2.1.22) form a *Fourier-transform pair*. By comparing these two equations it is seen that the value of $f(t)$ at, for example, $t = t_1$ depends on contributions from $F(\omega)$ at all frequencies and, conversely, that the value of the spectral density $F(\omega)$ at $\omega = \omega_1$ depends on contributions from $f(t)$ at all times. In both cases integrals with infinite limits are evaluated. In Figure 2.1.5, several Fourier-transform pairs are illustrated.

The relations (2.1.21) and (2.1.22) can be derived from the Fourier-analysis Equations (2.1.12) and (2.1.15b) by using a limiting process where the period T becomes infinite and the frequency interval between harmonics approaches zero.

The continuous Fourier transform (Equation 2.21) is often approximated or estimated by using the discrete Fourier transform (2.1.18) by assuming that $f(t)$ is periodic, with period T , and that its high-frequency content is zero beyond harmonic number $N/2$, where N equals the number of equally spaced sampled values of $f(t)$ within the assumed period.

The *Laplace transformation* $F(s)$ uses the complex variable $s = \sigma + j\omega$ in place of $j\omega$ in Equation (2.1.21) with the lower limit of integration set equal to zero. The quantity $\sigma > 0$ is a convergence factor that exponentially damps out $f(t)$ as $t \rightarrow \infty$. This transform is used in applications where finite-power signals (like sinusoids) that are zero for negative values of time are needed. While the inversion of the Laplace transform is more complicated than Equation (2.1.22), tables of transform pairs are available for commonly used signals.

2.1.2f Impulse Signal

The *unit impulse* is a very important finite-energy signal that is denoted by $\delta(t)$ and defined by the constant spectral density $F(\omega) = 1$ and zero phase for all frequencies. Substituting $F(\omega) = 1$ in Equation (2.1.22) gives

$$\delta(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{j\omega t} d\omega \quad (2.1.23)$$

Because $e^{j\omega t} = \cos(\omega t) + j\sin(\omega t)$ and the odd function $\sin(\omega t)$ integrates to zero, Equation (2.1.23) can be interpreted as summing (integrating) unit-amplitude cosine waves of every frequency. When $t = 0$, $\cos(\omega t) = 1$ regardless of the frequency ω so that all the cosine waves reinforce one another, giving $\delta(t)$ infinite value. At times other than $t = 0$, the cosine waves destructively interfere and cancel, so that $\delta(t) = 0$ for $t \neq 0$. Consider, for example, the last example given in Figure 2.1.4 as the number of harmonics and length of the period are increased without limit. By using Equation (2.1.21) with $f(t) = \delta(t)$ and the fact that $F(\omega) = 1$

$$1 = \int_{-\infty}^{\infty} \delta(t) e^{-j\omega t} dt \quad (2.1.24)$$

which implies that the value of the integral is simply the value of the integrand, in this case $e^{-j\omega t}$ evaluated or sampled at $t = 0$. If the numerical value 1 is substituted for $e^{-j\omega t}$ in Equation (2.1.24), then the integral of $\delta(t)$ itself results, which also equals 1. By interpreting this latter integral as the area underneath $\delta(t)$, the entire contribution to the value of the integral must come when $t = 0$ because $\delta(t)$ is zero at other times. The unit impulse $\delta(t)$, also called a *delta*

function, belongs to a class of *singularity functions*. It has many different interpretations in terms of limits of sequences of functions, one of which is the limit of a rectangularly shaped pulse whose width shrinks to zero and whose height increases to maintain constant area under the pulse so that the integrated value always is unity. The dimensions of $\delta(t)$ are 1/time or frequency.

The delta function provides an interesting link between line spectra and spectral density because finite-power signals can be interpreted as delta functions (in frequency) in the spectral density. For example, the finite-power signal $A \cos(\omega_l t)$ has spectral density

$$F(\omega) = 2\pi \left[\frac{A}{2} \delta(\omega - \omega_l) + \frac{A}{2} \delta(\omega + \omega_l) \right] \quad (2.1.25)$$

This can be verified by inserting Equation (2.1.25) into Equation (2.1.22) and evaluating the integral by using the sampling property of the delta function. The result is

$$f(t) = \frac{A}{2} (e^{j\omega_l t} + e^{-j\omega_l t}) = A \cos(\omega_l t) \quad (2.1.26)$$

From this example, it is clear that spectral density is a more general concept than line spectra.

2.1.2g Power Spectrum

In Fourier analysis, the product $f(t)e^{-j\omega t}$ is integrated with respect to time to resolve or extract the frequency components, either discrete as in Equation (2.1.12) or continuous as in Equation (2.1.21), from $f(t)$. These frequency components also manifest themselves as a function of τ in the product $f(t)f(t + \tau)$, where τ is a time-shift parameter. For example, if $f(t) = A \cos(\omega_l t)$, this latter product is

$$A^2 \cos(\omega_l t) \cos(\omega_l t + \omega_l \tau) = \frac{A^2}{2} \cos(2\omega_l t + \omega_l \tau) + \frac{A^2}{2} \cos(\omega_l \tau) \quad (2.1.27)$$

For a fixed value of the parameter τ , the first term on the right of Equation (2.1.27) oscillates in time t at $2\omega_l$ and has constant phase angle $\omega_l \tau$ (it is a sinusoid), while the second term does not depend on time t (it is a constant). By defining $r(\tau)$ as the time average of the product

$$r(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} f(t)f(t + \tau) dt \quad (2.1.28)$$

and inserting Equation (2.1.27) into Equation (2.1.28), the sinusoid at $2\omega_l$ with fixed phase integrates to zero for every choice of τ , and the second term contributes to yield

$$r(\tau) = \frac{A^2}{2} \cos(\omega_l \tau) \quad (2.1.29)$$

2-20 The Audio Spectrum

The coefficient $A^2/2$ equals the power of $f(t) = A \cos(\omega_1 t)$, and the function $r(\tau)$ oscillates at the same frequency ω_1 but as a function of τ rather than t . Because $f(t)$ generally would contain a multitude of different frequencies, it is worthwhile to examine the spectral properties of $r(\tau)$. The Fourier transform of $r(\tau)$ in Equation (2.1.28), using the variable τ instead of t in Equation (2.1.21), is

$$P(\omega) = \int_{-\infty}^{\infty} r(\tau) e^{-j\omega\tau} d\tau \quad (2.1.30)$$

and defined as the *power spectrum* or *power spectral density* of the finite-power signal $f(t)$. In Equation (2.1.28), $r(\tau)$ is called the *autocorrelation function* of $f(t)$, and its value for $\tau = 0$ is the average power of $f(t)$ as defined in Equation (2.1.2). The autocorrelation function also equals the inverse Fourier transform of the power spectral density, which, from Equation (2.1.22), is

$$r(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} P(\omega) e^{j\omega\tau} d\omega \quad (2.1.31)$$

Similarly to Equations (2.1.25) and (2.1.26) for spectral density, the power spectral density associated with Equation (2.1.29) would consist of delta functions of strength $2\pi(A^2/4)$ at frequencies $\pm\omega_1$.

Equations (2.1.28) and (2.1.30) are very useful when spectral properties of a non-deterministic, finite-power signal $f(t) = n(t)$, arising, for example, from a noise process, are studied. By definition, no functional relationship exists between $n(t)$ and t . The signal has finite average power but does not, in general, consist of pure sinusoids at fixed frequencies with constant phase that are representable by line spectra. Therefore, Fourier-analysis methods are not directly applicable. Normally, the autocorrelation function $r(\tau)$ in Equation (2.1.28) is a finite-energy signal which first can be estimated from a sufficient number of samples of $f(t)$, then Fourier-transformed as in Equation (2.1.30). The result is a power spectral density as a function of continuous frequency. The autocorrelation function contains oscillations at all the frequencies contained in $f(t)$ with amplitudes equal to their respective powers, as illustrated by Equation (2.1.29) for a single frequency ω_1 , and the power spectral density integral defined in Equation (2.1.30) extracts, via Fourier transformation, the distribution of power in frequency $P(\omega)$ of the underlying noise process. The power spectra density has zero phase because $r(\tau)$ is always an even function of τ .

White noise is a term used to characterize a noise process whose power spectral density is constant or flat as a function of frequency. In contrast to an impulse, whose amplitude spectral density is flat and is interpreted as a sum of cosine waves of all frequencies each with identically zero phase, the phases of the similar cosine waves in white noise would vary randomly in time. The former is a finite-energy signal existing for one instant in time and the latter a finite-power signal existing for all time; both have continuous, flat spectra.

2.1.2h Analytic Signal

The actual waveshapes of certain signals, viewed in the time domain, appear to have identifiable amplitude-modulation (AM) and frequency-modulation (FM) effects that may vary as time progresses. This is not easily recognized in the signal's spectrum unless the modulation and sig-

nal to be modulated are simple, such as that generated by AM or FM between two sinusoids. The *analytic signal* is one procedure used to define AM or FM effects and extract quantitative information about them from the spectrum $f(t)$. The analytic signal $f_A(t)$ is a linear transformation of the signal $f(t)$ which also has an interesting spectral interpretation.

Consider the signal

$$\begin{aligned} f(t) &= \cos(\omega t) \\ &= \frac{1}{2}(e^{j\omega t} + e^{-j\omega t}) \end{aligned} \quad (2.1.32)$$

and from it form a second signal $\hat{f}(t)$ by leading the phase of $e^{j\omega t}$ by 90° and leading the phase of $e^{-j\omega t}$ by 90° , which corresponds to multiplication by j and $-j$, respectively. The second signal is, by using Equation (2.1.3c)

$$\begin{aligned} \hat{f}(t) &= \frac{1}{2}(je^{j\omega t} - je^{-j\omega t}) \\ &= -\sin(\omega t) \end{aligned} \quad (2.1.33)$$

The analytic signal associated with (t) is defined as [see Eq. (2.4)],

$$\begin{aligned} f_A(t) &= f(t) - j\hat{f}(t) \\ &= e^{j\omega t} \end{aligned} \quad (2.1.34)$$

The analytic signal is a complex function of time, and its (line) spectrum contains no negative-frequency components. Figure 2.1.6 illustrates, with phasors, the successive operations in forming the analytic signal for $\cos(\omega t)$.

The amplitude or *envelope* $E(t)$ and phase $\theta(t)$ of the analytic signal are defined by

$$E(t)e^{j\theta(t)} = f(t) - j\hat{f}(t) \quad (2.1.35)$$

so that, in terms of $f(t)$ and $\hat{f}^2(t)$, they are

$$E(t) = \sqrt{f^2(t) + \hat{f}^2(t)} \quad (2.1.36a)$$

$$\theta(t) = \arctan [-\hat{f}(t)/f(t)] \quad (2.1.36b)$$

and, from Equation (2.1.36b), the *instantaneous frequency* is defined as

$$\omega_i(t) = \frac{d\theta(t)}{dt} \quad (2.1.37)$$

2-22 The Audio Spectrum

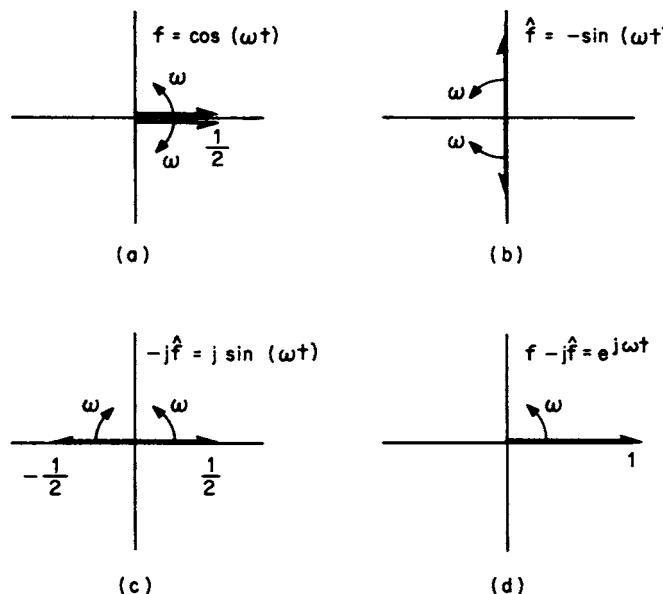


Figure 2.1.6 Formation of analytic signal $e^{j\omega t}$ of $\cos(\omega t)$: (a) counterrotating phasors representing $f = \cos(\omega t)$; (b) \hat{f} , the Hilbert transform of f ; (c) $-j$ times each phasor in b; (d) analytic signals given by sum of phasors in a and c.

In the example $f(t) = \cos(\omega t)$, the amplitude and instantaneous frequency of the associated analytic signal are 1 and ω respectively.

While the analytic signal is a linear transformation of the signal that, in effect, converts negative-frequency components in the spectrum of $f(t)$ to positive-frequency components, the envelope $E(t)$ and instantaneous frequency $\omega(t)$ are nonlinear functions of $f(t)$ and $\hat{f}(t)$. For more complicated signals, $\omega(t)$ is, in general, not the same as ω in the spectrum.

The spectral operation of advancing the phases of all positive-frequency components by 90° and advancing the phases of all negative-frequency components by 90° is called *Hilbert transformation*; that is, $\hat{f}(t)$ is the Hilbert transform of $f(t)$.

2.1.3 Bibliography

- Bracewell, R.: *The Fourier Integral and Its Applications*, McGraw-Hill, New York, N.Y., 1965.
- Childers, D. G.: *Modern Spectral Analysis*, IEEE, New York, N.Y., 1978.
- Connor, F. R.: *Signals*, Arnold, London, 1972.
- Jenkins, G. M., and D. G. Watts: *Spectral Analysis and Its Applications*, Holden-Day, San Francisco, Calif., 1968.

- Kharkevich, A. A.: *Spectra and Analysis*, English translation, Consultants Bureau, New York, N.Y., 1960.
- Lynn, P. A.: *An Introduction to the Analysis and Processing of Signals*, 2nd ed. Macmillan, London, 1982.
- Panter, P. F.: *Modulation, Noise and Spectral Analysis*, McGraw-Hill, New York, N.Y., 1965.
- Papoulis, A.: *The Fourier Integral and Its Applications*, McGraw-Hill, New York, N.Y., 1962.
- Rabiner, L. R., and C. M. Rader (eds.): *Digital Signal Processing*, IEEE, New York, N.Y., 1972.
- Schwartz, M.: *Information Transmission, Modulation and Noise*, McGraw-Hill, New York, N.Y., 1970.
- Westman, H. P. (ed.): *ITT Reference Data for Radio Engineers*, Howard W. Sams, New York, N.Y., 1973.

Chapter
2.2

Spectral Changes and Linear Distortion

Douglas Preis

2.2.1 Introduction

When a reproduced signal is not a replica of the original signal, it is *distorted*. Distortionless transmission of a time-varying signal through a system requires that the signal's shape be preserved. The two general mechanisms of signal distortion are *nonlinear* and *linear*.

2.2.2 Distortion Mechanisms

In a broad context, *nonlinear distortion* includes all forms of output-signal corruption that are not linearly related to (i.e., statistically linearly dependent on or correlated with) the input signal. *Modulation* (amplitude or frequency) of the signal (or even its time derivatives) by an imperfect system produces a certain amount of up conversion and down conversion of the signal's frequency components. For example, squaring or cubing of the signal resulting from a nonlinear transfer characteristic is a form of (self) amplitude modulation, whereas time-base errors, like speed variations, are equivalent to frequency modulation of the signal. These converted frequencies, like noise, are not linearly related to the input. The coherence function $\gamma^2(\omega)$ is a quantitative measure of the cumulative effect, at each frequency, of these various forms of signal corruption.

Linear distortion implies that even though the output signal is linearly related to the input, the shape of the output signal is different from that of the input signal. The system itself is linear and does process signals linearly (i.e., scale factors are preserved, and superposition is valid), but linear mathematical operations on the input signal such as differentiation or integration are permissible. Linear distortion changes the *relative* relationships among the existing constituents of the signal by altering either intensity or timing, or both, of its different frequency components. As a consequence, the output signal has a different shape. The system function (or complex frequency response) only predicts the spectral changes that the spectrum of the input signal will undergo and not the change of signal shape in time. The actual time-domain signal must be computed from direct convolution or inverse Fourier transformation.

When a single sinusoid is used as an input to a linear system, the corresponding steady-state output is also a sinusoid of the same frequency but, generally, with different amplitude and phase as prescribed by the complex frequency response. This single sinusoid is never linearly distorted

2-26 The Audio Spectrum

because it always is a replica of the input. In contrast, if the frequency content of the input signal is discrete (e.g., a square wave) or continuous (e.g., a single rectangular pulse), then linear distortion is observable as waveshape change. The extent to which an input signal will be linearly distorted depends on both its spectrum and the system function or complex frequency response of the linear system that processes it. Linear distortion encompasses what is sometimes called *transient distortion*, meaning the waveshape change of a finite-energy (transient) input signal (e.g., a short tone burst or pulse) under linear operating conditions. Finite-power input signals such as square waves, random noise, music, or speech also can be linearly distorted, however. Nonlinear effects such as clipping or slew-rate limiting of transient signals are a form of nonlinear distortion, which implies that transient distortion can be ambiguous terminology.

2.2.2a Linear Range

The range of linear operation of a system is usually established with sine-wave signals. The most frequently used procedure is to verify, using a single sine-wave input, whether the magnitude-scale factor is preserved as input amplitude is changed and/or to verify, using two sinusoids of different frequency simultaneously as an input, whether superposition is valid. Both procedures examine the linearity hypothesis under sinusoidal, steady-state operating conditions. Because the input spectrum is as narrow as possible (a line spectrum consisting of one or two discrete frequencies), the spectrum of the output can easily reveal the existence of other, or “new,” frequencies which would constitute nonlinear distortion corresponding to the chosen input signal. The ratio of total power contained in these other frequencies to the output power at the input frequency (or frequencies), expressed in percentage or decibels, is often used as both a measure and a specification of nonlinear distortion (e.g., harmonic distortion, intermodulation distortion, or *dynamic intermodulation distortion*, depending on the specific choice of input frequencies). This method of testing linearity is relatively simple and can be very sensitive. With modern, wide-dynamic-range, high-resolution spectrum analyzers and high-purity sine-wave generators, the effects of very small amounts of nonlinearity can be measured for sinusoidal steady-state operation.

Some aspects of this kind of spectral analysis are questionable, however. Incoherent (uncorrelated) power can exist at the test frequency itself (e.g., due to *cubic nonlinearity*, time-base errors, or noise), and, even more important, the system is never excited throughout its full operating bandwidth by the test signal. Because nonlinear effects do not superpose, the percentage of sine-wave nonlinear distortion measured, for example, as a function of test-signal frequency for fixed-output-power level, is not the same as the percentage of uncorrelated output power as a function of frequency for broadband operation with nonsinusoidal input signals. The latter can be expressed by using the coherence function to define a frequency-dependent signal-to-noise ratio as the ratio of coherent power to incoherent power, expressed in percentage or decibels.

2.2.2b Spectra Comparison

Comparison of input and output spectra of a system is worthwhile. Spectral changes can indicate, in the frequency domain, the existence of linear distortion as well as nonlinear distortion. For linear operation, the spectral changes that any input signal undergoes are predicted by the system function or complex frequency response. For nonlinear operation, the portion of the output sig-

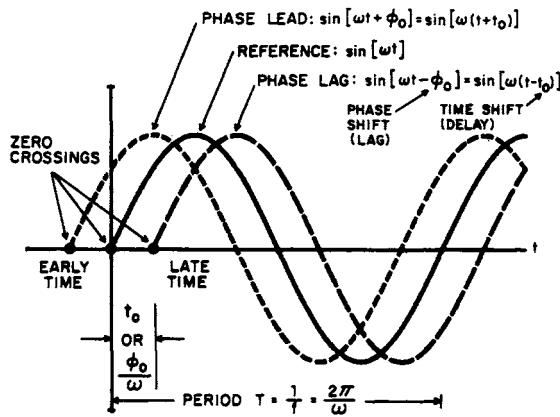


Figure 2.2.1 Standard conventions for sinusoids having steady-state phase shifts relative to a reference sinusoid (solid curve). The leading sinusoid (short dashes) has its first zero crossing at an earlier time than the reference, whereas the lagging sinusoid (long dashes) has its first zero crossing at a later time.

nal's spectrum not linearly related to the input signal depends, in detail, on each specific input signal and how it is nonlinearly processed by the system.

2.2.3 Sinusoidal Steady-State Measurements

The complex frequency response or system function

$$H(\omega) = |H(\omega)| e^{j\phi(\omega)} \quad (2.2.1)$$

of a linear system predicts the *magnitude* (or amplitude) $|H(\omega)|$ and *phase* $\phi(\omega)$ of an output sinusoid relative to an input sinusoid for steady-state operation at the frequency ω . These two frequency-domain quantities are fundamental and form the basis for the discussion of linear distortion of signals by a linear system.

The meaning of the *magnitude response* of a system is well understood and often displayed alone as the frequency response even though it is only the magnitude of $H(\omega)$. The phase-shift- $\phi(\omega)$ -versus-frequency characteristic is not commonly shown but is equally important. Some conventions, definitions, and properties of phase shift and the phase-shift characteristic merit discussion.

Figure 2.2.1 illustrates the convention for phase lead or lag between sinusoids of the same frequency relative to a reference. The reference sinusoid (solid line) has an upward-sloping zero crossing at zero time, and that leading (short dashes) has its corresponding zero crossing at an earlier time or is advanced in time, whereas that lagging (long dashes) has its corresponding zero crossing at a later time or is delayed in time.

2-28 The Audio Spectrum

Aside from this convention, Figure 2.2.1 has two ambiguities and raises a fundamental question about the nature of phase shift itself. First, a phase lead of π is indistinguishable from a phase lag of π ; also shifting any of these sinusoids by a multiple of $\pm 2\pi$ would not change the relative phase shifts, and consequently Figure 2.2.1 would not change. Thus, more information than Figure 2.2.1 shows is required to avoid ambiguity. Second, it appears that a phase shift and a time shift are equivalent, but they are not. This is so because $\phi_0 = \omega t_0$ and for a specific phase shift ϕ_0 the corresponding amount of time shift depends upon the sinusoid's frequency ω . To illustrate this point consider Figure 2.2.2a, wherein an approximate square wave is constructed from the first three nonzero harmonics of its Fourier series. Figures 2.2.2b and c show the difference between a constant phase shift for each harmonic and a constant time delay for each harmonic, respectively. The waveform of Figure 2.2.2b is severely linearly distorted, whereas that in Figure 2.2.2c is not. It is interesting to note that the amount of phase lag necessary to keep the waveform "together" is directly proportional to frequency (that is; the first, third, and fifth harmonics were lagged by $\pi/2$, $3\pi/2$, and $5\pi/2$, respectively). Therefore, in Figure 2.2.2c the amount of phase lag varies linearly with frequency, and this corresponds to a uniform time delay.

The fundamental question raised by Figure 2.2.1 is related to causality (i.e., cause and effect). How is it possible for a signal to be "ahead" of the reference signal in time, especially if the reference signal is the input or stimulus to a system and the phase shift of the output signal relative to the input is measured? Equivalently stated, how can the output-signal phase lead or be ahead of the input signal in time? Does this suggest that causality would permit only phase lags to occur in such a situation? It is indeed an interesting question in view of the fact that phase-shift measurements themselves can be somewhat ambiguous. The answer has to do with the fact that these are steady-state measurements, or more precisely, with how the steady state itself is achieved. Both phase lead or phase lag of a system output relative to its input are physically realizable without ambiguity or violation of causality. An example of each case is shown in Figure 2.2.3. Here the actual response of two different circuits to a tone-burst input was captured. In each case the input signal is also shown in time synchronization for reference purposes. It is during the initial transient state that either a phase lag or a phase lead is established. Note that in each case the output zero crossings are initially unequally spaced and that net phase lag or lead is gradually accumulated. Closer inspection of Figure 2.2.3 reveals that, from a mathematical viewpoint, the output in Figure 2.2.3a is approximately the integral with respect to time of the input, whereas in Figure 2.2.3b the output is nearly the derivative with respect to time of the input. In the steady state, the corresponding phase shifts are seen to be about $-\pi/2$ and $+\pi/2$, respectively. Circuits that have these properties are referred to as phase lag (integrators) or phase lead (differentiators). Another way of interpreting the results of Figure 2.2.3 then (referring also to Figure 2.2.1) is to note that the integral with respect to time of $\sin(\omega t)$ is

$$\frac{-1}{\omega} \cos(\omega t)$$

or

$$\frac{1}{\omega} \sin\left(\omega t - \frac{\pi}{2}\right)$$

whereas the time derivative of $\sin(\omega t)$ is $\omega \cos(\omega t)$ or

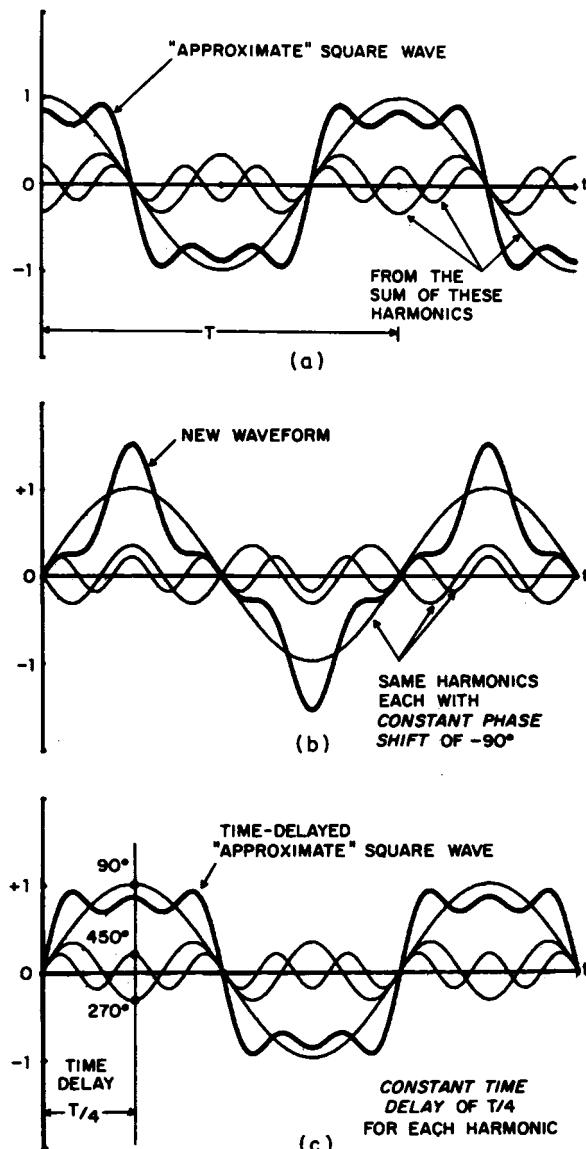


Figure 2.2.2 Characteristics of various waveforms: (a) an approximate square wave constructed from the first three (nonzero) Fourier harmonics, (b) constant phase shift for each harmonic yields a new waveform that is linearly distorted, (c) constant time delay for each harmonic uniformly delays the square wave while preserving its shape.

$$\omega \sin\left(\omega t + \frac{\pi}{2}\right)$$

2-30 The Audio Spectrum

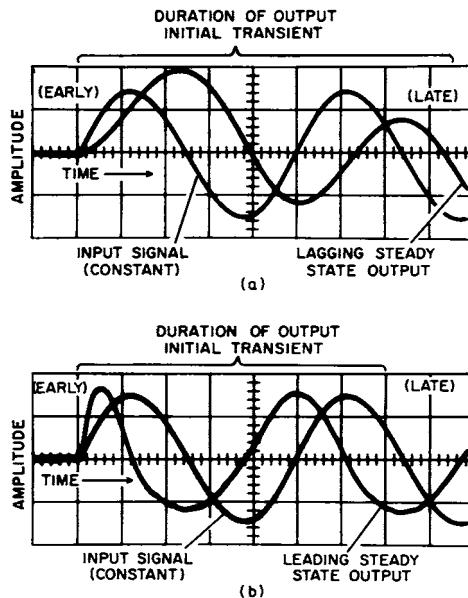


Figure 2.2.3 Steady-state phase lag or lead is established only after an initial transient period: (a) a phase lag of about 90° occurs, and the output signal is approximately the *integral* with respect to time of the input; (b) a phase lead of 90° occurs, and the output is nearly the time *derivative* of the input. The input sinusoid is the same in each case. The transient buildup of phase lag or lead illustrates that the phase shift of a system is remarkably a property of the steady state.

so that here the steady-state phase lag or lead of $\pi/2$ in the frequency domain corresponds to integration or differentiation, respectively, in the time domain.

An important conclusion to be drawn from Figure 2.2.3 is that the phase shift of a system is remarkably a property of the steady state.

Two significant aspects of a general phase-shift-versus-frequency characteristic $\phi(\omega)$ at a specific frequency ω_0 are its actual numerical value $\phi(\omega_0)$ (positive, zero, or negative) and its behavior in the vicinity of ω_0 (increasing, constant, or decreasing). The reason for this is that most useful signals passed through a system have finite spectral widths which are broad compared with that of any single frequency used to measure the phase characteristic itself.

The value of the phase shift at an arbitrary frequency that is close to a specific frequency ω_0 can be represented by

$$\phi(\omega) \approx \phi(\omega_0) + \text{correction terms} \quad (2.2.2)$$

and in Equation (2.2.2) the correction terms depend upon the difference $\Delta\omega = \omega - \omega_0$. To a first approximation and with reference to Figure 2.2.4

$$\phi(\omega) \approx \phi(\omega_0) + \Delta\phi \quad (2.2.3a)$$

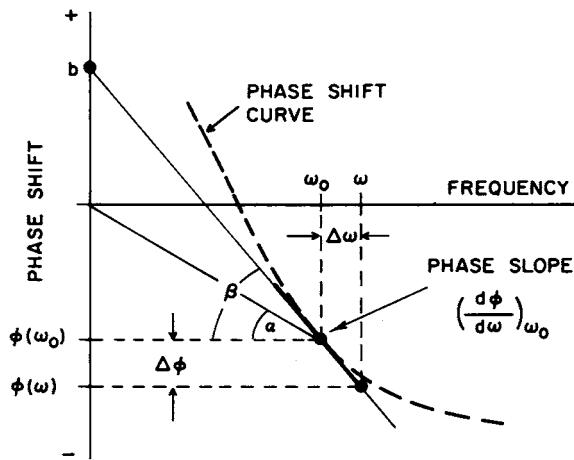


Figure 2.2.4 Arbitrary phase-shift-versus-frequency characteristic (long dashes). For frequencies ω near ω_0 the slope of the phase shift curve is nearly constant and can be approximated by the derivative $d\phi/d\omega$ evaluated at ω_0 . The numerical value of the phase slope (or derivative) indicates how the phase shift varies near ω_0 .

$$\phi(\omega) \approx \phi(\omega_0) + \frac{\Delta\phi}{\Delta\omega} \Delta\omega \quad (2.2.3b)$$

$$\phi(\omega) \approx \phi(\omega_0) + \left(\frac{d\phi}{d\omega} \right)_{\omega_0} [\omega - \omega_0] \quad (2.2.3c)$$

where

$$\left(\frac{d\phi}{d\omega} \right)_{\omega_0}$$

is the derivative of the phase shift at ω_0 or, equivalently, its slope there. If $\omega = \omega_0$, Equation (2.2.3) is exact, and when ω is near ω_0 , it is, in general, approximate. In this approximation it is the slope or first derivative of the phase characteristic at ω_0 that describes the behavior of $\phi(\omega)$ near ω_0 . Thus, for signals whose spectra lie in the neighborhood of ω_0 both

$$\phi(\omega_0) \text{ and } \left(\frac{d\phi}{d\omega} \right)_{\omega_0}$$

are important because $\phi(\omega_0)$ gives the steady-state absolute phase shift of the output relative to the input at ω_0 , whereas

2-32 The Audio Spectrum

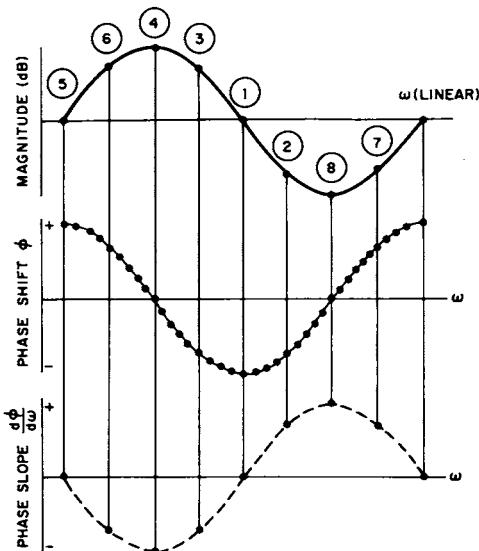


Figure 2.2.5 Possible relationships among magnitude, phase, and phase slope versus linear frequency for a general minimum-phase system. The numbered vertical-line segments each connect a different combination of these three quantities (based on algebraic signs) and indicate important test frequency ranges where transient response may differ.

$$\left(\frac{d\phi}{d\omega}\right)_{\omega_0} [\omega - \omega_0]$$

expresses the phase shift at ω relative to that at ω_0 .

Because no restrictions have been placed upon the behavior of $\phi(\omega)$ or its derivative, each may assume positive, zero, or negative values throughout the frequency range of interest. There are, therefore, several different possible combination pairs of phase shift and phase slope. Each of these presumably would affect the response to a transient signal differently. A preliminary discussion of the influence of these aspects of phase response on transient signals is given in the next section together with supporting experimental measurements.

2.2.3a Some Effects of Frequency Response on Transient Signals

An important class of linear systems used in a variety of applications, called *minimum-phase* systems, has mathematically interrelated magnitude and phase responses. These responses are not independent of one another. Specifying one determines the other.

Figure 2.2.5 summarizes the possible relationships between magnitude response, phase shift, and phase slope versus linear frequency for a general minimum-phase system. For example, a narrow bandpass system could have a magnitude response characteristic like that from 5-6-4-3-1, and the associated phase-shift and phase-slope characteristics would be the corresponding portions of those curves below. In a very wideband system the distance between points 5 and 1 on the frequency axis would be considerably greater and the curve 6-4-3 much flatter; also the corresponding sections of the phase and phase-slope curves would require appropriate modification. In a similar way, 1-2-8-7 could represent the magnitude response for a narrowband-reject filter which would have the appropriately corresponding phase and phase-slope curves as shown.

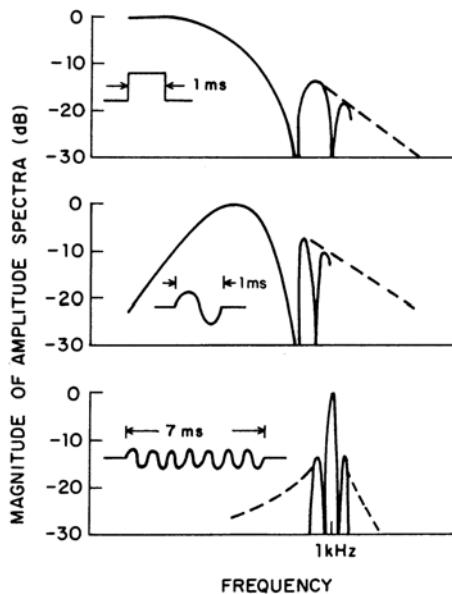


Figure 2.2.6 Magnitude of the spectral density in decibels versus frequency for test signals used to evaluate the response of a general minimum-phase system to transient signals. The dashed lines indicate the upper bound for the continuing spectral peaks.

The steady-state characteristics of the minimum-phase system shown in Figure 2.2.5 are also interconnected by eight numbered vertical lines. These numbers indicate important test points (or frequency ranges) where phase shift and phase slope are, in various combinations, positive, zero, and negative. Experimentally measured effects of each of these eight different combinations of $\phi(\omega)$ and $d\phi/d\omega$ on three different transient signals, illustrated along with their spectra in Figure 2.2.6, are shown in Figure 2.2.7. The three test signals were a 1-ms-wide rectangular pulse, a single cycle of a 1-kHz sine wave, and a 7-ms tone burst at 1 kHz.

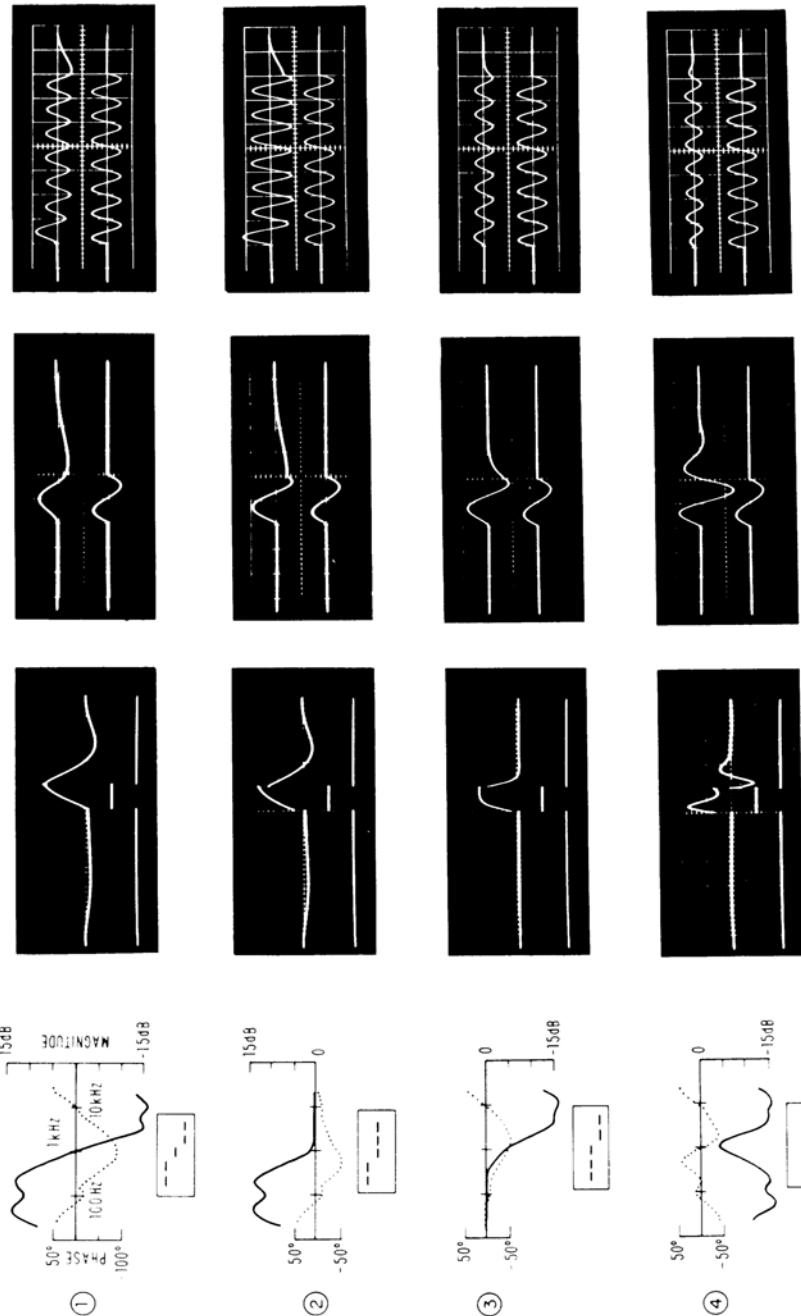
The eight different combinations of phase shift and phase slope in Figure 2.2.4 were simulated, in each case, near 1 kHz by using a five-band graphic equalizer as the minimum-phase system. In each of these eight cases, the linear distortion of transient signals is quite different. Here, the influence of complex frequency response on the system response to transient signals can be explained, qualitatively, in terms of the algebraic signs of phase shift and phase slope. Table 2.2.1 categorizes the eight test cases on this basis. Tables 2.2.2 and 2.2.3 summarize some general effects which the algebraic sign of the phase shift and the phase slope (at 1 kHz) has on response to transient signals.

For the tone bursts, it appears that the actual numerical value of ϕ influences the “inner” structure of the waveform, whereas $d\phi/d\omega$ mostly affects the envelope, or “outer” structure. This is consistent with the approximation in Equation. (2.2.3c)

$$\phi(\omega) = \phi(\omega_0) + \left(\frac{d\phi}{d\omega} \right)_{\omega_0} (\omega - \omega_0) \text{ with } \omega_0 / 2\pi = 1 \text{ kHz}$$

Positive values for both ϕ and $d\phi/d\omega$ imply that ϕ is positive and increasing near ω_0 , and this results in sharp and abrupt transient behavior. If ϕ and $d\phi/d\omega$ are both negative, then ϕ is

2-34 The Audio Spectrum



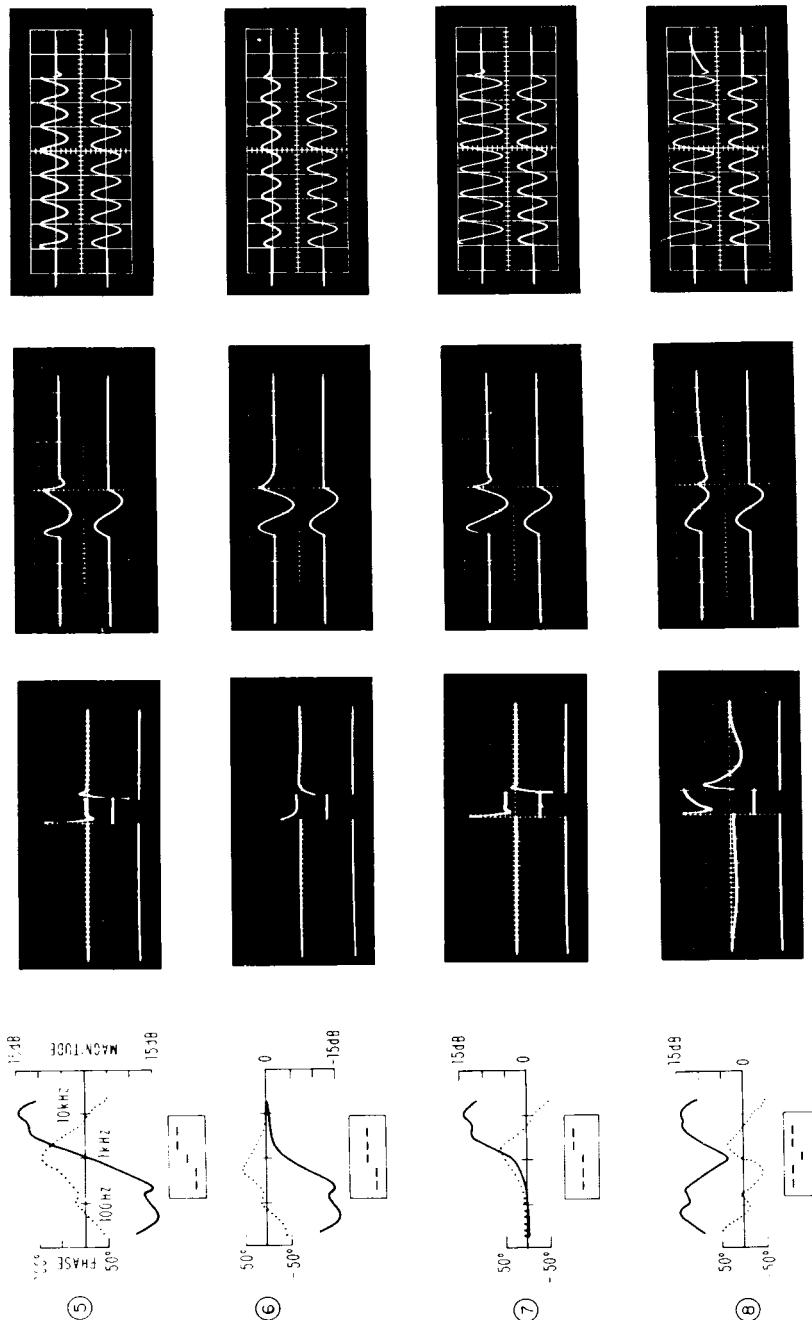


Figure 2.2.7 Comparison of steady-state magnitude and phase response measurements (of a five-band graphic equalizer) with responses to the transient signals whose spectra are given in Figure 2.2.6. These eight cases correspond to the eight test points for the minimum-phase system in Figure 2.2.5 normalized to a frequency range near approximately 1 kHz. (System output and input signals are shown in the upper and lower traces, respectively.)

2-36 The Audio Spectrum**Table 2.2.1 Categories of Test Cases**

		$\frac{d\phi}{d\omega}$ at 1 kHz		
		+	0	-
ϕ at 1 kHz	+	7	5	6
	0	8	*	4
	-	2	1	3

Table 2.2.2 Phase Shift Effects

ϕ (phase shift)	Effect
+	Totally or partially differentiates waveform Phase lead within first few cycles Cusplike transients
0	No phase shift at 1 kHz Tone-burst envelope shape subject to change
-	Totally or partially integrates waveform Phase lag within first few cycles Smooth transients

Table 2.2.3 Phase Shift Effects

$\frac{d\phi}{d\omega}$ (phase slope)	Effect
+	Very rapid tone-burst buildup Overshoot Some long decays
0	Little or no effect Mostly differentiation or integration due to $\phi \leq 0$
-	Slower tone-burst buildup Slight elongation of tone burst Smooth decays

becoming more negative, and this yields smoother, slower transients. When the phase shift and phase slope have opposite signs, these opposing effects are seen to combine.

Perhaps the most interesting case occurs when $\phi(\omega_0) = 0$ but $[d\phi/d\omega]_{\omega_0} \neq 0$ (see Figure 2.2.7, cases 4 and 8). For the very-narrow-spectrum tone burst there is no steady-state phase shift as expected; however, the response to the wide-spectrum rectangular pulse in these two cases is remarkable. These results may be interpreted in the following way. Because the rectangular pulse

has a considerably wider spectrum than the tone burst, it has many spectral components both above and below $\omega_0/2\pi = 1\text{kHz}$. In case 8 $[d\phi/d\omega]_{\omega_0} > 0$ and from the approximation (2.2.3c)

$$\phi(\omega) = [d\phi/d\omega]_{\omega_0} (\omega - \omega_0)$$

so that for $\omega > \omega_0$, $\phi(\omega)$ is positive, whereas for $\omega < \omega_0$, $\phi(\omega)$ is negative. Therefore, and in simple terms, higher frequencies tend to be differentiated in time, whereas lower ones are integrated. This is seen to occur. Just the opposite occurs in case 4 because $[d\phi/d\omega]_{\omega_0} < 0$.

From the foregoing examples it is clear that the response of a system to transient signals depends on its frequency response. For minimum-phase systems, the phase and derivative of phase with respect to frequency can be used to interpret, qualitatively, important aspects of linear distortion of signals in the time domain.

2.2.4 Phase Delay and Group Delay

Phase delay and group delay are useful quantities related to the phase shift $\phi(\omega)$ and defined as

$$\tau_p(\omega) = -\phi(\omega)/\omega \quad (2.2.4)$$

and

$$\tau_g(\omega) = -\frac{d\phi(\omega)}{d\omega} \quad (2.2.5)$$

respectively. The negative signs are required because, according to the conventions for sinusoids in Figure 2.2.1, negative values of phase shift correspond to positive time delays. At a specific frequency ω_0 , these two quantities are constants in the two-term Taylor-series expansion of $\phi(\omega)$. Equation (2.2.3c), valid near and at ω_0 , which can be rewritten as

$$\phi(\omega) \equiv -\omega_0 \tau_p(\omega_0) - \tau_g(\omega_0)[\omega - \omega_0] \quad (2.2.6)$$

Equation (2.2.6) restates the fact that the phase shift at ω is equal to the phase shift at ω_0 plus the phase shift at ω relative to ω_0 . The steady-state phase shift for the components of a narrowband signal near ω_0 is given by Equation (2.2.6), and the effect of the two terms in this equation can be interpreted in the following way. First, each component in the band undergoes a fixed phase shift

$$-\omega_0 \tau_p(\omega_0) = \phi(\omega_0)$$

then those components at frequencies different from ω_0 are subjected to additional phase shift $-\tau_g(\omega_0)[\omega - \omega_0]$. This additional phase shift is one which varies linearly with frequency, so it does not alter the waveshape (see Figure 2.2.2).

2-38 The Audio Spectrum

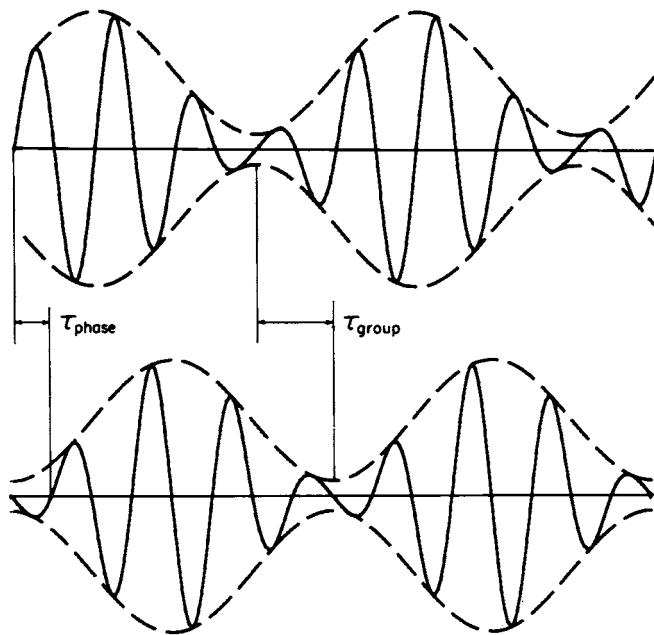


Figure 2.2.8 The difference between phase delay τ_{phase} and group delay τ_{group} is illustrated by comparing these two amplitude-modulated waveforms. The lower waveform has positive phase delay and positive group delay relative to the upper waveform. Because the envelope of the high-frequency oscillation is delayed by an amount of time τ_{group} group delay is sometimes referred to as *envelope delay*.

An amplitude-modulated (AM) sinusoid is a narrowband signal, and the effects of phase delay and group delay on such a signal are illustrated in Figure 2.2.8. Phase delay phase-lags (delays) the high-frequency carrier (inner structure), while the envelope (outer structure) is delayed by an amount equal to the group delay.

Geometrically, Figure 2.2.4 shows that

$$\tau_g(\omega_0) = \tan\beta \quad \text{and} \quad \tau_p(\omega_0) = \tan\alpha$$

Note that $\tau_g = \tau_p$ only when $\alpha = \beta$ and the intercept $b = 0$. In this special case $\phi(0) = 0$ and $\phi(\omega)$ varies linearly as a function of ω (that is, the entire phase-shift characteristic is a straight line which passes through the origin having slope $-\tau_g$).

Generally both τ_p and τ_g can assume positive, zero, or negative values depending upon the detailed behavior of the phase-shift characteristic. Referring to the special minimum-phase system in Figure 2.2.5 and in view of definition (2.2.5), the group-delay characteristic is the negative of the phase slope, and therefore its shape is the same as the magnitude characteristic for this special case. It is also interesting to note from the same figure that for the bandpass system 5-6-4-3-1

$$\tau_p(\omega) = -\phi(\omega)/\omega$$

can be positive, negative, or zero, but

$$\tau_g(\omega) = -d\phi/d\omega \geq 0$$

For the system 1-2-8-7, $\tau_g(\omega) \leq 0$.

The approximate nature of Equation (2.2.6) deserves particular emphasis because either is valid only over a narrow range of frequencies $\Delta\omega$ and outside this range the correction terms mentioned in Equation (2.2.2) contain higher-order derivatives in the Taylor series that generally cannot be neglected.

2.2.4a Distortionless Processing of Signals

In the time domain, the requirement for *distortionless* linear signal processing (i.e., no wave-shape change) is that the system impulse response $h(t)$ have the form

$$h(t) = K\delta(t-T) \quad (2.2.7a)$$

where $\delta(t)$ is the unit impulse, and the constants $K > 0$ and $T \geq 0$. Equation (2.2.7a) and the convolution theorem together imply that the output signal $g(t)$ is related to the input $f(t)$ by

$$g(t) = Kf(t-T) \quad (2.2.7b)$$

The distortionless system scales any input signal by a constant factor K and delays the signal as a whole by T seconds. The output is a delayed replica of the input. Through substitution, Equation (2.2.7b) gives the corresponding restrictions on the frequency response, namely

$$H(\omega) = Ke^{-j\omega T} \quad (2.2.8)$$

Comparison of Equation (2.2.8) with Equation (2.2.1) indicates that the frequency-domain requirements are twofold: constant magnitude response $|H(\omega)| = K$ and phase response proportional to frequency $\phi(\omega) = \omega T$. Waveform distortion or *linear distortion* is caused by deviations of $|H(\omega)|$ from a constant value K as well as departures of $\phi(\omega)$ from the linearly decreasing characteristic $-\omega T$. The former is called *amplitude distortion* and the latter *phase distortion*. From Equation (2.2.8) absence of phase distortion requires that the phase and group delays in Equations (2.2.4) and (2.2.5) each equal the overall time delay $T \geq 0$

$$\tau_p(\omega) = \tau_g(\omega) = T \quad (2.2.9)$$

Some experimentally measured effects of the deviations of $|H(\omega)|$ and $\tau_g(\omega)$ from a constant value are illustrated in Figure 2.15. In the experiment, four bandpass filters were connected in cascade to give the attenuation magnitude (reciprocal of gain magnitude) and group-delay characteristics shown in Figure 2.2.9a. The group delay is reasonably flat at midband, having a

2-40 The Audio Spectrum

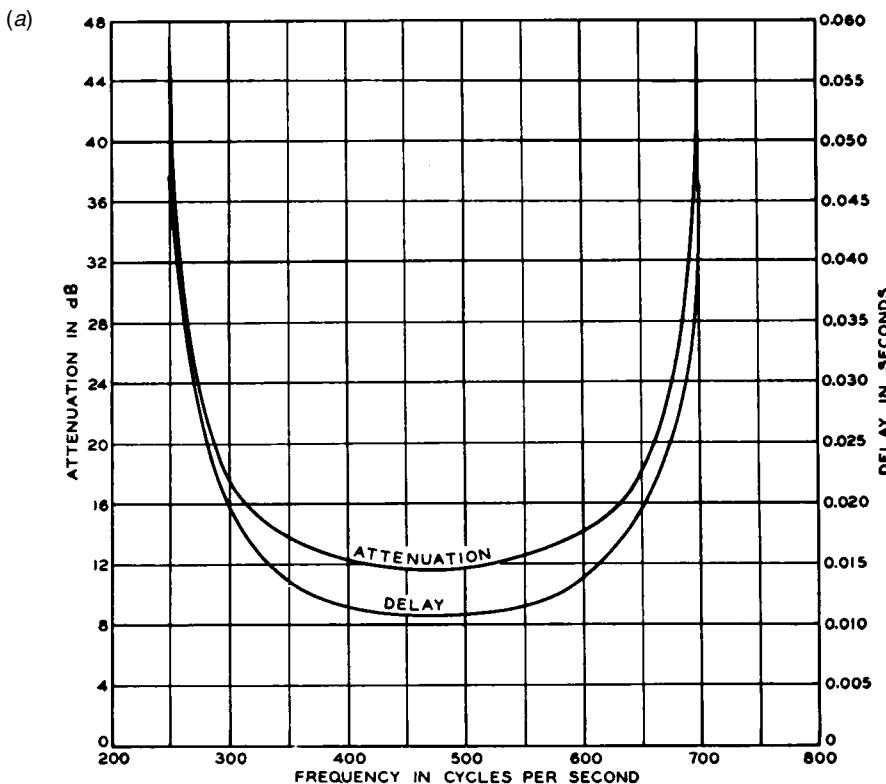
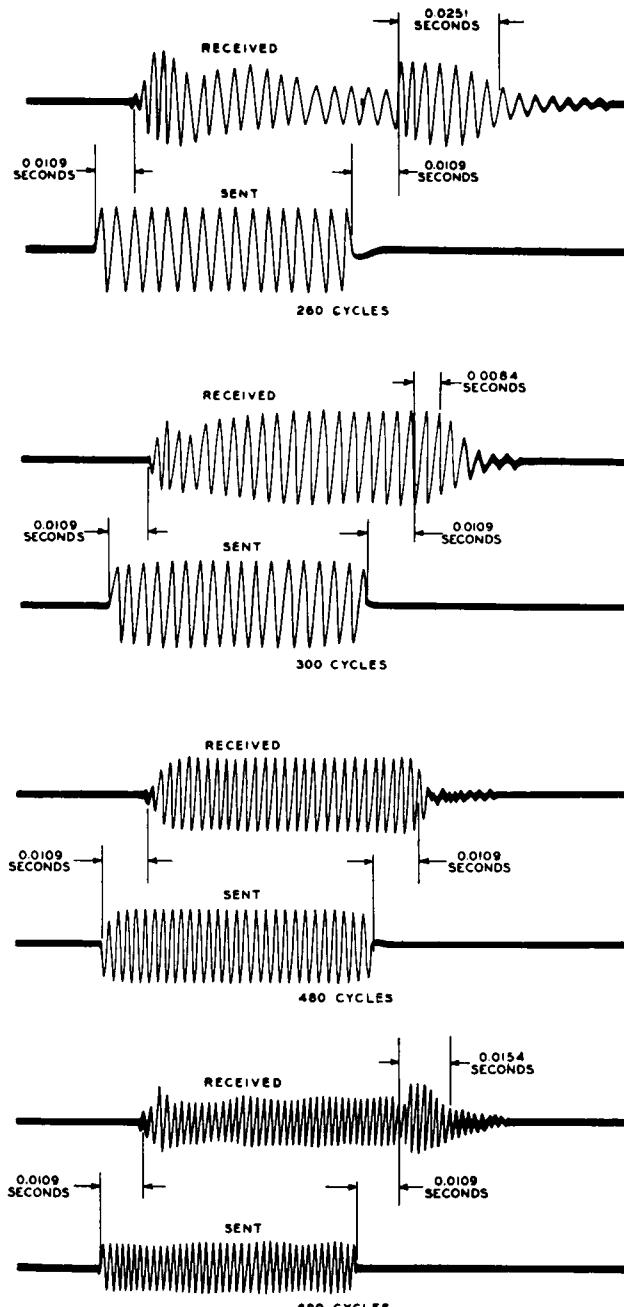


Figure 2.2.9 Attenuation and group delay characteristics: (a) attention (reciprocal of gain) magnitude in decibels and group delay in seconds versus frequency characteristics of four cascaded bandpass filters; (b) experimentally measured responses to transient input signal (tone burst) of filtering network whose steady-state characteristics are shown in a. For this bandpass system, each output signal is delayed by the minimum value of the group delay. When the tone-burst spectrum lies near either passband edge, significant amounts of linear distortion occur in the form of waveform elongation. This is due, for the most part, to the departure of the group-delay characteristic from its flat value in the midband and is called *group-delay distortion*.

minimum value there of $\tau_g = 10.9 \text{ ms}$. Near and at the passband edges τ_g deviates considerably from its minimum value. The effects in the time domain are shown in Figure 2.2.9b. Here, tone bursts at frequencies of 260, 300, 480, and 680 Hz were applied to the filter, and both input and output oscillographs were obtained. In each of these cases, the oscillations start to build up after a time equal to the minimum value of τ_g . There is significant linear distortion for the tone bursts whose spectra lie at the passband edges. Some of this distortion can be ascribed to nonconstant attenuation, but the waveform elongation is primarily due to the group delay $\tau_g(\omega)$ deviating from its minimum value.

These experimental results indicate that, for distortionless processing of signals, the band of frequencies throughout which both magnitude response and group delay of the system are con-

(b)

**Figure 2.2.9** Continued

2-42 The Audio Spectrum

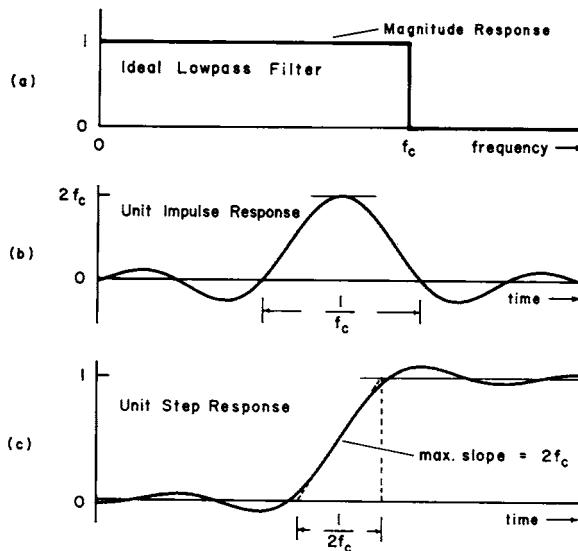


Figure 2.2.10 Response characteristics: (a) magnitude response; (b) impulse response; (c) step response of an ideal linear-phase, band-limited (low-pass) system with cutoff frequency f_c Hz.

stant or flat is important *relative* to the spectral bandwidth of signals to be processed by the system.

2.2.4b Linear Phase and Minimum Phase

The impulse response of a (distortionless) unity-gain, *linear-phase*, band-limited (low-pass) system is symmetrical in time about its central peak value and described mathematically by

$$h(t) = \frac{\sin(\omega_c t)}{\pi t} \quad (2.2.10)$$

where $\omega_c/2\pi = f_c$ is the cutoff frequency in hertz. The magnitude response is $|H(\omega)| = 1$, and the phase response $\phi(\omega) = 0$ (a special case of linear phase). This result can be interpreted as a cophase superposition of cosine waves up to frequency ω_c . In general, the group delay for such a linear-phase system is constant but otherwise arbitrary; that is, $\tau_g(\omega) = T$ s because the phase shift $\phi(\omega) = -\omega T$ is linear but can have arbitrary slope. Figure 2.10b illustrates $h(t)$ in Equation (2.2.10) delayed, shifted to the right in time, so that its peak value occurs at, say, a positive time $t = T$ rather than $t = 0$. Regardless of the value of T , the impulse response is not causal; that is, it will have finite values for negative time. So if an impulse excitation $\delta(t)$ were applied to the system when $t = 0$, the response to that impulse would exist for negative time. Such *anticipatory transients* violate cause (stimulus) and effect (response). In practice, a causal approximation to the ideal $h(t)$ in Equation (2.2.10) is realized by introducing sufficient delay T and truncating or

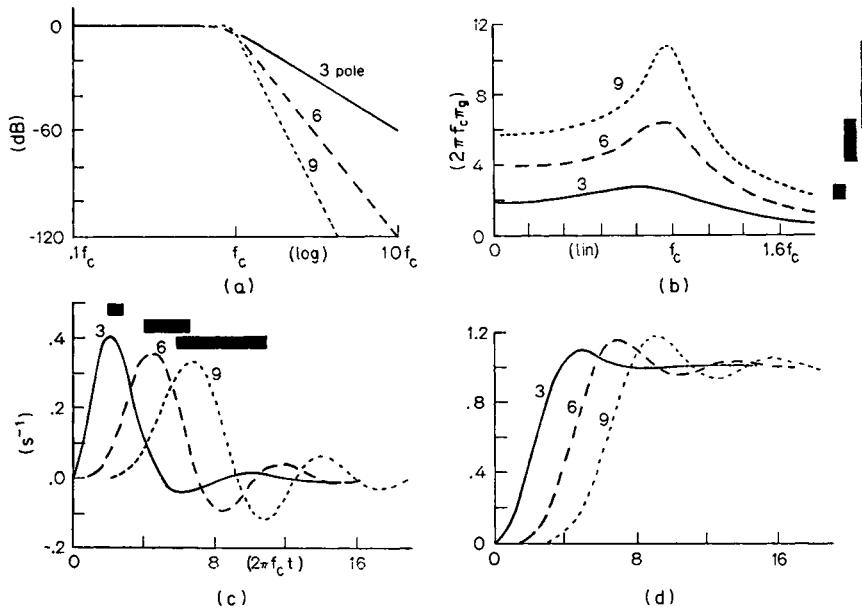


Figure 2.2.11 Response of maximally flat, minimum-phase, low-pass systems: (a) magnitude, (b) group delay, (c) impulse, (d) step.

“windowing” the response so that it is zero for negative time. This latter process would produce ripples in the magnitude response, however. Also shown in Figure 2.2.10 are the corresponding magnitude response (a) and step response (c), which is the integral with respect to time of $h(t)$. In principle, this ideal system would not linearly distort signals whose spectra are zero for $\omega > \omega_c$. In practice, only approximations to this ideal response are realizable.

A minimum-phase system is causal and has the least amount of phase shift possible corresponding to its specific magnitude response $|H(\omega)|$. The phase is given by the (Hilbert-transform) relationship

$$\phi_m(\omega) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\ln|H(\omega')|}{\omega' - \omega} d\omega' \quad (2.2.11)$$

and the minimum-phase group delay associated with Equation (2.2.11) is

$$\tau_{gm}(\omega) = - \frac{d\phi_m(\omega)}{d\omega} \quad (2.2.12)$$

While minimum-phase systems have impulse responses that are zero for negative time (causal), they are not distortionless. Because magnitude and phase responses are interrelated, the linear distortion they introduce can often be interpreted by using group delay—Equation (2.2.12).

2-44 The Audio Spectrum

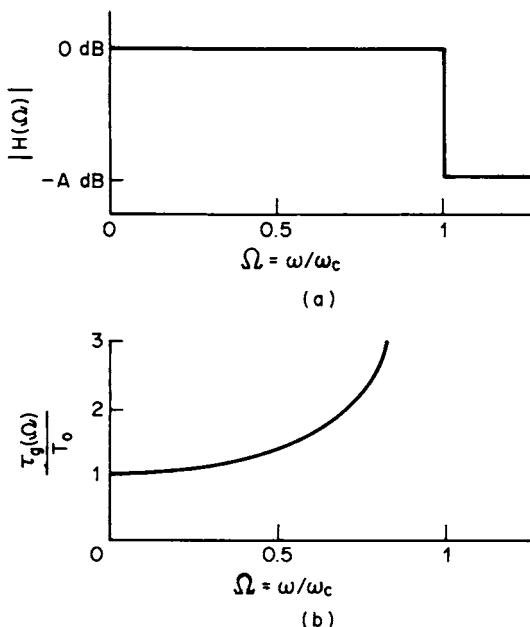


Figure 2.2.12 Magnitude and group delay characteristics: (a) magnitude response of a minimum-phase low-pass filter with stop-band attenuation of $-A$ db versus normalized frequency $\Omega = \omega/\omega_c$, (b) normalized group delay of minimum-phase filter in a.

In contrast to the preceding example, consider approximations to band limiting using realizable minimum-phase, maximally flat, low-pass systems of successively higher order. The frequency-domain and time-domain responses for three-, six-, and nine-pole systems are plotted in normalized form in Figure 2.2.11. Here the impulse responses are causal but not symmetrical. The loss of symmetry is due to group-delay distortion. The group delay changes as frequency increases, and this implies that phase shift is not proportional to frequency, especially near the cutoff frequency f_c . In this example, deviations of τ_g from its low-frequency value are a measure of group-delay distortion. Note that the maximum deviation of τ_g , indicated by the length of the solid vertical bars in Figure 2.2.11b, is quantitatively related to the broadening of the impulse response, while the low-frequency value of τ_g predicts the arrival time of the main portion of the impulse response, as indicated by the position and length of the corresponding solid horizontal bars in Figure 2.2.11c. Actually, the low-frequency value of τ_g equals the time delay of the center of gravity of the impulse response. By increasing the rate of attenuation above f_c , the overall delay of the impulse response increases, initial buildup is slower, ringing is more pronounced, and the response becomes more dispersed and less symmetrical in time.

The minimum-phase group delay $\tau_{gm}(\omega)$ can be evaluated, in theory, for an arbitrary magnitude response $|H(\omega)|$ by using Equations (2.2.11) and (2.2.12). Consider, for example, an interesting limiting case of the previous maximally flat low-pass system where an ideal “brick wall” magnitude response is assumed, as shown in Figure 2.2.12. Here the system has unity gain below the cutoff ω_c , and A dB of attenuation above ω_c . The normalized frequency $\Omega = \omega/\omega_c$. Although

this magnitude response cannot be realized exactly, it could be approximated closely with an elliptic filter. The group delay is

$$\tau_g(\Omega) = \frac{T_0}{1 - \Omega^2} \quad (2.2.13)$$

where the constant

$$T_0 = \frac{A \ln 10}{10\pi\omega_c} \quad (2.2.14)$$

As predicted by Equation (2.2.13) and seen in Figure 2.2.12, the group delay becomes infinitely large at the band edge where $\omega = \omega_c$. This is a consequence of the assumed rectangular magnitude response. As a specific numerical example, let $A = 80$ dB and $\omega_c/2\pi = 14$ kHz. Then $T_0 = 62$ μ s and $\tau_g(\omega) = 0.5$ ms when $\omega/(2\pi) \approx 14$ kHz. It is interesting to note that demanding greater (stop-band) attenuation for $\omega > \omega_c$, requires a larger value for the attenuation parameter A , and $\tau_g(0)$ increases, as does the deviation of the group delay within the passband.

A minimum-phase system is also a minimum-delay system since it has the least amount of phase change for a given magnitude response. (The group delay equals the negative rate of change of phase with respect to frequency.) Thus, signal energy is released as fast as is physically possible without violating causality. However, in doing so, certain frequency components are released sooner than others, and this constitutes a form of phase distortion, sometimes called *dispersion*. Its presence is indicated by deviations of the group delay from a constant value. A linear-phase system necessarily introduces greater delay than a minimum-phase system with the same magnitude response. However, it has the advantage that there is no dispersion. This is accomplished by delaying all frequency components by the same amount of time.

2.2.5 Bandwidth and Rise Time

An important parameter associated with a band-limited low-pass system is the *rise time*. As illustrated in Figures 2.2.10 and 2.2.11, eliminating high frequencies broadens signals in time and reduces transition times. The *step response* (response of the system to an input that changes from 0 to 1 when $t = 0$) shows, in the case of perfect band limiting illustrated in Figure 2.2.10, that the transition from 0 to 1 requires $\pi/\omega_c = 1/(2f_c)$ s. Defining this as the rise time gives the useful result that the product of bandwidth in hertz and rise time in seconds is $f_c \times 1/(2f_c) = 0.5$. For example, with $f_c = 20$ kHz the rise time is 25 μ s. The rise time equals half of the period of the cutoff frequency (for this perfectly band-limited system).

Practical low-pass systems, such as the minimum-phase systems shown in Figure 2.2.11, do not have a sharp cutoff frequency, nor do they have perfectly flat group delay like the ideal low-pass model in Figure 2.2.10. The product of the -3-dB bandwidth and the rise time for real systems usually lies within the range of 0.3 to 0.45. The reason that the rise time is somewhat shorter (faster) is twofold. Because the cutoff is more gradual, some frequencies beyond the -3-dB point contribute to the total output response. Also, the rise time to a unit step is conventionally defined as the time for the output to change from 10 to 90 percent of its final value.

2-46 The Audio Spectrum

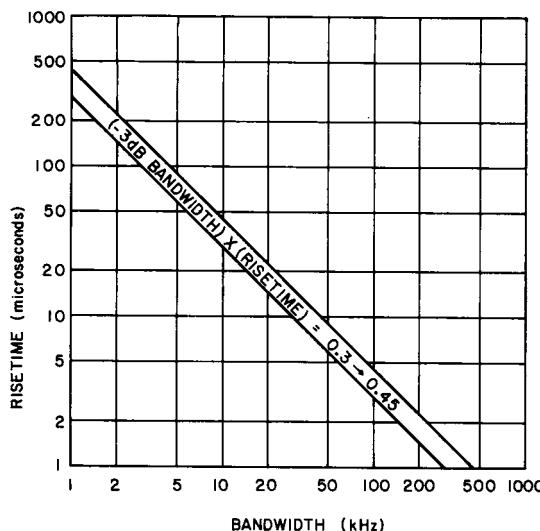


Figure 2.2.13 Relationship between rise time and bandwidth of practical linear systems. For a given bandwidth, the rise time will lie within the tolerance strip shown; conversely, the bandwidth requirements for a specific rise time also can be found.

Figure 2.2.13 displays this important relationship between rise time and bandwidth. Given the -3-dB bandwidth of a system, the corresponding range of typically expected rise times can be read. Conversely, knowing the rise time directly indicates nominal bandwidth requirements. Within the tolerance strip shown in Figure 2.2.13, fixing the bandwidth always determines the rise time, and conversely. This figure is a useful guide that relates a frequency-domain measurement (bandwidth) to a time-domain measurement (rise time). This fundamental relationship suggests that testing a practical band-limited linear system with signals having rise times significantly shorter than the rise time of the linear system itself cannot yield new information about its transient response. In fact, the system may not be able to process such signals linearly.

The *slew rate* of a system is the maximum time rate of change (output units/time) for large-signal nonlinear operation when the output is required to change between extreme minimum and maximum values. It is not the same as rise time, which is a parameter defined for linear operation.

2.2.5a Echo Distortion

Ripples in the magnitude and/or phase of the system function $H(\omega)$ produce an interesting form of linear distortion of pulse signals called *echo distortion*. Assuming that these ripples are small and sinusoidal, a model for the system function

$$H(\omega) = |H(\omega)| e^{j\phi(\omega)}$$

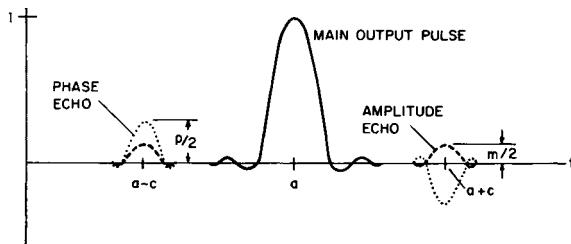


Figure 2.2.14 Small preechoes and postechoes are produced at the output of a linear system having linear distortion in response to a pulse-like input. The main output pulse is delayed by a seconds (minimum value of τ_g) and undistorted. Nonflat magnitude response produces symmetrical (dashed curves) "amplitude" echoes, whereas group-delay distortion produces unsymmetrical (dotted curves) "phase" echoes. These echoes are the linear distortion. In minimum-phase systems, the noncasual echoes at $t = a - c$ are equal and opposite and cancel one another. In practical systems, the echoes may overlap and change the shape of (linearly distort) the main output pulse.

is

$$|H(\omega)| = 1 - m \cos(\omega c) \quad (2.2.15a)$$

$$\phi(\omega) = -p \sin(\omega c) - \omega a \quad (2.2.15b)$$

where c is the number of ripples per unit bandwidth, in hertz, and m and p are the maximum values of the magnitude and phase ripples, respectively. If $m = p = 0$, then there is no linear distortion of signals, just unity gain, and uniform time delay of $T = \tau_g(\omega) = a$ due to the linear-phase term $-\omega a$. By using Fourier-transform methods, it can be shown that the output $g(t)$ corresponding to an arbitrary input $f(t)$ has the form

$$g(t) = f(t - a) + \frac{m - p}{2} f(t - a - c) + \frac{m + p}{2} f(t - a + c) \quad (2.2.16)$$

Equation (2.2.16) states that the main portion of the output signal is delayed by a seconds and is undistorted, but there are, in addition, small preechoes and postechoes (replicas) which flank it, being advanced and delayed in time (relative to $t = a$) by c seconds. This is shown in Figure 2.2.14. Amplitude echoes are symmetrical (+ + or --), but phase echoes are unsymmetrical (+ - or -+). These echoes are the linearly distorted portion of the output and are called echo distortion. The detection of linear distortion by observing paired echoes is possible when the echoes do not overlap and combine with the undistorted part of the signal to form a new (and linearly distorted) waveshape that may be unsymmetrical and have a shifted peak time.

In connection with minimum-phase systems, if the magnitude response varies in frequency as a cosine function, then the phase response varies as a negative sine function (see Figure 2.2.5 beginning at point 4) as the Hilbert-transform relationship—Equation (2.2.11)—would predict. (Also, the group delay varies, like the magnitude response, as a cosine function.) This result

2-48 The Audio Spectrum

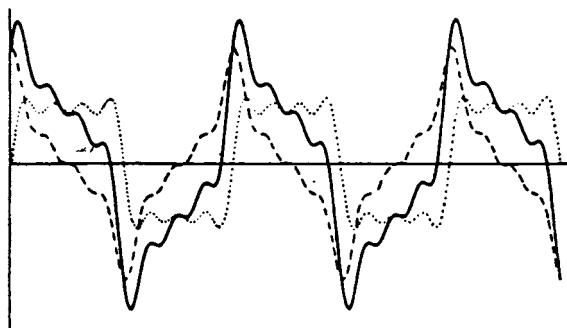


Figure 2.2.15 Band-limited square wave (dotted curve), its Hilbert transform (dashed curve), and the sum of dotted and dashed curves (solid curve).

implies that m and p in Equation (2.2.15) would be equal and opposite; that is, $m = -p$. In this case the preecho vanishes because the last term in Equation (2.2.16) is zero, but the postechoes reinforce. The impulse responses of many minimum-phase systems can be interpreted on this basis.

2.2.5b Classifications of Phase Distortion

When a system is causal, the minimum amount of phase shift $\phi_m(\omega)$ that it can have is prescribed by the Hilbert-transform relation, Equation (2.2.11). There can be additional or *excess* phase shift $\phi_x(\omega)$ as well, so that in general the total phase shift is the sum

$$\phi(\omega) = \phi_m(\omega) + \phi_x(\omega) \quad (2.2.17)$$

A practical definition for the excess phase is

$$\phi_x(\omega) = \theta_a(\omega) - (\omega T + \theta_0) \quad (2.2.18)$$

where θ_0 is a constant and $\theta_a(0) = 0$. In Equation (2.2.18) $-\omega T$ represents pure time delay, $\theta_a(\omega)$ is the frequency-dependent phase shift of an all-pass filter, and θ_0 represents a frequency-independent phase shift caused by, for example, polarity reversal between input and output or a Hilbert transformer which introduces a constant phase shift for all frequencies. The group delay, defined in Equation (2.2.5), is found by substituting Equation (2.2.18) into Equation (2.2.17) and differentiating. The result is

$$\tau_g(\omega) = T - \frac{d\phi_m(\omega)}{d\omega} - \frac{d\theta_a(\omega)}{d\omega} \quad (2.2.19a)$$

$$= T + \tau_{gm}(\omega) + \tau_{ga}(\omega) \quad (2.2.19b)$$

Because deviations of group delay from the constant value T indicate the presence of phase distortion, *group-delay distortion* is defined as

$$\Delta\tau_g(\omega) = \tau_{gm}(\omega) + \tau_{ga}(\omega) \quad (2.2.20)$$

This definition implies that $\Delta\tau_g(\omega) = 0$ is a *necessary* condition for no phase distortion and, furthermore, that the peak-to-peak excursions of $\Delta\tau_g(\omega)$ are both a useful indication and a quantitative measure of phase distortion. Although the all-pass group delay $\tau_{ga}(\omega) \geq 0$, the minimum-phase group delay $\tau_{mp}(\omega)$ can be negative, zero, or positive (as can be inferred from the phase responses in Figure 2.2.7 by examining their negative derivatives).

Note that when $\tau_g(\omega)$ is calculated from $\phi(\omega)$ by using Equation (2.2.5), only phase-slope information is preserved. The phase intercept

$$\phi(0) = \phi_m(0) + \theta_0$$

is lost through differentiation. This result implies that when $\Delta\tau_g(\omega) = 0$ in Equation (2.2.20), some phase distortion is possible if, for example

$$\phi(\omega) = \phi_m(\omega) = \frac{-\pi}{2}$$

[$H(\omega)$ is an ideal integrator] or

$$\phi(0) = \theta_0 = \frac{\pi}{2}$$

[$H(\omega)$ contains a Hilbert transformer]. Thus $\Delta\tau_g(\omega) = 0$ and $\phi(0) \neq 0$ (or a multiple of π) implies no group-delay distortion but a form of phase distortion known as *phase-intercept distortion*. With reference to Figure 2.2.4, the phase intercept b is zero when the phase delay and group delay are equal, as stated in Equation (2.2.9), which is the *sufficient* condition for no phase distortion. Generally, the total phase distortion produced by a linear system consists of both group-delay and phase-intercept distortion.

Figure 2.2.15 illustrates phase distortion caused by a frequency-independent phase shift or phase-intercept distortion. The dotted curve represents a band-limited square wave (sum of the first four nonzero harmonics), and the dashed curve is the Hilbert transform of the square wave obtained by shifting the phase of each harmonic $\pi/2$ rad, or 90° . This constant phase shift of each harmonic yields a linearly distorted waveshape having a significantly greater peak factor, as shown. The solid curve is the sum of the square wave and its Hilbert transform. Because corresponding harmonics in this sum are of equal amplitude and in phase quadrature, the solid curve could have been obtained by scaling the magnitude of the amplitude spectrum of the original square wave by $\sqrt{2}$ and rotating its phase spectrum by 45° . For this example

$$\phi_x(\omega) = \theta_0 = \pi/4 \text{ rad}$$

in Equation (2.2.18).

2-50 The Audio Spectrum

In summary, there are two classifications of phase distortion: 1) group-delay distortion, which is due to the minimum-phase response and/or the frequency-dependent all-pass portion of the excess phase response; and 2) phase-intercept distortion, which is caused by a fixed or constant (frequency-independent) phase shift for all frequencies.

2.2.6 Bibliography

- Bendat, J. S., and A. G. Riersol: *Engineering Applications of Correlation and Spectral Analysis*, Wiley, New York, 1980.
- Bendat, J. S., and A. G. Piersol: *Random Data: Analysis and Measurement Procedures*, Wiley-Interscience, New York, N.Y., 1971.
- Blinchikoff, H. J., and A. I. Zverev: *Filtering in the Time and Frequency Domains*, Wiley, New York, N.Y., 1976.
- Bloom, P. J., and D. Preis: "Perceptual Identification and Discrimination of Phase Distortions," *IEEE ICASSP Proc.*, pp. 1396–1399, April 1983.
- Bode, H. W.: *Network Analysis and Feedback Amplifier Design*, Van Nostrand, New York, N.Y., 1945.
- Cheng, D. K.: *Analysis of Linear Systems*, Addison-Wesley, Reading, Mass., 1961.
- Deer, J. A., P. J. Bloom, and D. Preis: "Perception of Phase Distortion in All-Pass Filters," *J. Audio Eng. Soc.*, vol. 33, no. 10, pp. 782–786, October 1985.
- Di Toro, M. J.: "Phase and Amplitude Distortion in Linear Networks," *Proc. IRE*, vol. 36, pp. 24–36, January 1948.
- Guillemin, E. A.: *Communication Networks*, vol. 11, Wiley, New York, N.Y., 1935.
- Henderson, K. W., and W. H. Kautz: "Transient Response of Conventional Filters," *IRE Trans. Circuit Theory*, CT-5, pp. 333–347, December 1958.
- Hewlett-Packard: "Application Note 63—Section II, Appendix A, 'Table of Important Transforms,'" Hewlett-Packard, Palo Alto, Calif, pp. 37, 38, 1954.
- Kupfmuller, K.: *Die Systemtheorie der elektrischen Nachrichtenübertragung*, S. Hirzel Verlag, Stuttgart, 1968.
- Lane, C. E.: "Phase Distortion in Telephone Apparatus," *Bell Syst. Tech. J.*, vol. 9, pp. 493–521, July 1930.
- Lathi, B. P.: *Signals, Systems and Communications*, Wiley, New York, N.Y., 1965.
- Mallinson, J. C.: "Tutorial Review of Magnetic Recording." *Proc. IEEE*, vol. 62, pp. 196–208, February 1976.
- Members of the Technical Staff of Bell Telephone Laboratories: *Transmission Systems for Communications*, 4th ed., Western Electric Company, Technical Publications, Winston-Salem, N.C., 1971.
- Oppenheim, A. V., and R. W. Schafer: *Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, N.J., 1975.

- Papoulis, A.: *Signal Analysis*, McGraw-Hill, New York, N.Y., 1977.
- Peus, S.: "Microphones and Transients," *db Mag.*, translated from *Radio Mentor* by S. Temmer, vol. 11, pp. 35–38, May 1977.
- Preis, D.: "A Catalog of Frequency and Transient Responses," *J. Audio Eng. Soc.*, vol. 25, no. 12, pp. 990–1007, December 1977.
- Pries, D.: "Audio Signal Processing with Transversal Filters," *IEEE Conf. Proc.*, 1979 ICASSP, pp. 310–313, April 1979.
- Pries, D.: "Hilbert-Transformer Side-Chain Phase Equalizer for Analogue Magnetic Recording," *Electron. Lett.*, vol. 13, pp. 616–617, September 1977.
- Pries, D.: "Impulse Testing and Peak Clipping," *J. Audio Eng. Soc.*, vol. 25, no. 1, pp. 2–14, January 1977.
- Pries, D.: "Least-Squares Time-Domain Deconvolution for Transversal-Filter Equalizers," *Electron. Lett.*, vol. 13, no. 12, pp. 356–357, June 1977.
- Pries, D.: "Linear Distortion," *J. Audio Eng. Soc.*, vol. 24, no. 5, pp. 346–367, June 1976.
- Pries, D.: "Measures and Perception of Phase Distortion in Electroacoustical Systems," *IEEE Conf. Proc.*, 1980 ICASSP, pp. 490–493, 1980.
- Pries, D.: "Phase Equalization for Analogue Magnetic Recorders by Transversal Filtering," *Electron. Lett.*, vol. 13, pp. 127–128, March 1977.
- Pries, D.: "Phase Equalization for Magnetic Recording," *IEEE Conf. Proc.*, 1981 ICASSP, pp. 790–795, March 1981.
- Pries, D.: "Phase Distortion and Phase Equalization in Audio Signal Processing—A Tutorial Review," *J. Audio Eng. Soc.*, vol. 30, no. 11, pp. 774–794, November 1982.
- Pries, D., and C. Bunks: "Three Algorithms for the Design of Transversal-Filter Equalizers," *Proc. 1981 IEEE Int. Symp. Circuits Sys.*, pp. 536–539, 1981.
- Pries, D., and P. J. Bloom: "Perception of Phase Distortion in Anti-Alias Filters," *J. Audio Eng. Soc.*, vol. 32, no. 11, pp. 842–848, November 1984.
- Pries, D., F. Hlawatsch, P. J. Bloom, and J. A. Deer: "Wigner Distribution Analysis of Filters with Perceptible Phase Distortion," *J. Audio Eng. Soc.*, December 1987.
- Small, R. H.: "Closed-Box Loudspeaker Systems, Part 1: Analysis," *J. Audio Eng. Soc.*, vol. 20, pp. 798–808, December 1972.
- Totzek, U., and D. Press: "How to Measure and Interpret Coherence Loss in Magnetic Recording," *J. Audio Eng. Soc.*, December 1987.
- Totzek, U., D. Preis, and J. F. Boebme: "A Spectral Model for Time-Base Distortions and Magnetic Recording," *Archiv. fur Elektronik und Ubertragungstechnik*, vol. 41, no. 4, pp. 223–231, July-August 1987.
- Wheeler, H. A.: "The Interpretation of Amplitude and Phase Distortion in Terms of Paired Echoes," *Proc. IRE*, vol. 27, pp. 359–385, June 1939.
- Williams, A. B.: *Active Filter Design*, Artech House, Dedham, Mass., 1975.

2-52 The Audio Spectrum

Zverev, A. I.: *Handbook of Filter Synthesis*, Wiley, New York, N.Y., 1967.

Section
3

Architectural Acoustic Principles and Design Techniques

Each person relates to sound in a unique way that depends not only on the individual's perception but also on the context of the sound. Audiences seek the best sound quality available. Those outside the audience often find that other people's sound is noise. Thus there is a need for quality sound as well as for isolation from another sound. A general introduction to the concepts is given in this section.

As with other engineering applications, the objective is to assess the potential acoustical problems in advance and engineer accordingly. Acoustical solutions that are applied after the fact are compromises at best, limited mostly by cost. Doing the job right the first time is less expensive and avoids loss of revenue during retrofit. For example, selecting the correct floor construction for preventing sound from traveling to an adjacent floor will avoid the difficult application of sound-barrier construction to the floor or ceiling soon after the building has been commissioned.

This section serves as a useful introduction to architectural acoustics, encouraging further reading. For those who do not require extensive knowledge in this field, this section will help communication with architects, engineers, and acoustical consultants.

If assistance is needed in acoustical design, various resources are available. Sales representatives for building materials may be able to help, but one should be prepared for narrow and occasionally inappropriate advice on single-product application. More extensive help may be obtained from active members of related professional societies, such as the Audio Engineering Society or the Acoustical Society of America. Eight years of experience specifically with noise control and a rigorous examination are requisites for membership in the Institute of Noise Control Engineering. There is also a professional group, the National Council of Acoustical Consultants, that can provide a directory of members.

In This Section: _____

Chapter 3.1: The Physical Nature of Sound	3-5
Introduction	3-5
The Hearing Process	3-5

3-2 Section Three

Computer Design of Acoustic Systems	3-5
Sound Characteristics	3-6
Sound Spectrum	3-7
Propagation	3-7
Sound Power	3-8
Directivity	3-8
Sound Buildup Within a Space	3-9
Frictional Absorbers	3-11
Resonant Panels	3-11
Cavity Resonators	3-12
Reverberation	3-12
Combination of Direct and Reverberant Sound	3-13
Composite Transmission Loss	3-16
Sound Transmission Class	3-17
Diffraction	3-20
References	3-21
Chapter 3.2: Criteria for Acceptability of Acoustical Performance	3-23
Introduction	3-23
Reverberation Time	3-23
Background Noise	3-24
Maximum Levels	3-26
Interference of Speech Communication	3-27
Exterior Noise	3-27
Mechanical Systems	3-27
Sound Generation by Fans	3-27
Turbulent Noise in Ducts	3-28
Attenuation of Noise by Ducts	3-30
Duct Silencers	3-33
Calculating Resultant Sound Levels	3-34
References	3-34
Chapter 3.3: Sound Isolation	3-35
Introduction	3-35
Sound Barriers	3-35
Partial Walls and Barriers	3-36
Doors	3-36
Ceilings and Floors	3-37
Floating Rooms	3-37
Windows	3-37
Duct Break-In Break-Out	3-37
Site Selection	3-38
Vibration	3-39
Driving Frequency	3-39
Vibration Transmission	3-40
Vibration Isolation	3-41
Interior Acoustics	3-41
Design Concerns of Spaces	3-45

Architectural Acoustic Principles and Design Techniques 3-3

Small Rooms	3-46
Masking	3-47
Annoyance	3-48
References	3-49
Bibliography	3-49

Reference Documents for this Section: _____

- ANSI: *American National Standard for Rating Noise with Respect to Speech Interference*, ANSI S3.14-1977, American National Standards Institute, New York, N.Y., 1977.
- ANSI: *Method for the Measurement of Monosyllabic Word Intelligibility*, ANSI S3.2-1960, rev. 1977, American National Standards Institute, New York, N.Y., 1976.
- ASA Standards Index 2*, Acoustical Society of America, New York, N.Y., 1980.
- ASHRAE: *ASHRAE Handbook—1984 Systems*, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, Ga., 1984.7.
- Beranek, L. L.: *Acoustics*, McGraw-Hill, New York, N.Y., 1954.
- Beranek, L. L.: *Noise and Vibration Control*, McGraw-Hill, New York, N.Y., 1971.
- Catalogue of STC and IIC Ratings for Wall and Floor/Ceiling Assemblies*, Office of Noise Control, Berkeley, Calif.
- Egan, M. D.: *Concepts in Architectural Acoustics*, McGraw-Hill, New York, N.Y., 1972.
- Huntington, W. C., R. A. Mickadeit, and W. Cavanaugh: *Building Construction Materials*, 5th ed., Wiley, New York, N.Y., 1981.
- Jones, Robert S.: *Noise and Vibration Control in Buildings*, McGraw-Hill, New York, N.Y., 1980.
- Kryter, K. D.: *The Effects of Noise on Man*, Academic, New York, N.Y., 1985.
- Lyon R. H., and R. G. Cann: *Acoustical Scale Modeling*, Grozier Technical Systems, Inc., Brookline, Mass.
- Marris, Cyril M.: *Handbook of Noise Control*, 2nd ed., McGraw-Hill, New York, N.Y., 1979.
- Marshall, Harold, and M. Barron: "Spatial Impression Due to Early Lateral Reflections in Concert Halls: The Derivation of the Physical Measure," *JSV*, vol. 77, no. 2, pp. 211–232, 1981.
- Morse, P. M.: *Vibration and Sound*, American Institute of Physics, New York, N.Y., 1981.
- Siebein, Gary W.: *Project Design Phase Analysis Techniques for Predicting the Acoustical Qualities of Buildings*, research report to the National Science Foundation, grant CEE8307948, Florida Architecture and Building Research Center, Gainesville, Fla., 1986.
- Talaske, Richard H., Ewart A. Wetherill, and William J. Cavanaugh (eds.): *Halls for Music Performance Two Decades of Experience, 1962-1982*, American Institute of Physics for the Acoustical Society of America, New York, N.Y., 1982.

Chapter
3.1
The Physical Nature of Sound

Richard G. Cann, Anthony Hoover

3.1.1 Introduction

This chapter encompasses the following basic aspects of sound propagation: the propagation of sound through air, the reflection of sound from a wall, and the transmission of sound through a wall. Also considered are the absorption of sound by materials, the criteria for desirable and/or undesirable sound, and methods for both improving the quality of desirable sound and reducing the impact of undesirable sound.

3.1.1a The Hearing Process

Although each listener is unique, there are bounds within which most listeners fall. Thus, over the years, standards have been developed for both measurement instrumentation and measurement procedures. The Acoustical Society of America publishes, on behalf of the American National Standards Institute (ANSI), a catalog that summarizes each standard. There is also an index that lists international standards [1]. In addition, there are many U.S. trade and professional societies that publish standards that relate to noise and their specific activities.

The manner in which the ear perceives sound is exceedingly complex. In some ways the ear is more sensitive to sound than acoustical instrumentation, being able to detect sound qualities that are extremely difficult to quantify. However, the hearing process may also interpret tonal sounds that in fact do not physically exist.

3.1.1b Computer Design of Acoustic Systems

A variety of computer programs are available to assist the designer concerned with architectural acoustics. Some programs are supplied either gratis or for a small fee by the manufacturers of building components. In addition, some programs may be purchased for generic applications. However, caution should be exercised when using these programs, for calculations performed by the computer may not be documented thoroughly and may not suit a particular application or on-the-job condition. Thus, it is possible to apply the programs improperly, resulting in substantial error.

3-6 Architectural Acoustic Principles and Design Techniques

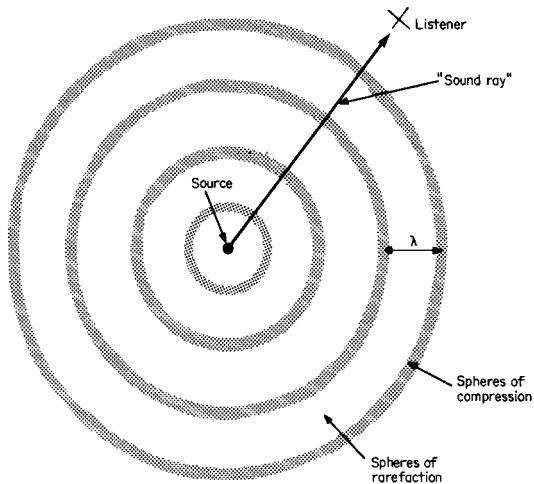


Figure 3.1.1 Simple sound source.

3.1.2 Sound Characteristics

The simplest source of sound expands and contracts equally in all directions as if a perfectly round balloon were rapidly inflated and deflated. The expansion and contraction of the source results in three-dimensional sound ripples that spread out unimpeded in all directions as ever-expanding spheres of compression and rarefaction at the velocity of sound. The rate at which the point source expands and contracts is the *frequency* in cycles per second, usually expressed numerically in hertz (Hz). The distance between consecutive spheres of either expansion or compression is identified as the *wavelength*, as shown in Figure 3.1.1. These three parameters are related by

$$c = f\lambda \quad (3.1.1)$$

where

c = velocity of sound, ft/s (m/s)

f = frequency, Hz

λ = wavelength, ft/Hz (m/Hz)

The speed of sound in air is approximately 1130 ft/s at normal room temperatures. For quick estimates, this may be rounded off to 1000 ft/s. For design surveys, it may be more convenient to use a simplification that sound travels about 1 ft/0.001 s.

Sound waves of all frequencies, whether from a low-frequency *woofer* or a high-frequency *tweeter*, travel at the same speed. An international standard (International Organization for Standardization, Recommendation R226, 1961) sets middle A (so-called tuning A) at 440 Hz. From Equation (3.1.1) this tone has a wavelength of 2.59 ft.

Table 3.1.1 Limits of Frequency Passbands

Band number	Nominal center frequency	One-third-octave passband	Octave passband
14	25	22.4–28.2	
15	31.5	28.2–35.5	22.4–44.7
16	40	35.5–44.7	
17	50	44.7–56.2	
18	63	56.2–70.8	44.7–89.1
19	80	70.8–89.1	
20	100	89.1–112	
21	125	112–141	89.1–178
22	160	141–178	
23	200	178–224	
24	250	224–282	178–355
25	315	282–355	
26	400	355–447	
27	500	447–562	355–708
28	630	562–708	
29	800	708–891	
30	1,000	891–1,120	708–1410
31	1,250	1,120–1,410	
32	1,600	1,410–1,780	
33	2,000	1,780–2,240	1,410–2,820
34	2,500	2,240–2,820	
35	3,150	2,820–3,550	
36	4,000	3,550–4,470	2,820–5,620
37	5,000	4,470–5,620	
38	6,300	5,620–7,080	
39	8,000	7,080–8,910	5,620–11,200
40	10,000	8,910–11,200	

3.1.2a Sound Spectrum

The audible spectrum of sound ranges from 20 Hz to 20 kHz. The fundamental tone or *pitch* of musical instruments ranges from piano at the lowest end of human hearing to about 4 kHz. However, every instrument also develops harmonics that are frequencies many times higher than the fundamental pitch. These harmonics are important in our ability to identify types of musical instruments.

For noise control applications, the frequency spectrum is conveniently divided into preferred octave bands, the frequencies of which are shown in Table 3.1.1. Each octave band encompasses the musical scale from F sharp to F. All noise control data are classified into these octave bands. One-third-octave bands may be used for more detailed work.

Propagation

Expanding sound waves are sometimes depicted in acoustical diagrams by using sound rays such as those shown in Figure 3.1.1. These rays are lines that are used to represent the radius of the spherical wave and are arrowed in a direction away from the source. They must not be interpreted as meaning “beams” of sound that travel only in the arrowed direction. Neither do they describe in any way the amplitude of the wave at any point. Their utility is limited to showing primary

3-8 Architectural Acoustic Principles and Design Techniques

sound propagation paths in environments that are dimensionally larger than the wavelength of the sound. This limitation is explained in more detail later in this section.

Sound waves travel away from a simple source in spheres of ever-increasing diameter. The sound pressure is reduced in amplitude by a factor of 4 each time that the radius is doubled, since the sound energy is distributed over the sphere's surface, which has become 4 times larger. In decibel terms, the new sound level is decreased by 20 log (ratio of distances).

Thus, when the radius or the distance that a sound wave travels has doubled, the sound level is reduced by 20 log (2), or 6 dB. Conversely, each time that a listener's distance from the source is halved, the sound level increases by 6 dB. This is not true once a listener is close to the source.

Most speaker cabinets have dimensions of less than 1 m; this is typically the minimum distance at which the rule of 6 dB per distance doubling can be applied. At a distance of less than 1 m the sound level increases asymptotically to a maximum value at the vibrating surface.

Sound Power

Because sound pressure and sound power levels are usually expressed in decibels, a logarithmic ratio, it is important to distinguish clearly between the two. *Sound power level* applies only to the source, whereas *sound pressure level* is also dependent on the environment and the distance from the source.

As an analogy, a common light bulb is rated in lumens to indicate how much light the bulb produces regardless of the kind of room it is in. But the amount of light perceived by an observer depends on such environmental factors as the distance from the bulb to the eye and the color of the wallpaper. Sound power level cannot be measured directly but is calculated from measurement of sound pressure level made with a sound-level meter.

Sound power is calculated from

$$L_w = L_p - 10 \log\left(\frac{Q}{4\pi r^2}\right) - 10.2 \quad (3.1.2a)$$

Where:

r = radius, ft

Q = directivity factor

$$L_w = L_p - 10 \log\left(\frac{Q}{4\pi r^2}\right) \quad (3.1.2b)$$

where r = radius, m.

3.1.2b Directivity

Most sound sources are not omnidirectional like the one described in the previous section. Instead, they emit sound more strongly in one direction than in another. The directivity characteristic can be specified by means of a *directivity factor*. If an omnidirectional source is placed against a large reflecting surface such as a floor, the sound will radiate only into a hemisphere, or half of the previous solid angle. The directivity factor Q of this source increases from 1 to 2. If the solid angle is again halved by another large plane, such as by placing the source on a floor

next to a wall, the directivity factor now increases to 4. When a source is placed in the corner of a rectangular room, the sound can radiate only into one-eighth of a sphere; so the directivity factor is now 8.

Loudspeakers and microphones also show directional characteristics. Their characteristics are usually given by the manufacturer in the form of graphical polar plots which compare the sound pressure level in all directions with that of the on-axis sound.

3.1.3 Sound Buildup Within a Space

When the spherical wavefront meets a large flat surface, some sound is reflected as a mirror image of the spherical wave with the angle of incidence of the wave equal to the angle of reflection. For most surfaces, sound is not totally reflected; some is absorbed. Regardless of the mechanism of absorption, the effectiveness of a surface material in reducing sound is given by its *absorption coefficient*. This is the fraction of the incident sound energy that is absorbed, with a value between 0 and 1. For example, if 25 percent of the sound is absorbed, then the coefficient is 0.25. The larger the coefficient, the more effective the absorber. Sound absorbers usually have different absorption coefficients at different frequencies. Examples of the performance of different materials are shown in Table 3.1.2.

It is to be noted that the coefficient of a highly effective absorber may be given as fractionally greater than 1. This is not an error but the result of the method used in testing the material.

In Equation (3.1.3) the sound absorption A of a surface is measured in *sabins*, a parameter of which the primary dimensional system is the British imperial foot. It is calculated by multiplying its area S by its sound absorption coefficient. The total absorption in sabins for several absorptive areas is calculated from

$$A = (S_1\alpha_1 + S_2\alpha_2 + \dots + S_n\alpha_n) \quad (3.1.3)$$

Where:

S = area, ft^2 (m^2)

A = total absorption, sabins (metric sabins)

For example, a 10- by 10-ft panel with an absorption coefficient of 0.68 in the 500-Hz band together with a 5- by 40-ft panel with an absorption coefficient of 0.79 in the 500-Hz band has $68 + 158$ sabins of absorption.

The values of absorption coefficient in the 250-, 500-, 1000-, and 2,000-Hz octave bands are often averaged to form a composite absorption coefficient called the *noise reduction coefficient* (NRC). Typical values are shown in the last column in Table 3.1.2. NRC numbers are primarily used in noise reduction computations applied to speech.

Absorption must not be confused with transmission loss or mechanical damping. The words *damping* and *deadening* are often inappropriately applied to mean the adding of sound-absorptive materials.

3-10 Architectural Acoustic Principles and Design Techniques

Table 3.1.2 Typical Absorption Coefficients

Material	Sound absorption coefficient				NRC number
	125 Hz	250 Hz	500 Hz	1000 Hz	
Concrete masonry units, painted	0.08	0.05	0.05	0.07	0.08
Gypsum wallboard, $\frac{1}{2}$ in thick, studs spaced 24 in on center	0.27	0.10	0.05	0.04	0.07
Typical window glass	0.30	0.22	0.17	0.13	0.15
Plaster on lath	0.15	0.10	0.06	0.05	0.07
Light fabric, flat against concrete wall	0.08	0.06	0.10	0.16	0.14
Thick drapery, draped to half area	0.15	0.36	0.55	0.70	0.59
Linoleum on concrete	0.02	0.03	0.03	0.03	0.03
Typical wood floor	0.15	0.12	0.10	0.06	0.09
Thin carpet on concrete	0.03	0.06	0.10	0.20	0.20
Thick carpet with underpadding	0.08	0.28	0.38	0.40	0.39
Typical $\frac{1}{8}$ -in-thick mineral-fiber acoustic ceiling tile	0.45	0.50	0.53	0.69	0.64
Typical $\frac{1}{8}$ -in-thick glass-fiber acoustic ceiling tile	0.44	0.65	0.90	0.92	0.85
					0.97
					0.85

3.1.3a Frictional Absorbers

Most of the commonly available materials intended for the absorption of sound, such as acoustic ceiling tile and acoustic foam, are *frictional absorbers*. They are porous materials that allow for the passage of air and, as a result, the passage of sound waves through them. The sound waves cause molecular motion within the narrow restrictions in the material which results in friction, converting a fraction of the sound energy to heat.

As an acoustical panel is increased in thickness or moved away from a solid backing surface, its absorption at low frequency improves. The application of a facing reduces the effect of high-frequency absorption; common facings are plastic membranes, wood slats, and woven fabric. Material manufacturers provide the absorption coefficient for various frequencies and for different styles of mounting.

Frictional absorbers by themselves are not very useful for reducing sound as it is transmitted from one side of the material to the other. Materials with a good transmission loss are used for this purpose.

3.1.3b Resonant Panels

Sound energy may also be reduced when reflected from impervious *resonant panels*, such as those made from gypsum board or plywood. Incident sound on the panel causes it to vibrate, the air and the material behind the panel dampen the movement, some of the sound is converted to heat, and the remainder is reradiated. For these panels to be effective, they must be large compared with the wavelength of the sound, be fully baffled at the sides and rear, and be tuned to the desired resonant frequency. The maximum absorption coefficient of such a panel is typically 0.5 over a frequency range of an octave. They are most usefully applied at resonant frequencies below 300 Hz. They are usually custom-designed for specific applications.

A typical absorber may be a 4- by 8-ft sheet of plywood 1/2 in thick held a distance d away from a solid wall by means of studs around its periphery. It is sealed to the wall and studs, and the cavity is lightly filled with sound-absorptive material. The resonance frequency is given by

$$f_r = K / \sqrt{md} \quad (3.1.4)$$

where:

$K = 174$ ($K = 60$)

m = surface weight, lb/ft (kg/m)

d = panel spacing from wall, in (m)

For this example, the plywood weighs 2 lb/ft and $d = 1.5$ in; then

$$f_r = 174 / \sqrt{2 \times 1.5} = 100\text{Hz} \quad (3.1.4a)$$

The maximum number of sabins that this typical panel can provide at this frequency can be calculated by multiplying its area by the absorption coefficient

$$A = 4 \times 8 \times 0.5 = 16 \quad (3.1.4b)$$

3-12 Architectural Acoustic Principles and Design Techniques

In this calculation, Equation (3.1.4) is a simplified version of a much more complex equation, and its application is limited to large flexible panels that have a low resonant frequency.

In some situations, construction using such materials as gypsum wallboard for a partition that is intended to be reflective may inadvertently be absorptive in one or two octave bands because of panel resonance. Two walls, perhaps on opposite sides of a studio, may appear identical, but if one is a freestanding partition and the other is mounted to a masonry wall, their absorption coefficients may be significantly different.

3.1.3c Cavity Resonators

The most popular form of *cavity resonator* is a cinder block manufactured with a slot formed through to its internal cavity. This is also most effective at lower frequencies, but it has an absorption coefficient close to 1 at the resonant frequency. The volume of the cavity and the dimensions of the slot determine this frequency. When the cavity of the block is stuffed with fiberglass, the range of frequencies over which the block is effective is increased from one or two octaves to two or three. When these blocks are assembled into large walls, the total number of sabins is obtained by multiplying the manufacturer's absorption coefficient by the wall area [see Equation (3.1.3)].

3.1.3d Reverberation

After the generation of sound within an enclosed space has ceased, the sound wave continues to travel, striking surfaces until it is entirely absorbed. The time taken for the sound pressure level to decay by 60 dB from its original level is called the *reverberation time*. There are two basic controlling factors: room volume and total sound absorption. The reverberation time T , in seconds is given by

$$T_{60} = 0.05V/A \quad (3.1.5a)$$

Where:

V = volume of room, ft^3

A = total area of absorption, sabins

or in metric sabins

$$T_{60} = 0.161V/A \quad (3.1.5b)$$

where V = volume of room, m^3 .

Equation (3.1.5) assumes that the sound pressure level is equally diffused throughout the room. In many actual situations, full diffusion does not exist because of large single areas of absorption. For example, the audience may provide most of the absorption in an auditorium, or a studio may be designed with a *dead end*. In some cases, two reverberation times may be exhibited simultaneously. The decay of sound may have an envelope of two line segments, one for each reverberation time, with the second segment being apparent only after the decay of the first. Thus, caution should be exercised when using a digital reverberation-time meter, for it may obscure valuable information or give erroneous data.

Equation (3.1.5) can be expressed in an alternative way to show the result of adding to a room volume or adding acoustically absorptive material to the material already in place. From Equation (3.1.5)

$$\text{Change in } T_{60} (\%) = (\text{change in \% of volume}) - (\text{change in \% total absorption}) \quad (3.1.6)$$

Thus, if the reverberation time of a space needs to be decreased, the percentage addition of absorptive material has to be significant. This can be understood intuitively from experience in which opening a door in a typical room that may have 200 sabins, and thus creating a totally absorptive area of an additional 20 ft^2 does not noticeably change the reverberation time. From Equation (3.1.6) the change in T_{60} is only 10 percent.

3.1.3e Combination of Direct and Reverberant Sound

The sound level within a space consists of two parts: 1) the sound that comes directly from the source and 2) the reverberant sound. Very close to the source the direct sound predominates. Further away, the direct sound decreases by 6 dB at each doubling of the distance while the reverberant-sound level stays almost constant. At a distant point, the direct sound contributes insignificantly to the total sound level, and no matter how much more distance from the source is increased, the sound level remains constant at the reverberant level.

At an intermediate point, at a distance from the source known as the *critical distance*, the direct sound is equal to the reverberant sound. This distance depends on the total absorption [see Equation (3.1.3) for total number of sabins] within the space and the directivity of the source.

If it is proposed to reduce sound within a space by means of absorption, applying an infinite amount will remove only the reverberant-sound contribution, not the direct sound. Of course, because of space constraints only limited amounts of absorption can be applied, with the result that not all the reverberant sound can be removed.

The critical distance is the closest distance at which any discernible sound-level reduction (3 dB) can be obtained by the application of absorptive materials. In addition, when taking into account the fact that a 3-dB reduction is just discernible but may not be significant, the cost of absorptive materials for noise reduction at the critical distance may not be justified. Noise control by means of absorption is usually practical only beyond 2 to 3 times the critical distance.

To calculate the critical distance, first calculate the room constant R

$$R = A \left[\frac{\Sigma S}{\Sigma S - A} \right] \quad (3.1.7a)$$

The critical distance d is

$$d = \sqrt{RQ/16\pi} \quad (3.7b)$$

The result of this calculation gives an immediate perspective on whether to control the sound through absorption within the space or to apply alternative means.

The parameters here are interrelated, as shown in Figure 3.1.2. The upper half of the figure shows a diagonal line indicating the direct sound falling at 6 dB of each doubling of distance.

3-14 Architectural Acoustic Principles and Design Techniques

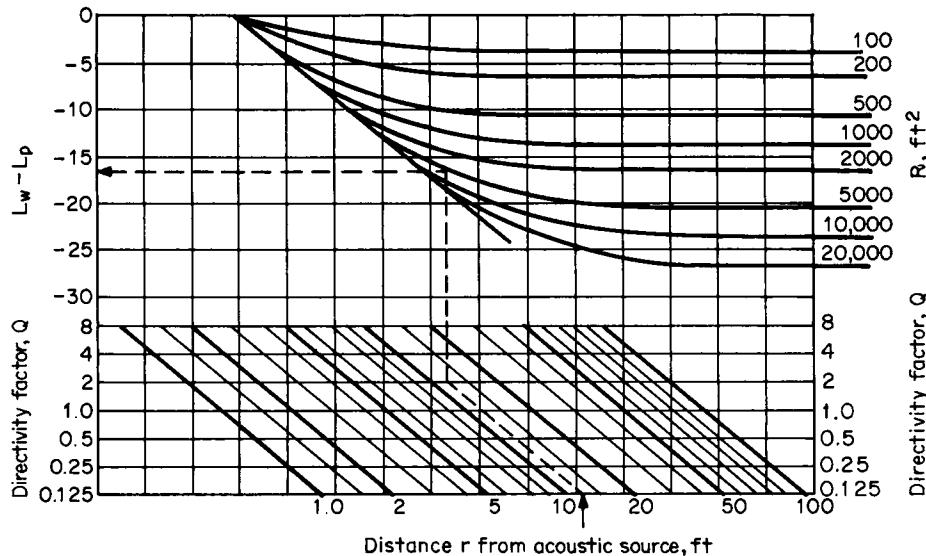


Figure 3.1.2 Relationship between distance from source, directivity factor, room constant, and $L_w - L_p$

The asymptotic lines show the contribution of reverberant sound. The lower half of the figure applies to source directivity. For example, at a distance r , 12 ft from a source with a directivity factor Q of 2, in a room with a room constant R of 5000 ft², the sound pressure level will be 17 dB less than the sound power level. Note that r also is approximately the critical distance where the sound level is 3 dB above the direct level.

Though the sound level within a space can be controlled to a limited degree through the application of absorptive materials, building a partition can be much more effective in separating a noise source from a listener, although it can never totally prevent all the sound from passing through.

Actually, sound does not pass “through” a typical wall. The sound pressure on one side of the wall results in a force that shakes the wall. The shaking wall in turn disturbs the air on the other side, causing sound pressure waves to spread again and thus to establish a new sound level in the receiving space.

The difference between the levels on either side of the wall is called the *noise reduction* of the wall. The noise reduction (NR) depends not only on the characteristics of the wall but on the total absorption in the receiving space. For example, a wall enclosing a highly reverberant room is significantly less effective than the same wall protecting a well-upholstered lounge. Thus, to help in defining the acoustical performance of the wall alone, a measure that is independent of the acoustical characteristics of the receiving space is required. This is termed *transmission loss* (TL).

In one typical situation, in which the sound travels from one reverberant space to another, the transmission loss of the wall is related to noise reduction by

$$TL = NR + 10 \log(S/A) \quad (3.1.9)$$

Where:

S = surface area of the common wall, ft (m)

A = total absorption of the receiving room, sabins (metric sabins)

In the other typical situation, in which the sound travels from a nonreverberant area to a reverberant space—for example, between a noisy highway and a studio—the noise reduction is 5 dB less

$$TL = NR + 10 \log(S/A) + 5 \quad (3.1.10)$$

The term $10 \log(S/A)$ is often called *room effect*.

In general terms, the transmission loss of a single wall is governed by its mass

$$TL = 20 \log(f) + 20 \log(M) - K \quad (3.1.11)$$

Where:

f = frequency, Hz

M = mass, lb/ft (kg/m)

K = 34 (48 for metric units)

This shows that every time that the mass of the wall is doubled, the transmission loss increases by 6 dB. However, in practical terms the mass of a wall cannot be doubled more than a few times before running into structural and space limitations. Also, Equation (3.1.11) shows that the transmission loss increases by 6 dB each time that the frequency is doubled. Thus, at high frequencies much more transmission loss is demonstrated than at bass frequencies.

Mass law essentially gives the maximum TL that can be expected from a homogenous wall. In fact, lower values are to be expected primarily as a result of *coincidence dip*. The frequency at which this dip occurs depends on the speed of sound within the wall material. Consequently, for each material the coincidence frequency occurs at a different frequency. For example, it occurs in the 2000-Hz band for gypsum wallboard. The dip may reduce the transmission loss by up to 15 dB. Figure 3.1.3 shows an example. Details of how and why coincidence dip occurs can be found in [2].

Because the transmission loss can be changed or improved by more complex methods of construction, perhaps incorporating double independent walls, it is good design practice to use certified transmission-loss test data to calculate noise reduction of specific design proposals [3].

Reputable salespersons of acoustical products should be able to provide certified sound-transmission-loss data for their products used in specific applications, but care should be exercised when using more than one of these products back to back because acoustic coupling and resonances between the products make resulting performance difficult to predict. Because the development of complex partitions is beyond the scope of this chapter, advice of an expert in acoustics should be sought in these situations.

3-16 Architectural Acoustic Principles and Design Techniques

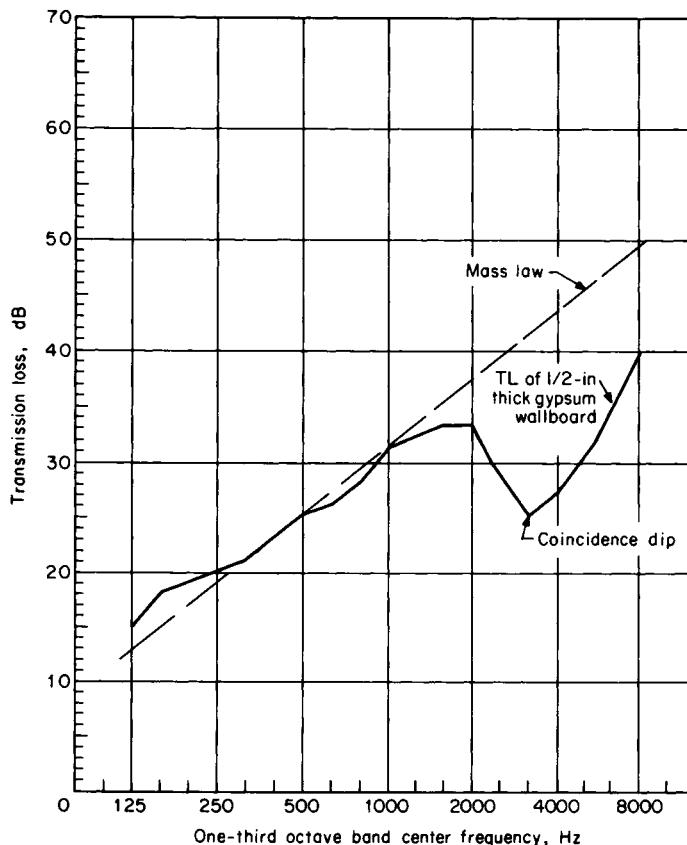


Figure 3.1.3 Example of coincidence dip reducing transmission loss.

3.1.3f Composite Transmission Loss

Often a wall is made up of several elements such as a wall, windows, doors, or even openings. The transmission loss of each of these elements can be combined into one *composite transmission loss* (TL_c). The procedure is best understood by defining a transmission-loss coefficient τ which is the fraction of the sound power passing through a unit area of the wall

$$TL = -10 \log(\tau) \quad \text{or} \quad \tau = 10^{-TL/10} \quad (3.1.12)$$

The fraction of sound passing through the composite wall τ_c made up of elements 1, 2, ... n , is then

$$\tau_c = \tau_1 W_1/W + \tau_2 W_2/W + \tau_n W_n/W \quad (3.1.13)$$

where W is the wall area, ft^2 (m^2). Then the composite transmission loss is

$$\text{TL}_c = -10 \log(\tau_c) \quad (3.1.14)$$

The formula is also very instructive in showing how apertures in a wall influence its performance. For example, if a wall with an area of 100 ft^2 with a TL of 40 dB has a 3-ft² hole, then

$$\tau_c = 0.0001\left(\frac{97}{100}\right) + 1\left(\frac{3}{100}\right) \text{ or } = 3.0097/100 \quad (3.1.14a)$$

and

$$\text{TL}_c = 10 \log \frac{100}{3.0097} = 15 \text{ dB} \quad (3.1.14b)$$

Thus, a 40-dB wall has had its transmission loss reduced from 40 to 15 dB by cutting the hole. Small holes, such as cracks and slits, let much more sound through than the equations presented above would predict. The value of τ for a crack may be up to 10 times greater than would be predicted by applying its area to these equations.

3.1.3g Sound Transmission Class

To simplify handling transmission-loss data in multiple-frequency bands, a single-number descriptor, called the *sound transmission class* (STC), is often used to rate a wall. STC and transmission-loss data of common materials are shown in Table 3.1.3.

However, in the process of condensing these multifrequency TL data, it is assumed that the general shape of the noise spectrum is similar to that of speech. Nevertheless, in spite of this process, STC ratings are often applied indiscriminately to the isolation of other types of sounds such as machinery or music.

The STC of a transmission-loss spectrum is determined by adjusting a fixed-shape contour over the plotted data according to the following predetermined criteria:

- Only one-third-octave bands from 125 to 4000 Hz are considered
- The sum of the deficiencies (the deviations below the contour) cannot exceed 32 dB
- Values above the contour are ignored
- The deficiency in any one band cannot exceed 8 dB

The STC rating is the ordinate of the contour at 500 Hz. Figure 3.1.4 shows the STC curves, and Figure 3.1.5 shows an example.

Transmission-loss curves are typically jagged and are not smooth and rising like the standardized STC curve. At the frequency where the STC curve may exceed the TL curve by up to 8 dB, insufficient noise reduction may be obtained. In addition, since the STC rating incorporates little transmission loss at low frequencies, it is most inappropriate to use STC ratings for the isolation of bass sounds.

3-18 Architectural Acoustic Principles and Design Techniques

Table 3.1.3 Typical STC and Transmission-Loss Data

Material	Transmission loss, dB						STC rating
	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	
Gypsum wallboard, $\frac{1}{2}$ in thick	14	20	24	30	30	27	27
Two layers in gypsum wallboard, both $\frac{1}{2}$ in thick	19	26	30	30	29	36	31
Flat concrete panel, medium weight, 6 in thick	37	43	51	59	67	73	55
One layer of $\frac{1}{2}$ -in-thick gypsum wallboard on each side of 2-by 4-in wood studs (16 in off center) with 2-in-thick glass-fiber batt in the cavity	20	28	33	43	43	40	38
Same as above, but with two layers of $\frac{1}{2}$ -in-thick gypsum wallboard on each side	24	37	44	49	50	50	46
Same as above, but with staggered studs	34	43	49	54	54	52	51
Same as above, but with double row of 2- by 4-in studs spaced 1 in apart on separate plates, using type X (fire-rated) gypsum wallboard, and two layers of 3-in glass-fiber batt in the cavity	45	54	63	66	66	64	63
Same as above, but with bracing across cavity at third points of studs	40	45	56	62	57	60	57
4-in face brick, mortared together	31	33	39	47	55	61	45
Two layers of mortared 4-in face brick separated by 2-in air space, with metal ties	36	36	46	54	61	66	50
Same as above, but air space filled with concrete grout	41	47	56	62	66	70	59
6-in-thick three-cell dense concrete masonry units, mortared together	36	38	42	49	53	60	48
2-in-thick hollow-core door, ungasketed	13	19	23	18	17	21	19
2-in-thick solid wood door with airtight gasketing and drop seal	29	31	31	31	39	43	35
Typical window glass, $\frac{1}{8}$ in thick, single plate	15	23	26	30	32	30	29
Typical thermal glazing window ($\frac{1}{8}$ -in glass, $\frac{1}{8}$ -in air space, $\frac{1}{8}$ -in glass)	22	21	29	34	30	32	30
$\frac{1}{8}$ -in-thick laminated glass	34	35	36	37	40	51	39
Composite window ($\frac{1}{8}$ -in laminated glass, 5-in air space, $\frac{1}{8}$ -in glass)	33	54	60	57	55	63	55
Typical $\frac{1}{8}$ -in-thick mineral-fiber acoustic ceiling tile	6	10	12	16	21	21	17

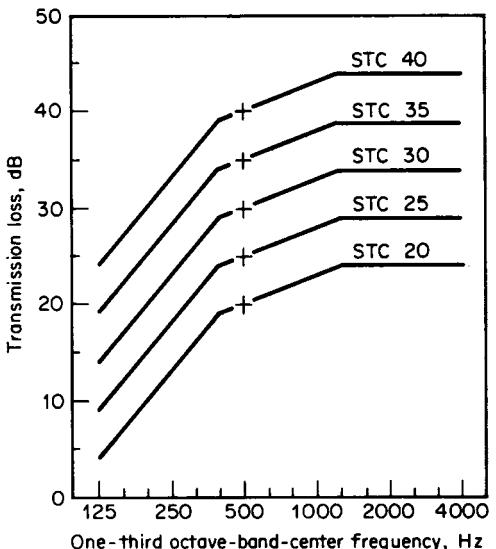


Figure 3.1.4 Example of sound-transmission-class (STC) curves. (From ASTM specification E 1413-73.)

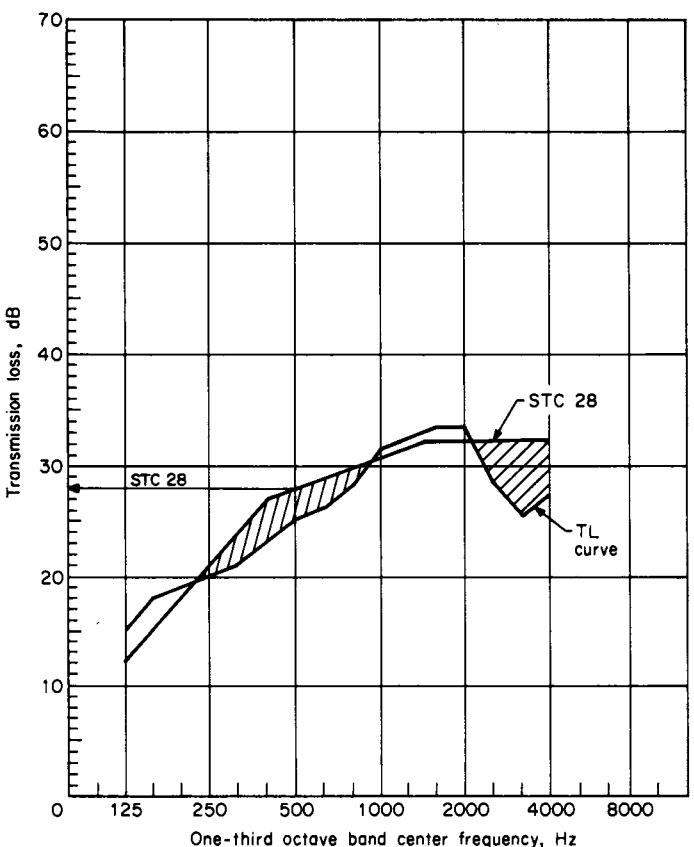


Figure 3.1.5 Sound transmission class (STC) for 1/2-in-thick gypsum wallboard. When a TL data point falls below the STC curve, the difference is termed a *deficiency*. The sum of all deficiencies must not be greater than 32 dB. No single deficiency shall be greater than 8 dB.

3-20 Architectural Acoustic Principles and Design Techniques

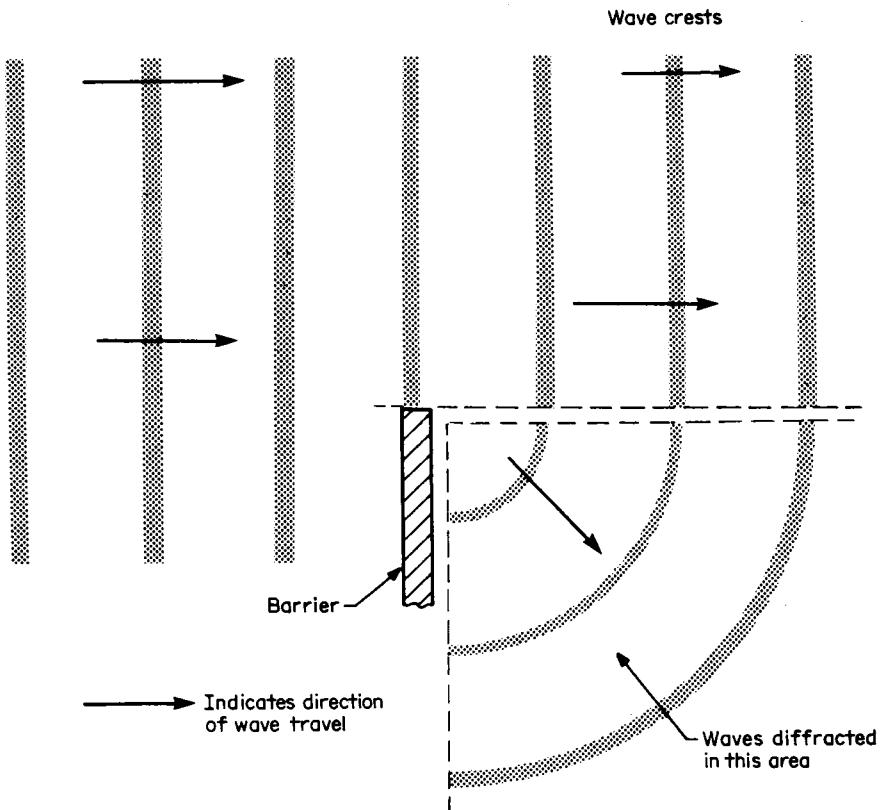


Figure 3.1.6 Diffraction of sound by a solid barrier.

3.1.3h Diffraction

In addition to being reflected or transmitted, sound waves can be diffracted. Figure 3.1.6 shows a barrier parallel to waves. At the free end of the barrier the waves spread around to the acoustical shadow area behind. The more the waves turn around the end of the barrier, the more the amplitude of the wave is decreased.

For example, for acoustically simple applications such as for a barrier built around rooftop equipment, the amount of noise reduction from a barrier depends on the increased distance that the sound has to travel over the top of the barrier to the listener caused by the insertion of the barrier. For the distance d , which is the additional distance that the sound must travel around the obstruction to the receiver, the Fresnel number N is defined as

$$N = 2d/\lambda \quad (3.1.15)$$

Where:

λ = wavelength of sound, ft (m)

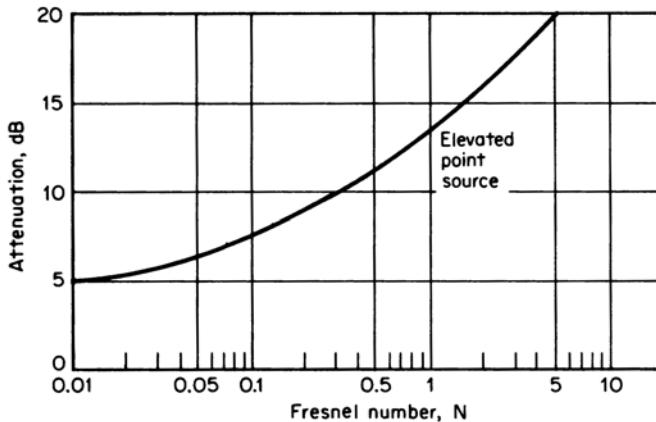


Figure 3.1.7 Noise reduction of a barrier for different Fresnel numbers.

d = total path length from the source over the barrier to the receiver less direct distance separating source and receiver, ft (m)

Thus, there is a different Fresnel number for each frequency.

From Figure 3.1.7, the noise reduction can be found for each frequency band of interest. It can be seen that low-frequency sound (with lower N) is attenuated less by diffraction than high-frequency sound.

3.1.4 References

1. *ASA Standards Index 2*, Acoustical Society of America, New York, N.Y., 1980.
2. Beranek, L. L.: *Noise and Vibration Control*, McGraw-Hill, New York, N.Y., 1971.
3. *Catalogue of STC and IIC Ratings for Wall and Floor/Ceiling Assemblies*, Office of Noise Control, Berkeley, Calif.

Criteria for Acceptability of Acoustical Performance

Richard G. Cann, Anthony Hoover

3.2.1 Introduction

Before beginning the acoustical design of any space, one of the first tasks is to establish criteria for its performance. Sometimes it might appear that this is superfluous, that the owner, architect, and engineer have an unwritten understanding of acoustical requirements, and that the design process can begin immediately. However, this is often not the case, and absence of thoughtfully written criteria may lead to fundamental design errors. The development of criteria is very important in defining just how spaces will be used and will ultimately determine just how well the spaces will function acoustically.

3.2.1a Reverberation Time

For example, if an auditorium will be used for both speech and music, basic design decisions will have to be made at the outset. Figure 3.2.1 shows a typical plot of reverberation time at 500 Hz versus room volume for auditoria used for different activities. The preferred reverberation time for music is approximately twice as long as it is for speech. Either the acoustical quality of some activities will have to be compromised by selecting a specific reverberation time, or provision will have to be made for adjusting the reverberation time for each activity.

Though there are generally accepted criteria for the reverberation time of a small auditorium, they are not nearly so clear-cut for recording studios. Some performers insist on live reverberant feedback from the space itself and directly from adjacent performers. Other artists, who work entirely through headphones, are much less concerned about the reverberant quality of a studio. The former scenario requires an architectural solution, while the latter requires none.

The design criteria for a control room often simply invoke, without acoustical reason, the currently fashionable proprietary design concept. It seems that no matter which concept is selected, there are always sound engineers who hate it or love it. There are no generally accepted design criteria even among those who record the same artist. Remembering that almost all the buying public never hear a recording with the fidelity available to sound engineers, what the engineers really want is sound quality that will allow them to project what they hear into the finished disk.

3-24 Architectural Acoustic Principles and Design Techniques

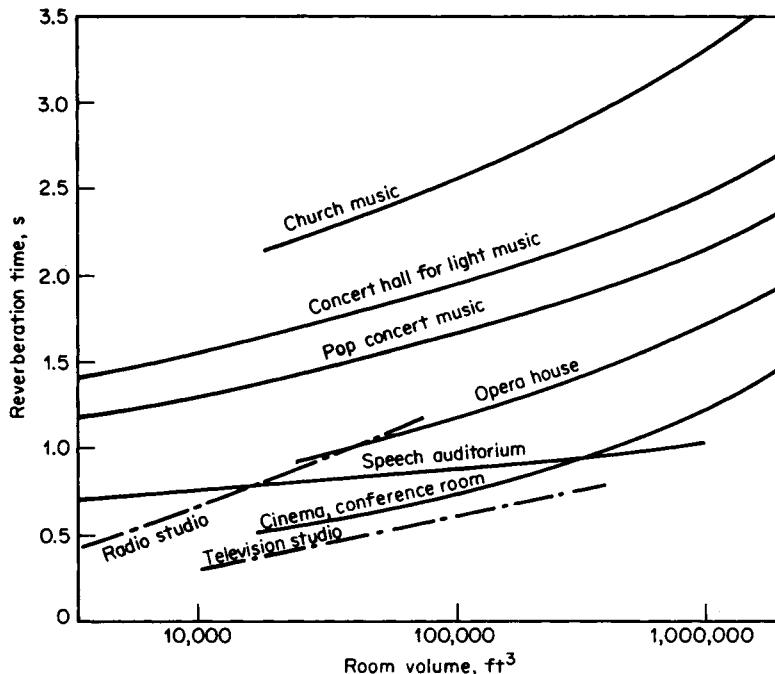


Figure 3.2.1 Typical reverberation times for different sizes of rooms and auditoria according to usage.

And, of course, that varies from one engineer to the next, according to their past experience. Therefore, it is wise to build flexibility into control-room design, allowing engineers to produce their own semicustom setup. No one way is best.

3.2.1b Background Noise

The background sound level should not interfere with the perception of the desired sound, especially during quiet periods, for music is a rhythm not only of sound but also of silence. It is important not to intrude upon the latter with a rumbling fan or traffic noise. The most frequently used criteria for background noise are the noise-criterion (NC) curves, which are classified according to space usage in Table 3.2.1. The interior octave-band sound levels are plotted on a standard graph as shown in Figure 3.2.2. Each NC curve is named according to its value at 500 Hz. The NC value of a plotted spectrum is usually defined by the highest NC value that is attained in any octave band. For example, the NC value of the spectrum in Figure 3.2.2 is defined by the sound pressure level at 250 Hz, which gives an NC value of 45.

However, NC curves are applicable only to essentially informationless sound, sound which is continuous and time-invariant. Where the intruding noise carries meaningful information such as intelligible speech or countermusical rhythms, more stringent criteria must be applied.

Criteria for Acceptability of Acoustical Performance 3-25**Table 3.2.1 Design Goals for Mechanical-System Noise Levels.**

Type of area	NC range
Residences	
Single-family homes	20–30
Apartments and condominiums	25–35
Hotels and motels	
Guest rooms	30–40
Meeting or banquet rooms	30–35
Corridors and lobbies	35–45
Offices	
Boardrooms	20–30
Executive offices	30–40
Open-plan areas	35–40
Public circulation	35–45
Hospitals	
Private rooms	25–35
Operating rooms	30–40
Wards	30–40
Laboratories	30–40
Public circulation	35–45
Churches	
Sanctuaries	20–30
Public circulation	30–40
Schools and universities	
Libraries	30–40
Lecture rooms and classrooms	20–35
Laboratories	35–45
Cafeterias and recreation halls	35–45
Public buildings	
Libraries and museums	30–40
Courtrooms	25–35
Post offices and banks	30–40
Auditoria, theaters, and studios	
Concert halls	15–25
Recording studios	15–25
Multipurpose halls	20–30
TV studios	20–30
Movie theaters	30–35

In some circumstances quieter is not necessarily better, for reducing noise levels may reveal other sounds that were previously hidden. Noise levels can often be increased to the NC value specified in Table 3.2.1 without causing noticeable intrusion. This masking of an intruding sound may be less expensive than controlling the sound itself.

It may be necessary to set special criteria for a particular activity. For example, for a recording studio, noise criteria may be set equal to the noise floor of the recording instrumentation. This level may be above the aural threshold of the performers, with the result that it is possible to

3-26 Architectural Acoustic Principles and Design Techniques

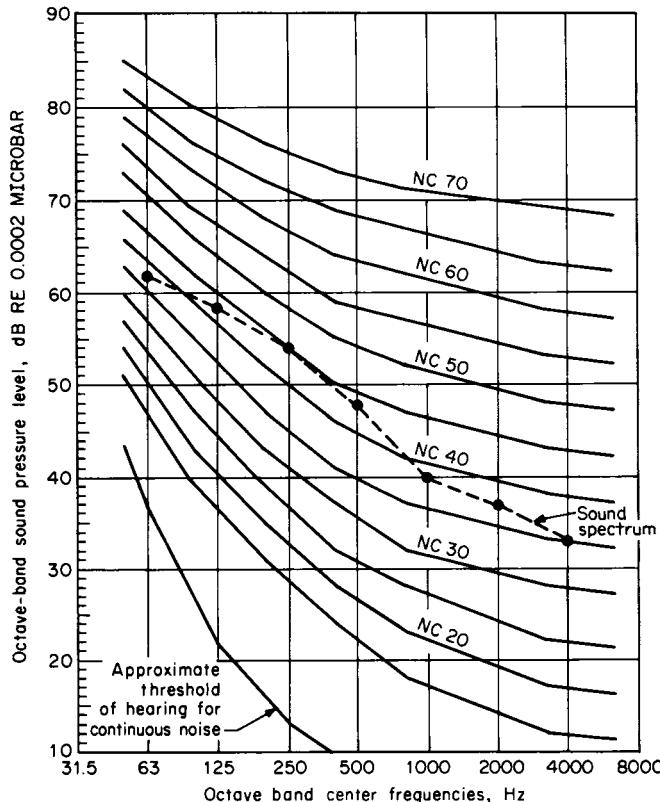


Figure 3.2.2 Noise-criterion (NC) curves. (From [4]. Used with permission.)

identify clearly a source of intruding noise before a recording session begins and yet not have the offending noise be audible on tape.

3.2.1c Maximum Levels

The maximum noise expected in a space has to be determined. For example, if a space is used primarily for dining at low sound levels but if it is anticipated that the space will also be used occasionally as a nightclub when a group will bring in large loudspeaker stacks, an estimate of these higher sound levels must be made so that intrusion into adjacent spaces can be calculated. The activities in an adjacent rehearsal room may be curtailed if the dining-room sound levels are 40 dB higher.

3.2.1d Interference of Speech Communication

Mechanical equipment, air-distribution louvers, and other sources of sound may generate noise that causes speech to be unintelligible. Thus person-to-person conversations and telephone conversations rapidly become exhausting, and the efficient functioning of an administrative office can be thoroughly impaired. Criteria for adequate speech intelligibility can be developed according to [1 and 2].

3.2.1e Exterior Noise

Concern need be given not only to sound levels within the space but also to those outside in the neighborhood. The community may have a specific bylaw limiting noise levels. If not, the criteria for maximum noise levels should be based upon available annoyance data. It is appropriate to establish these criteria up front in a permit application rather than hope that abutters will not stir up community objections at a later time.

3.2.2 Mechanical Systems

Mechanical systems are the source of ventilation for recording studios, control rooms, and associated spaces. The acoustical concerns generally are mechanical systems as a source of noise produced by such items as fans and airflow, and as a path for sound transmission between spaces as through ductwork. A more complete treatment of the mechanical system as a source of noise may be found [3].

3.2.2a Sound Generation by Fans

Fans that are required to move air through a ventilation system inherently generate noise. Many factors determine the amount of noise produced, including the type of fan used, the volume of air to be delivered, the static pressure against which the fan is forcing the air, the blade passage, and the efficiency of the fan system.

The most common type of fan used for ventilation systems is the centrifugal airfoil fan, although other types of system are not unusual. Each system tends to produce its own characteristic spectrum of frequency, but in general fans used for ventilation systems tend to produce more low-frequency noise energy than high-frequency noise energy. In most cases, these fans are contained within a prefabricated housing which in turn is connected to the supply-air ductwork system and to the return-air ductwork system. It is important to note that the sound generated by a fan propagates as easily through the return-air system as through the supply-air system because the speed of sound is so much faster than the speed of the air within the ductwork.

The sound power level generated by a fan may be calculated by the following

$$L_w = K_w + 10 \log(q) + 20 \log(p) + K \quad (3.2.1)$$

Where:

L_w = sound power level of fan, dB

K_w = specific sound-power-level factor for type of fan, dB

3-28 Architectural Acoustic Principles and Design Techniques

q = volume flow rate, ft³/min (m³/s)

p = static pressure, in of water (kPa)

$K = 0$ (metric $K = 45$)

K_w for centrifugal airflow types of fans tends to arrange itself in a smooth spectrum with values of approximately 35 to 40 dB in the 63-Hz octave band and transitioning down to 10 to 15 dB in the 8000-Hz band. It should be noted from Equation (3.2.1) that a doubling of the volume flow will add 3 dB of sound power level, whereas a doubling of pressure will add 6 dB to the sound power level. Therefore, it is important to design a ventilation system with adequately sized ductwork and with smooth transitions and bends to keep the static pressure as low as possible.

An additional pure-tone component to the noise, *blade-frequency increment* (BFI), is generated by each fan blade passing by an edge or obstruction, such as the discharge opening of the fan unit. The octave band in which it falls is determined by calculating the blade-passage frequency and referring to Table 3.2.2.

$$\text{Blade-passage frequency} = \text{r/min} \times \text{number of fan blades}/60 \quad (3.2.2)$$

Most fan systems are of a design that gives a BFI of 3 to 10 dB in the 125- or 250-Hz band. The result is that unless this BFI is adequately attenuated, the audible pure tone will be in the same range of frequency as the fundamentals of speech and many musical instruments.

Fans should be selected for a maximum efficiency rating. A decrease in efficiency results in an increase in the sound power level generated. Most systems operate at a reduced efficiency which adds approximately 5 dB to the fan power level, and poorly selected or improperly maintained fans have been known to add as much as 20 to 25 dB to the fan power level.

Noise generated by the fan not only travels down the supply and return ductwork systems but also is radiated off the fan housing. In general, the fan housing is a very poor isolator of sound and for most practical purposes, especially in lower frequencies, can be considered to provide no isolation whatsoever. Therefore, it is good practice to locate the fan assembly well removed from the recording studio and control room.

Most reputable fan-system manufacturers provide octave-band sound-power-level data for their systems. These numbers are obviously preferable to the generic methods of calculations described here. In addition, other devices within the mechanical system may incorporate smaller fans, such as fan-powered terminal boxes. These smaller devices are generally located in the ductwork closer to the specific rooms to be ventilated, providing an extra boost to the airflow as required by the system. The accuracy of these generic methods of calculation tends to decrease as the size of the fan decreases, but the octave-band sound power level of these smaller devices is also generally available through the manufacturer.

3.2.2b Turbulent Noise in Ducts

Airflow noise is generated by turbulence within the ductwork and at diffusers and dampers. Air turbulence and, therefore, airflow noise generally increase as the speed of airflow increases. Therefore, it is good practice to keep the speed of airflow low. Several rules for controlling airflow noise are:

Criteria for Acceptability of Acoustical Performance 3-29

Table 3.2.2 Limits of Frequency Passbands

Band number	Nominal center frequency	One-third-octave passband	Octave passband
14	25	22.4–28.2	
15	31.5	28.2–35.5	22.4–44.7
16	40	35.5–44.7	
17	50	44.7–56.2	
18	63	56.2–70.8	44.7–89.1
19	80	70.8–89.1	
20	100	89.1–112	
21	125	112–141	89.1–178
22	160	141–178	
23	200	178–224	
24	250	224–282	178–355
25	315	282–355	
26	400	355–447	
27	500	447–562	355–708
28	630	562–708	
29	800	708–891	
30	1,000	891–1,120	708–1410
31	1,250	1,120–1,410	
32	1,600	1,410–1,780	
33	2,000	1,780–2,240	1,410–2,820
34	2,500	2,240–2,820	
35	3,150	2,820–3,550	
36	4,000	3,550–4,470	2,820–5,620
37	5,000	4,470–5,620	
38	6,300	5,620–7,080	
39	8,000	7,080–8,910	5,620–11,200
40	10,000	8,910–11,200	

- Size ductwork so that the flow of air stays below 2000 ft/min and preferably below 1500 ft/min. The velocity of air in a duct may be calculated by dividing total cubic feet per minute in that duct by the cross-sectional area in square feet of the duct itself.
- Airflow velocities through diffusers should be kept below a maximum of 500 ft/min through all diffusers. For critical applications, lower speeds such as 200 to 300 ft/min are advisable.
- Air valves and dampers should be located so that the airflow noise that they generate does not contribute to the noise ducted from upstream sources.
- Splits and bends in the ductwork should be smooth. Abrupt corners and bends should be avoided, especially near the fan, near high-airflow-velocity locations, and near diffusers and grilles.

Airflow noise is typically a major component of mid- and high-frequency background noise in recording studios. However, when there are abrupt bends and turns in ductwork systems, especially with high airflow velocity, a considerable amount of low-frequency energy may be generated that is extremely difficult to control.

3-30 Architectural Acoustic Principles and Design Techniques

3.2.2c Attenuation of Noise by Ducts

Various elements within the ducted ventilation system inherently provide some attenuation of the noise as it travels through the ductwork, both down the supply-air system and up through the return-air system. Certain elements, such as internal duct lining and prefabricated silencers, can be added to the system as necessary to increase noise attenuation.

Bare ducts, that is, sheet-metal ducts which lack any added sound-absorptive lining, provide a minimal but measurable amount of attenuation to the noise. The amount of attenuation depends on such factors as width and height dimensions and the length of that section of ductwork under consideration. Such duct attenuation is approximately 0.1 dB/ft of duct length for frequencies of 250 Hz and above regardless of width and height dimensions. For example, the 1000-Hz-band noise level inside the end of a 10-ft length of bare rectangular duct should be approximately 1 dB less than inside the beginning. For lower frequencies, duct attenuation is approximately 0.2 dB/ft of duct length or even up to 0.3 dB/ft for ductwork as small as 5 to 15 in in either width or height. Typical data are shown in Table 3.2.3. It is important to note that duct attenuation decreases in the lower frequencies because the thin sheet metal of which most ductwork is constructed is a poor barrier for low-frequency sound transmission, and as a result these low frequencies “break out” of the ductwork and into the surrounding space. Thus, it is advisable to reroute ductwork which is known to contain high levels of sound energy, especially low-frequency sound energy, from spaces which require low background noise.

Sheet-metal duct may be lined with sound-absorptive material. This material is generally of about 1 1/2 lb/ft density and of either 1- or 2-in thickness (the 2-in thickness generally provides improved duct attenuation, which can be an important consideration, especially for low-frequency noise control), and it often has a mastic facing to reduce shredding and deterioration from high airflow velocities. For lined ductwork, duct attenuation is very much dependent on the width and height dimensions, and on the octave frequency band of interest. Sources such as [3] or specific manufacturers' data should be consulted for a detailed analysis of lined-ductwork attenuation. It should be noted that noise breakout, especially for low frequencies, is not significantly affected by lining.

Splits, divisions, and takeoffs in the ductwork represent further attenuation to the ducted noise. It is assumed that the amount of noise energy delivered is proportional to the amount of air delivered. For example, if a fan that provides a total of 10,000 ft³/min of air delivers only 1000 ft³/min of air to a particular room (the other 9000 ft³/min of air is delivered to other rooms by means of splits and divisions), the amount of air delivered to that room is reduced to 10 percent of the total, whereas the noise or power level is reduced by 10 log (0.1) of the total. In other words, the splits and divisions have reduced the amount of noise delivered to that room by 10 dB. Similarly, a split which sends half of the air down one duct and the other half down another duct has reduced the amount of noise entering each of these two ducts by 3 dB, which is derived from $10 \log (0.5) = 3$ dB.

Bends and elbows in ductwork are not very effective in attenuating low-frequency noise but can provide significant attenuation of higher-frequency noise. Lined elbows and bends provide better high-frequency attenuation than bare elbows and bends. Bare elbows may provide up to 3 dB of attenuation at 2000 Hz and above, and lined elbows can provide between 5 and 10 dB of attenuation in the higher frequencies, depending on the elbow radius and duct diameter. Typical data are shown in Table 3.2.4.

Table 3.2-3 Approximate Duct Attenuation

Duct type	Duct diameter, in	Approximate duct attenuation, dB/ft					
		63 Hz	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz
Bare, rectangular or square	5-15	0.3	0.3	0.2	0.1	0.1	0.1
	16-45	0.3	0.2	0.1	0.1	0.1	0.1
	46-90	0.2	0.1	0.1	0.1	0.1	0.1
Bare, oval or round	5-15	0.15	0.15	0.1	0.05	0.05	0.05
	16-45	0.15	0.1	0.05	0.05	0.05	0.05
	46-90	0.1	0.05	0.05	0.05	0.05	0.05
1-in-thick lining, all duct shapes	5-15	0.4	0.5	1.2	2.5	5.0	4.5
	16-30	0.4	0.3	0.7	1.3	2.5	2.0
	31-45	0.4	0.3	0.3	0.8	1.4	1.2
46-60	0.3	0.3	0.2	0.6	1.0	0.8	0.7
	61-75	0.3	0.3	0.2	0.4	0.8	0.6
	76-90	0.3	0.3	0.2	0.4	0.7	0.5
2-in-thick lining, all duct shapes	5-15	0.6	1.0	2.3	4.0	5.0	4.5
	16-30	0.4	0.4	1.0	2.0	2.5	2.1
	31-45	0.4	0.4	0.7	1.4	1.4	1.2
46-60	0.3	0.3	0.5	0.9	1.0	0.8	0.7
	61-75	0.3	0.3	0.4	0.8	0.8	0.7
	76-90	0.3	0.3	0.3	0.7	0.7	0.5

3-32 Architectural Acoustic Principles and Design Techniques

Table 3.2-4 approximate Elbow Attenuation

90° elbows and turns	Duct diameter, in	Approximate elbow attenuation, dB/elbow							
		63 Hz	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
Unlined round elbow or square turn	5-30	0	0	0	0	1	2	3	3
	31-45	0	0	0	0	1	2	3	3
Lined round elbow or square turn with turning vanes	46-90	0	0	0	0	2	3	3	3
	5-30	0	0	0	0	1	2	3	3
Lined square turn without turning vanes	5-30	0	0	0	0	1	2	3	4
	31-45	0	1	2	3	3	4	5	5
	46-90	1	2	3	5	4	5	5	6

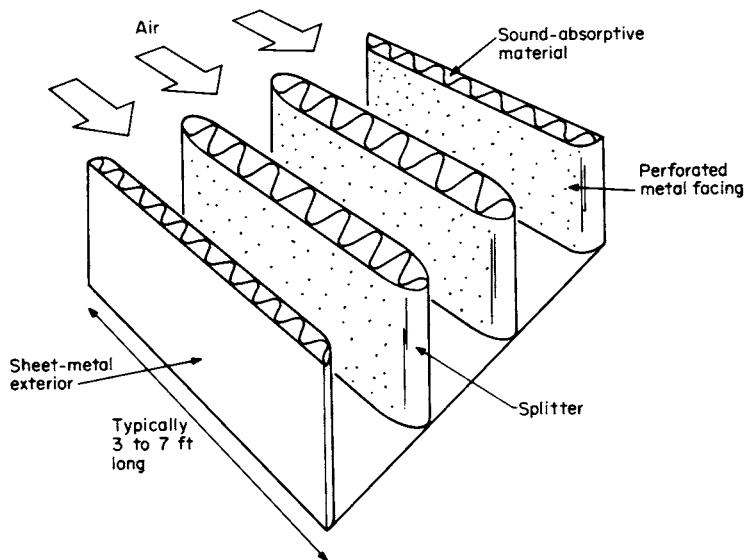


Figure 3.2.3 Cutaway view of a typical duct silencer.

3.2.2d Duct Silencers

Prefabricated duct silencers generally incorporate a system of parallel sound absorptive baffles between which the air must flow (Figure 3.2.3). Silencers are available in a wide range of sizes and duct attenuations which tend to vary with frequency. Manufacturers should be able to provide detailed data on the performance of their silencers. These data should include the octave-band *dynamic insertion loss* (DIL), the pressure drop across the silencer, and the self-generated noise of the silencer. The effective attenuation of a silencer in any given octave band can change with the airflow velocity through the silencer. This is measured in terms of the DIL, which is duct attenuation in octave bands at different airflow velocities, both positive and negative. Positive DILs rate the effectiveness of a silencer when the noise and the air both flow in the same direction, as in the case of a supply-air system, and negative DILs apply where noise flows in the opposite direction of the airflow as in a return-air system. Because the baffles in a silencer restrict the flow of air to a certain degree, the silencer can add to the static pressure against which the fan must work, so that the pressure-drop ratings of silencers can become an important consideration. Since the baffles generate a certain amount of turbulence in the airflow, silencers can generate a certain amount of noise. Silencers should be positioned so that the amount of attenuated noise leaving them is still higher than the generated noise of the silencers, which implies that the silencers should be placed relatively close to fans. On the other hand, it is good practice to locate silencers at least five duct diameters downstream of a fan in the supply-air system; otherwise noise generated by turbulent air, especially low-frequency noise, can greatly exceed the rated self-noise of the silencers. Placement of silencers in return-air systems is less critical, but a spacing of at least three duct diameters between fan and silencer is still advisable.

3-34 Architectural Acoustic Principles and Design Techniques

3.2.2e Calculating Resultant Sound Levels

A convenient method for determining the sound power levels that enter a room is to list the noise-generating and noise-attenuating devices in the ductwork system sequentially, beginning with the fan. This should be done for both supply-air systems and return-air systems. The following analysis should be performed octave band by octave band.

For each device, calculate either the sound power levels generated or the amount of attenuation provided by each element. Then sequentially subtract the attenuation provided by each element until a noise-generating device is encountered. At this point, the resultant sound power level which has made its way through the ductwork should be added to the sound power level generated by the appropriate device. Continue with this method until the octave-band sound power levels at the end of the run of ductwork (at the point at which the sound begins to enter the room) have been calculated. Then convert the sound power levels to sound pressure levels at various points of concern within the room.

The resultant octave-band sound pressure levels then may be plotted against NC curves, such as in Figure 3.2.2, in order to determine whether the ventilation-system noise satisfies the criterion decided upon for the appropriate space. If the levels are too high, noise attenuation devices may be incorporated in the ventilation system. The calculation procedure may then be repeated to take into account the effect of the attenuation devices. If an attenuation device is inserted into the system, it is important to delete the effect of the part of the system which has been omitted. For example, if a 5-ft-long attenuator is inserted in a length of lined duct, then the effect of 5 ft of lined duct should be eliminated from subsequent calculations. Otherwise the calculations may result in inappropriately low sound pressure levels.

3.2.3 References

1. ANSI: *Method for the Measurement of Monosyllabic Word Intelligibility*, ANSI S3.2-1960, rev. 1977, American National Standards Institute, New York, N.Y., 1976.
2. ANSI: *American National Standard for Rating Noise with Respect to Speech Interference*, ANSI S3.14-1977, American National Standards Institute, New York, N.Y., 1977.
3. ASHRAE: *ASHRAE Handbook—1984 Systems*, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, Ga., 1984.
4. Beranek, L. L.: *Noise and Vibration Control*, McGraw-Hill, New York, N.Y., 1971.

Chapter**3.3**
Sound Isolation**Richard G. Cann, Anthony Hoover****3.3.1 Introduction**

Sound may travel to an adjacent space by a multitude of paths. Obviously it can travel through apertures. But if all these holes are sealed, the sound is ultimately received after vibrations have been transmitted through the building structure. Because they travel readily in solid materials, the vibrations may take long, devious paths before arrival. If high noise isolation is required, all paths that the vibration may take need to be interrupted; there is little value in blocking one path when a significant amount of the sound travels through a flanking path.

3.3.2 Sound Barriers

In most building applications involving audio systems a very sizable transmission loss is often required. The performance of the single homogenous wall is inadequate; doubling the mass of the wall gains only 6 dB. Often space or floor-loading restrictions also limit this option. Alternatively, increased performance can be obtained from a two-wall system such as that shown in Figure 3.3.1, in which one wall is completely separated from the other, over its entire area, by an air gap. Reverberant sound within the interior cavity is absorbed by fiberglass. The coincidence dip is reduced by ensuring that the frequency at which it occurs is different for each wall.

Care must be taken not to inadvertently reduce the design transmission loss (TL) by tying the two walls together with a mechanical connection, such as a wall bolt or perhaps a pipe. Also, care must be taken with air leaks. Electrical outlet boxes on opposite sides of the wall must be staggered by at least 3 ft. Conduit and pipes should pass through the wall at its perimeter. The joints between the wall, ceiling, and floor must be grouted or caulked with an elastomeric compound. Never use foam caulk for acoustical isolation. If an air duct must pass through the wall, special arrangements must be made to ensure that sound does not travel into the duct wall, through the wall, and back out through the duct wall into the adjoining space.

3-36 Architectural Acoustic Principles and Design Techniques

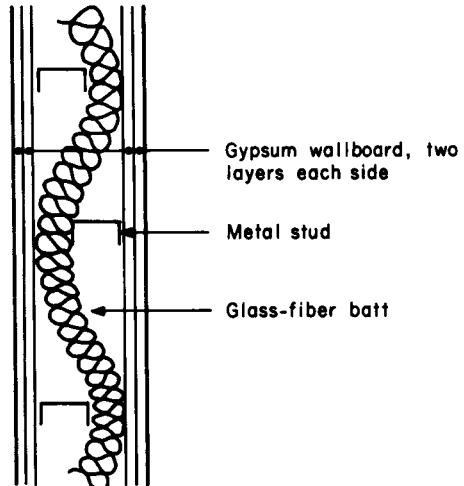


Figure 3.3.1 Section of a lightweight double wall.

3.3.2a Partial Walls and Barriers

The effectiveness of sound barriers depends partly on the amount to which sound traveling over the top or around the sides is diffracted. In practice, the effectiveness is limited, even under the most favorable geometric layout, to about 25 dB. The material of construction does not influence a barrier's performance provided that its transmission loss is sufficient to block significant sound from passing through. For many applications, 1/4-in plywood has sufficient mass and may be used effectively, but other materials are often preferred because they meet other criteria such as weatherability or structural strength.

Where barriers are also used in interior spaces, such as partitions, their performance is degraded by the reflection and reverberation of sound from the adjacent walls and ceiling. In many situations reverberant sound greatly exceeds the diffracted-sound level.

Where barriers fully or partially enclose a source, some additional performance can be gained by reducing the local reverberant field by applying absorptive materials to the inside surface of the partition and other adjacent surfaces.

3.3.2b Doors

The door is potentially the limiting element in noise isolation for an interior wall. Not only may the frame reduce the transmission loss of the mounting wall, but improper sealing of gaskets further reduces performance. Furthermore, a good noise control design does not assume that gaskets will remain good after abuse by substantial traffic. Alternatively, a design with two well-fitting ungasketed doors should be considered. A significant increase in performance can be obtained by placing the doors a minimum of 4 in apart. Further improvements can be made by increasing the separation between the two to form an *air lock* in which acoustically absorptive materials are applied. The close-separation arrangement is awkward to use since the two doors must open in opposite directions. The air-lock design may be more convenient, although it does occupy more space.

3.3.2c Ceilings and Floors

Much of that which applies to walls also applies to ceiling and floor elements, but there are additional precautions to be taken. The addition of carpet to a floor does not decrease the airborne sound transmitted to the space below. However, carpet does reduce the generation of impact noise. Impacts on a hard floor from shoe heels and rumblings from steel-tired hand trucks are readily discernible.

In renovations where it is often the preference to refinish old wooden floors, noise isolation is difficult. The floorboards often have significant cracks, and if—in addition—no solid ceiling is permitted below, the sound transmission loss may be unacceptable.

If the transmission loss of the floor structure is found to be inadequate, it can be increased by applying a ceiling mounted to resilient channels, which in turn are fastened to the joists. Better still, a complete lower ceiling may be supported by resilient hangers which penetrate the upper ceiling to attach to the joists. It may be possible to use the cavity for air ducts as long as duct breakout is not a problem. Otherwise additional noise isolation for the duct must be provided.

3.3.2d Floating Rooms

Where maximum noise isolation is required and cost is of little concern, a room may be “floated” within a building space on vibration isolation pads. All building services are supplied through flexible connections. All doors are double, with the inner door and frame attached only to the floating room. The cavity around the room and under the floor is filled with sound-absorptive material. This type of construction is little used in commercial applications.

3.3.2e Windows

Most window manufacturers can supply TL data on their products. These show that single-glazed windows typically have much less transmission loss than walls, and if they represent more than a small fraction of the wall area they are usually the controlling element when the composite transmission loss of wall and window is calculated. Data also show that the coincidence dip occurs in the midfrequency range, which could be of concern when controlling the compressor whine of a jet engine. However, for glass in which a plastic damping layer is sandwiched between two glass layers, the depth of the coincidence dip is reduced.

Thermal glazing, two glazings with an air space, is also commonly used. However, because the panes are close together, cavity resonance restricts any improved performance. But when the panes are separated by several inches and sound-absorptive material is applied around the cavity perimeter, there is a marked improvement in performance.

The local building code may require that a window with a specified sound transmission class (STC) be installed, but for noise control applications STC values are insufficient. The full spectral data should be used in any computations.

3.3.2f Duct Break-In Break-Out

Mechanical-system ductwork has the potential for reducing the overall transmission-loss integrity of a partition as a result of sound in one space *breaking in* through the sides of the ductwork, traveling through the duct, and *breaking out* of the ductwork into another room. Detailed calcula-

3-38 Architectural Acoustic Principles and Design Techniques

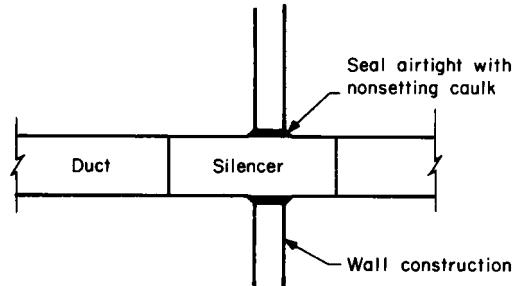


Figure 3.3.2 Installation of a silencer between two rooms.

tions of this effect can be quite complicated, but the effect should be accounted for. The calculation procedure for duct break-in and break-out can be found in [1].

It is good practice to avoid running a duct through a partition which is common to two spaces that require a high degree of sound isolation between them. It is generally advisable to run the main branch of ductwork outside either room under consideration and branch ductwork from the main duct into the appropriate rooms; then, to determine the resultant sound isolation, the maximum sound pressure level expected in one room should be converted to sound power level at the point at which the ductwork serving that room begins. Then the attenuation provided by the ductwork elements between the rooms may be calculated as described above, and the sound power level may be converted to sound pressure level in the other room in question. If it is determined that the resultant sound pressure level will be too high, attenuation devices can be provided in the ductwork. For example, a silencer should be inserted into the ductwork in the manner shown in Figure 3.3.2. Sometimes it is also advisable to box in the ductwork with a construction that consists of several layers of gypsum wallboard to reduce the amount of sound which breaks into a piece of ductwork. It is almost never sufficient to wrap the ductwork with fiberglass materials, and it is rarely sufficient to wrap the ductwork with lead sheet or vinyl sheet.

3.3.2g Site Selection

When building a new facility, it is worthwhile to evaluate in detail the suitability of the site itself for ambient and interfering noise as well as for production and aesthetic requirements. The cost of construction to meet the specified criteria may be much greater at one location than another. For example, the interior noise might be dominated by ground vibration from an adjacent railroad yard. Breaking this vibration path can be very expensive.

An industrial site might appear to be an excellent location to avoid the complaints of neighbors late at night, but the facility may not be usable if punch presses are installed next door. If jet aircraft fly close by, special construction may be required to protect the facility from low-frequency noise. Consequently, it is imperative to check with the local airport authority about flight patterns that are likely with different prevailing winds.

3.3.3 Vibration

Excessive vibration can cause several problems for audio engineers, including reradiation of vibration-induced noise from walls and direct vibration of microphone elements through mike stands. There exist quite a variety of vibration sources, such as mechanical equipment, automobile and truck traffic, and even pedestrian traffic within a building. Once vibration has entered the building structure, it may be difficult to control, and this problem can be exaggerated by resonances found in all buildings, which commonly occur in the range of 5 to 25 Hz. Thus, it is important to try to determine which sources of vibration may be problematical and to isolate them before vibration enters the building structure.

3.3.3a Driving Frequency

The driving frequency of the vibration source is the most important consideration in trying to develop a vibration isolation system. It is not unusual for a source to have several driving frequencies, but it is the lowest driving frequency that is of primary concern. The lowest driving frequency of most electrical and mechanical equipment can be determined from the lowest rotational or vibrational motion. For example, a fan that operates at 1200 rpm has a lowest driving frequency of 20 Hz (this same fan may have a harmonic at 200 Hz if the fan has 10 blades).

To isolate mechanical equipment, some sort of vibration isolation device is placed between all the supports of the piece of equipment and the building structure, such as underneath each of four legs. The vibration isolation elements generally consist of steel springs, some sort of resilient material such as neoprene rubber, or a combination of the two. As the machine is installed, it compresses the mounts by an amount known as the *static deflection*. When installation is complete, it can also be seen to have its own natural frequency. The natural frequency can be thought of as the frequency in hertz at which the machine would oscillate after it was deflected from rest. For example, an automobile with poor shock absorbers (shock absorbers add damping to the vibration isolation system) would continue to bounce up and down at a couple of hertz after it had been driven through a pothole.

The frequency ratio f_r is defined by

$$f_r = f_d/f_n \quad (3.3.1)$$

Where:

f_d = the driving frequency, Hz

f_n = the natural frequency, Hz

The static deflection d and the frequency ratio f_r are related in the simplest case by

$$d = 9.8(f_r / f_d)^2 \quad (3.3.2a)$$

where d is deflection, in, or

$$d = 250(f_r / f_d)^2 \quad (3.3.2b)$$

3-40 Architectural Acoustic Principles and Design Techniques

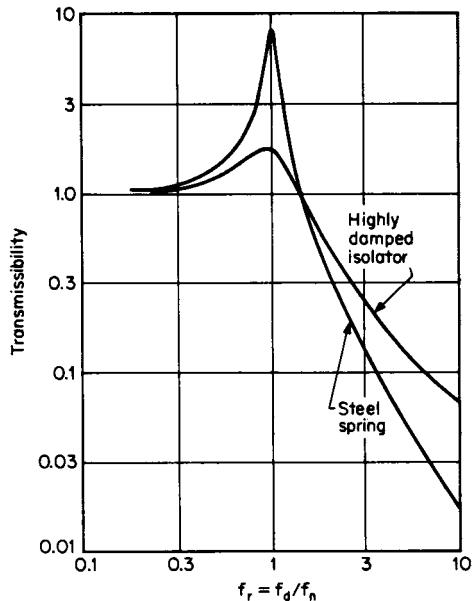


Figure 3.3.3 Transmissibility of a steel spring and a highly damped isolator.

where d is deflection, mm.

3.3.3b Vibration Transmission

Transmissibility rates the effectiveness of the vibration isolation system. Figure 3.3.3 shows a plot of transmissibility against frequency ratio for a steel spring and a highly damped isolator. Steel springs are used where the driving frequency is below 30 Hz. Neoprene is used for higher driving frequencies.

The vibration isolation system is selected according to the static deflection necessary to provide adequate transmissibility. The procedure is as follows:

- Select the material of construction of the mounts for transmissibility according to the application. A transmissibility of 0.03 is generally acceptable. Also, if the equipment to be isolated is to be located in close proximity to a critical space, special designs should be considered.
- Find the value of f_r , the ratio of driving frequency to the natural frequency, which intersects the transmissibility curve at the proper value of transmissibility. For example, in Figure 3.3.3, which shows the transmissibility curve of a steel spring, a minimum value of 8 for f_r is necessary to achieve a transmissibility of 0.03.
- Determine the static deflection from Equation (3.3.2).
- Select from a manufacturer's catalog mounts of the required material of a size that will deflect the calculated amount under the static load of the machine.

3.3.3c Vibration Isolation

It is important to note that several factors can significantly reduce the effectiveness of isolators. Any item that short-circuits the isolator, such as rigid conduit or solid refuse lodged under the equipment, can carry substantial vibrations to the building structure and seriously degrade the performance of an isolation system. In addition, if the isolators stand on a flexible floor span rather than a basement slab, reduced isolation must be expected. In this special situation more sophisticated methods than those outlined here must be used to ensure adequate performance [1].

3.3.4 Interior Acoustics

Previous sections have primarily addressed issues relating to the amplitude of sound, or sound pressure level. Here the quality of sound is considered with primary emphasis on the natural acoustics of an interior space.

Interior acoustics are graphically illustrated here with actual displays from an acoustical scale model in which wall components are progressively assembled to form a complete room. The noise source is a small electric spark that produces a short impulsive sound, and the listener is represented by a special high-frequency microphone. For each model configuration an echogram is made in which the vertical scale is in decibels and the horizontal scale is in milliseconds, with zero time being the instant when the sound is generated at its source. Though a 1:20-scale model was used, all illustrations are indicated for full scale. Similar echo-grams may be made in a building by using a bursting balloon as a sound source, but, of course, it is not as easy to change the room shape rapidly.

When source and microphone are located above a very reflective floor with the angles of incidence and reflection at 45°, the echogram can be seen as the spiky trace shown by the solid line in Figure 3.3.4. The floor echo arrives 10 ms after the direct sound, traveling 2 times farther and also spreading 6 dB per doubling of the distance. With the two impulses arriving at less than 50 ms apart, the ear is unable to resolve the sound into two separate sounds, just hearing one. The second trace, a steplike waveform in the same figure, is calculated in the signal analyzer and is a summation of all the sound energy in arriving echoes. At any point of time it shows the total sound energy that has been received up to that moment. This display is the equivalent steady-state sound level of the impulsive sound. It can be seen from this line that the reflection adds only a step of 2 dB to the level of the direct sound. Subjectively, the echo makes the direct sound seem only just a little louder, for the ear cannot easily detect a 2-dB increase in sound level.

With the erection of a single wall, another impulse appears on the echogram, the reflection from the wall. But Figure 3.3.5 shows that there is yet another echo, one that is the result of the sound traveling in sequence from floor to wall. The three echoes are not equal in strength to the direct sound. Instead, since they are attenuated by spherical spreading, they add merely 4 dB to the direct sound.

One wall may be added to the first wall at right angles as shown in Figure 3.3.6. Now it becomes more complicated to calculate the number of echoes that will appear. In fact, eight spikes are now shown. They can be traced to reveal the direct sound, three single reflections, three double reflections, and one triple reflection. From the steplike trace it can be seen that they add 6 dB to the direct-sound level. An increase in level of this magnitude is readily detected by the ear.

3-42 Architectural Acoustic Principles and Design Techniques

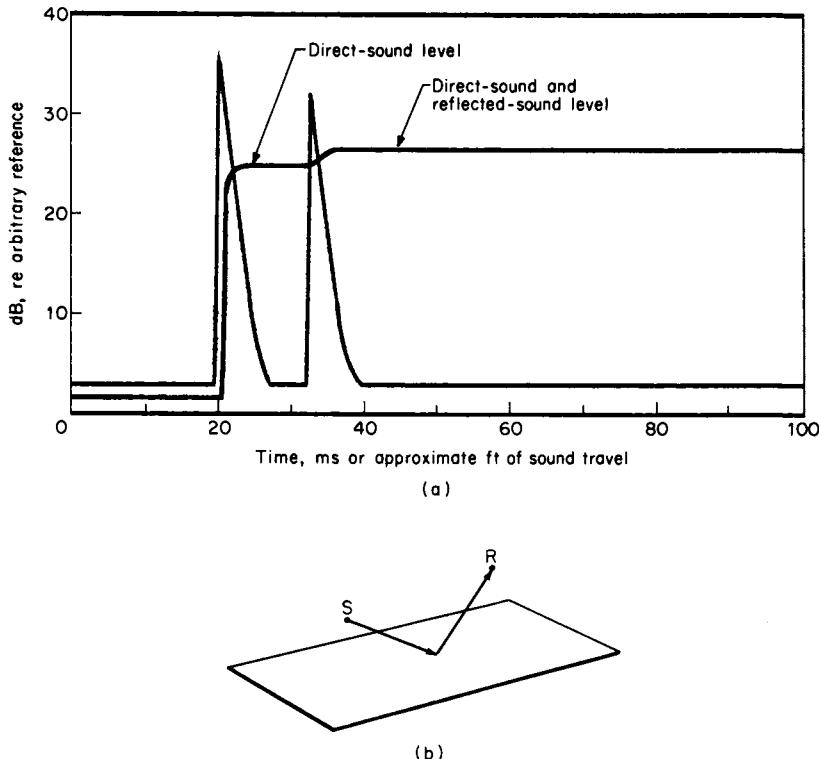


Figure 3.3.4 Echogram of sound reflecting from a hard surface.

As can be seen from the echogram in Figure 3.3.6, this initial package of sound echoes typically constitutes the bulk of the sound energy that arrives at the listener in a completed room. The initial echo structure is important in the way in which the listener perceives the sound. If this package of early arrivals is no more than 50 ms wide, it will be perceived by the listener as being the direct sound. If it is stretched out, the music will lose precision and definition. In smaller, narrower halls, the sidewalls provide early reflections, but in a large fan-shaped hall lateral reflections from walls arrive too late to help. Likewise the ceiling often cannot contribute because it is too far away. Thus to provide these early reflections, “clouds” are often hung above the stage and angled for sound dispersion.

Alternatively a parallel wall may be added as shown in Figure 3.3.7. This gives an echogram reminiscent of sneaking a sideways view between parallel mirrors. The images extend with decreasing strength to infinity. They are not evenly spaced but paired because the sound travels farther in its reflection from one wall than the other. This echogram depicts the phenomenon known as *flutter*, in which sound most unpleasantly zings between parallel walls. Care must be taken in room design to prevent flutter between windows and the opposite wall or between a hard ceiling and floor. One of the pair of reflective surfaces must be made absorptive or be angled to diffuse the sound.

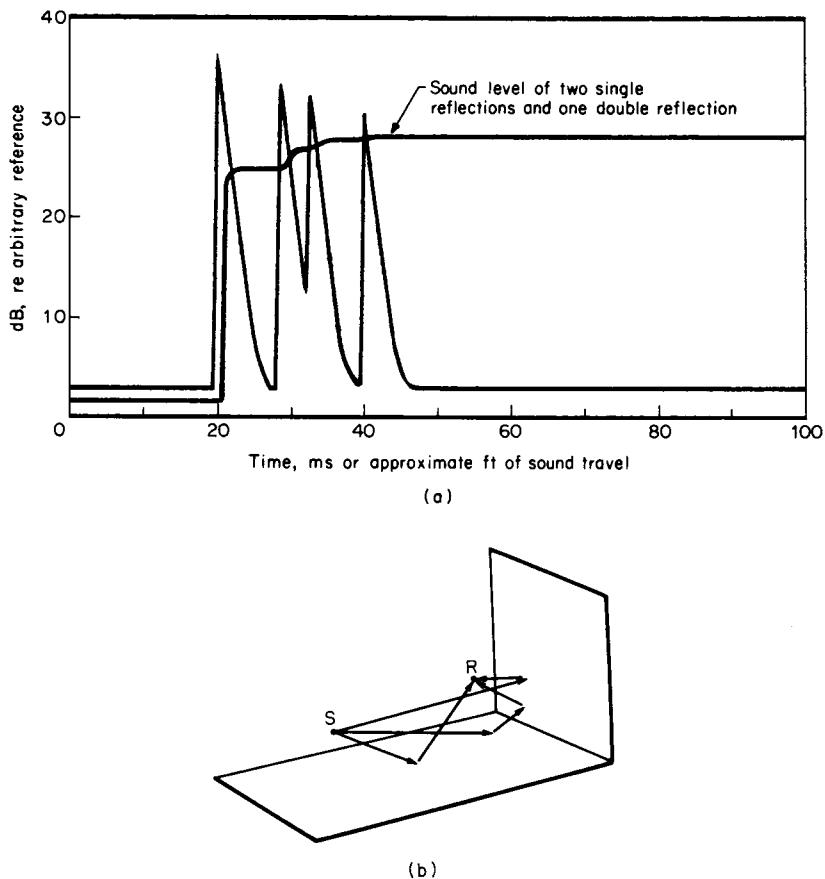


Figure 3.3.5 Echogram of sound reflecting from one wall and a floor.

With the first wall replaced so that three walls are now standing, the configuration in Figure 3.3.8 is similar to a complete room with an acoustically dead end, a live end, and a dead ceiling. The previous clear infinite string of images is now concealed among many other echoes, making it almost impossible to identify the path of each, particularly as many echoes overlap each other. The echogram is reminiscent of a reverberation decay curve. However, its unevenness reinforces the concern that reverberation-time measurements within such a room can lead to erroneous results.

As the remaining wall and ceiling are added, the echogram becomes more continuous until it is filled in with echoes. Figure 3.3.9 shows the fully reverberant echogram of the complete space. The reverberation time is 3 s. Note that the final sound level is 34 dB. This sound level is reached substantially in just 10 percent of the reverberation time.

It is important that the later-arriving echoes do not “muddy” or mask the earlier sound. Thus, to maintain good sound definition for music, the sound energy arriving after the first 50 ms should increase the level by approximately 3 dB. In Figure 3.3.9 note that the sound level

3-44 Architectural Acoustic Principles and Design Techniques

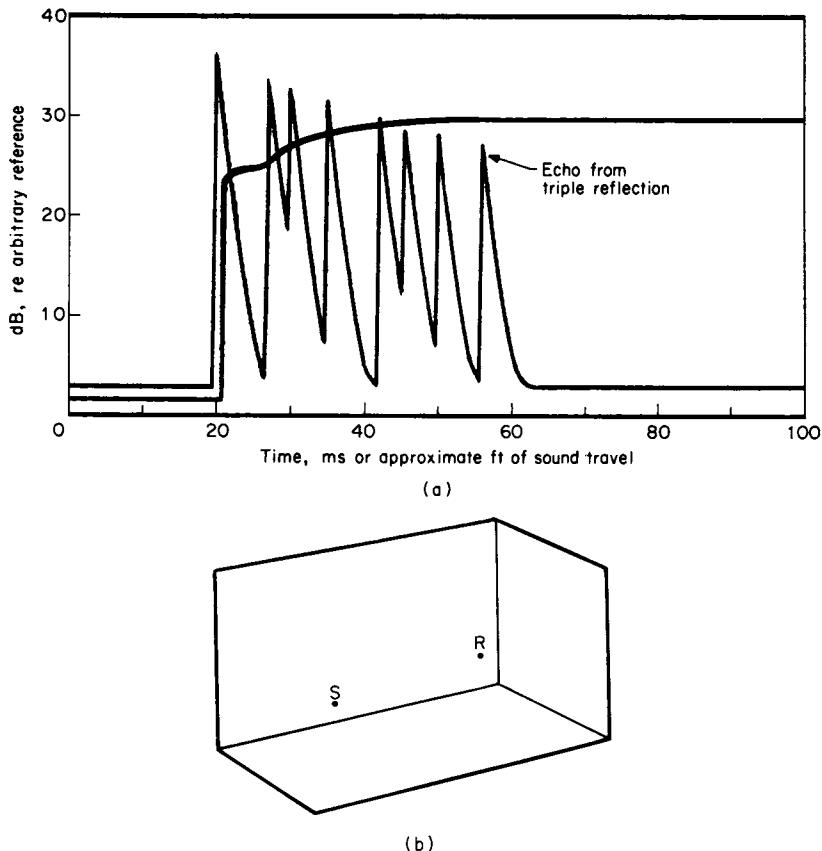


Figure 3.3.6 Echogram of sound reflecting from a floor and two walls at right angles.

increases by 7 dB after the early arrivals, showing too much reverberation for music. For clear speech, the increase should be approximately 1 to 2 dB above the earlier sound energy. A more in-depth treatment of preferred relationships between early- and late-arriving sound energy can be found in [2 and 3].

It is most desirable to have the echoes decaying with gradually decreasing amplitude. A discrete echo can be caused by *focusing*, as when sound reflects from a curved wall in the rear of a fan-shaped room. For some listeners the focused sound may arrive well after the early arrivals and be clearly identifiable. For a rear wall, this effect is usually controlled by absorbing the sound at the curved wall. However, the sound may be diffused by installing angled reflecting surfaces which each have dimensions greater than the wavelength of the sound. The placement and angles will need careful determination so that other focusing is not formed.

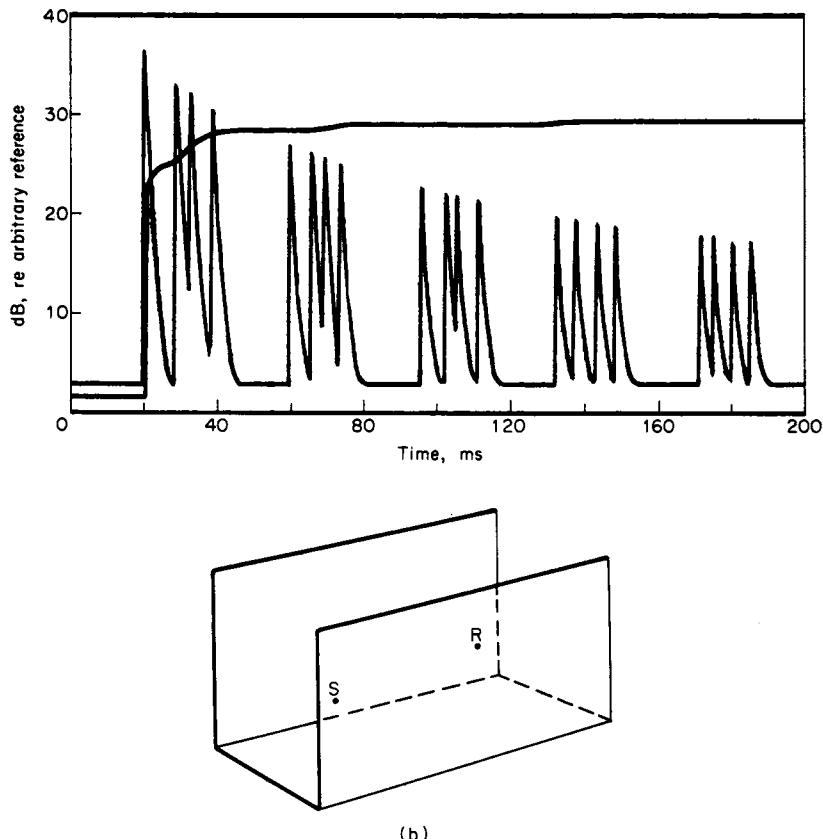


Figure 3.3.7 Echogram of flutter between parallel walls.

3.3.4a Design Concerns of Spaces

It is rare that an auditorium is used for a single purpose; optimum acoustical requirements vary widely for each type of usage. For example, reverberation time may require changing over a range of 1 to 2 s. Some hall designs have used revealable acoustical absorption to attempt this change; solid wall panels open up so that fabric panels can be extended. This may be an unwieldy solution, for not only does it require large dedicated wall spaces, but it may also kill desirable lateral reflections that enhance the early-arriving sound.

Alternatively, significant gains can be made through the application of directional loudspeakers for those activities that require a shorter reverberation time, such as speech. By aiming the majority of the sound energy at the audience, the proportion of the energy in the early-arriving sound is substantially increased compared with the sound energy entering the reverberant field, thus increasing intelligibility. Be warned that directional speakers cannot be used to full effectiveness in spaces where the basic sound absorption has not been applied.

3-46 Architectural Acoustic Principles and Design Techniques

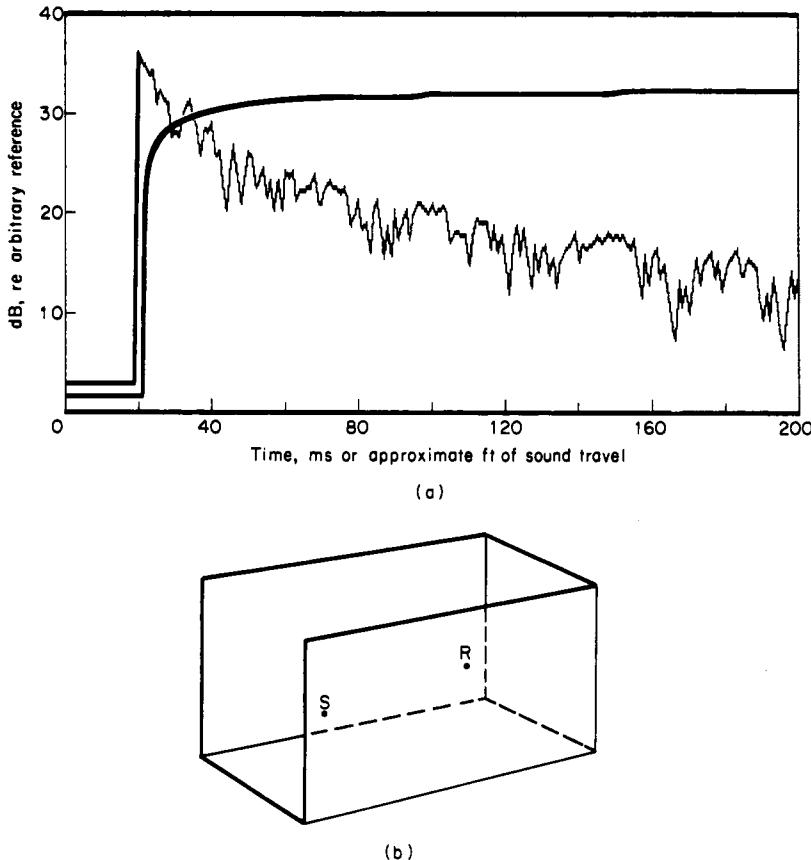


Figure 3.3.8 Echogram of partial reverberation in an incomplete room.

Another choice is to build a short reverberation time into the basic acoustics of the space; then as each application warrants, the reverberant field is reinforced with a sound system that amplifies the reverberant-sound field. One advantage of these electronic methods is that sound can be readily equalized for each application over the full frequency range; in contrast, this is almost impossible with mechanical devices. These electronic techniques may be applied strategically only to selected areas of a hall, such as areas deep under balcony seating, where local reverberation is restricted.

Small Rooms

When the dimensions of a space become comparable with the wavelength of the sound, simple acoustic methods are inappropriate because it no longer is possible to access the properties of the whole space by using the available properties of component parts. More confidence can be placed in the design of acoustic properties of an interior space when the space has dimensions

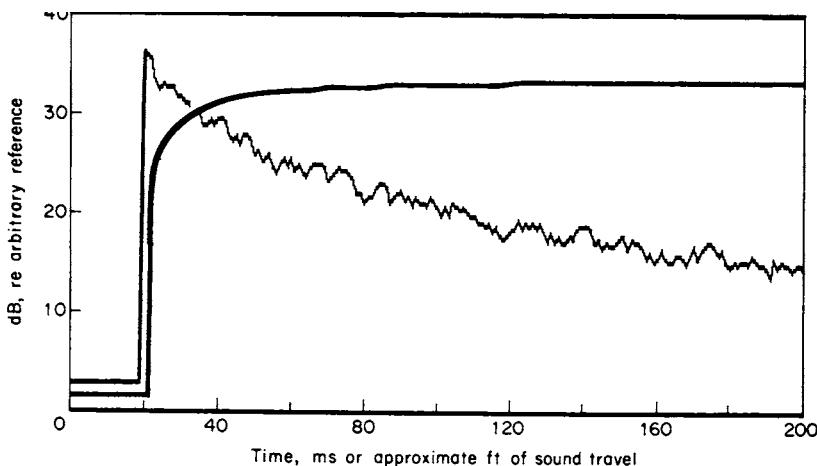


Figure 3.3.9 Echogram of reverberation in a complete room.

much larger than the wavelength of the sound, where, for example, flat surfaces reflect sound with the angle of incidence equal to the angle of reflection. Sound energy pumped into a small space may influence the production of the immediately following sound. Thus, as the room becomes smaller, it becomes less possible to predict the outcome accurately. Surfaces are smaller, and the popular sound-ray diagram is entirely inappropriate, especially for low frequencies. Also, the smaller-angled surfaces can become stiffer and the construction becomes less uniform, making any mathematical description of the environment very difficult.

For these reasons, no specific recommendations can be made, for each component is an integral part of a close-coupled system in which each component reacts with all others. However, two useful observations may be made:

- If an acoustically symmetrical room is required at low frequencies, it should have symmetrical construction well beyond the interior walls. A room with a gypsum wall tied to a cinder-block wall on one side is not the same as a freestanding wall on the other. They may appear to be the same, but they are not the same acoustically.
- A variety of construction techniques, each of a different surface weight and stiffness, can improve the uniformity of the response of the room throughout its frequency range.

When the acoustics of a space needs correction, many complex techniques are available to track down the problem. Besides microphone systems, accelerometers can be used to analyze wall vibration levels in conjunction with the sound pressure level.

3.3.4b Masking

Masking is used in spaces in which sound sources need concealing. For example, in an open administrative office where the conversations of others are readily overheard, the noise criterion may be less than 35. But rather than applying noise controlling devices and destroying the func-

3-48 Architectural Acoustic Principles and Design Techniques

tionality of the space, the interior noise level may be raised to mask the conversations. This is achieved by installing an array of small loudspeakers inside the drop ceiling that uniformly distribute broadband noise throughout the plenum to the space below. Provision is made within the master broadband noise generator for adjustment of the spectral shape so that maximum speech masking can be obtained with a minimum amount of obvious sound intrusion. Even so, it is preferable to introduce such systems to the office environment slowly by increasing the sound level gradually over a period of weeks until the goal is reached. These units may also be installed above corridors to prohibit eavesdropping on conversations in adjacent rooms. In sleeping areas masking will blot out intrusive sounds.

There is a limit to the application of masking systems. The masking noise levels cannot exceed NC 40 without being regarded as a noise intrusion.

3.3.5 Annoyance

Annoyance is very difficult to quantify. But often when sound becomes noise, pleasure becomes annoyance. Each person has his or her own way of reacting to the wide variety of sounds. Here are some factors that influence the degree to which people are annoyed.

- *Source of sound:* An individual will react differently to the same noise coming from different sources. The noise of a neighbor cutting a lawn is annoying. On the other hand, the noise from cutting one's own lawn is acceptable, while the noise of a neighbor cutting your lawn is music.
- *Benefit of sound:* If the sound is the result of an activity that brings economic benefit to the listener, tolerance increases. A town official supports renting the local stadium to a rock group because it will provide significant income, but residents fear outsiders' causing damage and congestion.
- *Adaptation:* People become used to certain noises. A freight train passes each night 50 ft from a sleeping family. It passes undetected. For the same family, a night in the country is disturbingly quiet.
- *Impulsive noise:* Impulsive noise is more annoying than steady noise, particularly if it occurs at unpredictable intervals. The slamming of automobile doors as people arrive and depart often causes complaints.
- *Tonal noise:* If the sound contains a tonal component such as a whine, buzz, or hum, it is more annoying than broadband noise of the same loudness. Many state and local governments recognize this by lowering the permissible noise limits for such sounds. A rooftop transformer and an exhaust air blower are typical offenders.
- *Variability:* Sounds which vary more in amplitude with time are more annoying than steady sounds. The greater the statistical standard deviation of the sound, the more annoying it is.
- *Speech interference:* Typical speech has a level of 60 dBA at the listener's ear. Other sounds may mask speech so that it becomes unintelligible. For example, the noise of a nearby heat pump may make speech unintelligible on the neighbors' patio. Special procedures to determine the intelligibility of speech may be found in [5].

In some instances, however, evaluation of a noise complaint in terms of the listed items may show minimal impact. But careful investigation may further show that other important issues

hide behind protests. For example, residents who live near an outdoor amphitheater may be more concerned about property damage, trespassing, and parking-lot activities than about noise. It is therefore important to consider all aspects of the operation of a facility.

3.3.6 References

1. ANSI: *American National Standard for Rating Noise with Respect to Speech Interference*, ANSI S3.14-1977, American National Standards Institute, New York, N.Y., 1977.
2. ASHRAE: *ASHRAE Handbook—1984 Systems*, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, Ga., 1984.7.Marshall, Harold, and M. Barron: "Spatial Impression Due to Early Lateral Reflections in Concert Halls: The Derivation of the Physical Measure," *JSV*, vol.77, no. 2, pp. 211-232, 1981.
3. Siebein, Gary W.: *Project Design Phase Analysis Techniques for Predicting the Acoustical Qualities of Buildings*, research report to the National Science Foundation, grant CEE8307948, Florida Architecture and Building Research Center, Gainesville, Fla., 1986.
4. ANSI: *Method for the Measurement of Monosyllabic Word Intelligibility*, ANSI S3.2-1960, rev. 1977, American National Standards Institute, New York, N.Y., 1976.

3.3.7 Bibliography

- Beranek, L. L.: *Acoustics*, McGraw-Hill, New York, N.Y., 1954.
- Egan, M. D.: *Concepts in Architectural Acoustics*, McGraw-Hill, New York, N.Y., 1972.
- Huntington, W. C., R. A. Mickadeit, and W. Cavanaugh: *Building Construction Materials*, 5th ed., Wiley, New York, N.Y., 1981.
- Jones, Robert S.: *Noise and Vibration Control in Buildings*, McGraw-Hill, New York, N.Y., 1980.
- Kryter, K. D.: *The Effects of Noise on Man*, Academic, New York, N.Y., 1985.
- Lyon R. H., and R. G. Cann: *Acoustical Scale Modeling*, Grozier Technical Systems, Inc., Brookline, Mass.
- Marris, Cyril M.: *Handbook of Noise Control*, 2nd ed., McGraw-Hill, New York, N.Y., 1979.
- Morse, P. M.: *Vibration and Sound*, American Institute of Physics, New York, N.Y., 1981.
- Talaske, Richard H., Ewart A. Wetherill, and William J. Cavanaugh (eds.): *Halls for Music Performance Two Decades of Experience, 1962-1982*, American Institute of Physics for the Acoustical Society of America, New York, N.Y., 1982.

Section
4

Microphone Devices and Systems

Every electronic system has a starting point, and in the case of audio, that point is usually a microphone. The source origination equipment available for a well-equipped audio studio range from traditional mics to special-purpose devices intended to capture sounds in difficult environments. Microphones are transducers—nothing more, nothing less. No matter how large or small, elaborate or simple, expensive or economical a microphone might be, it has only one basic function: to convert acoustical energy to electrical energy.

With this fundamental point clearly established, you might wonder why microphones exist in such a mind-boggling array of sizes, shapes, frequency response characteristics, and element types. The answer is simple. Although the basic function of all microphones is the same, they have to work in many different applications and under various conditions.

Choosing the right microphone for a particular application might seem as easy as falling off a log, but it is a decision that deserves considerable thought. Just as no two production sessions are alike, the microphone requirements are varied also.

Microphone manufacturers offer a selection of units to match almost any application. With a good working knowledge of the various microphone designs, choosing the right mic for the job becomes a much simpler task. The education process begins with a look at some of the microphones commonly in use today.

In This Section:

Chapter 4.1: Microphones	4-7
Introduction	4-7
Pressure Microphones	4-7
Piezoelectric Microphone	4-9
Electrostatic (Condenser) Microphones	4-10
Electret Microphone	4-12
Boundary Microphone	4-14
Electrodynamic Microphones	4-15
Pressure-Gradient (Velocity) Microphones	4-17
Combination Pressure and Pressure-Gradient Microphones	4-20
Frequency Response as a Function of Distance	4-21

4-2 Section Four

Dual-Diaphragm Condenser Polydirectional Microphone	4-21
Twin-Cardioid-Element Polydirectional Condenser Microphone	4-23
Single-Element Unidirectional Microphones	4-24
Unidirectional Condenser Microphone	4-26
Moving-Coil Unidirectional Microphone	4-27
Variable-Distance Unidirectional Microphone	4-27
Ultra-Directional Microphones	4-29
Line Microphone	4-29
Wave Microphones	4-31
Miscellaneous Types of Microphones	4-31
Sound-Field Microphone	4-31
Lavalier Microphone	4-33
Wireless Microphone	4-33
Selecting Microphone Types	4-33
References	4-34
Bibliography	4-35
Chapter 4.2: Stereophonic Techniques	4-37
Introduction	4-37
Two-Microphone Coincident Techniques	4-37
XY Cardioids and Hypercardioids	4-37
Blumlein	4-38
MS Stereo	4-39
Two-Microphone Near-Coincident Techniques	4-40
Two-Microphone Spaced Techniques	4-41
Spaced Omnidirectional Microphones	4-43
Spaced Cardioid Microphones	4-43
Spaced Bidirectional Microphones	4-44
Spaced Hypercardioid Microphones	4-44
Performance Considerations	4-44
References	4-45
Chapter 4.3: Low Power Amplifiers	4-47
Introduction	4-47
Single-Stage Transistor or FET Amplifier	4-47
DC Conditions	4-47
Input and Output Impedance, Voltage, and Current Gain	4-49
AC Gain	4-49
Common-Base or Common-Gate Connection	4-50
Common-Collector or Common-Drain Connection	4-51
Bias and Large Signals	4-52
Multistage Amplifiers	4-53
DC-Coupled Multistage Amplifiers	4-53
Cascaded Transistors	4-55
Parallel-Connected Devices for High Currents	4-56
Series-Connected Devices for High Voltage	4-57
AC-Coupled Multistage Amplifiers	4-57
Power Output Stages	4-58

Microphone Devices and Systems 4-3

Single-Ended Amplifiers	4-58
Push-Pull Amplifiers	4-58
Parallel-DC, Series-AC Amplifiers	4-59
Series-DC, Parallel-AC Amplifiers	4-59
Full-Bridge Amplifiers	4-60
Classes of Amplifiers	4-60
Gain Block and the Operational Amplifier	4-62
Feedback and Feed Forward	4-63
Linear Feedback	4-64
Feed Forward	4-65
Nonlinear Feedback	4-65
Voltage Feedback	4-65
Current Feedback	4-66
Output and Input Impedance	4-66
Feed Forward and Correction of Estimated Errors	4-67
Differential Amplifier	4-67
References	4-68
Bibliography	4-68

Reference Documents for this Section: _____

- “A Phased Array,” *Hi Fi News Record Rev.*, July 1981.
- Bevan, W. R., R. B. Schulein, and C. E. Seeler: “Design of a Studio-Quality Condenser Microphone Using Electret Technology,” *J. Audio Eng Soc. (Engineering Reports)*, vol. 26, pp. 947–957, December 1978.
- Black, H. S.: U.S. Patent 2,102,671.
- Blumlein, A.: British Patent 394,325, December 14, 1931; reprinted in *J. Audio Eng. Soc.*, vol. 6, April 1958.
- Dooley, W. L., and R. D. Streicher: “M-S Stereo: A Powerful Technique for Working in Stereo,” *J. Audio Eng. Soc.*, vol. 30, pp. 707–718, October 1982.
- Eargle, J.: *Sound Recording*, Van Nostrand Reinhold, New York, 1976.
- Eargle, J.: *The Microphone Handbook*, Elar Publishing, Plainview, N.Y., 1981.
- Fewer, D. R.: “Design Principles for Junction Transistor Audio Power Amplifiers,” *Trans. IRE PGA*, AU-3(6), November–December 1955.
- Fredericksen, E., N. Firby, and H. Mathiasen: “Prepolarized Condenser Microphones for Measurement Purposes,” *Tech. Rev.*, Brüel & Kjaer, Copenhagen, no.4, 1979.
- Garner, L. H.: “High-Power Solid State Amplifiers,” *Trans. IRE PGA*, 15(4), December 1967.
- Gordon, J.: “Recording in 2 and 4 Channels,” *Audio*, pp. 36–38, December 1973.
- Harper, C. A. (ed.): *Handbook of Components for Electronics*, McGraw-Hill, New York, N.Y., 1977.
- Instruction Book for RCA BK-16A Dynamic Microphone*, IB-24898, Radio Corporation of America, Camden, N.J.

4-4 Section Four

- Killion, M. C., and E. V. Carlson: "A Subminiature Electret-Condenser Microphone of New Design," *J. Audio Eng. Soc.*, vol. 22, pg. 237–243, May 1974.
- Kirchner, R. J.: "Properties of Junction Transistors," *Trans. IRE PGA*, AU-3(4), July-August 1955.
- Kishi, K., N. Tsuchiya, and K. Shimura: "Unidirectional Microphone," U.S. Patent 3,581,012, May 25, 1971.
- Kubota, H.: "Back Electret Microphones," presented at the 55th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 24, no. 862, preprint 1157, December 1976.
- Lipshitz, S. P.: "Stereo Microphone Techniques: Are the Purists Wrong?" presented at the 78th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 33, pg. 594, preprint 2261, July-August 1985.
- Long, E. M., and R. J. Wickersham: "Pressure Recording Process and Device," U.S. Patent 4,361,736, November 30, 1982.
- Lynn, D. K., C. S. Meyer, and D. C. Hamilton (eds.): *Analysis and Design of Integrated Circuits*, McGraw-Hill, New York, N.Y., 1967.
- Microphones—Anthology*, Audio Engineering Society, New York, 1979.
- "Miking with the 3-Point System," *Audio*, pp. 28–36, December 1975.
- Nisbett, A.: *The Technique of the Sound Studio*, Hastings House, New York, N.Y., 1974.
- Olson, H. F.: *Acoustical Engineering*, Van Nostrand, Princeton, N.J., 1957.
- Olson, H. F.: "Directional Electrostatic Microphone," U.S. Patent 3,007,012, October 31, 1961.
- Olson, H. F. (ed.): *McGraw-Hill Encyclopedia of Science and Technology*, 5th ed., vol. 18, McGraw-Hill, New York, N.Y., pg. 506, 1982.
- Olson, H. F.: *Music, Physics, and Engineering*, 2d ed., Dover, New York, N.Y., 1967.
- Olson, H. F.: "Ribbon Velocity Microphones," *J. Audio Eng. Soc.*, vol. 18, pp. 263–268, June 1970.
- Petersen, A., and D. B. Sinclair: "A Singled-Ended Push-Pull Audio Amplifier," *Proc. IRE*, vol. 40, January 1952.
- Rasmussen, G.: "A New Condenser Microphone," *Tech. Rev.*, Brüel & Kjaer, Copenhagen, no. 1, 1959.
- Sank, J. R.: "Equipment Profile-Nakamichi CM-700 Electret Condenser Microphone System." *Audio*, September 1978.
- Shockley, W.: "A Unipolar Field-Effect Transistor," *Proc. IRE*, vol. 40, November 1952.
- Shockley, W.: "The Theory of P-N Junctions in Semiconductors and P-N Junction Transistors," *Proc. IRE*, vol. 41, June 1953.
- Trent, R. L.: "Design Principles for Transistor Audio Amplifiers," *Trans. IRE PGA*, AU-3(5), September–October 1955.
- Walker, P. J.: "A Current Dumping Audio Power Amplifier," *Wireless World*, December 1975.

Microphone Devices and Systems 4-5

Weinberg, L.: *Network Analysis and Synthesis*, McGraw-Hill, New York, N.Y., 1967.

Widlar, R. J.: "A Unique Current Design for a High Performance Operational Amplifier Especially Suited to Monolithic Construction," *Proc. NEC*, 1965.

Woszczyk, W. R.: "A Microphone Technique Employing the Principle of Second-Order Gradient Unidirectionality," presented at the 69th Convention of the Audio Engineering Society., *J. Audio Eng. Soc. (Abstracts)*, vol. 29, pg. 550, preprint 1800, July-August 1981.

Chapter**4.1**
Microphones

Jon R. Sank, Ronald D. Streicher, Wesley L. Dooley

4.1.1 Introduction

A microphone is an electroacoustic device containing a transducer which is actuated by sound waves and delivers essentially equivalent electric waves. The classes of microphones include pressure, pressure-gradient (velocity), combination pressure and pressure-gradient, and wave-interference. The electrical response of a pressure microphone results from pressure variations in the air. The directional (polar) pickup pattern is omnidirectional (nondirectional) because sound pressure is a scalar quantity which possesses magnitude but no direction. The electrical response of a velocity microphone results from variations in the particle velocity of the air. The polar pattern is bidirectional (cosine or figure-of-eight) because particle velocity is a vector quantity which possesses magnitude and direction. The electrical response of the combination pressure and pressure-gradient microphone is also proportional to the particle velocity. The polar pattern may be cardioid, hypercardioid, or of a similar cosine-function limacon shape and may be fixed or adjustable.

A particular class of microphones may include one of the following types of transducers: carbon, ceramic, condenser, moving-coil, inductor, ribbon, magnetic, electronic, or semiconductor.

The functioning of various types of microphones is described in this chapter by reference to the equivalent circuits of the acoustical and mechanical systems. The mechanical equivalent circuit is considered, for simplicity, when the discussion involves mathematical equations. In other instances, the discussion omits mathematics, and the acoustical network affords the clearest illustration of operating principles.

4.1.2 Pressure Microphones

A carbon microphone depends for its operation on the variation of resistance of carbon contacts. The high sensitivity of this microphone is due to the relay action of the carbon contacts. It is widely used in telephone communications. This is true because the high sensitivity eliminates the need for audio amplification in a telephone set. Restricted frequency range, distortion, and carbon noise limit the application of the carbon microphone in other than voice-communications applications.

4-8 Microphone Devices and Systems

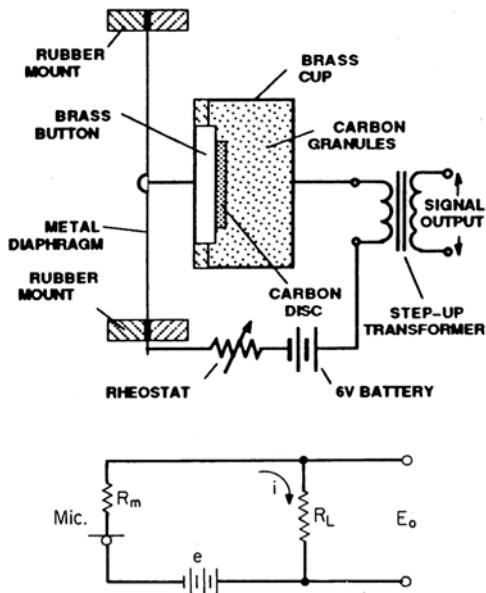


Figure 4.1.1 Carbon microphone and equivalent electric circuit.

A typical single-button carbon microphone and its electric circuit are shown in Figure 4.1.1. The carbon transducer consists of a contact cup filled with carbon granules, which are usually made from anthracite coal [1]. The granules make contact with the electrically conductive diaphragm via the contact button on the diaphragm. The diaphragm is frequently made from a thin sheet of aluminum alloy. The periodic displacement of the diaphragm causes a variation in mechanical pressure applied to the carbon granules. This results in a periodic variation in electric resistance from the diaphragm to the contact cup. For small displacements, the variation in resistance is proportional to the displacement.

The output voltage is given by

$$E_0 = \frac{eR_L}{(R_m + R_L) + (hx \sin \omega t)} \quad (4.1.1)$$

Where:

e = dc voltage of bias source

h = constant of carbon element, Ω/cm

x = amplitude of diaphragm, cm

$\omega = 2\pi f$

f = frequency, Hz

The useful audio output is, of course, the ac component of E_0 . Equation (4.1.1) may be expanded ([2], pg. 248) to show that the ac component consists of harmonics at $f, 2f, \dots$, which means that the carbon transducer has intrinsic distortion. For a limited frequency range of reproduction, the distortion is not objectionable.

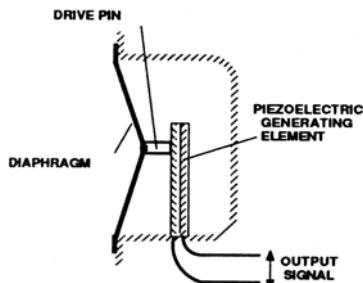


Figure 4.1.2 Typical construction of a ceramic microphone.

The large second-harmonic distortion component can be eliminated through use of two carbon buttons in push-pull. It was used in the 1920s for broadcasting but was replaced by condenser, ribbon, and dynamic microphones. Although the double-button microphone has a wide-range frequency response and low distortion, it and the single-button types suffer from carbon compaction and carbon noise. These effects mean that the signal-to-noise ratio or dynamic range of the microphone is variable. Repeatability of frequency response, sensitivity, and noise measurements of carbon microphones are very poor.

For improved performance in telephone and speech communications, carbon microphones have largely been replaced by dynamic, magnetic, and electret condenser microphones, which have built-in amplifiers. These amplifiers are powered by the direct current normally provided by the communications equipment for carbon microphones. These *carbon replacements* may offer noise-canceling features as well as improved frequency response and low distortion and noise. They are offered as replacement cartridges for telephone handsets, in replacement handsets, in hand-held microphones, and in headsets.

4.1.2a Piezoelectric Microphone

The *piezoelectric microphone* contains a transducer element that generates a voltage when mechanically deformed. The voltage is proportional to the displacement in the frequency range below the resonance of the element. Rochelle salt crystals were used prior to 1960 but were sensitive to humidity and heat. Newer ceramic materials such as barium titanate and lead zirconate titanate are more resistant to environmental extremes and have replaced the Rochelle salt crystals. There are two general classifications of ceramic microphones: direct-actuated and diaphragm-actuated. Directly actuated transducers consist of stacked arrays of bimorph crystals or *sound cells*.

Figure 4.1.2 shows the most common construction in use today for a ceramic microphone. The element is mounted as a cantilever and actuated by the diaphragm via the drive pin. The diaphragm is frequently made from thin aluminum sheet, although polyester film may also be used. The impedance of the ceramic microphone is capacitive on the order of 500 to 1000 pF. This permits use of a short length of cable with only a small loss in output level. The advantage of the ceramic microphone is that the output voltage is sufficient to drive a high-impedance input of an amplifier directly. The frequency response (with a very high input resistance) is uniform from a very low frequency up to the transducer resonance, which may be situated at 10,000 Hz or higher. The sensitivity and the frequency response are stable with time and over a wide range of

4-10 Microphone Devices and Systems

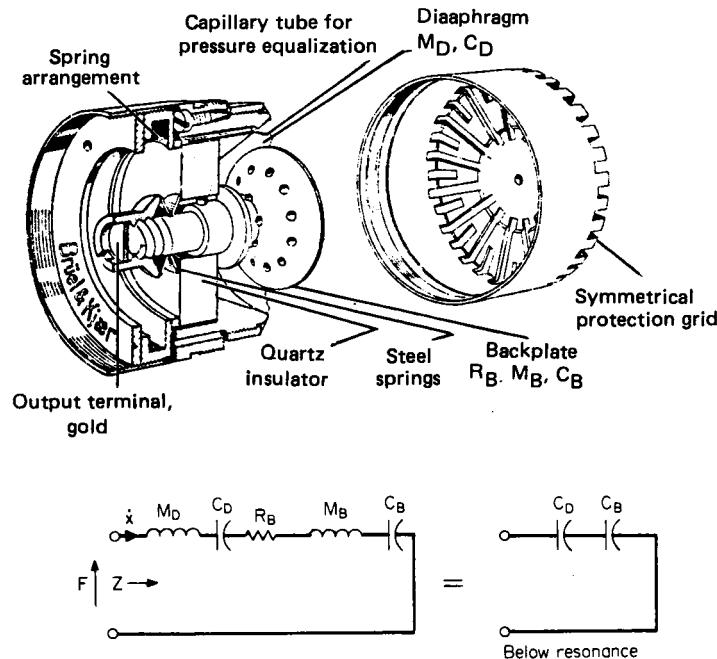


Figure 4.1.3 Condenser pressure microphone and mechanical network. (From [3]. Used with permission.)

temperature and humidity. The cost is relatively low. Therefore, the ceramic microphone was widely used with tube-type home tape recorders and low-cost communications equipment. With the advent of solid-state equipment, low-impedance microphones were needed and the ceramic microphone has since been replaced by inexpensive moving-coil (dynamic) microphones or electret condenser microphones, which typically include integral field-effect transistor (FET) preamplifiers that convert their output to low impedance.

The piezoelectric diaphragm transducer is a variation on the basic theme. A thick or thin film of the polymer polyvinylidene fluoride (PVF_2) may be processed to form a piezoelectric element. As with the ceramic element, it must be provided with plated-on output terminals.

4.1.2b Electrostatic (Condenser) Microphones

A *condenser microphone* depends for its operation on variations in its internal capacitance. Figure 4.1.3 shows the capsule of an omni-directional pressure-sensing condenser microphone [3]. Condenser microphones are divided into two classes: externally polarized (air condenser) and prepolarized (electret condenser). The function of the polarizing voltage or its equivalent is to translate the diaphragm motion into a linearly related audio output voltage, which is amplified by a very-high-impedance FET preamplifier, which must be located close to the capsule. Alter-

nately, the capacitance variation can be used to frequency-modulate a radio-frequency (RF) oscillator.

The diaphragm of this microphone is a thin membrane of nickel that is spaced about 0.001 in (25 μm) from the backplate. Because the electroacoustical sensitivity is inversely proportional to the spacing d , special measures must be taken to prevent this distance from changing because of temperature. The laboratory-grade microphone of Figure 4.1.3 is made almost entirely of nickel and nickel alloys and has nearly constant sensitivity from 20 to 150° C.

The performance may be determined by consideration of the mechanical network (Figure 4.1.3). The resonance is placed at the high end of the usable frequency range. The backplate air load includes mass M_B , compliance C_B , and resistance R_B . M_B and C_B plus the diaphragm mass M_D and compliance D_D determine the resonance frequency. R_B provides damping of the resonance. Below the resonance frequency, the microphone is stiffness-controlled (reciprocal of compliance) and only C_D and C_B appear in the circuit. The open-circuit output voltage E is given by [2] and [4].

$$E = \frac{E_0}{d}x \quad x = \frac{\dot{x}}{j\omega} \quad (4.1.2)$$

Where:

E_0 = polarizing voltage (or equivalent voltage for electrets)

d = spacing from diaphragm to backplate, m

x = diaphragm displacement, m

\dot{x} = diaphragm velocity, m/s

$\omega = 2\pi f$

f = frequency, Hz

The velocity is given by

$$\dot{x} = \frac{F}{Z} = \frac{PA}{(1/j\omega)(1/C_D + 1/C_M)} \quad (4.1.3)$$

Where:

F = force on diaphragm, N

P = sound pressure on diaphragm, N/m²

A = area of diaphragm, m²

Z = mechanical impedance system, mechanical ohms

The output voltage is obtained by combining Equations (4.1.2) and (4.1.3).

$$E = \frac{E_0 PA}{d \left(\frac{1}{C_D} + \frac{1}{C_B} \right)} \quad (4.1.4)$$

This means that below resonance the response is independent of frequency.

4-12 Microphone Devices and Systems

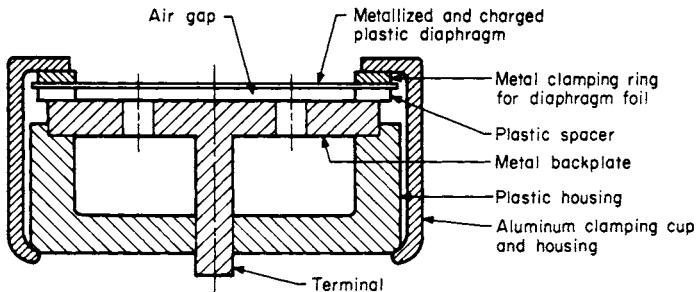


Figure 4.1.4 Typical design of an electret capsule with a charged foil diaphragm. (From [7]. Used with permission.)

The polarization field strength for most condenser microphones, independent of the polarization principle, is on the order of $100,000 \text{ V/cm}^3$ so that the slightest bit of contamination between diaphragm and backplate will cause impulsive noise due to arcing. Microphones used in corrosive environments may develop pinholes in the diaphragm, and the resulting corrosion behind the diaphragm eventually may short-circuit the transducer. Normally, impulsive noise is caused by humidity, which can be eliminated by desiccation.

4.1.2c Electret Microphone

The simplest type of electret microphone is the charged diaphragm type. This is illustrated in Figure 4.1.4. The spacing between diaphragm and backplate is exaggerated for clarity. Figure 4.1.5 shows a schematic of the foil electret with the electric charge distribution illustrated. The electret foil is selected as a compromise between good electret properties and good mechanical properties as a diaphragm. Polymer materials such as polyacrylonitrile, polycarbonate, and some fluoroc resins are examples of suitable plastic films used as electret diaphragms.

There are several methods of making an electret. Typically, one side of the plastic film is coated by vacuum sputtering a conductive metal such as aluminum, gold, or nickel. The thickness of the coating is about 500 \AA (50 nm). The film is then heated and charged with a high dc potential, with the electret-forming electrode facing the nonconductive side of the film [5]. A well-designed electret capsule will retain its charge and exhibit nearly constant sensitivity for 10 years, and it is predicted that it will take 30 to 100 years before the sensitivity is reduced by 3 dB.

These plastic-film electrets generally will not stand the tension required to obtain the high resonant frequencies commonly employed in externally polarized microphones. One solution is to reduce tension and support the diaphragm at many points by means of a grooved backplate (Figure 4.1.6). This and other schemes used to increase stiffness can lead to short-term instability [6]. Therefore, the charged-diaphragm electret generally does not possess the extended high-frequency response and stability of the air-condenser microphone. Its great advantage is that it can be made very cheaply by automated manufacturing methods.

An improved form of electret transducer is the *back electret*, or charged back-plate design [7]. Figure 4.1.7 shows a simplified cross section of a typical design. (Dimensions are exaggerated for clarity. (This is a pressure-gradient microphone, to be discussed later.) The diaphragm is a

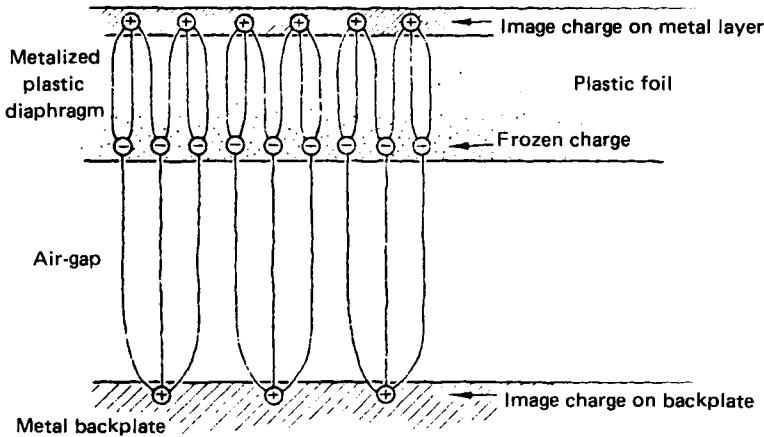


Figure 4.1.5 Positions of charges for space-charge electret when the electret is an integral part of the diaphragm. The frozen charge and the charge on the backplate produce the field in the air that is necessary for microphone operation. (From [7]. Used with permission.)

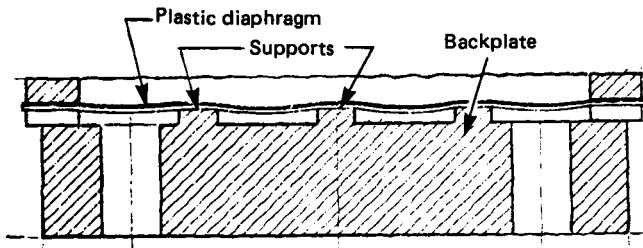


Figure 4.1.6 Principle used by some manufacturers to obtain a sufficiently high resonance frequency of plastic diaphragms having low creep stability. (From [7]. Used with permission.)

polyester film such as Mylar, approximately 0.0002 in (5 μm) thick. This is an ideal material and thickness for a diaphragm. The diaphragm is coated on one or both sides with a thick film of gold or other metal. The electret is made of a fluoric film such as Teflon, which must be at least 0.001 in (25 μm) thick to form a stable electret. This electret is placed on the backplate, which must have a conducting surface to form the "high" output terminal. The electret element is charged similarly to the charged-diaphragm electret. Since the electret does not function as a diaphragm, the material and thickness are chosen as optimal for high sensitivity and stability. The diaphragm-to-back-plate (electret) spacing is the same as for the air condenser, approximately 0.001 in (25 μm). The equivalent polarization potential is 100 to 200 V, which is the same as that used in high-quality air-condenser microphones. (Teflon and Mylar are trademarks of E. I. du Pont de Nemours and Co., Inc.)

4-14 Microphone Devices and Systems

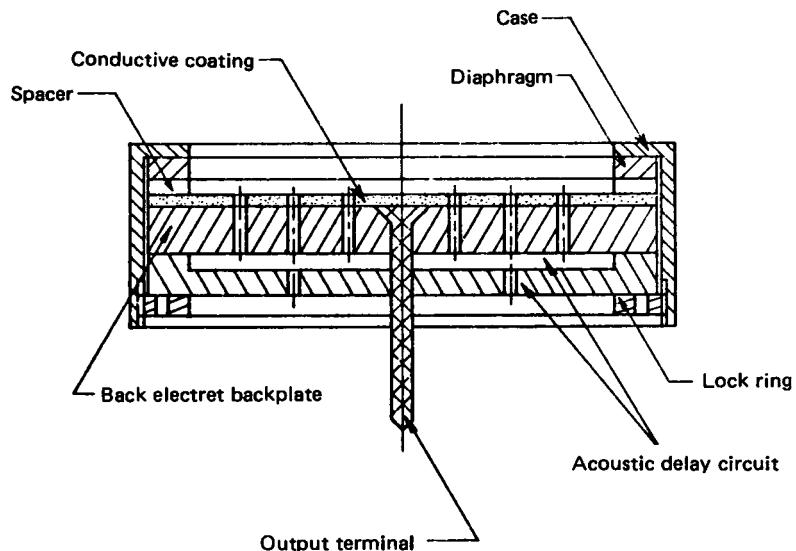


Figure 4.1.7 The back-electret capsule. (From [5]. Used with permission.)

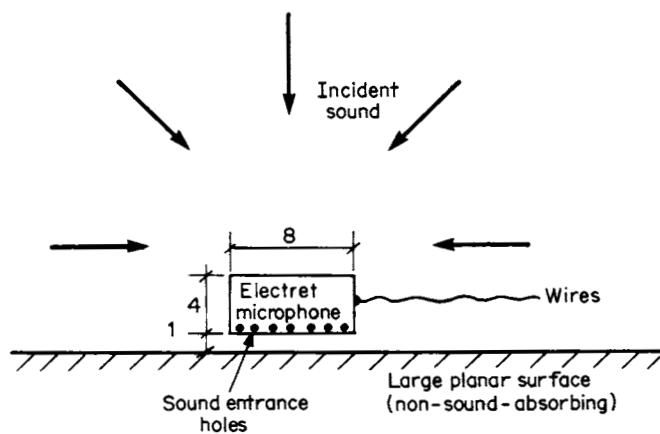


Figure 4.1.8 Boundary-microphone principle. Dimensions are in millimeters.

4.1.2d Boundary Microphone

The boundary microphone involves a *pressure-recording process* in which a conventional microphone is placed very close to a plane surface such as a floor [8]. This has given rise to a number of products which basically function as shown in Figure 4.1.8. A miniature electret microphone is spaced about 0.04 in (1 mm) from a large reflecting plane. A conventional microphone, which is situated above the floor, receives the direct sound wave plus a reflected wave from the floor. It

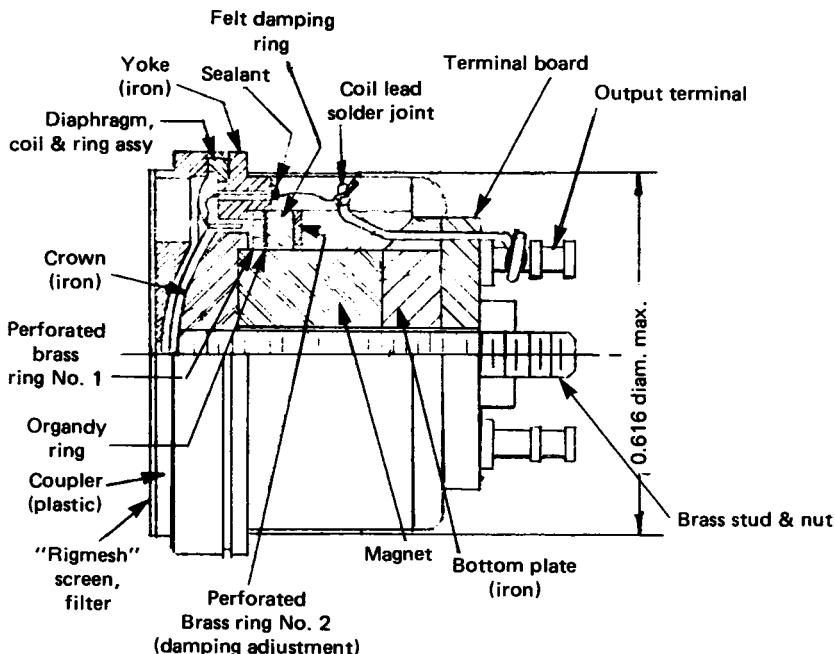


Figure 4.1.9 Dynamic moving-coil pressure-microphone cartridge.

suffers from dips in frequency response at the frequency where the spacing is one-quarter wavelength and its harmonics, as the reflected sound wave interferes with the direct sound wave. When the spacing is reduced to about 0.04 in (1 mm), the null frequency moves far above the audible range. Therefore, in actual use the boundary microphone does not suffer from the comb-filter series of dips in frequency response. The system has, in essence, a directional gain of 6 dB due to pressure doubling at the reflecting plane; for example, the reflected wave is in phase and adds to the amplitude of the direct wave. This results in a hemispheric pickup pattern where the 90° response (direction parallel to the plane) is 6 dB down with respect to the 0° or perpendicular incidence response.

4.1.2e Electrodynamic Microphones

A cross section of a moving-coil-microphone cartridge is shown in Figure 4.1.9, and the complete microphone assembly in Figure 4.1.10 [9]. The diaphragm, which is made of Mylar polyester film 0.00035 in (9 µm) thick, is glued to a voice coil, which moves in the magnetic air gap. The flux density is 10,000 G (1 Wb/m²). The self-supporting coil is wound with four layers of no. 50 AWG copper wire, which results in a dc resistance of 220 Ω. The ac impedance of 200 to 250 Ω is suitable for standard low-impedance microphone inputs of 150 to 600 Ω. Older microphone coils were on the order of 5- to 20-Ω resistance and required a step-up matching transformer in the microphone case. Thus the modern moving-coil microphone will drive standard

4-16 Microphone Devices and Systems

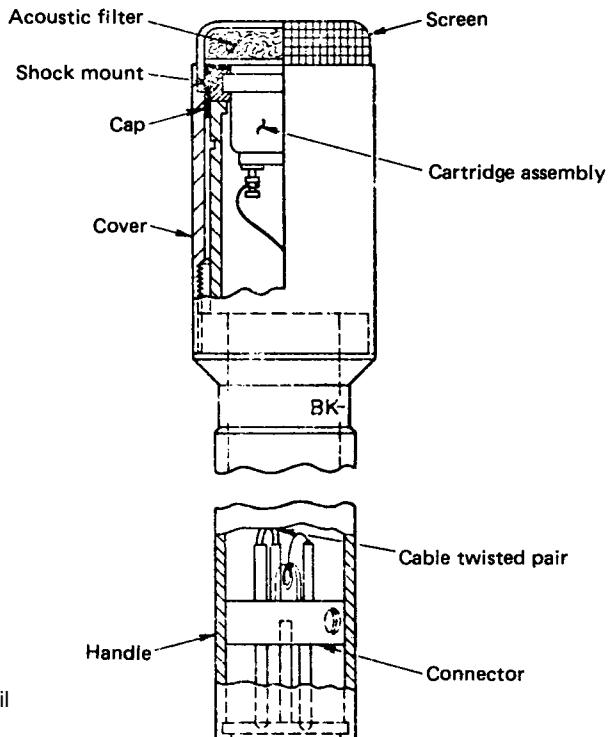


Figure 4.1.10 Dynamic moving-coil microphone.

bipolar integrated circuits directly. The coupler (Figure 4.1.9) fits closely to the diaphragm to provide mechanical protection without frequency discrimination. The cartridge is shock-mounted in the case of Figure 4.1.10, which includes a foam filter screen for dirt and breath “pop” protection.

The voltage induced in the voice coil is given by

$$E = Bl\dot{x} \quad (4.1.5)$$

Where:

- E = open-circuit voltage, V
- B = air-gap flux density, Wb/m²
- l = length of conductor in air gap, m
- \dot{x} = velocity of coil, m/s

This shows that the microphone will have uniform E with respect to frequency if the coil velocity is uniform with frequency. The mechanical resonance of the coil and diaphragm (measured in a vacuum) is about 800 Hz. If the resonance is not well damped, the coil velocity will peak at 800 Hz. This resonance is heavily damped by the acoustic resistance of the felt damping ring so that the resulting response is uniform from 40 to 20,000 Hz. The coil motion is then said to be *resistance-controlled*. The case volume is sufficient to support this extended low-frequency

response. In older microphones, it was necessary to add a vent tube inside the case, possibly as long as 4 in (10 cm). This provided a form of bass-reflex action in which the mass of the air in the tube resonated with the compliance of the air in the case.

4.1.3 Pressure-Gradient (Velocity) Microphones

A sectional view of a classic ribbon velocity microphone (RCA type BK-11A) is shown in Figure 4.1.11. This microphone has an air gap 0.125 in (3.2 mm) wide with a flux density of 6500 G (0.65 Wb/m^2). The ribbon is made of pure aluminum foil weighing 0.56 mg/cm^2 . This corresponds to a thickness of 0.000082 in ($2 \mu\text{m}$). The ribbon is 1.4 in (36 mm) long and corrugated transversely, as shown. Magnetic fine-mesh steel screens are on both sides of the ribbon to provide resistance damping of the ribbon and dirt protection. The ribbon resonance is approximately 30 Hz. The ribbon is soldered to the clamp after assembly and tuning. Soldering has no effect on tuning when done properly. Without soldering, in several years microphone impedance may rise and eventually result in an open circuit at the ribbon. The $0.2\text{-}\Omega$ ribbon impedance is stepped up to $30/150/250 \Omega$ by the transformer. The reactor and switch provide low-frequency rolloff for the proximity effect. The frequency response is $+2 \text{ dB}$, 30 to 15,000 Hz.

The elements of the complete equivalent mechanical circuit (Figure 4.1.11) are R_L and M_L , the mechanical resistance and mass of the air load on the ribbon, imposed by the damping screens; M_R and C_R , the mass and compliance of the ribbon, and M_S and R_S , the mass and mechanical resistance of the slits formed by the ribbon to pole-piece clearance, which is nominally 0.005 in (125 μm). Above resonance, the circuit is simplified as shown, and the ribbon velocity is given by

$$\dot{x} = \frac{(P_1 - P_2)A_R}{j\omega(M_R + M_L)} \quad (4.1.6)$$

Where:

\dot{x} = ribbon velocity, m/s

$(P_1 - P_2)$ = difference in sound pressure (pressure gradient) between two sides of ribbon, N/m^2

A_R = area of ribbon, m^2

M_R = mass of ribbon, kg

M_L = mass of air load acting on ribbon, kg

$\omega = 2\pi f$

f = frequency, Hz

The driving sound pressure gradient $(P_1 - P_2)$ at a given frequency is proportional to the size of the baffle formed by the magnet structure. The ribbon-to-polepiece clearance forms a *leak* which, if excessive, will reduce sensitivity. To maintain a constant ribbon velocity with mass control per Equation (4.1.6), the pressure gradient must increase linearly with frequency. The open-circuit ribbon voltage is given by

$$E = Bl\dot{x} \quad (4.1.7)$$

4-18 Microphone Devices and Systems

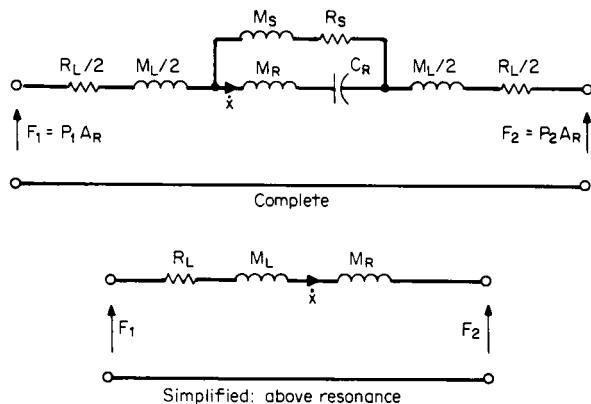
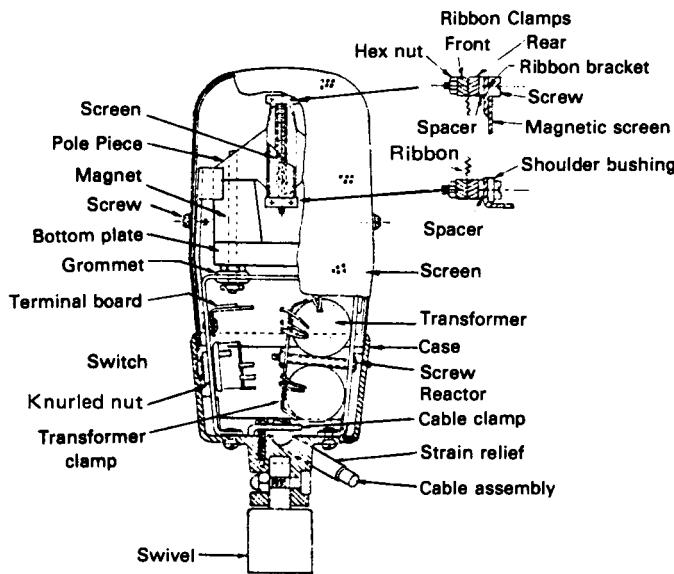


Figure 4.1.11 Classic ribbon velocity microphone (RCA type BK-11A) and mechanical networks.

Where:

E = open-circuit voltage, V

B = air-gap flux density, Wb/m²

l = length of ribbon, m

\dot{x} = ribbon velocity, m/s

At zero frequency the pressure gradient is zero. At the frequency where the path length around the baffle, from the front to back of the ribbon, corresponds to one-half of the sound

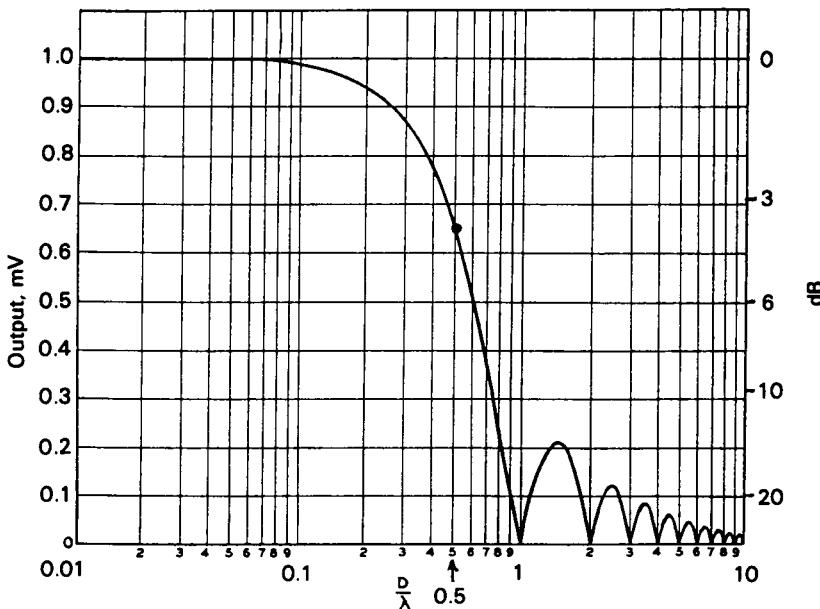


Figure 4.1.12 Computed open-circuit voltage response frequency characteristic of a pressure-gradient mass-controlled electrodynamic microphone. (From [10]. Used with permission.)

wavelength, the pressure gradient departs from a linear characteristic to 65 percent of the value needed for a constant ribbon velocity. At the frequency where the path length equals one wavelength, the pressure gradient is zero. Figure 4.1.12 shows the resulting E versus frequency for an ideal microphone, applicable to the region well above ribbon resonance. A practical microphone may have small ripples in response in the region just above resonance frequency, plus dips or peaks at high frequencies due to pole-piece shape or transverse resonances of the ribbon.

Figure 4.1.13 shows how the figure-of-eight polar pattern becomes severely distorted above the half-wavelength frequency (D equals the path length). Below this frequency, the patterns are essentially perfect cosines.

A compromise solution is found in the contemporary ribbon velocity microphone. The head diameter is typically on the order of 1.5 in (38 mm). The magnetic assembly is extremely small but efficient. The two ribbons are electrically in parallel and make use of most of the space and magnetic flux available in the air gap. They are usually corrugated longitudinally for most of their length, but a few conventional transverse corrugations may be formed near the ends to provide compliance. This type of ribbon, while difficult to make, can potentially solve several problems as compared with the conventional ribbons with transverse corrugations:

- The rigid central portion resists twisting, sagging, and scraping along the pole pieces.
- With the more rigid ribbon, the pole-piece-to-ribbon clearance may be reduced, thus increasing sensitivity.

4-20 Microphone Devices and Systems

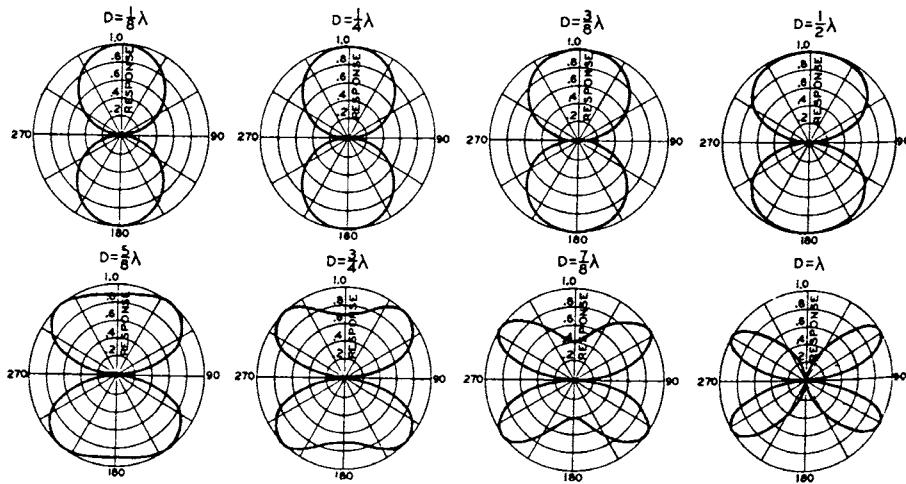


Figure 4.1.13 Directional characteristics of a pressure-gradient microphone as a function of dimensions and wavelength. The polar graph depicts output, in volts, as a function of angle, in degrees. The maximum response is arbitrarily chosen as unity. (From [10]. Used with permission.)

- The short length of transverse corrugations may reduce the need for laborious manual stretching and tuning, and may greatly reduce the downward drift of tuning with time.
- The longitudinal corrugations may reduce or eliminate transverse resonances, which produce small dips and peaks in frequency response above 8000 Hz.
- The short length of the ribbon makes the polar pattern in the vertical plane more uniform with frequency.

Most ribbon microphones have low magnetic-hum sensitivity because the ribbon circuit is easily designed to be *hum-bucking*. Ribbon microphones have low vibration sensitivity because the moving mass is very low.

4.1.3a Combination Pressure and Pressure-Gradient Microphones

Figure 4.1.14 illustrates graphically how the outputs of a bidirectional and a nondirectional microphone transducer can be mixed to obtain three unidirectional polar patterns. Actually, there are an infinite number of unidirectional patterns that may be obtained. The three patterns shown are hypercardioid, cardioid, and limacon, from left to right. The energy responses to random sounds (such as room noise and reverberant sound) are also shown relative to the nondirectional, which is assigned a value of unity. Note that the bidirectional and the cardioid have exactly the same response, but the hypercardioid is superior to both of them in discrimination against random sound. A number of unidirectional microphones produced today are hypercardioids, but the cardioid remains the most popular. The limacon is not as popular, and so to obtain this pattern a microphone with variable directivity is needed. An alternate way to obtain a unidirectional pattern is by using a single transducer with an appropriate acoustical phase-shifting system. Some

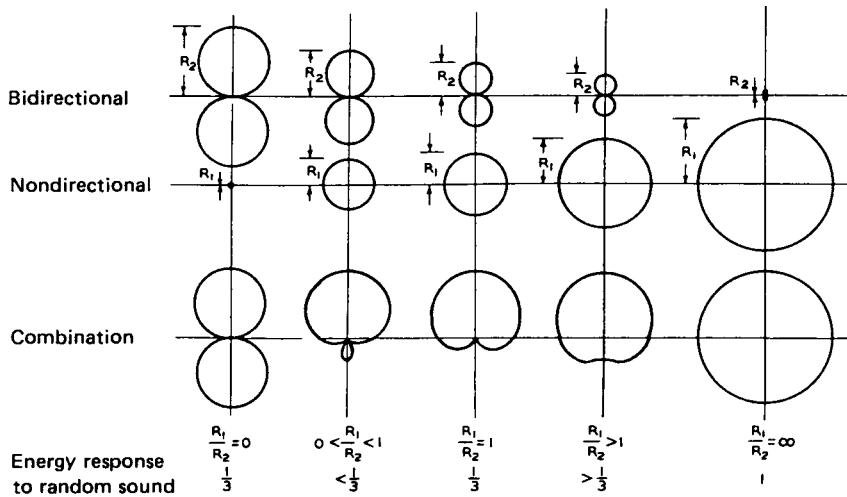


Figure 4.1.14 Directional diagrams of various combinations of bidirectional and nondirectional microphones and energy response to random sounds. (From [2]. Used with permission.)

single-transducer microphones have a mechanically variable delay system so that the pattern can be varied from bidirectional to cardioid to nondirectional.

Frequency Response as a Function of Distance

The low-frequency response of the velocity microphone is accentuated when the distance between source and microphone is less than a wavelength. This happens to a lesser degree with the unidirectional microphone [2]. Figure 4.1.15 gives curves for velocity and unidirectional microphones. If the curves for 0° are plotted to a decibel scale, the slopes follow linear 6-dB-per-octave characteristics. The unidirectional curves exhibit a corner (+3-dB) frequency that is one octave higher than those of the velocity microphone. The +3-dB frequencies rise one octave when the distance is halved. Therefore, for each distance a simple resistance-capacitance rolloff equalizer can be designed to provide flat response. This so-called proximity effect pertains to all pressure-gradient (velocity) and combination pressure and pressure-gradient (unidirectional cardioid) microphones to the same degree. These characteristics are essentially invariant between models of microphones. The exception to these rules is the variable-distance unidirectional microphone, which has a reduced proximity effect.

Dual-Diaphragm Condenser Polydirectional Microphone

The dual-diaphragm microphone vibrating system consists of a pair of diaphragms, each spaced a small distance from the backplate, as in the pressure microphones described previously [11]. The space behind each diaphragm provides acoustical resistance damping as well as acoustical capacitance (stiffness). The cavities behind the diaphragms are interconnected by small holes in the backplate. The phase shift in this system plus the variable electrical polarizing system make possible a variety of directional patterns.

4-22 Microphone Devices and Systems

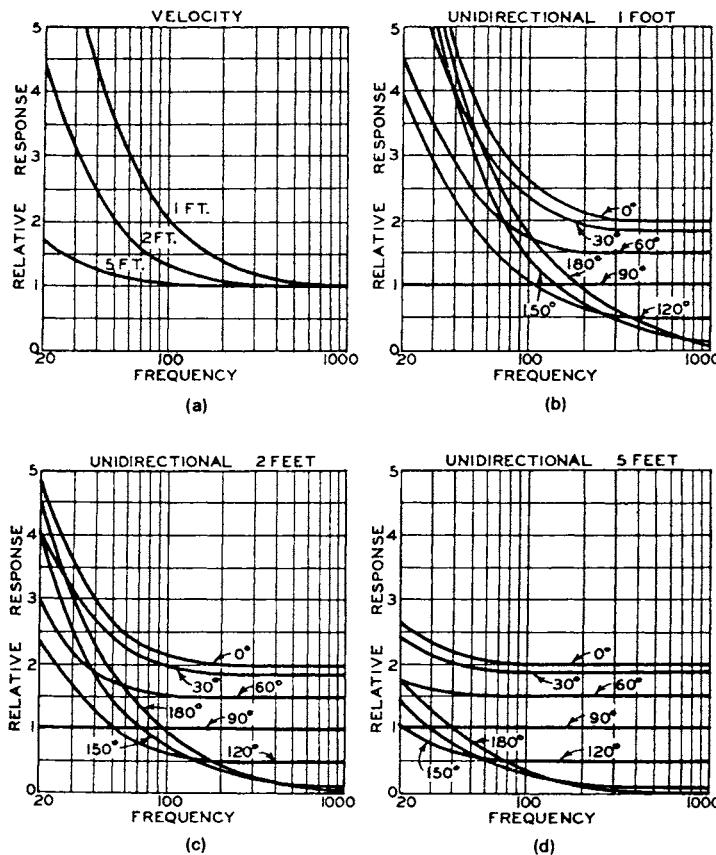


Figure 4.1.15 Microphone characteristics: (a) relative voltage output of a velocity (or pressure-gradient) microphone as compared with a nondirectional pressure microphone for distances of 1, 2, and 5 ft; (b-d) relative voltage output of a unidirectional microphone as compared with a nondirectional pressure microphone for distances of 1, 2, and 5 ft and for various angles of incident sound. (From [2]. Used with permission.)

With switch position 1, the diaphragms are oppositely polarized, and the transducer has a bidirectional pattern. This may be deduced by observing that sound incident at 90° or 270° will produce equal but oppositely phased outputs from each diaphragm, and thus the net voltage output is a null.

With the switch at position 5, the diaphragms are similarly polarized and the outputs are in phase at all angles of incidence, resulting in an omnidirectional pattern. At intermediate switch settings, a variety of unidirectional patterns are obtained. Note that at switch setting 3 a cardioid pattern is obtained with maximum polarizing voltage E_0 on the front diaphragm and 0 V on the back diaphragm. The unenergized diaphragm and the acoustical capacitance and resistance of

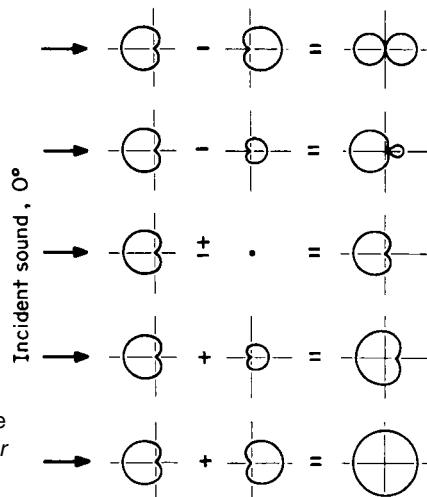


Figure 4.1.16 Condenser polydirectional microphone using two cardioid transducers back to back. (After [11].)

the backplate form a phase-shift network similar to the rear sound aperture of a single-element unidirectional microphone.

The frequency response of the polydirectional microphone will be flat, and the polar pattern uniform with frequency, if the diaphragms are carefully matched and the resistance elements are the controlling acoustical impedances. As in the case of the velocity microphone, acoustical characteristics deteriorate as the frequency approaches that where the path length from front to back approaches a wavelength of sound. A diameter of 0.5 in (12.5 mm) maximum is required for uniform directional characteristics to 15,000 Hz. However, the axial frequency response of a 1-in- (25-mm-) diameter polydirectional microphone can be made uniform to 20,000 Hz, so some uniformity of polar pattern is often traded for the higher sensitivity and lower noise level obtained with the larger-diaphragm transducers.

Twin-Cardioid-Element Polydirectional Condenser Microphone

The dual-diaphragm polydirectional condenser microphone may be thought of as a superposition of two single-diaphragm cardioid microphones back to back. Figure 4.1.16 shows how two cardioid capsules placed back to back will function as a polydirectional microphone. As in the case of the dual-diaphragm transducer, the front transducer has maximum polarizing voltage E_0 at all times and maintains cardioid response with maximum sensitivity. The voltage on the rear transducer is varied down to zero and up to $+E_0$, the same as in the dual-diaphragm transducer. The same polar patterns are obtained. Likewise, the same effect can be obtained by mixing the individual audio outputs in the various amplitude ratios and polarities.

This polydirectional microphone obviously has the most uniform acoustical properties in the cardioid mode because only one transducer is involved. In the other modes, the spacing between capsules, which may be 0.4 to 1.2 in (10 to 30 mm), comes into play, and the polar characteristics at high frequencies become nonuniform.

4-24 Microphone Devices and Systems

4.1.3b Single-Element Unidirectional Microphones

The classic single-element ribbon polydirectional microphone (RCA type 77-DX) is shown in Figure 4.1.17. The ribbon is located between the pole pieces of a relatively large horseshoe magnet. The flux density is 13,000 G (13 Wb/m^2), which results in high sensitivity in all modes of operation. The vertical tube behind the magnet leads to a damped pipe (acoustic line) in the central body of the microphone. The acoustic line has a developed length of about 3 ft (1 m) and is lightly packed with ozite so as to provide a constant acoustical resistance to the ribbon over a wide frequency range. The vertical connector tube is D-shaped in cross section and has a long, narrow slot that opens to the rear. This slot is covered with an organdy screen, which is inside the tube. The rotary shutter varies the effective size of the slot or rear sound opening. This provides six polar patterns by means of a detent, but the actual number of available patterns is infinite. The shutter is shown at the bidirectional setting with the slot fully uncovered. When the shutter is rotated 60° counterclockwise, the slot is fully covered and a nondirectional pattern is obtained. An additional 60° rotation results in the slot being about 10 percent uncovered, which yields a cardioid pattern.

The simplified acoustical equivalent circuit of the microphone (Figure 4.1.17) consists of the following elements:

- M_R = inertance (acoustical mass) of ribbon plus air load on ribbon
- R_L = acoustical resistance of air load on ribbon
- M_S = inertance of air in slot, including screens
- R_S = acoustical resistance of air in slot, including screens
- R_P = acoustical resistance of acoustic line
- P_1 = front sound pressure
- P_2 = rear sound pressure

The circuit applies to the frequency range above ribbon resonance, where the acoustical capacitive reactance of the ribbon is negligible. When the shutter fully uncovers the slot, the impedance of $M_S + R_S$ becomes very small and short-circuits R_P . Then the circuit becomes that of a pressure-gradient (velocity) microphone. The quantity $(P_1 - P_2)$ is the input pressure gradient. The acoustical circuit impedance is that of the ribbon plus air load and is inductive or mass-controlled. This results in a constant volume current U in $(M_R + R_L)$, constant ribbon velocity versus frequency, and uniform ribbon output voltage. The polar pattern is bidirectional or figure-eight.

With the shutter fully closed, the impedance of $M_S + R_S$ becomes very large; so P_2 no longer drives the ribbon circuit. The acoustic line resistance R_P is large compared with the impedance of $(M_R + R_L)$; so the volume current U is given by

$$U = \frac{P_1}{R_p} \quad (4.1.8)$$

This means that the microphone is pressure-responsive and has a nondirectional polar pattern.

With the shutter set for a cardioid pattern, part of the ribbon volume current U flows through R_P and part through $(M_S + R_S)$. Thus, the ribbon is partly controlled by P_1 and the line resistance

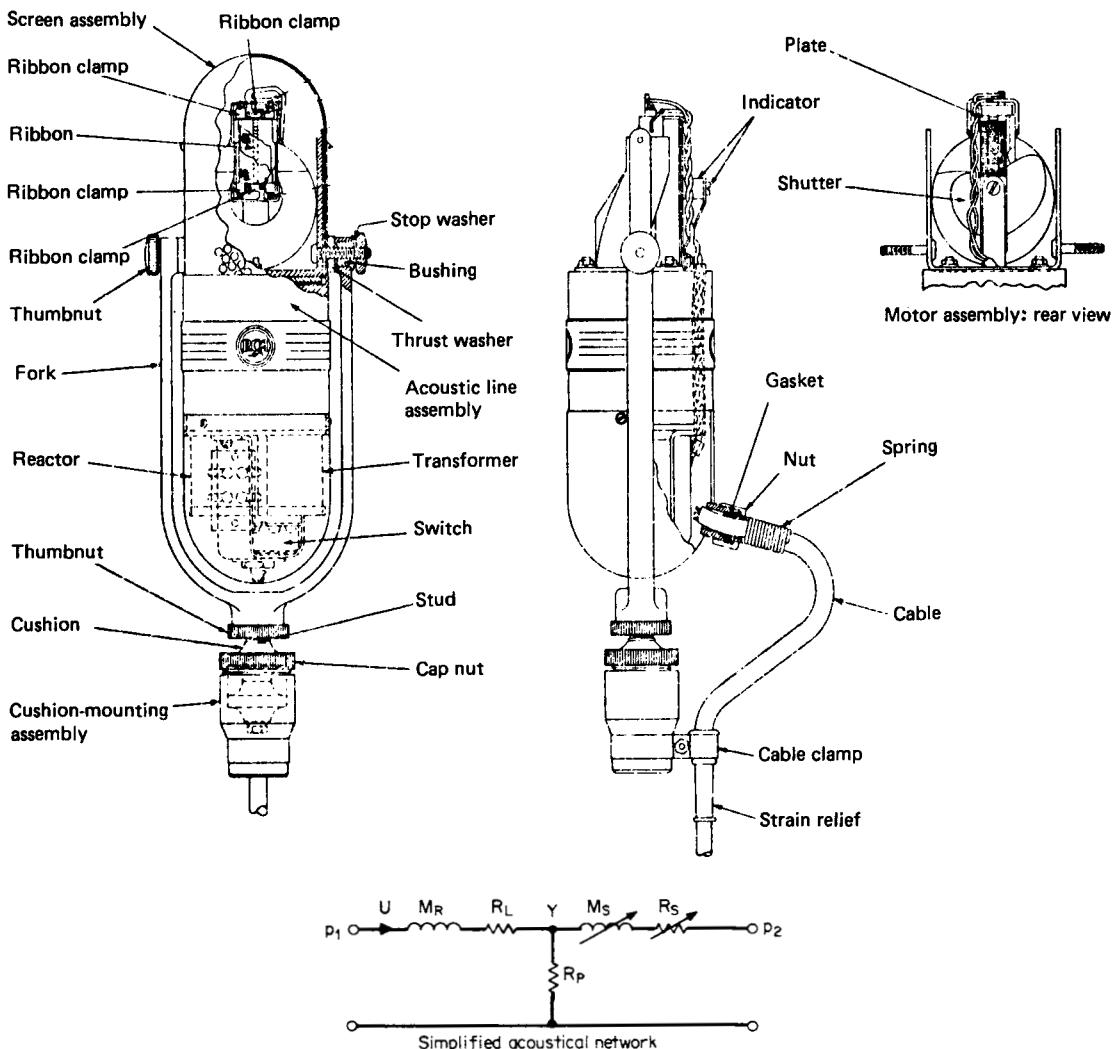


Figure 4.1.17 Ribbon polydirectional microphone and acoustical network (RCA type 77-DX).

R_P and is pressure-responsive. The balance of the ribbon volume current U flows through $(M_S + R_S)$; so the transducer is partly velocity-responsive. The shutter setting for a cardioid pattern is at a critical point where the phase shift through $(M_S + R_S)$ is such that sound incident from 180° arrives at point Y somewhat delayed in time so as to match the phase of sound at P_1 . Thus $U = 0$, a null in response occurs at 180° , and a cardioid pattern is obtained. This is the principle by which single-element unidirectional electrodynamic microphones operate.

Three additional directional patterns are detent-selectable. The axial frequency response at the cardioid setting is reasonably flat from 30 to 15,000 Hz. The response at the bidirectional set-

4-26 Microphone Devices and Systems

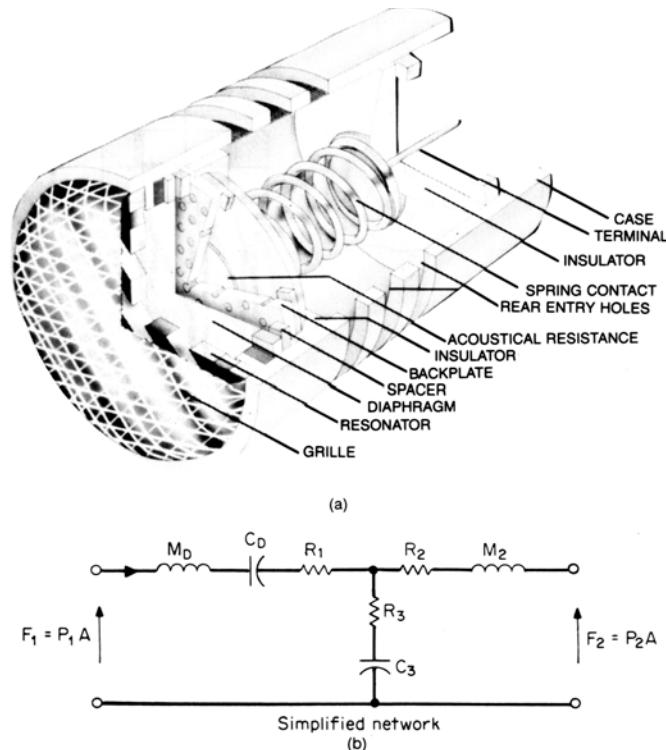


Figure 4.1.18 Unidirectional condenser microphone: (a) mechanical layout, (b) simplified mechanical network. (From [4]. Used with permission.)

ting slopes downward with frequency, whereas the response at the nondirectional setting slopes upward. This is a limitation of the ribbon polydirectional microphone.

Unidirectional Condenser Microphone

The unidirectional condenser microphone incorporates a prepolarized capsule where the electret is on the backplate [4] and [12]. The unidirectional capsule backplate has holes which communicate through an acoustic resistance screen into the case volume (normally having a closed bottom end) and to the atmosphere through resistance screens and rear entry ports.

The operation of the microphone can be determined from a consideration of the simplified mechanical network. (See Figure 4.1.18.) M_D and C_D are the mass and compliance of the diaphragm; R_1 is the resistance of the air film between diaphragm and backplate; R_3 is the resistance of the screen which connects to the case volume C_3 ; and R_2 and M_2 represent the holes and screens at the rear sound entry.

The velocity \dot{x} of the diaphragm is given by

$$\dot{x} = \frac{F_D}{Z_M} = \frac{j\omega KPA}{Z_M} \quad (4.1.9)$$

Where:

Z_M = mechanical impedance of vibrating system, mechanical ohms

F_D = force on diaphragm, N

K = transducer

P = sound pressure, N/m²

A = area of diaphragm, m²

$\omega = 2\pi f$

f = frequency, Hz

and the displacement is given by

$$x = \frac{\dot{x}}{j\omega} = \frac{KPA}{Z_M} \quad (4.1.10)$$

The output voltage is given by Equation (4.1.2).

Thus, for the displacement (and output voltage) to be uniform with frequency, Z_M must be resistive. The resistance elements R_1 , R_2 , and R_3 are the controlling elements.

The phase-shift network R_2 , M_2 , R_3 , and C_3 may take on a variety of configurations similar to the various networks in ribbon and dynamic microphones.

Moving-Coil Unidirectional Microphone

Figure 4.1.19 shows the basic mechanical cross section and acoustical network of the moving-coil unidirectional microphone. The resonance of M_1 and C_{A1} , the diaphragm-and-coil-assembly inertance and acoustical capacitance, is at the low end of the usable audio-frequency range. Depending on the application of the microphone, this may be anywhere from approximately 70 to 140 Hz. The lowest attainable resonance is limited by the stiffness of the plastic-film diaphragm material.

The moving-coil system is mass-controlled above resonance as in the ribbon transducer. Therefore, the difference in sound pressure between the two sides of the diaphragm must be proportional to frequency so as to maintain a constant volume current and a constant diaphragm and coil velocity throughout the useful audio-frequency range. This is done by selection of the parameter values of the phase-shift network. Also, the network values must provide for the correct delay time versus frequency such that a null is maintained at 180° for a cardioid pattern. Alternately, the network values may be adjusted for a hypercardioid pattern.

Variable-Distance Unidirectional Microphone

Figure 4.1.20 shows a sectional view and the acoustical network of the variable-distance unidirectional microphone. The distance from front to rear sound entry varies approximately inversely with frequency [2]. Sound pressure P_1 acts on the front of the diaphragm. Pressures P_2 , P_3 , and P_4 act on the back of the diaphragm through suitable acoustic impedance. P_2 acts in the high-frequency region, P_3 at middle frequencies, and P_4 at low frequencies. The advantage of this design

4-28 Microphone Devices and Systems

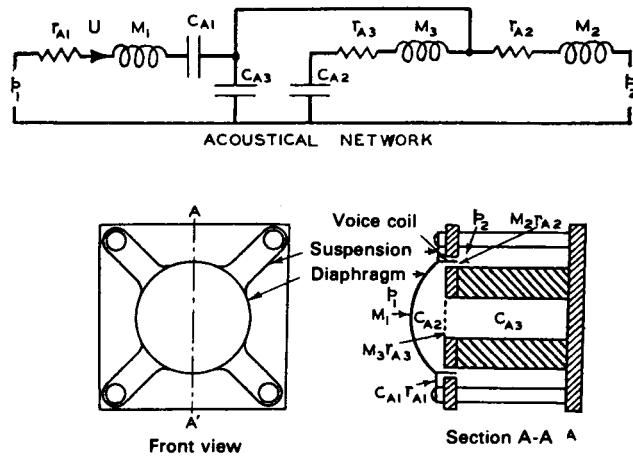


Figure 4.1.19 The basic acoustical network and mechanical construction of the moving-coil unidirectional microphone. (From [2]. Used with permission.)

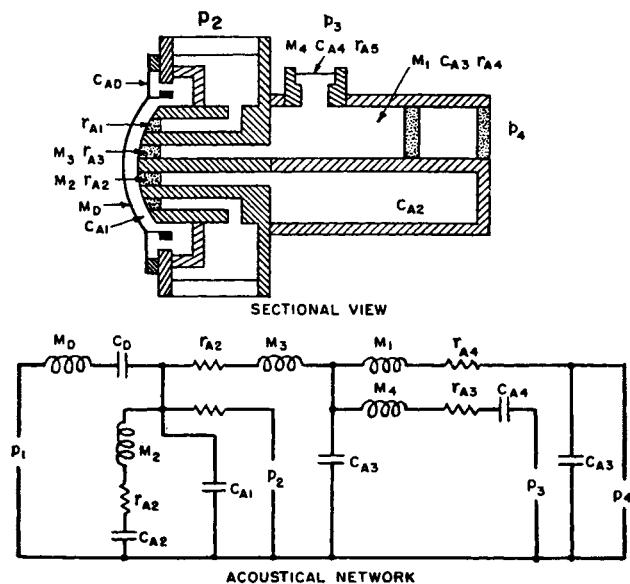


Figure 4.1.20 Sectional view and acoustical network of the variable-distance unidirectional microphone. (From [2]. Used with permission.)

is that accentuation of low frequencies due to the proximity effect is reduced. As with the moving-coil unidirectional microphone, the moving-system resonance is in the region of 100 Hz and is mass-controlled at higher frequencies.

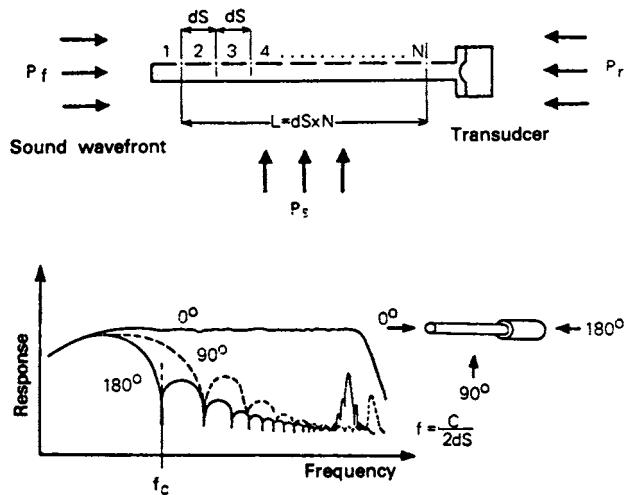


Figure 4.1.21 Operating principles of the line microphone. (Courtesy of Sony.)

4.1.4 Ultra-Directional Microphones

For the purpose of this discussion, an *ultradirectional microphone* is defined as one that has an energy response to random sound of less than 0.25, relative to an omnidirectional microphone, over a major portion of its useful audio-frequency range. The value of 0.25 is the random energy efficiency of a hypercardioid, which represents the highest directivity obtainable with a first-order gradient microphone [13]. This category includes higher-order pressure-gradient microphones and wave-interference types of microphones. The applications of ultradirectional microphones include long-distance pickup of sound in the presence of random noise and/or reverberant sound or close talking in high-noise environments.

Of the many types of ultradirectional microphones developed, only the line-type microphone remains in common use. It employs high-sensitivity condenser or moving-coil electrodynamic transducers.

4.1.4a Line Microphone

A simple line microphone is shown in Figure 4.1.21. An acoustic line (pipe) with equally spaced sound openings along its entire length is connected to a pressure microphone element. The transducer element may be of the electrostatic or electrodynamic varieties. A high order of directivity is indicated by the frequency-response curves in the mid- and high-frequency region where the 90° and 180° responses are far below the 0° curve. The low-frequency limit of the useful range of ultradirectional characteristics is given by [14]

$$f_c = \frac{C}{2L} \quad (4.1.11)$$

4-30 Microphone Devices and Systems

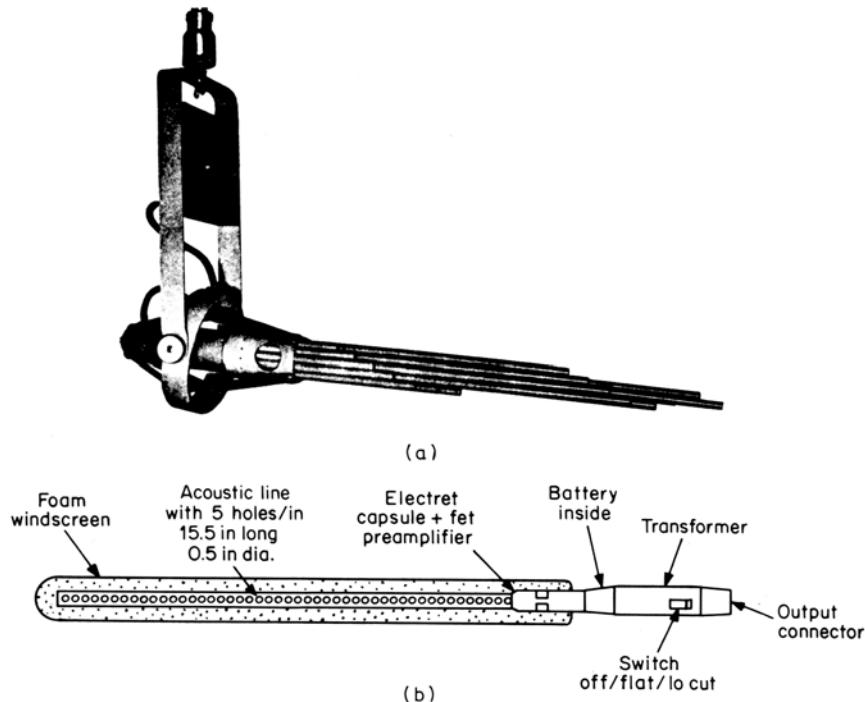


Figure 4.1.22 Line microphones: (a) bundled pipes, (b) single pipe with holes and electret condenser.

Where:

f_c = frequency, Hz

c = velocity of sound = 331 m/s

L = total length of line

The high-frequency limit of the ultradirectional region is determined by the hole spacing dS

$$f_n = \frac{C}{2dS} \quad (4.1.12)$$

where dS is the hole spacing, m.

If f_c is chosen to be 100 Hz, then L must equal 65 in (1.66 m), which is too long for most practical applications. However, this requirement may be eased by substituting a pressure-gradient cardioid element. This provides good 180° rejection below f_c , and with careful optimization of parameters a microphone of practical length can have good rejection at 90°, well below f_c . It is relatively easy to achieve $f_n = 10,000$ Hz or higher with practical hole spacings.

Alternately, the line may consist of a bundle of small tubes of lengths which vary from dS to L in even steps of dS . Similarly, a single pipe with a series of slots may be used. With modern small-diaphragm condenser transducers, the single pipe is appropriate because the diameters of

the tubes in a bundle would be so small that the acoustic resistance (viscosity) loss would reduce sensitivity and roll off the high-frequency response.

Figure 4.1.22 shows an electret condenser line microphone with a small-diameter line and a transducer capsule 0.6 in (16 mm) in diameter. The capsule and line are made as an assembly that is interchangeable with standard cardioid and pressure elements. Although f_c is 420 Hz, 15-dB rejection is maintained at 90° down to 100 Hz [15].

4.1.4b Wave Microphones

A parabolic reflector may be used to concentrate distant parallel rays of sound at a microphone placed at the focus. This concept is illustrated in Figure 4.1.23a. As in all wave-type microphones, the reflector must be large compared with a wavelength of sound to obtain a high order of directivity.

An acoustic lens microphone is a lens-like device made of sheet metal that can focus sound waves onto a microphone in a manner similar to the parabolic reflector (Figure 4.1.23b). The directivity follows the laws of wave-type microphones in much the same way as the parabola [2].

A large-surface microphone consisting of a large number of pressure-microphone elements arranged on a spherical surface is shown in Figure 4.1.23c. The polar pattern is similar to that of a curved-surface sound source, which emits uniformly over a solid angle subtended by the surface at the center of curvature. The microphone shown in Figure 4.1.23c is 4 ft (1.22 m) in diameter and has an angular spread of 50°. The pattern is reasonably uniform above 300 Hz [2].

4.1.5 Miscellaneous Types of Microphones

A two-channel microphone such as the one shown in Figure 4.1.24 is a convenient tool for sound pickup in the x - y or M - S stereophonic modes where coincident microphone transducers are required. The example device shown utilizes two dual-diaphragm condenser transducers, which are mounted on top of each other and in adjacent capsules sharing a common axis; the capsules may be rotated with respect to each other. A remote-control unit permits any one of nine polar patterns to be selected for each channel.

4.1.5a Sound-Field Microphone

The original sound-field microphone was developed for the *ambisonic* surround system patented by the United Kingdom National Research Corporation and was produced by Calrec Audio Limited. This system was a form of quadraphonic sound. A later version of the device became essentially an electronically steerable stereophonic microphone. Four single-diaphragm cardioid condenser capsules are mounted in a tetrahedral array and connected to an electronic control unit. This unit permits selection of cardioid, figure-of-eight, and omnidirectional patterns for each stereo output. In addition, the sound pickup axes may be electronically steered in azimuth and elevation. By processing the pressure and pressure-gradient components of the audio signal, the microphone may be moved fore and aft as the ratio of direct to reverberant sound is varied. The electronic steering may be done before or after the audio is recorded, allowing flexibility in the postproduction phase of sound recording.

4-32 Microphone Devices and Systems

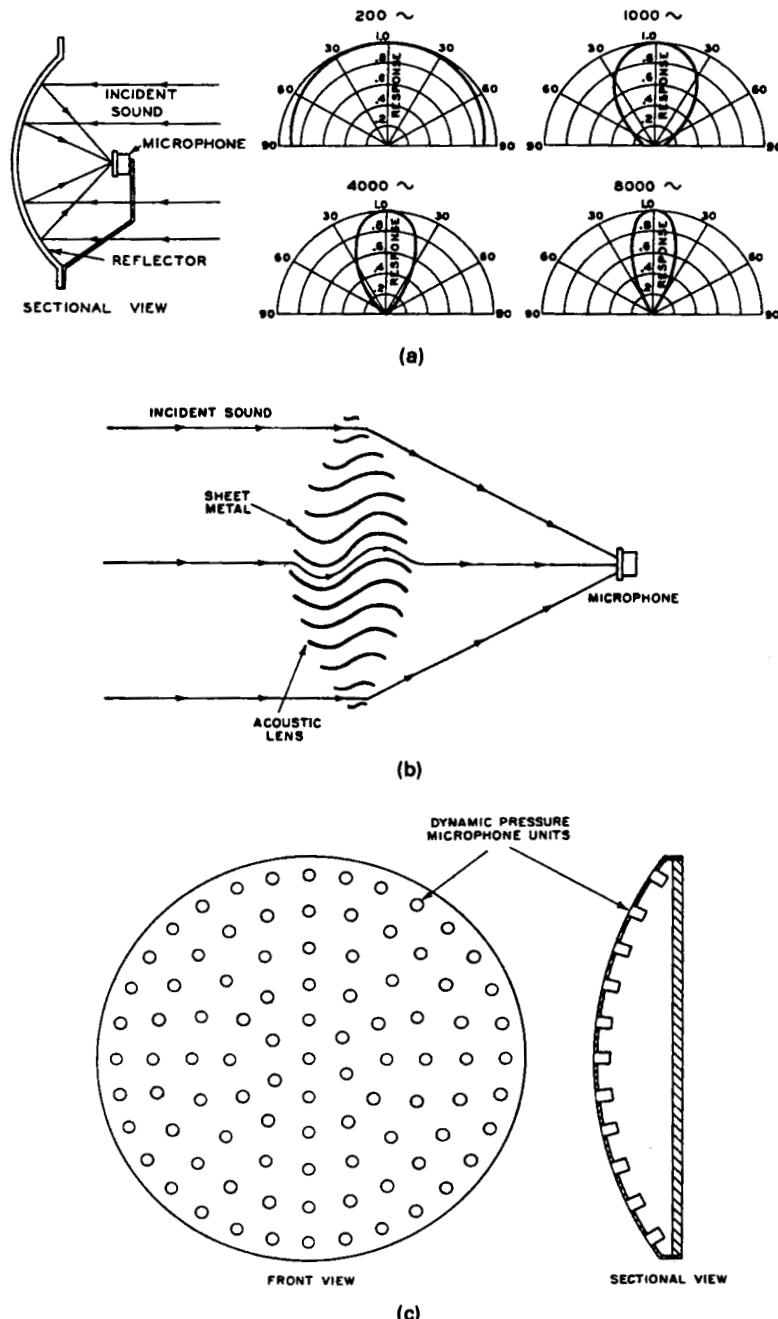


Figure 4.1.23 Wave microphones: (a) parabolic reflector, (b) lens, (c) large-surface. (From [3]. Used with permission.)

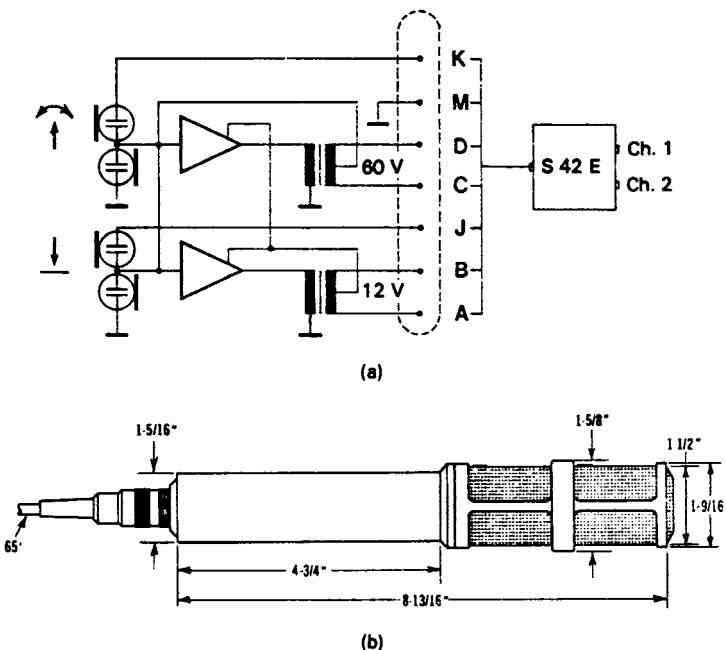


Figure 4.1.24 Stereophonic condenser microphone: (a) wiring diagram, (b) shell construction.

4.1.5b Lavalier Microphone

The term *lavalier microphone* refers to a small microphone that is typically fastened to the clothing of the speaker. When resting on the chest, the microphone requires rising high-frequency response compensation to adjust for the loss in response due to its location off the axis of the mouth. Very small electret condenser models available today utilize a subminiature pickup element. They are light enough so that they may be fastened to the clothing by means of a small clip attached to the cable below the microphone.

4.1.5c Wireless Microphone

A variety of wireless microphones are available today, usually either in a hand-held style or as a lavalier microphone connected to a separate body-pack transmitter. These systems are widely used in television broadcasting and in professional entertainment.

4.1.6 Selecting Microphone Types

The hand-held microphone, probably the most popular type of mic, is available in many shapes and sizes. Manufactured in both directional and non-directional versions, the hand-held mic provides wide frequency response, low handling-noise and a wide choice of characteristic “sounds.”

4-34 Microphone Devices and Systems

Because adequate space is available, a shock-mount system is incorporated into most professional hand-held microphones. Just holding a microphone or dragging the cable across a floor can cause low-frequency noise to be transmitted to the pickup element. A shock-mount system minimizes such noise.

The lavalier microphone is also in high demand today. Its small size and wide frequency response offer professional users what appears to be the best of all worlds. Small size translates into minimum visual distraction on camera or before an audience. Wide frequency response assures good audio quality. There are other points to consider, however, before a lavalier microphone is chosen for a particular application.

The smallest lavalières available are omnidirectional. This makes the talent's job easier because staying *on mic* is less of a problem. However, extraneous noise from the surrounding area can result in a generally poor pickup. The omnidirectional lavalier microphone can pick up unwanted sounds just as easily as it captures the talent's voice. In an indoor, controlled environment, this is usually not a problem. However, outside the ambient noise can make the audio track unusable.

Directional lavalières are available, but they too have performance tradeoffs. The most obvious is size. In order to make a lavalier directional, a back entry usually must be added to the housing so that sound can reach the back of the microphone. This translates into a larger housing for the microphone capsule. Although not as large as a hand-held microphone, a unidirectional lavalier is noticeably larger than its omnidirectional counterpart.

In order to keep the size under control, shock-mounting of the directional capsule is usually kept to a minimum. This results in a microphone that exhibit more handling noise than a comparable omni.

Windscreens for lavalier microphones are a must on any outdoor shoot. Even a soft breeze can cause the audio track to sound as if it was recorded in a wind tunnel. The culprit is turbulence, caused by wind hitting the grille or case of the microphone. The sharper the edges, the greater the turbulence. A good windscreen helps to break up the flow of air around the microphone and reduce turbulence.

Windscreens work best when fitted loosely around the grille of the microphone. A windscreen that has been jammed down on a mic only serves to close off part of the normal acoustic path from the sound source to the diaphragm. The end result is attenuated high-frequency response and reduced wind protection.

4.1.7 References

1. Olson, H. F. (ed.): *McGraw-Hill Encyclopedia of Science and Technology*, 5th ed., vol. 18, McGraw-Hill, New York, N.Y., pg. 506, 1982.
2. Olson, H. F.: *Acoustical Engineering*, Van Nostrand, Princeton, N.J., 1957.
3. Rasmussen, G.: "A New Condenser Microphone," *Tech. Rev.*, Brüel & Kjaer, Copenhagen, no.1, 1959.
4. Olson, H. F.: "Directional Electrostatic Microphone," U.S. Patent 3,007,012, October 31, 1961.

5. Kubota, H.: "Back Electret Microphones," presented at the 55th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 24, no. 862, preprint 1157, December 1976.
6. Killion, M. C., and E. V. Carlson: "A Subminiature Electret-Condenser Microphone of New Design," *J. Audio Eng. Soc.*, vol. 22, pg. 237–243, May 1974.
7. Fredericksen, E., N. Firby, and H. Mathiasen: "Prepolarized Condenser Microphones for Measurement Purposes," *Tech. Rev.*, Brüel & Kjaer, Copenhagen, no. 4, 1979.
8. Long, E. M., and R. J. Wickersham: "Pressure Recording Process and Device," U.S. Patent 4,361,736, November 30, 1982.
9. *Instruction Book for RCA BK-16A Dynamic Microphone*, IB-24898, Radio Corporation of America, Camden, N.J.
10. Olson, H. F.: "Ribbon Velocity Microphones," *J. Audio Eng. Soc.*, vol. 18, pp. 263–268, June 1970.
11. Eargle, J.: *The Microphone Handbook*, Elar Publishing, Plainview, N.Y., 1981.
12. Bevan, W. R., R. B. Schulein, and C. E. Seeler: "Design of a Studio-Quality Condenser Microphone Using Electret Technology," *J. Audio Eng. Soc. (Engineering Reports)*, vol. 26, pp. 947–957, December 1978.
13. Woszczyk, W. R.: "A Microphone Technique Employing the Principle of Second-Order Gradient Unidirectionality," presented at the 69th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 29, pg. 550, preprint 1800, July-August 1981.
14. Kishi, K., N. Tsuchiya, and K. Shimura: "Unidirectional Microphone," U.S. Patent 3,581,012, May 25, 1971.
15. Sank, J. R.: "Equipment Profile-Nakamichi CM-700 Electret Condenser Microphone System." *Audio*, September 1978.

4.1.8 Bibliography

Eargle, J.: *Sound Recording*, Van Nostrand Reinhold, New York, 1976.

Eargle, J.: *The Microphone Handbook*, Elar Publishing, Plainview, N.Y., 1982.

Lipshitz, S. P.: "Stereo Microphone Techniques: Are the Purists Wrong?" presented at the 78th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 33, pg. 594, preprint 2261, July-August 1985.

Microphones—Anthology, Audio Engineering Society, New York, 1979.

Nisbett, A.: *The Technique of the Sound Studio*, Hastings House, New York, N.Y., 1974.

Olson, H.: *Music, Physics, and Engineering*, 2d ed., Dover, New York, N.Y., 1967.

Chapter
4.2

Stereophonic Techniques

Jon R. Sank, Ronald D. Streicher, Wesley L. Dooley

4.2.1 Introduction

Accurate stereo imaging is the foundation for the art of stereo recording. Experience with the basic techniques and knowledge of their attributes are essential for anyone working in stereo formats. The *art* of recording lies in manipulating illusions. The *science* of recording involves the tools and techniques used to create these illusions.

4.2.2 Two-Microphone Coincident Techniques

Coincident or *intensity stereo* techniques are achieved with a pair of directional microphones, most often vertically aligned on a common axis and set at an angle to each other in the horizontal plane. (See Figure 4.2.1.) Thus, there is minimum time (phase) difference between the two capsules for sound sources on the horizontal plane. Properly done, this style relies solely on intensity differences between the two signals for directional cues. The choice of the microphone pair's polar pattern can vary from subcardioid to bidirectional, depending on the specific technique being implemented. The angles formed by the microphone pair are typically symmetrical about the centerline of the sound source, and the included angles discussed in this section are the total angles between the axes of the microphones.

An advantage of intensity stereo is that the angular accuracy of the stereo imaging is unaffected by the distance of the microphone pair from the sound source. A disadvantage is that without the interchannel time delay common to other miking techniques, the stereo image sometimes seems to lack a sense of *space*.

4.2.2a XY Cardioids and Hypercardioids

The microphone pair is typically set at an included angle of between 60 and 120°. The specific angle chosen determines the *apparent width* of the stereo image, and the choice of this angle is subjective, with consideration given to the distance of the microphone pair from the sound source, the actual width of that source, and the polar pattern of the microphone. A critical factor to consider when using these techniques is this polar response. As the individual microphones

4-38 Microphone Devices and Systems

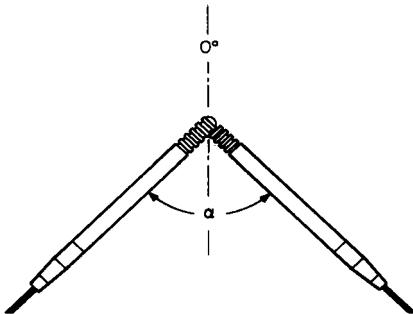


Figure 4.2.1 Coincident XY microphone pair.

are oriented at an angle to most of the sound source, considerable off-axis coloration is possible. As with any stereo technique, the microphones comprising the pair should have as good a polar response as possible. Furthermore, they should be closely matched with regard to polar and frequency response, since any differences will cause the image to wander with changes in pitch.

The use of cardioid microphones is common in coincident techniques, typically with an included angle of 90 to 120° and placed fairly close to the sound source (Figure 4.2.2). Often the axes of the microphones are aimed at a point near the extremes of the sound source. As the direct-to-reverberant-sound ratio of this approach is high, this can offer some rejection of unwanted sound from the rear of the pair. Sometimes a distant pickup with a large reverberation component is desired. In such circumstances, included angles as large as 180° may be employed.

Using a hypercardioid pair is similar to using cardioids except that the included angle is typically narrower to preserve a solid center image (Figure 4.2.3). The increased reach of the hypercardioid allows a more distant placement for a given direct-to-reverberant-sound ratio. With their small reverse-polarity lobes, using hypercardioids is a good compromise between implementing XY with cardioids and the Blumlein technique.

Blumlein

The crossed pair of figure of eights is the earliest of the XY techniques and is configured with two bidirectional microphones oriented at an included angle of 90° (Figure 4.2.4). It was developed in the early 1930s by British scientist Alan Blumlein and was presented in his seminal patent [1].

One attribute of this technique is that the rear lobes of these microphones record the rear 90° quadrant in phase but out of polarity and place this into the stereo image (cross-channeled) together with the front quadrant. Signals from the two side quadrants are picked up out of phase. Placement is therefore critical in order to maintain a proper direct-to-reverberant-sound ratio and to avoid strong out-of-phase components. Typically, this technique works very well in a wide room or one with minimal sidewall reflections, where strong signals are not presented to the side quadrants of the stereo pair. It is often commented that this configuration produces a very natural sound.

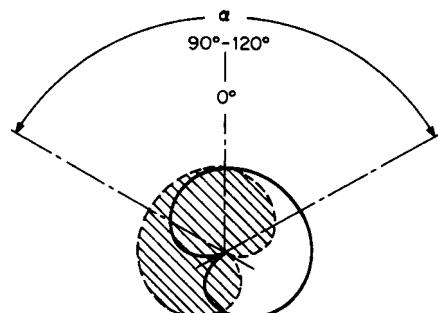


Figure 4.2.2 XY cardioid pair.

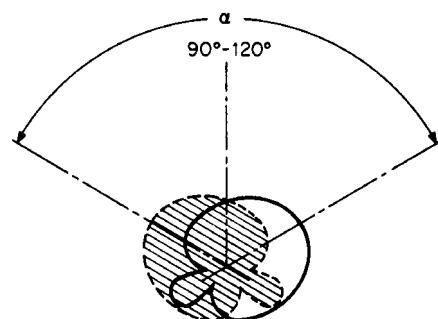


Figure 4.2.3 XY hypercardioid pair.

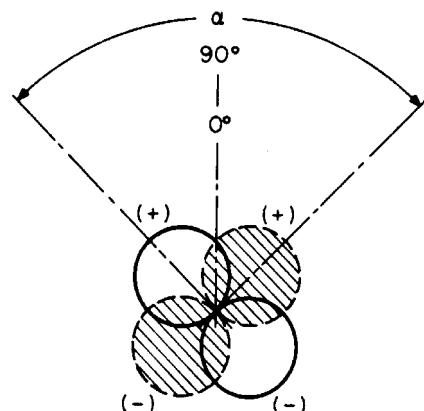


Figure 4.2.4 XY crossed figure-of-eights. (From [1]. Used with permission.)

MS Stereo

This form of intensity stereo uses one microphone (the M or midcomponent) aimed directly at the centerline of the sound source and another, a bidirectional microphone (the S or side component), oriented laterally (Figure 4.2.5). Their outputs are processed by a sum-and-difference matrix network, which resolves them into conventional XY stereo signals, $(M + S)$ and $(M - S)$. The left-right orientation is determined by the direction of the positive lobe of the S microphone.

4-40 Microphone Devices and Systems

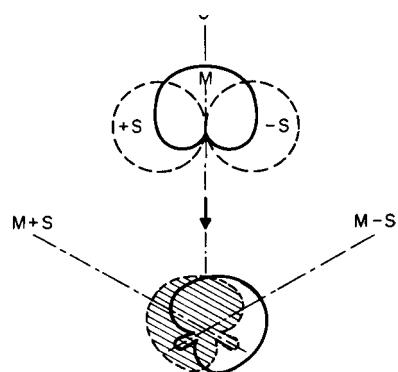


Figure 4.2.5 MS conversion to XY.

An advantage of this system is that it provides absolute monaural predictability: when the left and right signals are combined, the sum is solely the output of the M component: $(M + S) + (M - S) = 2M$. By judicious choice of polar pattern and placement of the M microphone, the monaural signal can be optimized. Conveniently, this pickup is by definition on axis to the midline of the sound source and suffers minimally from off-axis coloration. The S component (the bidirectional microphone), with its null plane bisecting the sound source, provides more reverberant information than the M component. As it is generally desirable that there be less reverberation in a monaural signal than in stereo, there is a built-in advantage to MS in that it automatically has a less reverberant character when summed to mono than in its stereo image.

Finally, the MS technique offers the mixing engineer greater control of the stereo image from the mixing desk than available with any other technique. By changing the pattern of the M pickup (using a remote-pattern microphone), the apparent distance from the sound source and the amount of ambience inherent in the M signal can be adjusted. Furthermore, by varying the M/S ratio in the sum-and-difference matrix, the apparent width of the stereo stage can also be adjusted (Figure 4.2.6). This adjustment can be made either during the original recording session or later, during a postproduction session [2].

4.2.2b Two-Microphone Near-Coincident Techniques

This term is used to describe that class of techniques in which a microphone pair is placed close enough together to be substantially coincident for low frequencies yet is far enough apart to have an appreciable time delay between channels for sound sources located to the far right and left. Such techniques otherwise differ little from coincident microphone configurations, except that the stereo imaging results from differences in both intensity and time (phase).

The value of these techniques is that they exhibit good localization combined with a sense of depth. Close miking is not recommended when using these techniques, since small movements of the sound source can produce large image shifts. Sounds arriving from the far left or far right can also create problems for disk cutting or monaural summation owing to interchannel time delay.

The following examples illustrate some of the common configurations and techniques:

- Figure 4.2.7 shows two cardioid microphones oriented outward from the centerline of the sound source with an included angle of 110° and with a capsule spacing of 17 cm.

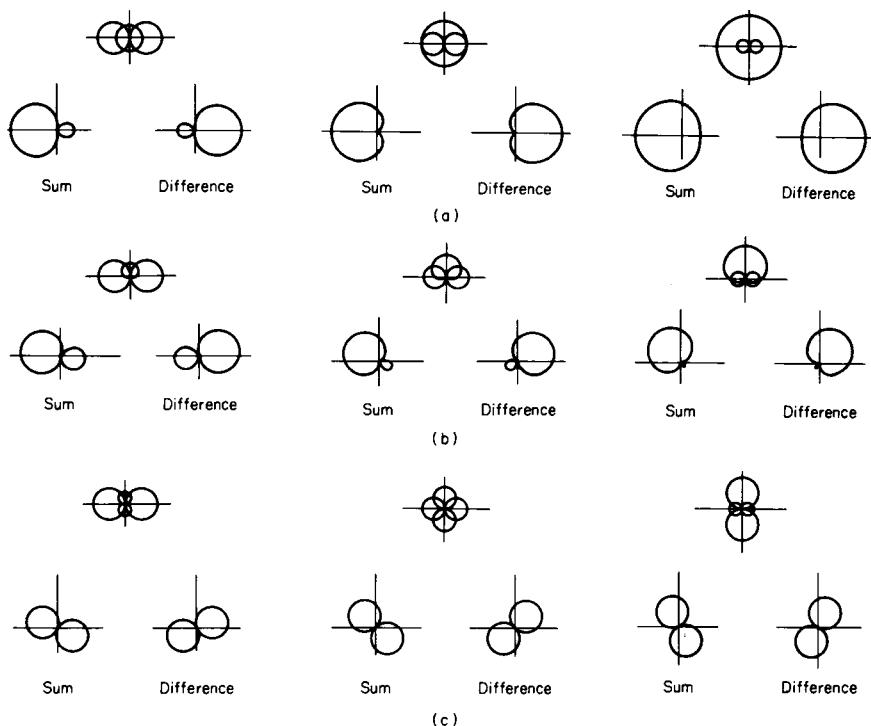


Figure 4.2.6 MS to equivalent XY transformations for M/S ratios of 30:70, 50:50, and 70:30: (a) omnidirectional M component, (b) cardioid M component, (c) bidirectional M component. (From [2]. Used with permission.)

- Figure 4.2.8 shows two cardioid microphones oriented outward from the centerline with an included angle of 90° and a capsule spacing of 30 cm.
- Figure 4.2.9 shows two bidirectional microphones facing directly forward toward the sound source, spaced 20 cm apart.

4.2.2c Two-Microphone Spaced Techniques

Spaced microphones were the first configuration known to relate a stereo image [3]. Generally these techniques employ two or more microphones set symmetrically along a line that is perpendicular to and bisected by the midline of the sound source. The polar pattern of the stereo pair, their spacing, and their distance from the sound source are all variables within this style. Stereo information in these configurations is created by the differences in both amplitude and time of arrival of the sound wave. A characteristic of this approach is that positional information will radically change as the distance to the sound source varies. Extremely distant sounds can present negligible directional cues to the listener.

4-42 Microphone Devices and Systems

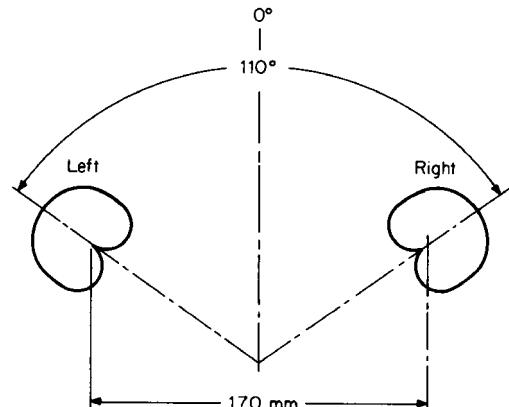


Figure 4.2.7 Two-microphone near-coincident technique known as the ORTF configuration.

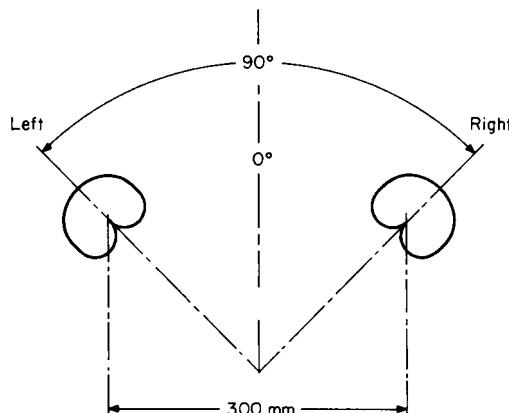


Figure 4.2.8 Two-microphone near-coincident technique known as the NOS configuration.

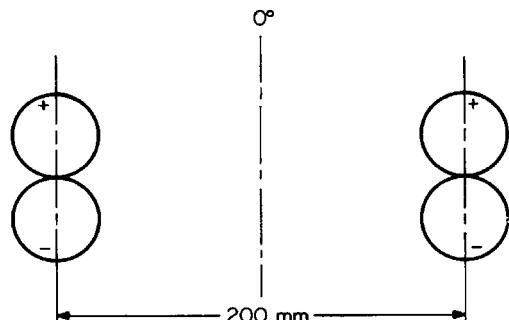


Figure 4.2.9 Two-microphone near-coincident technique known as the Faulkner configuration.

When using spaced microphone configurations, special attention must be given to the following potential problems: 1) low-frequency comb-filter effects on sound sources to the extreme left or right of the sound stage, 2) vague center imaging, and 3) erratic monaural compatibility.

With these techniques, placement and aiming are the essential elements of the art, and as with all stereo recording, a stereo phase-monitor oscilloscope is a useful setup and monitoring tool.

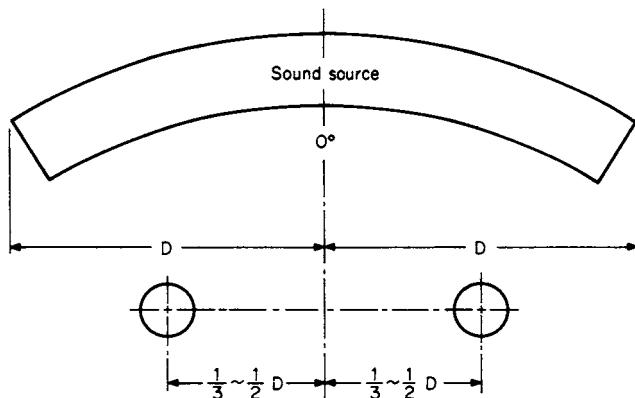


Figure 4.2.10 Spaced omnidirectional pair.

Formulas for spaced microphone layouts have been widely published, and variations on these recommendations are often employed, necessitated by the physical or aesthetic needs of the recording environment [4].

Spaced Omnidirectional Microphones

Typically this style is realized with two (or three) microphones. Common spacings are from 2 to 10 ft (0.6 to 1 m) on either side of the centerline. The spacing is determined by the width of the sound source and the distance of the microphone pair from it. A general rule is that the microphones should be placed one-third to one-half of the distance from the centerline to the outer edge of the sound stage (Figure 4.2.10).

When omnidirectional microphones are used, there is a good general sense of the acoustic space, coupled with the pressure pickup's outstanding, if sometimes overpowering, low-frequency response. Wind-noise problems are generally eliminated, although very-low-frequency sounds, such as air conditioning or traffic noise, are well recorded.

Omnidirectional microphones are designed to be either flat to an on-axis sound source (free-field) or flat to a reverberant sound field (random-incidence). In the latter case, the on-axis frequency response will be tipped up at the high end. Experimentation with the microphones' axial orientation to the sound field can therefore be productive. Omnidirectional microphones require the closest placement to the sound source for a given direct-to-reverberant-sound ratio of any polar pattern and have the maximum potential for pickup of undesirable sounds from the environment.

Spaced Cardioid Microphones

This style is similar to spaced omnidirectional microphones as described previously. Because these microphones are directional, they will tend to favor that segment of the sound source which is more on axis. For reverberation, audience response, and other off-axis sources, they will exhibit the effects of off-axis coloration. Thus, their orientation and placement can sometimes be more critical than with omnidirectional microphones.

4-44 Microphone Devices and Systems

Spaced Bidirectional Microphones

Bidirectional microphones have more reach to the front than do cardioids, but they also have an equal, though reverse-polarity, pickup lobe to the rear. Thus, they must be placed farther back from the sound source than either omnidirectionals or cardioid microphones in order to achieve the same coverage. The rear lobe provides that the reverberation components and audience response will have the same sonic characteristics as the front lobe (that is, there will be little off-axis coloration of these sounds).

One advantage of this technique is provided by the null plane of these microphones. Proper orientation of this plane can reduce the pickup of unwanted sounds quite effectively. However, care must be taken that the most desired sound not be placed in the out-of-polarity rear lobes.

Spaced Hypercardioid Microphones

This polar pattern is midway between cardioid and bidirectional types. The front lobe has more reach (that is, narrower) than that of a cardioid, while the small rear lobe has the reverse-polarity aspect of the bidirectional microphone. The null area is generally a cone, somewhere between the 90° null plane of the bidirectional and the 180° null point of the cardioid. The exact null cone angle, the amount of rear-lobe pickup, and the coloration of sound arriving from off axis will depend on the particular design of the microphone being used. Considerations involved in using such a spaced pair would be an amalgam of those for spaced cardioid and spaced bidirectional microphones.

4.2.3 Performance Considerations

Numerous factors must be carefully considered when planning a stereo recording. The sonic and technical characteristics of the microphones are important, and so are the visual aesthetics of their placement. During a recording session without an audience, there are few concerns other than the obvious rules of safety for both the microphones and the performers. When an audience will view the performance, the mixing engineer must also consider appearance. Live, telecast, or filmed performances all demand compromises between conflicting requirements of sight lines and microphone placement. This is particularly true with staged performances such as opera, musicals, or dramatic theater.

The discrete use of single-point coincident stereo microphones flown from above can often prove beneficial. In addition, the use of boundary-surface techniques will provide a good, clear pickup of stage activity and still be quite invisible to the audience.

The use of a single-point remote-control stereo microphone can also offer the engineer the added flexibility of making alterations in the stereo perspective if or when the performance or sound source dictates without the need for changing the physical position of the microphone.

The final consideration in any mixing situation is, of course, the sound: does it adequately represent the original sound source (if, indeed, it is supposed to)? Such aspects as localization, depth, presence, clarity of individual components, and lack of unnatural coloration are primary things to consider. Equally important: does the pickup adequately avoid the unwanted sounds in the environment?

There is no magic answer, no one right way to accomplish the task. What works well today may not suffice tomorrow. Thus it is imperative that mixing engineers learn as many approaches

as they can, so that when that “impossible situation” does present itself, the relevant knowledge and the tools will be at hand to meet the situation.

4.2.4 References

1. Blumlein, A.: British Patent 394,325, December 14, 1931; reprinted in *J. Audio Eng. Soc.*, vol. 6, April 1958.
2. Dooley, W. L., and R. D. Streicher: “M-S Stereo: A Powerful Technique for Working in Stereo,” *J. Audio Eng. Soc.*, vol. 30, pp. 707–718, October 1982.
3. “A Phased Array,” *Hi Fi News Record Rev.*, July 1981.
4. Gordon, J.: “Recording in 2 and 4 Channels,” *Audio*, pp. 36–38, December 1973.
5. “Miking with the 3-Point System,” *Audio*, pp. 28–36, December 1975.

Chapter
4.3
Low Power Amplifiers

Daniel R. von Recklinghausen

4.3.1 Introduction

Amplifiers are the functional building blocks of audio and video systems, and each of those building blocks contains several amplifier stages coupled together. An amplifier may contain its own power supply, while an amplifier stage needs one or more external sources of power. The active component of each amplifier stage is usually a transistor or a field-effect transistor (FET). Other amplifying components, such as vacuum tubes, can also be used in amplifier circuits if the general principles of small- and large-signal voltages and current flows are followed [1–4].

4.3.2 Single-Stage Transistor or FET Amplifier

The single-stage amplifier can best be described as using a single transistor or FET connected as a common emitter or common-source amplifier, using an NPN transistor (Figure 4.3.1a) or an N-channel FET (Figure 4.3.1b) and treating PNP transistors or P-channel FET circuits by simply reversing the current flow and the polarity of the voltages.

4.3.2a DC Conditions

At zero frequency or dc and also at low frequencies, the transistor or FET amplifier stage requires an input voltage E_1 equal to the sum of the input voltage of the device (the transistor V_{be} or FET V_{gs}) and the voltage across the resistance R_e or R_s between the common node (ground) and the emitter or source terminal. The input current I_1 to the amplifier stage is equal to the sum of the current through the external resistor connected between ground and the base or gate and the base current I_b or gate current I_g drawn by the device. In most FET circuits the gate current may be so small that it can be neglected, while in transistor circuits the base current I_b is equal to the collector current I_c divided by the current gain beta of the transistor. The input resistance R_1 to the amplifier stage is equal to the ratio of input voltage E_1 to input current I_1 .

The input voltage and the input resistance of an amplifier stage increase as the value of the emitter or source resistor becomes larger.

4-48 Microphone Devices and Systems

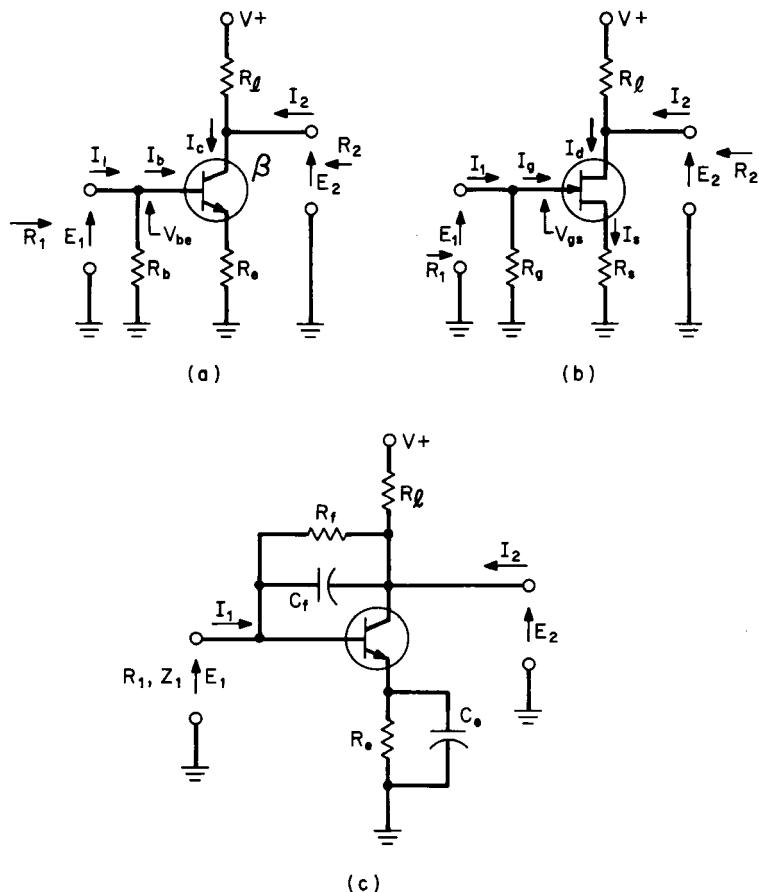


Figure 4.3.1 Single-stage amplifier circuits: (a) common-emitter NPN, (b) common-source N-channel FET, (c) single stage with current and voltage feedback.

The output voltage E_2 of the amplifier stage, operating without any external load, is equal to the difference of supply voltage $V+$ and the product of collector or drain load resistor R_1 and collector current I_c or drain current I_d . An external load will cause the device to draw an additional current I_2 , which increases the device output current.

As long as the collector-to-emitter voltage is larger than the *saturation voltage* of the transistor, collector current will be nearly independent of supply voltage. Similarly, the drain current of an FET will be nearly independent of drain-to-source voltage as long as this voltage is greater than an equivalent saturation voltage. This saturation voltage is approximately equal to the difference between gate-to-source voltage and *pinch-off voltage*, the latter voltage being the bias voltage which causes nearly zero drain current. In some data sheets for FETs, the pinch-off voltage is given under a different name as *threshold voltage*. At lower supply voltages the collector

or drain current will become less until it reaches zero when the drain-to-source voltage is zero or the collector-to-emitter voltage has a very small reverse value.

The output resistance R_2 of a transistor or FET amplifier stage is—in effect—the parallel combination of the collector or drain load resistance and the series connection of two resistors, consisting of R_e or R_s and the ratio of collector-to-emitter voltage and collector current or the equivalent drain-to-source voltage and drain current. In actual devices an additional resistor, the relatively large output resistance of the device, is connected in parallel with the output resistance of the amplifier stage.

The collector current of a single-stage transistor amplifier is equal to the base current multiplied by the current gain of the transistor. Because the current gain of a transistor may be specified as tightly as a 2:1 range at one value of collector current or may have just a minimum value, knowledge of the input current is usually not quite sufficient to specify the output current of a transistor.

4.3.2b Input and Output Impedance, Voltage, and Current Gain

As derived above for a common-emitter or common-source single amplifier stage, input impedance is the ratio of input voltage to input current and output impedance is the ratio of output voltage to output current. As the input current increases, the output current into the external output load resistor will increase by the current-amplification factor of the stage. The output voltage will decrease because the increased current flows from the collector or drain voltage supply source into the collector or drain of the device. Therefore, the voltage amplification is a negative number having the magnitude of the ratio of output-voltage change to input-voltage change.

The magnitude of voltage amplification is often calculated as the product of *transconductance* G_m of the device and load resistance value. This can be done as long as the emitter or source resistor is zero or the resistor is bypassed with a capacitor that effectively acts as a short circuit for all signal changes of interest but allows the desired bias currents to flow through the resistor. In a bipolar transistor the transconductance is approximately equal to the emitter current multiplied by 39, which is the charge of a single electron divided by the product of Boltzmann's constant and absolute temperature in kelvins. In an FET this value will be less and usually is proportional to input bias voltage with reference to the pinch-off voltage.

The power gain of the device is the ratio of output power to input power, often expressed in decibels. Voltage gain or current gain may be stated in decibels but must be so marked.

AC Gain

The resistor in series with the emitter or source causes negative feedback of most of the output current, which reduces the voltage gain of the single amplifier stage and raises its input impedance (Figures 4.3.1a and b). When this resistor R_e is bypassed with a capacitor C_e (Figure 4.3.1c), the amplification factor will be high at high frequencies and will be reduced by approximately 3 dB at the frequency where the impedance of capacitor C_e is equal to the emitter or source input impedance of the device, which in turn is approximately equal to the inverse of the transconductance G_m of the device (Figure 4.3.2a). The gain of the stage will be approximately 3 dB higher than the dc gain at the frequency where the impedance of the capacitor is equal to the emitter or source resistor. These simplifications hold in cases where the product of transconductance and resistance value is much larger than 1.

4-50 Microphone Devices and Systems

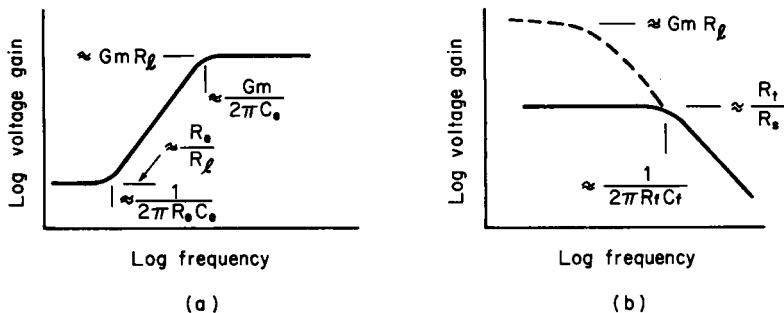


Figure 4.3.2 Feedback amplifier voltage gains: (a) current feedback, (b) voltage feedback.

A portion of the output voltage may also be fed back to the input, which is the base or gate terminal. This resistor R_f will lower the input impedance of the single amplifier stage, reduce current amplification, reduce output impedance of the stage, and act as a supply voltage source for the base or gate. This method is used when the source of input signals and internal resistance R_s is coupled with a capacitor to the base or gate and a group of devices with a spread of current gains, transconductances, or pinch-off voltages must operate with similar amplification in the same circuit. If the feedback element is also a capacitor C_f , high-frequency current amplification of the stage will be reduced by approximately 3 dB when the impedance of the capacitor is equal to the feedback resistor R_f and voltage gain of the stage is high (Figure 4.3.2b). At still higher frequencies amplification will decrease at the rate of 6 dB per octave of frequency. It should be noted at this point that the base-collector or gate-drain capacitance of the device has the same effect of limiting high-frequency amplification of the stage, but this capacitor becomes larger as collector-base or drain-gate voltage decreases.

Feedback of the output voltage through an impedance lowers the input impedance of an amplifier stage. Voltage amplification of the stage will be affected only as this lowered input impedance loads the source of input voltage. If the source of input voltage has a finite source impedance and the amplifier stage has very high voltage amplification and reversed phase, the effective amplification for this stage will approach the ratio of feedback impedance to source impedance and also have reversed phase.

Common-Base or Common-Gate Connection

For this configuration (Figure 4.3.3a), voltage amplification is the same as in the common-emitter or common-source connection, but input impedance is approximately the inverse of the transconductance of the device. As a benefit, high-frequency amplification will be less affected because of the relatively lower emitter-collector or source-drain capacitance and the relatively low input impedance. This is the reason why the cascade connection (Figure 4.3.3b) of a common-emitter amplifier stage driving a common-base amplifier stage exhibits nearly the dc amplification of a common-emitter stage with the wide bandwidth of a common-base stage. The other advantage of a common-base or common-gate amplifier stage is stable amplification at very high frequencies (VHF) and ease of matching to transmission-line impedances, usually 50 to 75Ω .

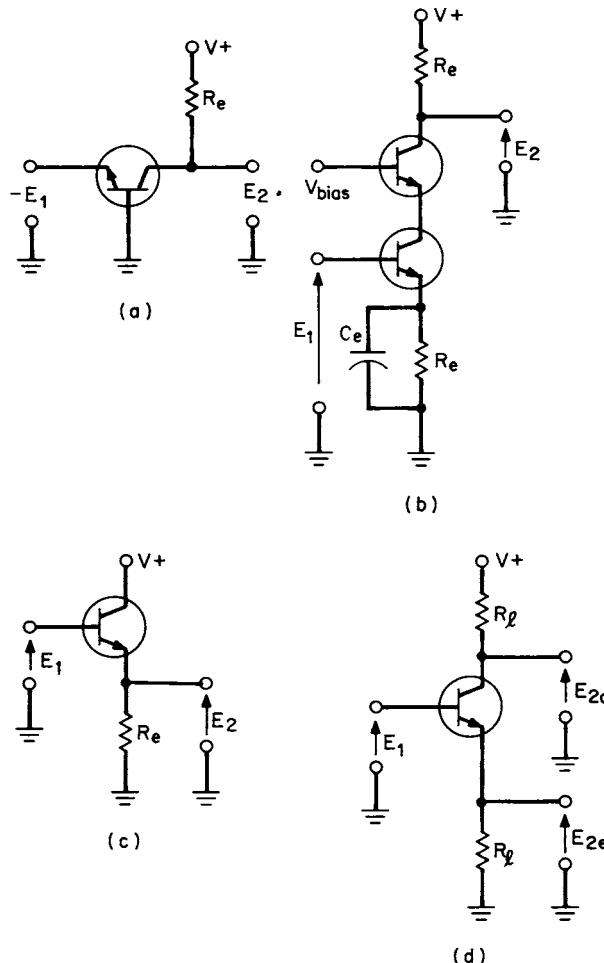


Figure 4.3.3 Transistor amplifier circuits: (a) common-base NPN, (b) cascade NPN, (c) common-collector NPN emitter follower, (d) split-load phase inverter.

Common-Collector or Common-Drain Connection of a Transistor or FET

For this case, voltage gain is slightly below 1.000, but the input impedance of a transistor so connected will be equal to the value of the load impedance multiplied by the current gain of the device plus the inverse of the transconductance of the device (Figure 4.3.3c). Similarly, the output impedance of the stage will be the impedance of the source of signals divided by the current gain of the transistor plus the inverse of the transconductance of the device.

When identical resistors are connected between the collector or drain and the supply voltage and the emitter or source and ground, an increase in base or gate voltage will result in an increase of emitter or source voltage that is nearly equal to the decrease in collector or drain voltage. This

4-52 Microphone Devices and Systems

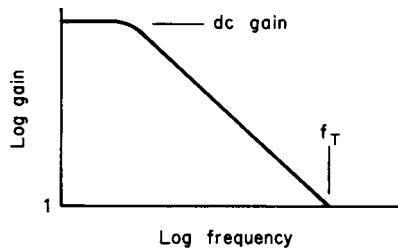


Figure 4.3.4 Amplitude-frequency response of a common-emitter or common-source amplifier.

type of connection is known as the *split-load phase inverter*, useful for driving push-pull amplifiers, although the output impedances at the two output terminals are unequal (Figure 4.3.3d).

The current gain of a transistor decreases at high frequencies as the emitter-base capacitance shunts a portion of the transconductance, thereby reducing current gain until it reaches a value of 1 at the transition frequency of the transistor (Figure 4.3.4). From this it can be seen that the output impedance of an emitter-follower or common-collector stage will increase with frequency, having the effect of an inductive source impedance when the input source to the stage is resistive. If the source impedance is inductive, as it might be with cascaded emitter followers, the output impedance of such a combination can be a negative value at certain high frequencies and be a possible cause of amplifier oscillation. Similar considerations also apply to common-drain FET stages.

Bias and Large Signals

When large signals have to be handled by a single-stage amplifier, distortion of the signals introduced by the amplifier must be considered. Although feedback can reduce distortion, it is necessary to ensure that each stage of amplification operates in a region where normal signals will not cause the amplifier stage to operate with nearly zero voltage drop across the device or to operate the device with nearly zero current during a portion of the cycle of the signal. Although the amplifier is described primarily with respect to a single-device amplifier stage, the same holds true for any amplifier stage with multiple devices, except that here at least one device must be able to control current flow in the load without being *saturated* (nearly zero voltage drop) or *cut off* (nearly zero current):

If the single-device amplifier load consists of the collector or drain load resistor only, the best operating point should be chosen so that in the absence of a signal, one-half of the supply voltage appears as a quiescent voltage across the load resistor R_L . If an additional resistive load R_L is connected to the output through a coupling capacitor C_c (Figure 4.3.5a), the maximum peak load current I_L in one direction is equal to the difference between quiescent current I_q of the stage and the current that would flow if the collector resistor and the external load resistor were connected in series across the supply voltage. In the other direction maximum load current is limited by the quiescent voltage across the device divided by the load resistance. The quiescent current flows in the absence of an alternating signal and is caused by bias voltage or current only. For signals with an equal probability of positive and negative peak excursions, such as audio-frequency waveforms, it is advisable to have the two peak currents equal. This can be accomplished by increasing the quiescent current as the external load resistance decreases.

When several devices contribute current into an external load resistor (Figure 4.3.5b), one useful strategy is to set bias currents so that the sum of all transconductances remains as constant

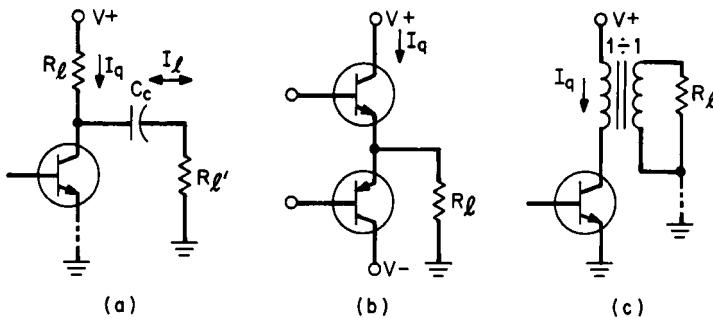


Figure 4.3.5 Output load-coupling circuits: (a) ac-coupled, (b) series-dc parallel-ac push-pull half bridge, (c) single-ended, transformer-coupled.

as practical, which means a design for minimum distortion. This operating point for one device is near one-fourth of the peak device current for push-pull FET stages and at a lesser value for bipolar push-pull amplifiers.

When the load resistance is coupled to the single-device amplifier stage with a transformer (Figure 4.3.5c), the optimum bias current should be nearly equal to the peak current that would flow through the load impedance at the primary of the transformer with a voltage drop equal to the supply voltage.

4.3.2c Multistage Amplifiers

All practical audio, video, and radio-frequency amplifiers are multistage amplifiers in which cascaded single-stage amplifiers are connected together. Overall feedback then is used to stabilize amplification and quiescent operating points [5].

DC-Coupled Multistage Amplifiers

Commonly, amplifier stages enclosed in an overall feedback loop are direct-coupled so that the quiescent operating point is determined primarily by the bias of the first stage.

Two cascaded common-emitter amplifier stages can form a gain block useful as a low-cost preamplifier (Figure 4.3.6a). Here, the collector of the first stage is connected to the base of the second stage and to a resistor that supplies collector current to the first stage and base current to the second stage from the supply voltage. Both stages have an emitter resistor connected to the common ground, with the second resistor bypassed with a large capacitor. Base current to the first stage is supplied from the emitter of the second stage through a pair of voltage-divider resistors connected to ground, with audio input signals fed to the base of the first stage through a coupling capacitor. The audio-signal output is taken from the collector of the second stage through a coupling capacitor. The second collector receives its operating current through a resistor and supplies feedback current through a resistor-capacitor network to the emitter of the first stage. When using two NPN transistors and a single positive supply voltage, a low-cost preamplifier with the appropriate feedback network is constructed. The maximum no-feedback voltage ampli-

4-54 Microphone Devices and Systems

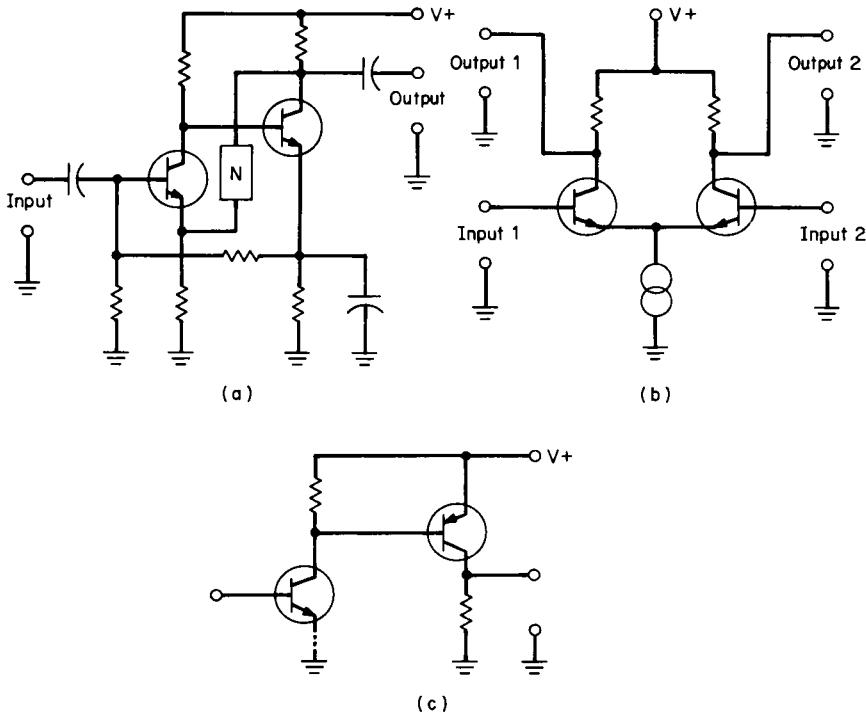


Figure 4.3.6 Multistage dc-coupled circuits: (a) cascaded two-transistor amplifier, (b) emitter-coupled phase inverter, (c) cascade NPN and PNP transistors.

fication is approximately equal to the transconductance of the first stage multiplied by the current gain of the second stage and the load resistor value of the second stage.

When two transistors or FETs have their emitters or sources connected together and that junction is supplied with a constant current and output voltages are obtained from identical-value resistors connected between the supply voltage and the two collectors or drains, the long-tailed-pair or emitter-coupled or source-coupled phase-inverter stage is described (Figure 4.3.6b). Here the emitter or source of one device acts as the emitter or source resistor for the other device, and an alternating signal impressed on one input will be amplified with a phase reversal in the same stage while this reversal is not experienced at the other output. This stage, then, is capable of taking one single-ended signal and transforming it into a push-pull signal. The signal applied to the other input now will arrive at the first output without a phase reversal. Thus, the emitter- or source-coupled amplifier is able to amplify the difference between two signals, where one may be an input signal and the other a feedback signal. Its voltage amplification is the same as the normal grounded emitter or grounded source stage, except that the output voltage should now be measured between the two outputs, with the voltage from either output to ground being one-half of that value. This type of input stage is the almost universal input stage of operational amplifiers.

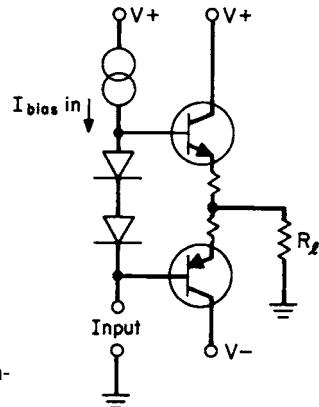


Figure 4.3.7 Input signal and bias circuit in a single-ended push-pull amplifier.

Almost the same performance can be expected when the commonly connected emitters or sources are supplied by a resistor instead of a current source or the two input circuits have differing impedances. The long-tailed pair performs best when the two devices are matched and operate at the same temperature. For this reason, matched monolithic transistors and FET provide the best performance, with the additional benefit that, with equally shared current, even-harmonic distortion is substantially reduced when compared with single-ended amplifiers.

Two-device amplifier stages are not restricted to a construction using two devices of the same type (bipolar versus FET) or the same polarity (NPN versus PNP). Using stages of opposite polarity often results in a higher available amplification factor and output voltage when limitations of supply voltage are considered (Figure 4.3.6c).

A common use of transistors or FETs of differing polarity is in series-connected push-pull amplifiers (Figure 4.3.7), where one stage supplies current of one polarity from one supply voltage to a grounded load and the other stage supplies opposite-polarity current from a supply voltage of opposite polarity. The bias current flows from one supply to the other through the two devices without passing through the load. Here, the two stages may have a relatively low bias current that is often stabilized by a diode per device connected in shunt with the base circuit and kept at the same temperature as the amplifying transistors. As the transistors temperature increases with heat, the base-emitter voltage decreases and the diode forward voltage decreases also, thereby keeping the quiescent current in the transistors within much smaller limits than without diode compensation.

Cascaded Transistors

Transistors connected in cascade with overall voltage feedback are basic building blocks of amplifiers. When the collectors of two NPN transistors are connected together with the base signal of the second transistor derived entirely from the emitter of the first, an NPN Darlington transistor (Figure 4.3.8a) is described that has only three terminals: the emitter of the second, the base of the first, and the common collector of both. Additional internal base-emitter resistors ensure that leakage currents cannot cause conduction in the last transistor. Such a cascaded transistor may have a current gain nearly equal to the product of the two current gain factors, requires an input voltage which is two base-emitter voltages higher than the emitter voltage, and can

4-56 Microphone Devices and Systems

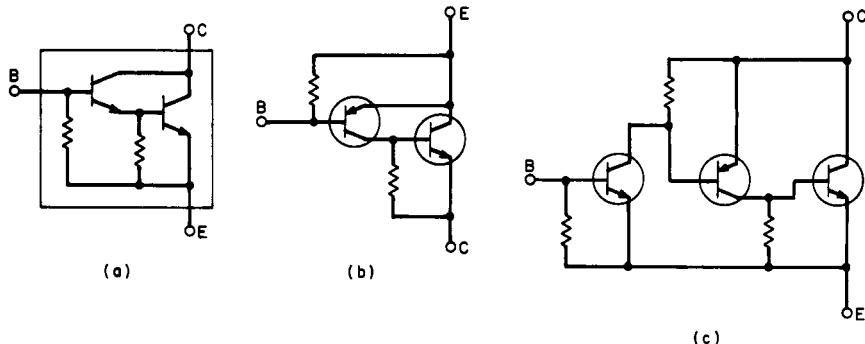


Figure 4.3.8 Cascaded circuits: (a) compound Darlington or coupled emitter follower, (b) compound PNP-NPN transistors, (c) triple compound NPN-PNP-NPN transistors.

reach in normal operation a minimum voltage drop between the external collector-emitter terminals equal to one base-emitter voltage. Two PNP transistors can be similarly interconnected.

The collector of a first PNP transistor can be connected to supply the entire base current of a second NPN transistor, with the emitter of the first transistor connected to the collector of the second transistor (Figure 4.3.8b). The entire assembly now functions as a compound PNP transistor having as its base terminal the base of the first PNP transistor, as its collector the emitter of the second NPN transistor, and as its emitter the emitter of the first PNP and the collector of the second NPN transistor. The input voltage now must be one base-emitter voltage lower than the emitter voltage because of the reversed current flow in PNP transistors, and the minimum voltage drop between the collector-emitter terminals is now one base-emitter voltage. The two compound transistors described here are the compound output devices in quasi-complementary push-pull amplifiers.

Three or more transistors of like or mixed polarity may be cascaded, such as PNP-NPN-NPN or NPN-PNP-NPN (Figure 4.3.8c), to form compound PNP or NPN transistors, respectively. Here, the polarity of the input transistor defines the polarity of the compound transistor. The minimum voltage drop and the required input voltage may be different in each connection, becoming highest when only devices of the same polarity are used.

Parallel-Connected Devices for High Currents

When high currents have to be delivered to a load, several transistors or FETS are often connected in parallel, with each device sharing a portion of the output current (Figure 4.3.9a). Nearly equal current sharing can be achieved when all devices are matched to each other as much as possible. Current sharing can be improved when each of the devices has local current feedback with equal separate emitter or source resistors connected to the common emitter or source connection and all devices share a common heat sink. The emitter resistor for bipolar transistors is typically a fraction of $1\ \Omega$ which allows current sharing to currents as low as a fraction of $1\ A$.

The circuit layout for a parallel connection must be done carefully to avoid constructing an oscillator circuit at very high frequencies. When using power FETs, it is necessary to connect

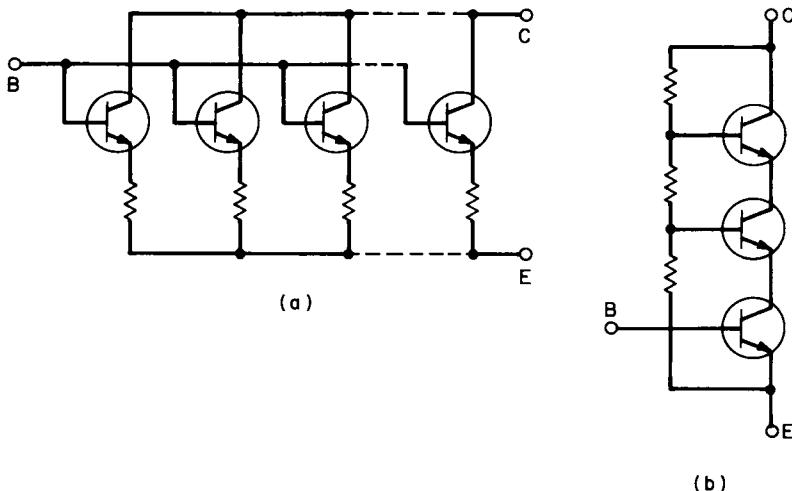


Figure 4.3.9 High-voltage and high-current circuits: (a) current-sharing parallel transistors, (b) totem-pole series-connected transistors.

series resistors of a few ohms in series with each gate lead and, perhaps, to have a ferrite bead in each gate lead to avoid oscillation.

Series-Connected Devices for High Voltage

When high voltages have to be delivered to a load and single devices are incapable of operating at the maximum peak voltage, several devices can be connected in series to share the voltage while conducting nearly the same current. The resulting *totem-pole connection* for transistors involves connecting the transistors so that the emitter of the second transistor is tied to the collector of the first and the emitter of the third transistor is tied to the collector of the second, and so on (Figure 4.3.9b). A series string of as many equal resistors as there are transistors has its ends connected to the collector of the last transistor and the emitter of the first transistor, and each junction is connected in the same sequence to the base of the same transistor in the sequence, except to the first transistor, whose base receives the input signal. The object of this circuit is to have the first transistor operate as a grounded-emitter device, driving all the others as grounded-base devices. This goal is not perfectly achieved, particularly at high frequencies and at high-output currents where voltage division in the resistor string under load departs from uniformity.

AC-Coupled Multistage Amplifiers

Amplification of a given signal usually does not include amplification of the dc component of the source of signals. One or more coupling capacitors between stages of amplification reduce the low-frequency response of the system and prevent the dc offset voltages from being propagated to the output (Figure 4.3.10a). When using transistor amplifiers in this fashion, the input impedance of a single-ended input stage after each capacitor may act as a partial rectifier diode for the pulsating signal and produce a low-frequency transient for each pulse. The solution to this

4-58 Microphone Devices and Systems

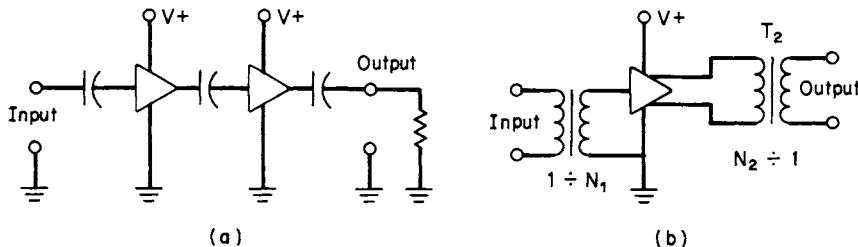


Figure 4.3.10 AC amplifier circuits: (a) cascaded, capacitor-coupled, (b) transformer-coupled.

problem may be, in part, the use of FET circuits, the use of push-pull circuits, and the selection of low-frequency time constants in the amplifier and power supply filtering circuits.

Amplifier stages may be coupled with transformers to the signal source and load and to each other (Figure 4.3.10b). Transformers are excellent devices that can reject common-mode interfering signals which may appear on a program line, and they can also match the source and load impedances to the amplifier circuit. Such an impedance match is needed, for example, in amplifiers operating from a power supply directly connected to the power line where connection of external loudspeakers or headphones would present a shock hazard. Transformers are the only practical components that can match the devices in broadband radio-frequency power amplifiers to source and load.

4.3.2d Power Output Stages

Power output stages of audio or video amplifiers usually are called upon to drive a variety of loads, which may or may not be connected when the signal or the power supply is turned on. Consequently, not only must power amplifiers be stable with any load, but they must be tolerant of excessive signals or loads unless such conditions are prevented from occurring [6].

Single-Ended Amplifiers

A single-ended amplifier has only one single or compound transistor or FET acting as a variable controlled resistor between power supply and load. The load may be coupled to the output stage through a capacitor or a transformer, which must also return the average direct current to the power supply. Single-ended amplifiers intended for audio-frequency amplification are usually of low power output capability and generally operate from a single power supply voltage. In a single-ended amplifier, transformer- or choke-coupled to the load, the bias current through the device must be at least equal to the peak current through the load and the peak voltage across the load must be less than the supply voltage when the turns ratio of the transformer is 1:1 between the primary and secondary windings.

Push-Pull Amplifiers

A push-pull amplifier has at least one pair of single or compound output devices that act as variable resistors between supply and load, with the first device *pushing* load current in one direction

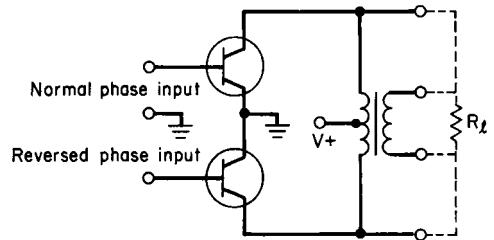


Figure 4.3.11 Series-ac, parallel-dc push-pull amplifier.

while the other *pulls* load current in the opposite direction under control of the input signal. When the two devices have the same characteristics and the driving signal is equally balanced, the direct current and all even-harmonic distortion in the load current are canceled [7].

Parallel-DC, Series-AC Amplifiers

The usual transformer-coupled amplifier (Figure 4.3.11) has two like devices connected between ground and the end of the primary winding of the transformer, with the supply voltage fed to the center tap of the same winding. Signal voltage is fed to the two devices with opposed phase so that one device increases conduction of current while the other decreases conduction.

The load may be connected to a secondary winding or between the ends of the primary winding. When it is connected in the latter way, maximum conduction of one device, resulting in nearly 0 V at that point, will raise the voltage at the opposite device to almost twice the power supply voltage, which then becomes the peak voltage across the load.

The peak-to-peak voltage across the load then becomes nearly 4 times the power supply voltage, and the peak load current becomes nearly 2 times the power supply voltage divided by the load resistance. The average power supply current is equal to the sum of the average current drawn by each device. Thus, the dc supply load is in parallel, while the ac load signal is in series between the two devices.

The parallel-dc, series-ac push-pull amplifier provides a very high relative power output when supply voltage is low, as in the 12-V automotive electrical system. The transformer- or choke-coupled amplifier makes use of two like devices and is therefore the preferred connection in radio-frequency power amplifiers.

Series-DC, Parallel-AC Amplifiers

With the availability of complementary transistors as amplifiers, the single-ended or half-bridge amplifier became practical as a transformerless power amplifier in the early 1960s. Prior to that time such amplifiers were constructed by using driver transformers or floating phase inverter amplifiers.

A half-bridge amplifier that is fully balanced has one device connected between the load and one power supply and a second complementary device connected between the load and a second power supply of opposite polarity but the same voltage. The load and the two power supplies are connected to a common ground. The driving voltage is fed to both devices without phase inversion, decreasing conduction in one device while increasing conduction of an opposite current in the other direction.

4-60 Microphone Devices and Systems

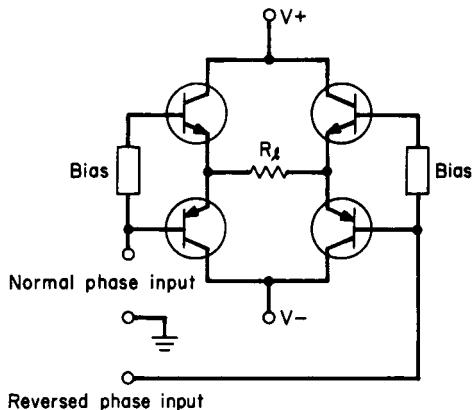


Figure 4.3.12 Full-bridge amplifier coupled to a load.

The maximum peak voltage across the load will be slightly less than one supply voltage. The maximum peak-to-peak voltage thus cannot exceed the total of the two supply voltages, which are series-connected. The two devices operate in parallel for ac signals, where one device increases a current of one polarity while the other decreases a current of opposite polarity.

Full-Bridge Amplifiers

Full-bridge amplifiers are constructed by using two half-bridge amplifiers with the load connected between the two output terminals and the two input terminals driven by signals of opposite polarity from a phase inverter circuit (Figure 4.3.12). Peak voltage across the load then becomes nearly equal to the total supply voltage, and peak-to-peak load voltage becomes nearly twice the total supply voltage. This type of amplifier connection is preferred over the totem-pole-transistor connection when high-voltage limitations of power devices restrict total available output power into a fixed load resistance without using a transformer.

4.3.3 Classes of Amplifiers

Amplifiers are described as classes depending on the angle of conduction of signal current and voltage-current relationships in the load.

Class A amplifiers (Figure 4.3.13a) conduct signal current throughout the cycle of the signal waveform. They have the lowest distortion before feedback and may be single-ended or push-pull. An ideal Class A amplifier can have a sine-wave output efficiency not exceeding 50 percent at full output.

Class B amplifiers (Figure 4.3.13b) conduct signal current exactly for one-half of the cycle of the input-signal waveform. In a push-pull Class B amplifier, one device conducts for one half-cycle, and the other device conducts for the remaining half-cycle. Linear Class B radio-frequency amplifiers may have only one device, since the second-harmonic components are filtered out in the narrowband matching network. An ideal Class B amplifier can have a maximum sine-wave efficiency not exceeding 78 percent at full power output.

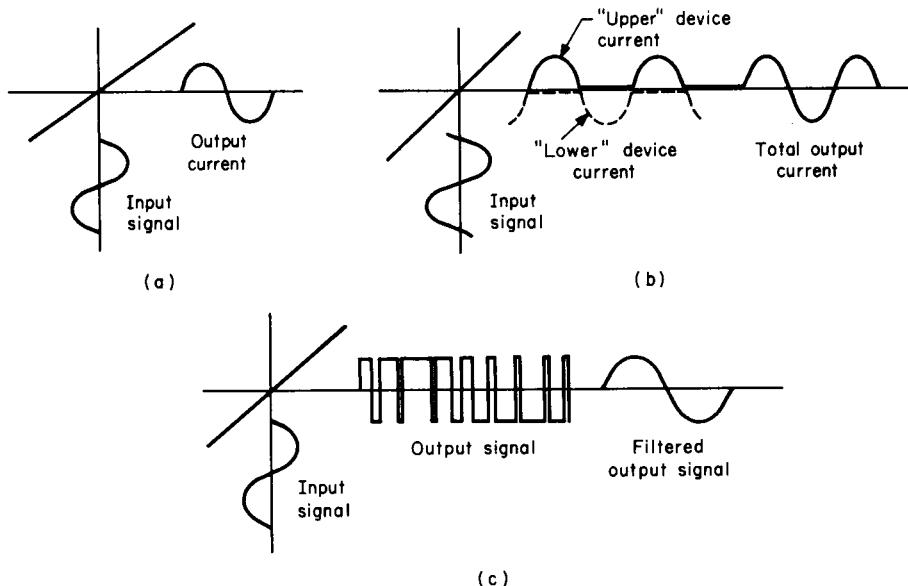


Figure 4.3.13 Output current as a function of input signal for amplifiers of various classes: (a) Class A, (b) Class B, (c) Class D.

Class AB amplifiers have a conduction angle between full conduction and half-cycle conduction and efficiencies between Classes A and B. Most conventional audio-frequency amplifiers are adjusted in this way. As lower power outputs are needed with variations in signal amplitude, the efficiency of Class A, AB, and B amplifiers will decrease proportionally to output voltage, decreasing toward zero at very low output. When the load of such an amplifier is a reactive impedance, such as a loudspeaker, efficiency will decrease still further, since any voltampere energy sent to a reactance in one part of a cycle will be returned to the source and the resistance in the circuit in the other part of the cycle.

Class C amplifiers conduct for less than one-half of a complete signal cycle. These amplifiers are used primarily as radio-frequency amplifiers with the load tuned to the signal frequency.

Class D amplifiers (Figure 4.3.13c) are switching amplifiers using a high-frequency carrier signal where the positive pulse on time is proportional to the modulation amplitude. The negative pulse on time completes the rest of the cycle as with Class A amplifiers. In other designs, separate circuits control positive and negative pulses as with Class B amplifiers. The load is isolated from the amplifier output stage with a low-pass filter that does not consume the high-frequency pulse energy. Class D amplifiers have a theoretical efficiency of 100 percent at all signal levels but are difficult to design for wideband low-distortion operation because of the short switching transition times required of the final high-power output stages and the difficulty of design of feedback loops.

Class E amplifiers have as input signals rectangular pulses. The output load is tuned, but the output voltage resembles a damped single pulse.

4-62 Microphone Devices and Systems

There is no generally accepted agreement on naming amplifiers of classes above E. Several types of amplifiers have been described with differing letters. These include amplifiers in which several stages are connected in series, drawing power from several power supplies through isolation diodes to provide output signal to a load. For small signals, only the stages connected to the low-voltage supplies conduct current. As these stages saturate, the next stages then conduct current from the next higher power supplies through the saturated stages into the load, and so on.

In a different version, a normal Class B amplifier obtains supply voltage from a high-efficiency switching power supply, the output voltage of which is raised as output voltage demands are increased. In all these amplifier designs attempts are made to improve efficiency with varying signal levels and with program-signal waveforms, which are nonsinusoidal.

4.3.4 Gain Block and the Operational Amplifier

A large number of audio- and video-frequency circuits are constructed by using operational amplifiers because they permit these circuits to be designed with minimum complexity of components. In most applications, the open-circuit voltage gain of an operational amplifier will be much larger than the gain of the amplifier.

An amplifier gain block is *matched* to the output of the previous gain block or other circuit when it is able to extract maximum power from the previous circuit. When the amplifier gain block draws little current from the preceding circuit, the gain block is said to be *bridged* across that circuit and must have a relatively high input impedance.

An operational amplifier connected as a noninverting amplifier can have a high input impedance at its positive input while receiving feedback voltage from a voltage divider connected between output and ground, with the voltage-divider junction connected to the negative input (Figure 4.3.14a). The voltage gain is the voltage ratio of the divider.

An inverting operational amplifier can be used as a gain block, with the input resistor connected between the source and the negative input and matching the desired load of the previous stage. A feedback resistor connected between the output and the same negative input then sets voltage gain equal to the ratio between the two resistors; the positive input is grounded (Figure 4.3.14b). Because there exists only a very low voltage at the negative input, several input resistors can be connected between various input sources and the negative input. In this fashion, these input signals can be mixed together with little danger of feedthrough between signals at each source.

If resistance-capacitance networks are used in place of resistors, equalizer blocks can be designed. With more complex networks, high-pass, low-pass, bandpass, and phase-shifting all-pass blocks result.

These circuits make use of the common-ground mode and thus are unbalanced circuits. An operational amplifier can also amplify the voltage difference between the two wires of a balanced program line while having only little sensitivity to common-mode signals arriving in phase, thereby reducing ground-loop voltages.

The simplest connection involves the use of two identical voltage-divider resistor pairs having their junctions connected to the positive and negative inputs (Figure 4.3.14c). The input terminals of the two networks are then connected to the two wires of the signal line. The return terminal of the network connected to the positive input is grounded, and the other return terminal is connected to the output of the operational amplifier. The *differential voltage gain* of such a cir-

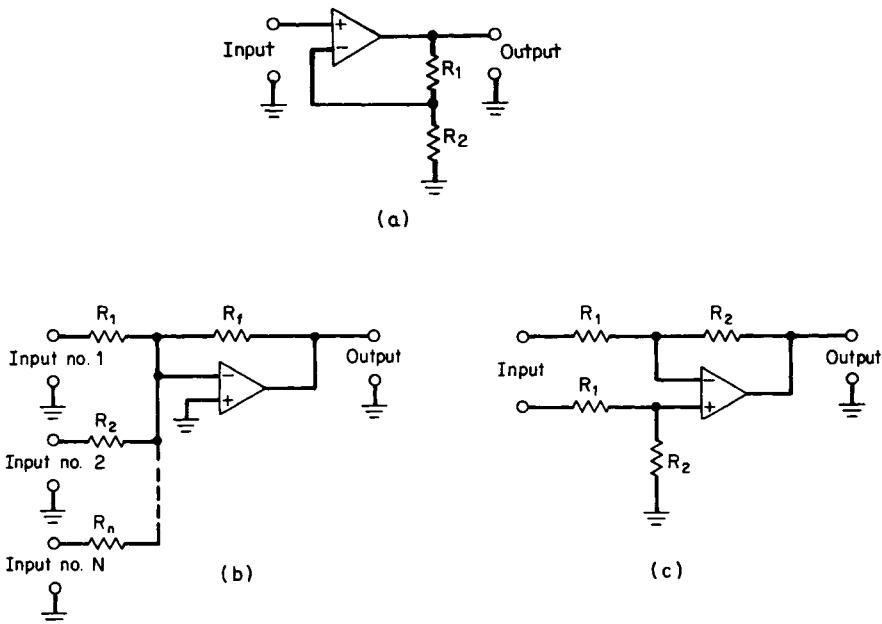


Figure 4.3.14 Operational amplifier circuits: (a) the gain-block voltage gain = $(1 + R_1/R_2)$ with the same polarity of the input and output signals, (b) the gain-block voltage gain = (R_f/R_1) , $1 (R_f/R_2)$ or $-(R_f/R_n)$ with opposite polarity of input and output signals, (c) the gain-block differential voltage gain = (R_2/R_1) and low common-mode gain, limited by resistor matching and loop gain.

cuit is equal to the ratio of the resistance values, while the *common-mode gain* is limited by resistor accuracy and the residual errors of the operational amplifier, and particularly its common-mode rejection. The input impedances of the positive and negative inputs to the circuit are not equal. A number of gain blocks can be interconnected to become a more complex amplifier system [8].

4.3.4a Feedback and Feed Forward

Feedback is the return of a fraction of the output signal to the input (Figure 4.3.15). The returned fraction is added to the input signal at the feedback node in the feedback loop of the system. The input signal to the system with feedback for the same output as before feedback is now the vector sum of the original input signal and the feedback signal. Feedback is negative when the new required input signal is larger than the signal without feedback and positive when it is smaller.

Feedback may be acoustic, mechanical, or electronic, depending upon the type of signal amplified. In amplifiers for audio- and video-frequency signals the feedback signal is usually a portion of the output voltage or output current. When the returned fraction is negative and is obtained through a linear network, the reduction ratio of amplifier errors, such as distortion or phase shift, is proportional to the reduction in amplifier gain due to feedback. In the limit, with

4-64 Microphone Devices and Systems

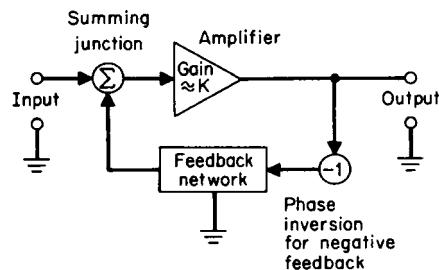


Figure 4.3.15 Negative-feedback amplifier. Voltage gain = $K/[1 + (Kb)]$, where K = amplifier voltage gain without feedback and b = gain of the feedback network, usually 1 on less.

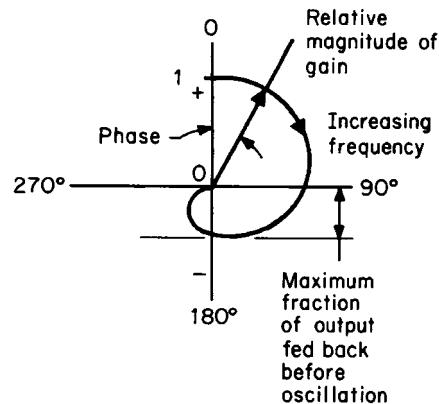


Figure 4.3.16 Polar plot of amplifier gain and phase from dc relative to frequency. At higher frequencies, a negative component of the gain characteristics limits the maximum usable feedback before oscillation occurs.

very high amplifier gain before feedback, the response of the amplifier with respect to frequency will nearly equal the reciprocal of the loss of the feedback network as measured from output to input [9].

Linear Feedback

Linear feedback exists when the feedback signal is only a level-independent portion of the output signal. Negative feedback cannot be applied in ever-increasing amounts because all amplifiers have increasing amounts of phase shift as the limits of the frequency range are approached, particularly at high frequencies (Figure 4.3.16). Whenever phase shift of the feedback signal is between 90 and 270° of phase with respect to the input signal at the feedback node, amplification with feedback will be greater with feedback than without. In the limit, no more positive feedback signal than the original input signal can be returned to the amplifier input before oscillation starts at the frequency where the returned feedback signal is equal to the original input signal in both amplitude and phase. This condition is desirable only in oscillators, not in amplifiers.

Another reason for using little negative feedback is that some intermediate stage of the amplifier may current-limit feeding a capacitor before the output stage is overloaded by input signals. This distortion is known as *slew-rate limiting* and is a cause of transient intermodulation distortion.

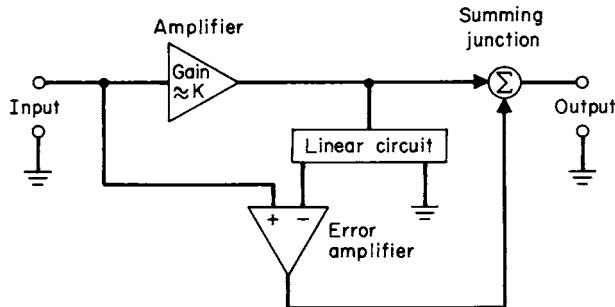


Figure 4.3.17 Feed-forward amplifier phase or gain error correction by negative feedback.

Feed Forward

Feedback is a method of correcting amplifier errors after they have occurred and have been compared with the input signal (Figure 4.3.17). One feed-forward method is to measure the errors that an amplifier will introduce into the output signal and then to feed these errors, inverted in phase, directly to the output summing junction through a separate path, which may also include an amplifier of lesser output range, since only the error signal will have to be supplied. A portion of the resulting output signal may also be fed back for further error correction.

In feed-forward circuits, the error signal is handled by an amplifier, separate from the amplifier whose errors need correcting. In feedback circuits, the error signal is handled by the amplifier causing the errors.

Nonlinear Feedback

Precision rectifier circuits make use of nonlinear feedback. Here, one or more diodes in the feedback loop of an operational amplifier result in an output signal which is the half-wave or full-wave rectified signal originally present at the input of the circuit (Figure 4.3.18). Rectification of signals is a function needed in signal-processing circuits, such as compressors, expanders, meters, and noise reduction circuits for pulse or random noise.

Voltage Feedback

The output voltage of an amplifier, attenuated in a voltage-divider network, is subtracted from the input voltage, resulting in the amplifier input voltage with feedback. The gain reduction ratio is equal to the reduction ratio of the output impedance of the amplifier, equal to the distortion reduction ratio for signals of the same output voltage, and inversely proportional to the input impedance increase ratio of the amplifier. When the loop gain of the amplifier is very high, as is normal in operational amplifiers, the voltage gain of the amplifier with feedback is nearly equal to the inverse of the loss of the feedback attenuator, and the output impedance of the amplifier becomes very low. Therefore, any variation in amplifier load will have little effect on output voltage until the maximum output-current capability of the amplifier is reached.

In an alternate circuit, the amplifier need not add signal and feedback voltages because only one input terminal is required. Here, the output voltage is converted to a current using an imped-

4-66 Microphone Devices and Systems

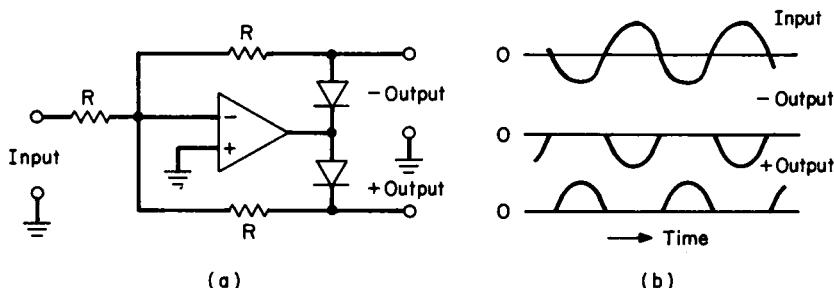


Figure 4.3.18 Nonlinear precision-rectifier feedback: (a) operational amplifier circuit, (b) sine-wave input-output waveforms.

ance and fed to the input terminal, which also receives input current from the signal source, perhaps through a second impedance. The amplifier now amplifies the difference between the two currents. Output impedance and distortion are affected similarly.

Voltage feedback causes the amplifier to become a nearly constant voltage source with a fixed input signal [10].

Current Feedback

The output current of an amplifier may conveniently be converted to a voltage by passing this current through an impedance connected in series with the load impedance. The resulting voltage is then applied as a feedback voltage to the amplifier. With such a circuit, the internal output impedance ratio of the amplifier will be equal to the inverse of the current gain reduction ratio achieved, and the distortion reduction ratio will be the same as the output-current reduction ratio for the same input signal. Again, the feedback signal may be a feedback voltage, with the amplifier utilizing the difference in feedback and signal voltages, or a feedback current, with the amplifier supplying current gain.

Current feedback causes the amplifier to become a nearly constant current source with a fixed input signal.

4.3.4b Output and Input Impedance

The output impedance of an amplifier usually varies with frequency and is mostly resistive for an amplifier that has a constant fraction of the output fed back to the input or an intermediate stage. The output impedance is sometimes expressed as the *damping factor* of the amplifier, defined as the ratio of nominal load resistance to amplifier internal output impedance. The value of output impedance also includes any impedances connected between the output wiring terminals and the actual output and ground nodes of the circuit.

Another measure of output impedance of an amplifier is *regulation*, usually measured in percentages and defined as the change in output voltage as the nominal load is changed from open-circuit to rated load. Damping factor and regulation are normally rated at mid-frequencies. At the extremes of the frequency range, the output impedance of an amplifier will be different from its midfrequency value because the loop gain of an amplifier decreases, particularly at high frequencies.

cies. The output impedance of an amplifier can then be described as a two-terminal network of resistors and reactive elements.

Input impedance of an amplifier may often be set by a physical resistor connected across the input terminals that shunts the input impedance of the circuit. An additional component of the input impedance sometimes rated is the *input capacitance*, partially composed of the wiring capacitance and any capacitors which are part of the radio-frequency filters of the input circuit or are capacitors designed to give the desired termination to certain source signals. In low-noise amplifiers, the input resistance component will be largely determined by feedback to the input circuit.

When a number of amplifiers and other circuits are connected in series so that each amplifier amplifies the output signal from the previous circuit, the connections are often made on a voltage basis, in which each amplifier has a relatively high input impedance and a relatively low output impedance. Here, the amplifier takes very little of the load current that could be provided by the circuit at its input, and its own output voltage changes very little, whether supplying full-load current or not. This type of design is used most often in self-contained equipment or in pieces of equipment operated in close proximity to each other.

Equipment used as part of large distributed systems or with program transmission lines is often designed to present a constant output impedance and input impedance to match the nominal impedance of transmission lines. Audio line impedances of 150 or 600 Ω are common values. At video and radio frequencies, transmission-line impedances of 50, 75, and 300 Ω are preferred.

Equipment designed to operate in constant-impedance circuits is often rated in decibels with respect to 1 mW (dBm) of output power into a matched load. Gain or loss are given in decibels, and systems are designed on a power gain or loss basis. The advantage of operating circuits at matched transmission-line impedances is that reflections or echoes of signals will not be generated at the receiving end of a traveling signal.

4.3.4c Feed Forward and Correction of Estimated Errors

Some error correction can be accomplished by making a good estimate of the error and then pre-distorting the signal with opposing distortion. This model may contain amplitude or phase nonlinearities, or both. Such correction can be quite complex and requires considerable accuracy in modeling of distortion components.

4.3.4d Differential Amplifier

Amplifiers that allow the measurement or use of signals generated remote from the point of equipment location are called *instrumentation amplifiers*. These amplifiers have controlled amplification for the difference in voltage between two signals and very low amplification for the sum of the two signals, the measurement of the voltages made with the local ground reference. In general terms, these amplifiers have a controlled differential- or transverse-mode gain and a low common- or longitudinal-mode gain. The common instrumentation amplifier connection uses three operational amplifiers, with the first two circuits amplifying the two signals equally and sharing a common feedback resistor R_1 between the two negative inputs. The output signals then pass through two identical resistor attenuators R_3 and R_4 , with the resistor junctions connected to the positive and negative inputs of the third amplifier and resistor R_4 completing the connection between positive input and ground and negative input and output. A simplified-ver-

4-68 Microphone Devices and Systems

sion instrumentation amplifier omits the two input stages and connects signal wires directly to the two resistor networks, which now present unequal loads to the two signal wires.

The performance of a differential or instrumentation amplifier is measured by its *common-mode rejection*, which is the ratio of differential- to common-mode gain. This value is largely determined by the accuracy of resistor matching and the ratio of amplifier loop gain to circuit gain. Common signals must not exceed the maximum allowable common-mode input signal for the amplifier.

Differential amplifiers are used in audio, video, and computer equipment when signal sources are widely separated or when ground-loop signals may exist. The two input-signal leads of a differential amplifier are brought directly to the source of signals, often as a pair of twisted wires

4.3.5 References

1. Shockley, W: "The Theory of P-N Junctions in Semiconductors and P-N Junction Transistors," *Proc. JRE*, vol. 41, June 1953.
2. Shockley, W: "A Unipolar Field-Effect Transistor," *Proc. IRE*, vol. 40, November 1952.
3. Kirchner, R. J.: "Properties of Junction Transistors," *Trans. IRE PGA*, AU-3(4), July-August 1955.
4. Trent, R. L.: "Design Principles for Transistor Audio Amplifiers," *Trans. IRE PGA*, AU-3(5), September-October 1955.
5. Garner, L. H.: "High-Power Solid State Amplifiers," *Trans. IRE PGA*, 15(4), December 1967.
6. Fewer, D. R.: "Design Principles for Junction Transistor Audio Power Amplifiers," *Trans. IRE PGA*, AU-3(6), November-December 1955.
7. Petersen, A., and D. B. Sinclair: "A Singled-Ended Push-Pull Audio Amplifier," *Proc. IRE*, vol. 40, January 1952.
8. Widlar, R. J.: "A Unique Current Design for a High Performance Operational Amplifier Especially Suited to Monolithic Construction," *Proc. NEC*, 1965.
9. Black, H. S.: U.S. Patent 2,102,671.
10. Walker, P. J.: "A Current Dumping Audio Power Amplifier," *Wireless World*, December 1975.

4.3.6 Bibliography

Harper, C. A. (ed.): *Handbook of Components for Electronics*, McGraw-Hill, New York, N.Y., 1977.

Lynn, D. K., C. S. Meyer, and D. C. Hamilton (eds.): *Analysis and Design of Integrated Circuits*, McGraw-Hill, New York, N.Y., 1967.

Weinberg, L.: *Network Analysis and Synthesis*, McGraw-Hill, New York, N.Y., 1967.

Section **5**

Sound Reproduction Devices and Systems

One of the problems with selecting a high-quality monitor system lies in the difficulty of defining quality. Some not-too-scientific descriptions may develop from discussions with other users, and even with knowledgeable authorities. Terms such as "solid bass," "smooth highs," "tight," or "clean" all may be mentioned as monitor system requirements. Trying to incorporate these subjective requirements into a working system is almost impossible. On the other hand, selecting a speaker solely on the basis of frequency response and harmonic distortion is likewise inappropriate. It is difficult to equate either scientific measurements or subjective considerations with how "good" or "bad" a particular speaker system sounds.

For the purposes of this examination, the term *speaker* will refer to a single transducer. Monitor or monitor system will refer to an assembly of speaker(s), enclosures and, where appropriate, crossovers and amplifiers. In fact, it might be best to consider a monitor system as having at least three major components: source driver (amplifier), transducer (speaker) and mounting assembly (cabinet).

A speaker cannot produce acoustic energy without being driven by an electronic source. It likewise requires an enclosure to properly couple acoustic energy into the listening environment in a controlled manner. Even the most expensive speaker, if set on a shelf without an appropriate enclosure, will perform poorly.

In This Section:

Chapter 5.1: Electroacoustic Transducers	5-5
Introduction	5-5
Basic Equations and Features of Dynamic Transducers	5-5
Basic Equations and Features of Electromagnetic Transducers	5-7
Basic Equations and Features of Electrostatic Transducers	5-9
Basic Equations and Features of Piezoelectric Transducers	5-11
Control System and Its Acoustic Characteristics	5-14
Bibliography	5-15

5-2 Section Five

Chapter 5.2: Direct-Radiator Loudspeakers	5-17
Introduction	5-17
Piston Source in an Infinite-Plane Baffle	5-17
Baffle Shape and Acoustic Characteristics	5-20
Acoustic Characteristics of Rigid Disk with Constant-Force Drive	5-22
Bibliography	5-25
Chapter 5.3: Dynamic-Type Direct-Radiation Speakers	5-27
Introduction	5-27
Operational Details	5-27
Equivalent Circuit and Frequency Response	5-28
Efficiency	5-32
Nonlinear Distortion	5-34
Driving-Force Distortion	5-34
Support-System Distortion	5-35
Air Distortion	5-38
Frequency-Modulation Distortion	5-39
Diaphragm and Support System	5-41
Diaphragm	5-41
Support System	5-41
Bibliography	5-44

On the CD-ROM

- “Sound Reproduction Devices and Systems” by Katsuaki Satoh, an archive chapter from the first edition of the *Audio Engineering Handbook*. This material provides an excellent, in-depth examination of a wide variety of loudspeakers.

References for this Section:

- Allison, R., et al.: “On the Magnitude and Audibility of FM Distortion in Loudspeakers,” *J. Audio Eng Soc.*, vol. 30, no. 10, pg. 694, 1982.
- Beranek, L. L.: *Acoustics*, McGraw-Hill, New York, N.Y., pg. 183–185, 1954.
- Hayasaka, T., et al.: *Onkyo-Kogaku Gairon (An Introduction to Sound and Vibration)*, Nikkan Kogyo Shinbunshya, pg. 67, 1973 (in Japanese).
- Hayasaka, T., et al.: *Onkyo-Shindo Ron (Sound and Vibration)*, Maruzen Kabushikigaishya, pg. 201, 1974 (in Japanese).
- Hirata, Y.: “Study of Nonlinear Distortion in Audio Instruments,” *J. Audio Eng. Soc.*, vol. 29, no. 9, pg. 607, 1981.
- Kinsler, L. E., et al: *Fundamentals of Acoustics*, Wiley, New York, N.Y., 1982.
- Melillo, L., et al.: “Ferrolluids as a Means of Controlling Woofer Design Parameters,” presented at the 63d Convention of the Audio Engineering Society, vol. 1, pg. 177, 1979.
- Morse, P. M.: *Vibration and Sound*, McGraw-Hill, New York, N.Y., pg. 326, 1948.

Sound Reproduction Devices and Systems 5-3

- Morse, P. M., and K. U. Ingard: Theoretical Acoustics, McGraw-Hill, New York, N.Y., pg. 366, 1968.
- Niguchi, H., et al.: "Reinforced Olefin Polymer Diaphragm for Loudspeakers," *J. Audio Eng. Soc.*, vol. 29, no. 11, pg. 808, 1981.
- Okahara, M., et al: *Audio Handbook*, Ohm Sya, pg. 285, 1978 (in Japanese).
- Olson, H. F.: *Elements of Acoustical Engineering*, Van Nostrand, Princeton, N.J., 1957.
- Rayleigh, J. W. S.: *The Theory of Sound*, Dover, New York, N.Y., pg 162, 1945.
- Sakamoto, N.: *Loudspeaker and Loudspeaker Systems*, Nikkan Kogyo Shinbunshya, pg. 36, 1967 (in Japanese).
- Sakamotoet, N., et. al.: "Loudspeaker with Honeycomb Disk Diaphragm," *J. Audio Eng. Soc.*, vol. 29, no. 10, pg. 711, 1981.
- Shindo, T., et al: "Effect of Voice-Coil and Surround on Vibration and Sound Pressure Response of Loudspeaker Cones," *J. Audio Eng. Soc.*, vol. 28, no 7-8, pg. 490, 1980.
- Suwa, H., et al.: "Heat Pipe Cooling Enables Loudspeakers to Handle Higher Power," presented at the 63d Convention of the Audio Engineering Society, vol. 1, pg. 213, 1979.
- Suzuki, H., et al.: "Radiation and Diffraction Effects by Convex and Concave Domes," *J. Audio Eng Soc.*, vol. 29, no. 12, pg. 873, 1981.
- Takahashi, S., et al.: "Glass-Fiber and Graphite-Flake Reinforced Polyimide Composite Diaphragm for Loudspeakers," *J. Audio Eng. Soc.*, vol. 31, no. 10, pg. 723, 1983.
- Tsuchiya, H., et al.: "Reducing Harmonic Distortion in Loudspeakers," presented at the 63d Convention of the Audio Engineering Society, vol. 2, pg. 1, 1979.
- Yamamoto, T., et al.: "High-Fidelity Loudspeakers with Boronized Titanium Diaphragm," *J. Audio Eng. Soc.*, vol. 28, no. 12, pg. 868, 1980.
- Yoshihisa, N., et al.: "Nonlinear Distortion in Cone Loudspeakers," Chuyu-Ou Univ. Rep., vol. 23, pg. 271, 1980.

Chapter
5.1
Electroacoustic Transducers

Katsuaki Satoh

5.1.1 Introduction

Conversion from electrical signals to acoustic signals ordinarily does not involve direct electroacoustic transformation; the electrical signal is transformed into mechanical vibration, which then is transformed into an acoustic signal.

The following transducers are used in the audio field generally as electromechanical transducers: electrodynamic transducers, electromagnetic transducers, electrostatic transducers, and piezoelectric transducers.

5.1.2 Basic Equations and Features of Dynamic Transducers

Among the various forms of transducers listed above, the electrodynamic type is the basis for the design of the majority of loudspeakers in use today. Invented by C. W. Rice and E. W. Kellogg in 1925, when combined with the vacuum-tube amplifier, it provided the means for the use of audio technology in applications far greater than the telephone, introduced 50 years earlier by Alexander Graham Bell. Figure 5.1.1 shows the principle of operation. A permanent magnet and magnetic-pole pieces form a uniform magnetic field in the gap. The coil vibrating direction is at right angles to the magnetic field so that the force acts on the coil in accordance with the Fleming rule. This relationship is expressed by the following equation

$$F_d = BIl \quad (5.1.1)$$

Where:

F_d = driving force, N

B = flux density, Wb/m²

l = total length of coil, m

I = current flowing into coil, A

5-6 Sound Reproduction Devices and Systems

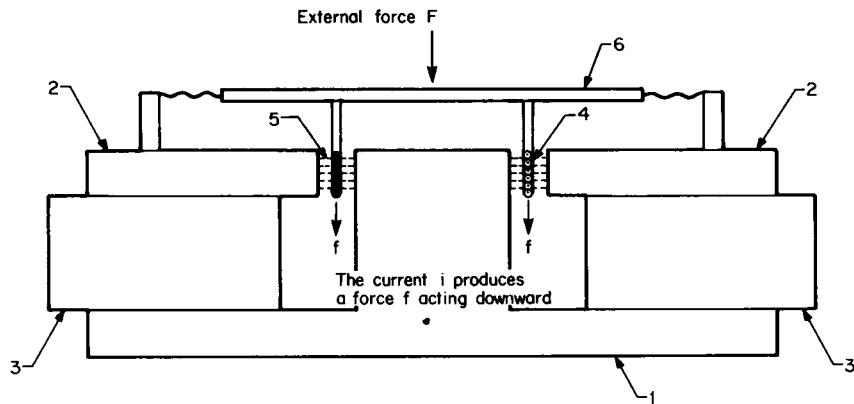


Figure 5.1.1 Simplified form of a moving-coil transducer consisting of a voice coil cutting a magnetic field of a flux density B . 1, 2 = pole pieces; 3 = permanent magnet; 4 = voice coil; 5 = magnetic flux; 6 = diaphragm.

Assuming the velocity at which a coil moves by means of driving force F_d to be v , the electromotive force E_d arising from this movement is in the opposite direction to the direction of current I . Therefore, E_d is determined by

$$E_d = -Blv \quad (5.1.2)$$

Where:

E_d = counterelectromotive force (V)

v = moving-coil velocity (m/s)

BI in Equations (5.1.1) and (5.1.2) is called the *power coefficient A*, which shows the conversion efficiency of a dynamic transducer. Assuming the mechanical impedance of the vibrating system as viewed from the coil side to be Z_m , the force acting on the coil corresponds to a summation of external forces F and driving forces F_d which is balanced with drag $Z_m v$.

$$F + F_d = Z_m v \quad (5.1.3)$$

Where:

F = external force, N

F_d = driving force, N

Z_m = mechanical impedance of the vibrating system, mechanical ohms

By substituting Equation (5.1.1), F is found as follows

$$F = Z_m v - AI \quad (5.1.4)$$

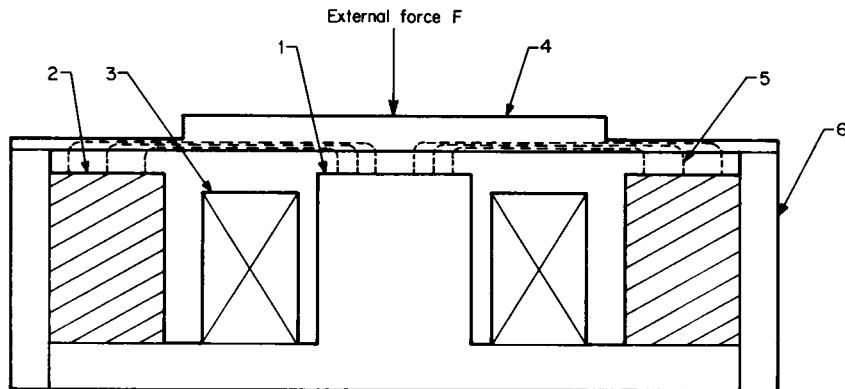


Figure 5.1.2 Simplified form of an electromagnetic transducer. 1 = pole piece; 2 = permanent magnet; 3 = drive coil; 4 = diaphragm; 5 = magnet flux; 6 = frame.

In the electrical system, assuming the electrical impedance of the driving coil to be Z_e the total voltage at the coil terminals corresponds to a summation of E and E_d whereby the following equation is obtained

$$E + E_d = Z_e I \quad (5.1.5)$$

Where:

E = voltage applied across coil terminals (V)

Z_e = electrical impedance of coil (Ω)

When Equation (5.1.2) is substituted, E is determined by

$$E = Z_e I + A v \quad (5.1.6)$$

Thus, Equations (5.1.4) and (5.1.6) are basic equations of the dynamic mechanical-electrical systems.

5.1.2a Basic Equations and Features of Electromagnetic Transducers

For an electromagnetic transducer, a magnetic diaphragm placed in a static magnetic field, in which a permanent magnet supplies the steady magnetic flux, is vibrated in an ac magnetic field formed by signal current flowing into a coil, thus generating a sound. This principle is shown in Figure 5.1.2. In this figure, assume that the diaphragm is subjected to attraction force F_m by the static magnetic field and the external force F . At this time, the diaphragm vibrates from a summation of static displacement ξ_s by the attraction force in the static magnetic field and by the dynamic displacement generated by an ac magnetic field and external force F . Assuming this to be ξ , ξ is expressed by

5-8 Sound Reproduction Devices and Systems

$$\xi = \xi_s + \xi_d \quad (5.1.7)$$

Where:

ξ = total displacement, m

ξ_s = static displacement, m

ξ_d = dynamic displacement, m

Assuming the equivalent circuit of the mechanical system of the diaphragm to be a single-resonance circuit with the number of degrees of freedom equal to 1, it may be regarded as being composed of the lumped constant of equivalent mass, the mechanical resistance, and the stiffness s . Therefore, from force-balanced conditions, the following is established

$$F + F_m = m \frac{\partial^2 \xi}{\partial t^2} + r \frac{\partial \xi}{\partial t} + s \xi \quad (5.1.8)$$

Where:

F = external force, N

F_m = attraction force by static magnetic field, N

m = equivalent mass, kg

r = mechanical resistance, N/m

s = stiffness, Ns/m

If the resistance is ignored, since it is quite negligible compared with magnetic resistance in the air space, the following relation is obtained

$$Z_m = r + j\omega m - j(s - s_n)/\omega \quad (5.1.9)$$

$$A = \mu_0 s_n U_0 / g_0^2 \quad (5.1.10)$$

$$s_n = \mu_0 S U_0^2 / g_0^2 \quad (5.1.11)$$

$$Z_e = Z_c + j\omega L_m \quad (5.1.12)$$

$$L_m = \mu_0 n^2 S / g_0 \quad (5.1.13)$$

$$F = Z_{m^v} - AI \quad (5.1.14)$$

$$E = Z_e I + A v \quad (5.1.15)$$

Where:

Z_m = mechanical impedance of the vibrating system, mechanical ohms

ω = angular frequency, rad/s

A = force factor, N/A

s_n = negative stiffness, Ns/m

L_m = inductance, H

Φ = total magnetic flux in space, Wb

B = flux density, Wb/m²

μ_0 = magnetic permeability in space, H/m

U_m = magnetic motive force of magnet, A/m

S = magnetic-pole area, m²

g_0 = quiescent space length in magnetic-force-free conditions, m

n = number of coil windings, turns

I = current flowing into coil, A

Z_c = coil electrical impedance, Ω

The difference between this transducer and the magnetic or dynamic transducer, in addition to the gap, is that negative stiffness in Equation (5.1.11) is generated. This stable condition is as follows

$$s < U_0^2 / 2\mu_0 S_0 g_0 R_{\text{air}}^2 \quad (5.1.16)$$

where R_{air}^2 = magnetic resistance out of the air space, A/m. This relationship is shown in Figure 5.1.3. Other differences are that because the coil is fixed, reliability is high and construction is simple, and that if the frequency is high, the force factor becomes small because of the coil inductance, thereby reducing efficiency.

5.1.2b Basic Equations and Features of Electrostatic Transducers

In the electrostatic transducer, when voltage is applied to two opposite conductive electrodes, an electrostatic attraction force is generated between them, and the action of this force causes a conductive diaphragm to be vibrated, thereby emitting sound. Figure 5.1.4 shows the construction. Electrostatic attraction force F_s , when signal voltage E is applied to polarized E_0 , is

$$F = \frac{\epsilon_0 S (E_0 + E)^2}{2(g_0 - \xi_0 - \xi_d)^2} \quad (5.1.17)$$

Where:

F = static attraction force, N

ϵ_0 = dielectric constant, F/m

S = electrode area, m²

E_0 = polarized voltage, V

E = signal voltage, V

g_0 = interelectrode distance, m

5-10 Sound Reproduction Devices and Systems

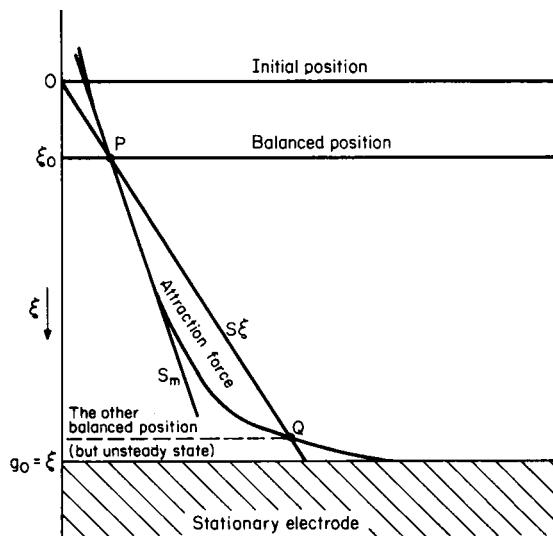


Figure 5.1.3 Static displacement shows balancing the attraction and the recover force.

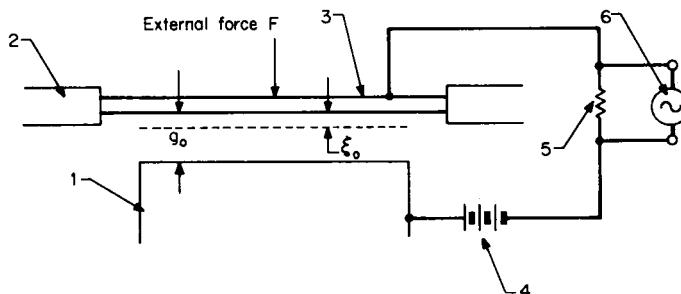


Figure 5.1.4 Cross-sectional view of an electroacoustic transducer. 1 = back electrode; 2 = clamping ring; 3 = diaphragm with electrode; 4 = polarizing power supply; 5 = polarizing electrical resistance; 6 = signal source.

$$\xi_s = \text{static displacement, m}$$

$$\xi_d = \text{signal displacement, m}$$

Considering the correspondence between electromagnetic and electrostatic types, Equation (5.1.17) is as shown in Table 5.1.1.

The basic equations of the electrostatic type are

$$s_n = \epsilon_0 S E_0^2 / g_0^3 \quad (5.1.18)$$

Table 5.1.1 Correspondence Between Electromagnetic and Electrostatic Types

Electromagnetic	nl	U_0	μ_0	$F_m\Phi$
Electrostatic	E	E_0	ε_0	F_sq

$$A = \varepsilon_0 S E_0 / g_0^2 \quad (5.1.19)$$

$$Z_m = r + j\omega m - j(s - s_n)/\omega \quad (5.1.20)$$

$$Y_s = j\omega(\varepsilon_0 S / g_0) \quad (5.1.21)$$

$$F = Z_m v - AE \quad (5.1.22)$$

$$I = Y_s E + A v \quad (5.1.23)$$

Where:

Z_m = mechanical impedance of the vibrating system, mechanical ohms

Y_s = electrical admittance of electrostatic capacity before displacement

F = external force, N

I = current, A

s_n = negative stiffness, N/m

A = force factor, N/V

r = mechanical resistance, Ns/m

m = mass, kg

s = diaphragm stiffness, N/m

ω = angular frequency, rad/s

Equations (5.1.22) and (5.1.23) are basic equations of the electrostatic transducer. Sensitivity of this transducer can be obtained by increasing the polarized voltage and reducing the distance between electrodes. Since the electrostatic type, unlike the electromagnetic type, has nothing to restrict attraction force, the force of the diaphragm to stick to the electrode is infinite. Therefore, the diaphragm requires a very large stiffness. Electrical impedance decreases inversely proportionally to the frequency since it is quantitative. This type is simply constructed, and since it has relatively good characteristics, it is used for high-range speakers and headphones.

5.1.2c Basic Equations and Features of Piezoelectric Transducers

If a crystal section is distorted with a force applied in one direction, positive and negative charges appear on the opposite surfaces of the crystal. This is called the piezoelectric direct effect. When a field is applied to the crystal section from the outside, a mechanically distorted force is generated. This is a piezoelectric counter effect. Ferrodielectric substances, which exhibit such a phenomenon, are polarized. These include crystal, piezoelectric crystals such as Rochelle salts, titanium oxide, and lead zirconate titanate (PZT). In general, PZT, having high

5-12 Sound Reproduction Devices and Systems

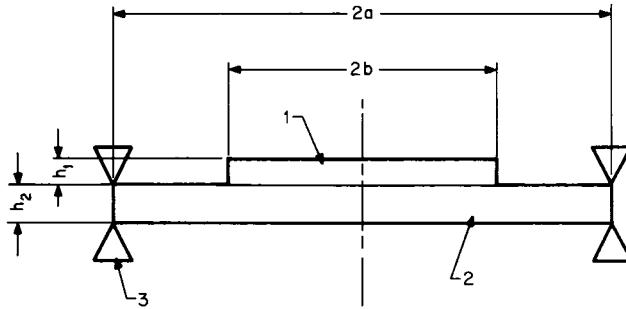


Figure 5.1.5 Simplified form of a monomorphic piezoelectric transducer. 1 = piezoelectric element ($E_1, P_1 \mu_1$); 2 = metal plate ($E_2, P_2 \mu_2$); 3 = supporting ring.

reliability and a reasonable price, is used as the piezoelectric element for the speaker. By using a configuration such as shown in Figure 5.1.5, the output sound level and resonance frequency can be determined. Power sensitivity q , when radian frequency $\omega \rightarrow 0$ is calculated by

$$q_0 = 20\log \left| \frac{K_1 U_0 Z_e}{E_0} \right| \quad (5.1.24)$$

Where:

q_0 = power sensitivity

U_0 = volume velocity, m^3/s

Z_e = electrical impedance of piezoelectric element, Ω

E_0 = input voltage, V

K_1 = constant

Assuming displacement at the piezoelectric element and laminated metal sheet to be ξ' and displacement at the peripheral metal part to be ξ , U_0 is found as follows

$$U_0 = \int_0^b 2\pi r \xi' dr + \int_b^a 2\pi r \xi dr \quad (5.1.25)$$

Z , which is mainly a qualitative component, is determined by

$$Z = \frac{K_2}{\pi \omega \epsilon_{33} T} \times \frac{h_1}{a^2 \eta^2} \quad (5.1.26)$$

Where:

ϵ = dielectric constant of piezoelectric element

$\eta = b/a$

$K_2 = \text{constant}$

To find the optimum condition of η if $\mu_1 = \mu_2 = \mu$ with radius a , material thickness $h = h_1 + h_2$, and the piezoelectric constant d_{31} , ε^T_{33} constant, the following is obtained

$$\frac{U_0\sqrt{Z}}{E} \propto \frac{\alpha(1+\beta)\sqrt{\beta}}{1+\alpha\beta} \times \frac{\mu[3+\mu-\eta(1+\mu)]}{(1+\mu)C + \eta^2 \left[(1-\mu)C + 2(1-\mu^2) \left(1 - \frac{3}{2}\zeta + \frac{3}{4}\zeta^2 \right) \right]} \quad (5.1.27)$$

$$C = (1-\mu_2) \left(\beta_2 + \frac{3}{2}\beta\zeta + \frac{3}{4}\zeta^2 \right) \alpha\beta + 2\mu(1-\mu) \left(1 - \frac{3}{2}\zeta + \frac{3}{4}\zeta^2 \right) \quad (5.1.28)$$

$$\zeta = (1 - \alpha\beta^2)(1 + \alpha\beta) \quad (5.1.29)$$

Where:

$$\alpha = Q_1/Q_2$$

μ = Poisson ratio, defined as the charge density at any point divided by the absolute capacitativity of the medium

From the above, it is found that $\eta = 0.5$ to 0.8 is better. β is dependent on α , but when the relative sensitivity of various metals is compared, $0.2 < \beta < 1.0$; therefore, aluminum is the best. The primary resonance frequency of the vibrator is

$$f_1 = \frac{2.22^2 h_1}{2\pi a^2 \beta} \sqrt{\frac{Q^2}{3p_2(1-\mu_2)} \left(1 - \frac{3}{2}\zeta + \frac{3}{4}\zeta^2 \right)} \quad (5.1.30)$$

Where:

$$f_1 = \text{primary resonance frequency, Hz}$$

$$Q_2 = \text{Young's modulus, N/m}^2$$

$$P_2 = \text{density, kg/m}^3$$

Assuming radius α , thickness h , and Poisson's ratio to be constant, C is determined by

$$f_1 \propto C \sqrt{\left(1 - \frac{3}{2}\zeta + \frac{3}{4}\zeta^2 \right) / \beta}$$

$$C = \sqrt{Q_2 / P_2} \quad (5.1.31)$$

where C = sound velocity, m/s. Furthermore, the resonance frequency of the vibrator is expressed as follows:

5-14 Sound Reproduction Devices and Systems

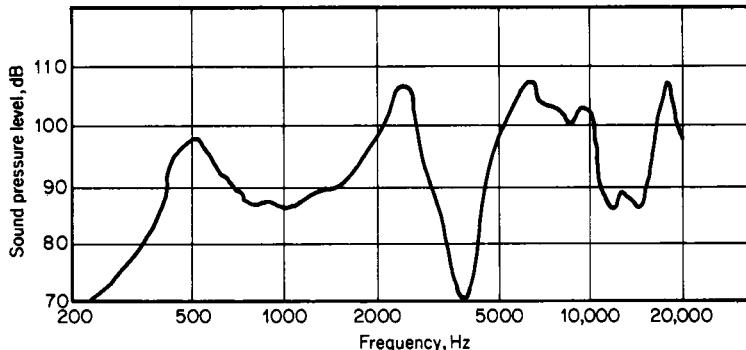


Figure 5.1.6 Frequency characteristics of a typical monomorphic piezoelectric transducer.

$$f_1 = \frac{1}{2\pi} \sqrt{\frac{s_o}{m_o}} \quad (5.1.32)$$

Where:

m_o = vibrator mass, kg

s_o = vibrator stiffness, N/m

However, to reduce mechanical Q , a small $s_o \times m_o$ is preferable, and therefore aluminum is the best material. Figure 5.1.6 shows the sound-pressure-frequency characteristics of a speaker with this construction.

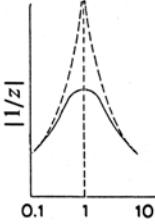
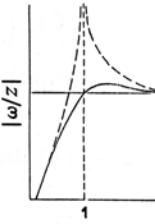
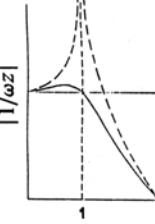
5.1.3 Control System and Its Acoustic Characteristics

For acoustic equipment, in the process of transforming electrical energy to acoustic energy, conversion from the electrical system to the mechanical system and from the mechanical system to the acoustic system is performed. The conversion process is expressed approximately by the equation

$$\frac{P}{E} = \frac{F}{E} \times \frac{V}{F} \times \frac{P}{V} \quad (5.1.33)$$

The left-hand term shows the ratio of electrical input to sound pressure, which should be kept constant regardless of frequency. However, the first term, the ratio of electrical input to driving force, and the third term, the ratio of diaphragm velocity V to sound pressure P on the right, are fixed by the conversion and radiation systems in the relationship with frequency. For example, the sound pressure of a direct-radiation type of speaker increases in proportion to frequency if the velocity V is constant. Consequently, if V/F decreases with frequency, the ratio is not related to frequency as a whole even when F/E is constant. This corresponds to a mass when the vibrat-

Table 5.1.2 Three Control Systems

	Resistance control	Mass control	Stiffness control
Z_m approximation $\nu\nu/F $	r $1/r$	ωm $1/\omega m$	s/ω ω/s
Characteristics			
Applications	Horn speaker	Direct radiant-type speaker	Headphone

ing system is regarded as a single resonance system, which is called *mass control*. Likewise, when V/F becomes unrelated to frequency, both the resistance control and the frequency increase; this is called *stiffness control*. Table 5.1.2 summarizes these characteristics.

5.1.4 Bibliography

- Beranek, L. L.: *Acoustics*, McGraw-Hill, New York, N.Y., pg. 183, 1954.
- Hayasaka, T., et al.: *Onkyo-Kogaku Gairon (An Introduction to Sound and Vibration)*, Nikkan Kogyo Shinbunshya, pg. 67, 1973 (in Japanese).
- Hayasaka, T., et al.: *Onkyo-Shindo Ron (Sound and Vibration)*, Maruzen Kabushikigaisha, pg. 201, 1974 (in Japanese).
- Kinsler, L. E., et al.: *Fundamentals of Acoustics*, Wiley, New York, N.Y., pg. 348, 1982.
- Olson, H. F.: *Elements of Acoustical Engineering*, Van Nostrand, Princeton, N.J., pg. 124, 1957.

Chapter
5.2

Direct-Radiator Loudspeakers

Katsuaki Satoh

5.2.1 Introduction

The diameter of a speaker diaphragm normally ranges between a few centimeters and dozens of centimeters when high-amplitude sound must be produced. The following sections outline the basic principles involved in direct-radiator loudspeakers.

5.2.2 Piston Source in an Infinite-Plane Baffle

An actual diaphragm has many different oscillation modes, and its motion is complicated. On the assumption—for easier analysis—that the diaphragm is rigid, radiation impedance and directivity are considered for typical circular and rectangular shapes. As shown in Figure 5.2.1, part of a circular rigid wall is oscillating at a given velocity $v \exp(j\omega t)$. The upper part of this circular piston is subdivided into the micro area d_s , and when a micro part is oscillated by the piston, the total reaction force subjected from the medium side is calculated. Thus, the radiation impedance Z_R of the diaphragm is found from the ratio of this reaction force to the diaphragm's oscillating speed. This shows how effectively sound energy from the diaphragm is used. Radiation impedance in the circular diaphragm is shown in the following equation, and the results in Figure 5.2.2.

$$Z_R = (\pi a^2 p C) \left[1 - \left(\frac{J_1(2ka)}{ka} \right) + j \frac{S_1(2ka)}{ka} \right] \quad (5.2.1)$$

Where:

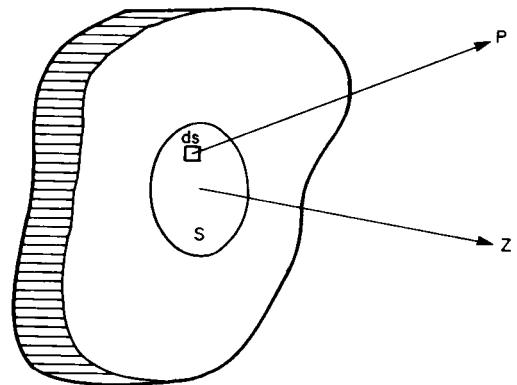
J_1 = Bessel function of the first order

S_1 = Struve function

Directional characteristics of the circular diaphragm are shown in the following equation, and the results in Figure 5.2.3.

$$D(\theta) = \left| \frac{2J_1(ka \sin \theta)}{ka \sin \theta} \right| \quad (5.2.2)$$

5-18 Sound Reproduction Devices and Systems

**Figure 5.2.1** Piston on an infinite rigid wall.

Where:

$D(\theta)$ = ratio between sound pressures whose angles θ are in 0 and θ directions

θ = perpendicular on the surface center

k = number of waves

a = radius, m

Rectangular impedance is shown in Equation (5.2.3), directional characteristics in Equation (5.2.4), and the respective calculation results in Figure 5.2.4.

$$R(v, \sigma) = 1 - (2/\pi v^2)[1 + \cos(vq) + vq \sin(vq) - \cos(vp) - \cos(v/p)] \\ + (2/\pi)[pI_1(v, \sigma)] + I_1(v, 1/\sigma)/p$$

$$X(v, \sigma) = (2/\pi v^2)[\sin(vq) - vq \cos(vq) + v(p + 1/p) - \sin(vp) - \sin(v/p)] \\ - (2/\pi)[pI_2(v, \sigma)] + I_2(v, 1/v)/p$$

$$v = k\sqrt{S}$$

$$q = (\sigma + 1/\sigma) \quad (5.2.3)$$

Where:

v = nondimensional frequency

$$p = \sqrt{\sigma}$$

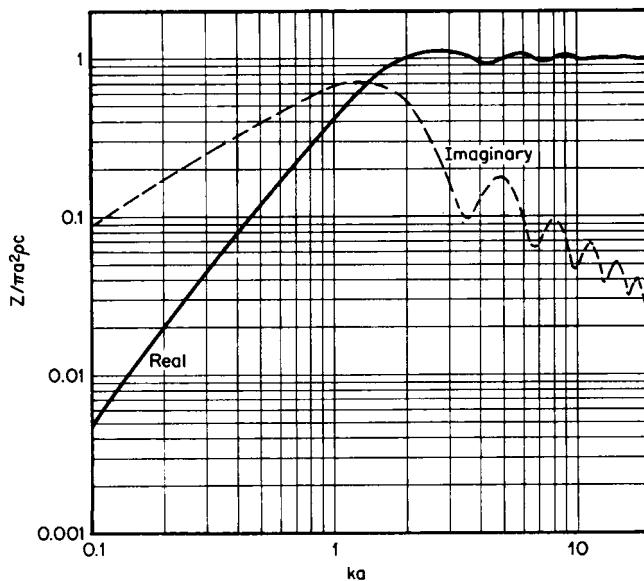


Figure 5.2.2 Radiation impedance for a rigid circular diaphragm in an infinite baffle as a function of $ka = 2\pi a/\lambda$.

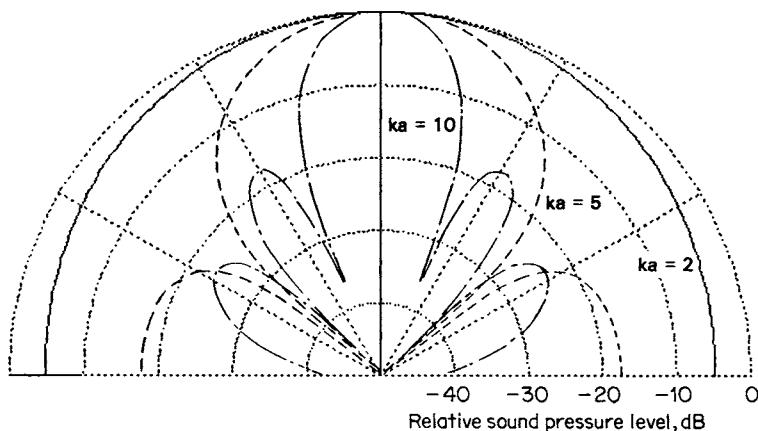


Figure 5.2.3 Directional characteristics of a circular diaphragm.

5-20 Sound Reproduction Devices and Systems

$$I_{1,2} = \int_{(\zeta - 1/\sqrt{2})}^{(\zeta + 1/\sqrt{2})} (1 - 1/\xi t^2)^{1/\sqrt{2}} \cos(vt) \sin(\phi_1) dt$$

$$\xi = \infty r - \sigma$$

1,2, subscripts of I, = cos for 1 and sin for 2

$$D(\theta_1, \theta_2) = \frac{\sin \phi_1}{\phi_1} \cdot \frac{\sin \phi_2}{\phi_2}$$

$$\phi_{1,2} = \frac{\pi d_{1,2}}{\lambda} \sin \phi_{1,2}$$
(5.2.4)

Where:

$D(\theta_1, \theta_2)$ = ratio between sound pressures in 0 and θ_1/θ_2 directions ($\theta_1 = \theta_2 = 0$ is a perpendicular of the center on the rectangular surface)

λ = wavelength, m

$d_{1,2}$ = length of each side of rectangle, m

Radiation impedance shows how effectively sound energy is radiated, while directional gain is used to show how expanding sound energy is radiated in space. The ratio of total acoustic energy W is found by integrating the sound strength from that on a spherical surface a distance r from the sound source with the sound strength that exists on the same point from the nondirectional sound source that emits the same energy. This is expressed in decibels:

$$V = \frac{r^2}{pC} \int_0^{2\pi} \int_0^\pi |\dot{P}(r, \theta, \phi)|^2 \sin^2 \theta \, (d\theta)(d\phi)$$
(5.2.5)

$$DI = 10 \log \left(\frac{4\pi r^2}{W} \cdot \frac{|\dot{P}_{max}|^2}{pC} \right)$$
(5.2.6)

Where:

W = total acoustic energy, W

r = distance in the maximum sound pressure direction for standardization, m

P_{max} = sound pressure at distance r , N/m²

DI = directivity index (directional gain), dB

5.2.2a Baffle Shape and Acoustic Characteristics

In the preceding section an infinite baffle was discussed, but such a baffle cannot be put to practical use. Consequently, it is necessary to precheck the types of characteristics that can be obtained when a definite baffle is installed in a speaker. Because the sounds radiated to the front baffle and reflected to the rear are opposite in phase, the difference in distance between the

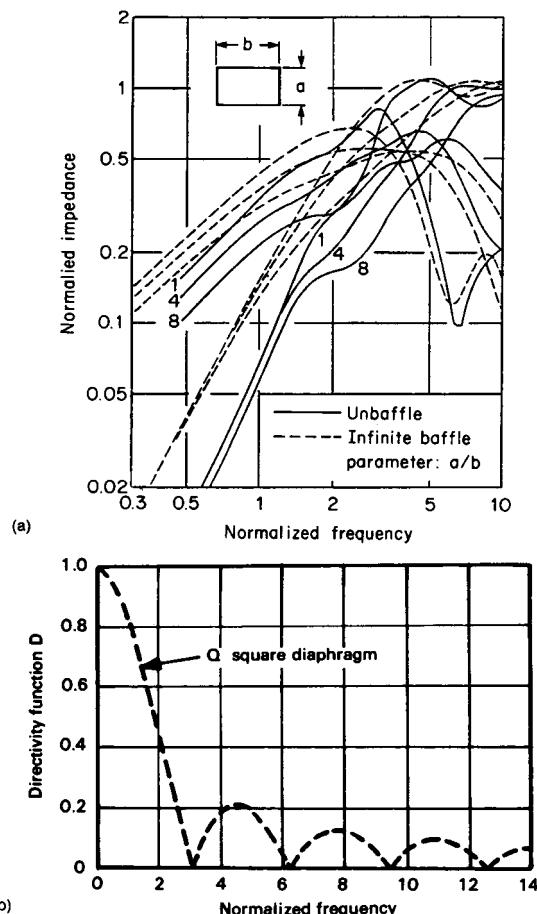


Figure 5.2.4 Diaphragm characteristics: (a) radiation impedance for a rigid rectangular diaphragm, (b) directivity function for a rigid square diaphragm. Note that in (a) solid lines, which have been calculated by using the finite element method (FEM), are instructive for practical designs.

passes of sound through the rear and front baffles from a speaker is canceled by the front and rear sounds of one-half even multiples and added to each other by the sounds of odd multiples.

Therefore, high and low sound pressures occur. To avoid this, the speaker is installed off center, resulting in a baffle with a complicated shape. One side should be a few times longer than the wavelength. However, this shape does not produce favorable characteristics, and this type of baffle is not often used in practical applications. Typical baffle characteristics are shown in Figures 5.2.5 and 5.2.6.

5-22 Sound Reproduction Devices and Systems

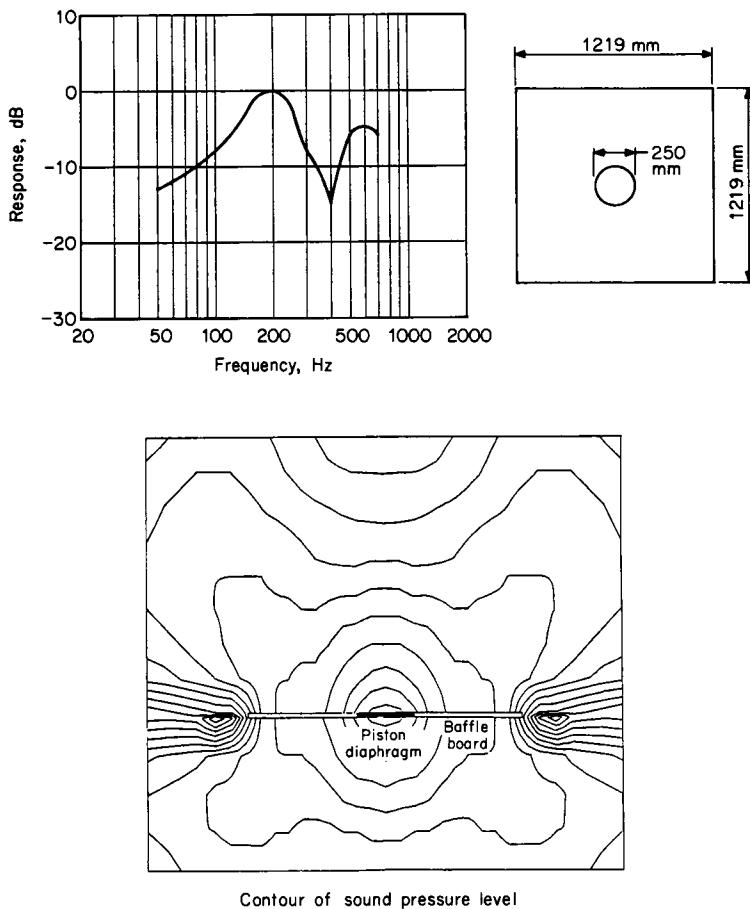


Figure 5.2.5 Pressure-response-frequency characteristics for a direct radiator installed in the center of a finite baffle, estimated by FEM.

5.2.2b Acoustic Characteristics of Rigid Disk with Constant-Force Drive

This section comments on the types of sound-pressure-frequency characteristics produced at a remote distance on the center axis of a diaphragm and the acoustic output obtained therefrom when a circular piston diaphragm is placed in an infinite rigid wall and driven at a given force. When a circular diaphragm with radius α is subjected to a constant force F moving in the axial direction, sound pressure P is determined by

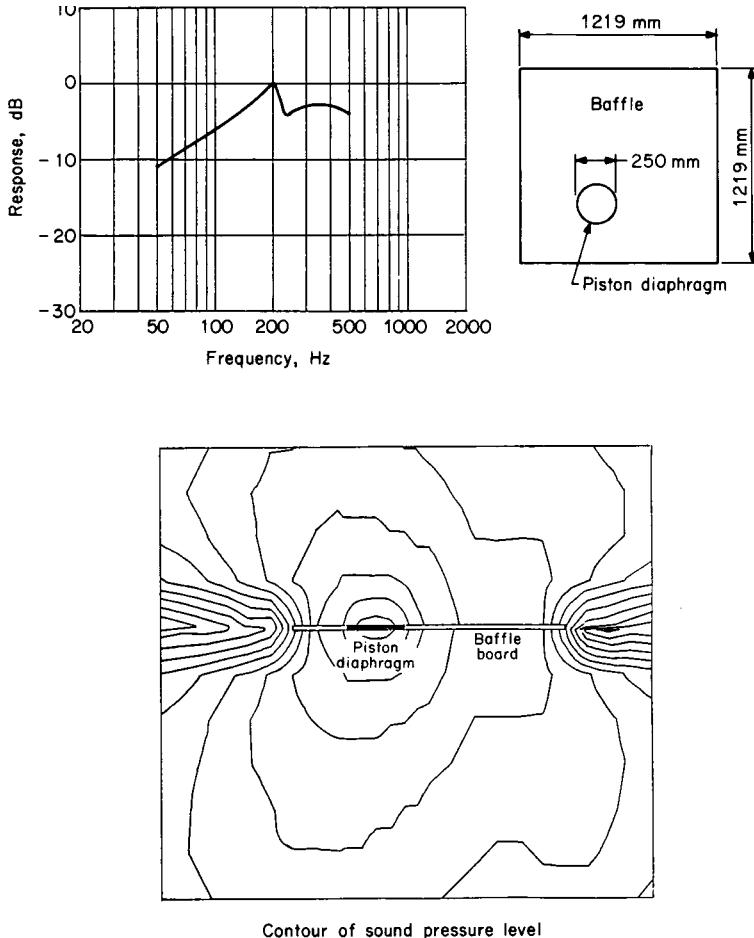


Figure 5.2.6 Pressure-response-frequency characteristics of a direct radiator installed off center, estimated by FEM.

$$\begin{aligned}
 \dot{P} &= p_\theta \frac{\partial \phi}{\partial t} = j \frac{\omega p_\theta a^2}{2r} \exp(-jkr) \cdot \dot{v} \\
 &= j \frac{\omega p_\theta a^2}{2r} \exp(-jkr) \cdot \frac{\dot{F}}{Z}
 \end{aligned} \tag{5.2.7}$$

The absolute value $|\dot{P}|$ of sound pressure is shown in the equation

5-24 Sound Reproduction Devices and Systems

$$|\dot{P}| = \frac{\omega p_0 a^2}{2r} \left| \frac{\dot{F}}{\dot{Z}} \right| \quad (5.2.8)$$

Where:

P = sound pressure on the axis, N/m²

p_0 = gas density, kg/m³

θ = velocity potential

a = diaphragm radius, m

r = distance from diaphragm on the axis, m

\dot{P} = driving force, N

ω = angular frequency, rad/s

When the oscillation system is regarded as a single resonance system, \dot{Z} is obtained as follows

$$\dot{Z} = r_m + j\omega n + \frac{1}{j\omega C_m} \quad (5.2.9)$$

Where:

\dot{Z} = mechanical impedance of oscillation system, mechanical ohms

r_m = mechanical resistance of oscillation system, N/m

C_m = oscillation-system compliance, m/N

m = mass of oscillation system, kg

Therefore, sound pressure $|\dot{P}|$ is determined by

$$|\dot{P}| = \frac{\omega^2 p_0 a^2 C_m}{2r} |\dot{F}| \quad \omega < \omega_0 \quad (5.2.10)$$

$$|\dot{P}| = \frac{p_0 a^2}{2rm} |\dot{F}| \quad \omega > \omega_0 \quad (5.2.11)$$

This is shown in Figure 5.2.7.

$$W_a = \frac{\pi p_0 a^4 \omega^4 C_m}{2c_2} |\dot{F}|^2 \quad ka < 1 \quad \omega < \omega_0 \quad (5.2.12)$$

$$W_a = \frac{\pi p_0 a^4 \omega^2}{2cr^2} |\dot{F}|^2 \quad ka < 1 \quad \omega = \omega_0 \quad (5.2.13)$$

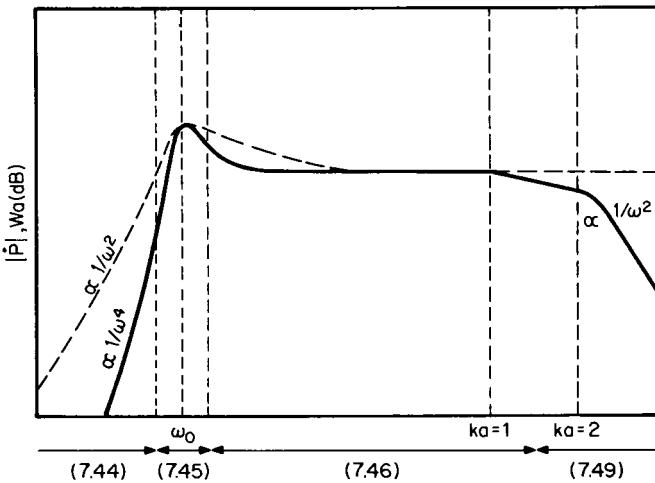


Figure 5.2.7 Acoustic power and pressure-response-frequency characteristics of a piston source in an infinite-plane baffle.

$$W_a = \frac{\pi p a^4}{2cm} |\dot{F}|^2 \quad ka < 1 \quad \omega > \omega_0 \quad (5.2.14)$$

$$W_a = \pi p a^2 c \omega^2 C_m^2 |\dot{F}|^2 \quad ka > 1 \quad \omega > \omega_0 \quad (5.2.15)$$

$$W_a = \frac{\pi p a^2 c}{r^2} |\dot{F}|^2 \quad ka > 1 \quad \omega = \omega_0 \quad (5.2.16)$$

$$W_a = \frac{\pi p a^2 c}{\omega_m^2} |\dot{F}|^2 \quad ka > 1 \quad \omega > \omega_0 \quad (5.2.17)$$

Where:

C = sound velocity, m/s

k = number of waves

ω = resonance angular frequency, rad/s

5.2.3 Bibliography

Beranek, L. L.: *Acoustics*, McGraw-Hill, New York, N.Y., pg. 185, 1954.

5-26 Sound Reproduction Devices and Systems

- Hayasaka, T., and S. Yoshikawa: *Onkyo-Kogaku Gairon* (An Introduction to Sound and Vibration), 1983.
- Morse, P. M.: *Vibration and Sound*, McGraw-Hill, New York, N.Y., pg. 326, 1948.
- Morse, P. M., and K. U. Ingard: Theoretical Acoustics, McGraw-Hill, New York, N.Y., pg. 366, 1968.
- Olson, H. F.: *Elements of Acoustical Engineering*, Van Nostrand, Princeton, N.J., pg. 38, 1957.
- Rayleigh, J. W. S.: *The Theory of Sound*, Dover, New York, N.Y., pg 162, 1945.
- Sakamoto, N.: *Loudspeaker and Loudspeaker Systems*, Nikkan Kogyo Shinbunshya, pg. 18, 1967.

Chapter

5.3

Dynamic-Type Direct-Radiation Speakers

Katsuaki Satoh

5.3.1 Introduction

The dynamic direct-radiation loudspeaker is divided broadly into the following components:

- Magnetic circuit
- Drive coil
- Diaphragm
- Support system
- Frame

Although the construction of each speaker design is unique, the underlying fundamentals are the same.

5.3.2 Operational Details

The typical configuration of a dynamic-type direct-radiation speaker is shown in Figure 5.3.1. Most magnetic circuits are of the external type, using a ferrite magnet designed to generate a magnetic-flux density of a few thousand to a few ten thousand G in an approximately 1- to 2-mm air gap formed by the north and south poles. To control distortion, the drive coil provided in the air gap is designed so that it does not move out of the uniform magnetic field formed by the magnetic pole because of vibration. Thus, the drive coil used has approximately 0.1-mm-diameter windings of several turns. The impedance normally is a multiple of 4Ω . The diaphragm is available in a variety of shapes and materials, as described later. The dust cap is used to prevent dust from intruding into the magnetic air gap; when the cap must function as a damper, a permeable material is used. Thus, the centering suspension and cone suspension function to: 1) support these vibration systems, 2) hold the drive coil in the magnetic air gap, and 3) generate deemphasis in the axial direction.

5-28 Sound Reproduction Devices and Systems

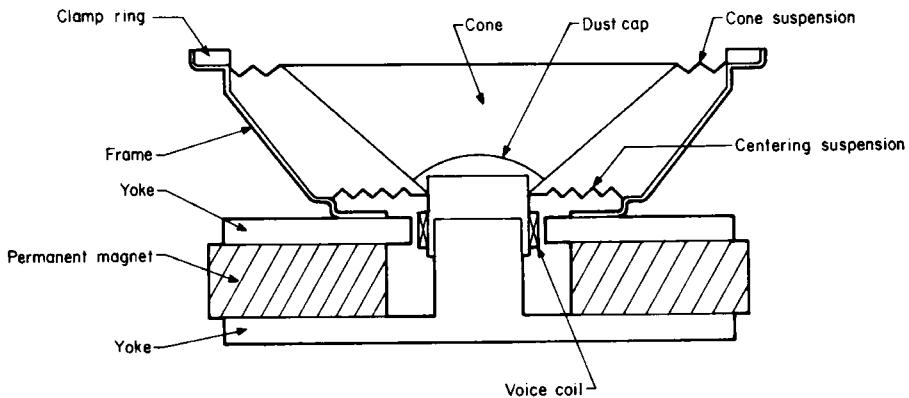


Figure 5.3.1 Structure of the dynamic direct-radiator loudspeaker.

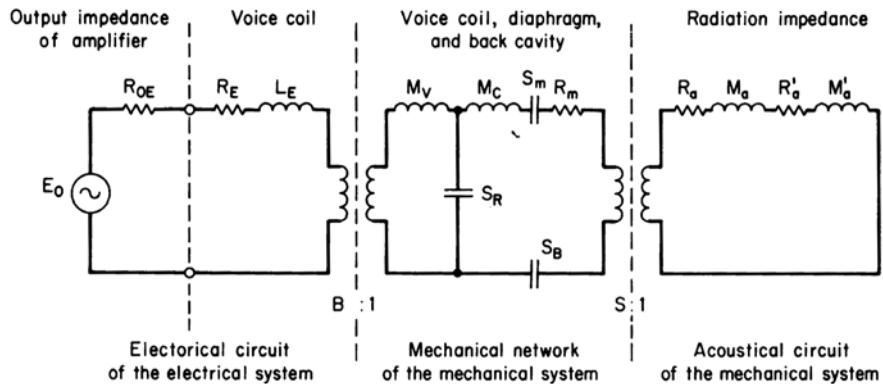


Figure 5.3.2 Electromechanical equivalent circuit. R_{OE} = output impedance of amplifier, Ω ; R_E = resistance of voice coil, Ω ; L_E = inductance of voice coil, H; M_V = mass of voice coil, kg; S_R = stiffness between cone and voice coil, N/m; M_C = mass of cone, kg; S_B = stiffness of back cavity, N/m; R_a , R'_a = radiation resistance of diaphragm, mechanical ohms; M_a , M'_a = radiation mass of diaphragm, kg; B : 1 = force factor; S = area of diaphragm, m^2 .

5.3.2a Equivalent Circuit and Frequency Response

Figure 5.3.2 shows the equivalent circuit of a dynamic type of speaker. The sound-pressure-frequency characteristics of the equivalent circuit are shown in Figure 5.3.3. An examination of these characteristics divided by frequency bands follows.

In low ranges, the diaphragm and support system are free from split vibration, but they are considered to be a single resonance system. Thus, an equivalent circuit as shown in Figure 5.3.4a is produced. The velocity, amplitude characteristics, and sound pressure characteristics on the axis are as shown in Figure 5.3.4b. As can be seen from this figure, Q_0 determines sound pressure characteristics near the resonance frequency. If all element constants are found,

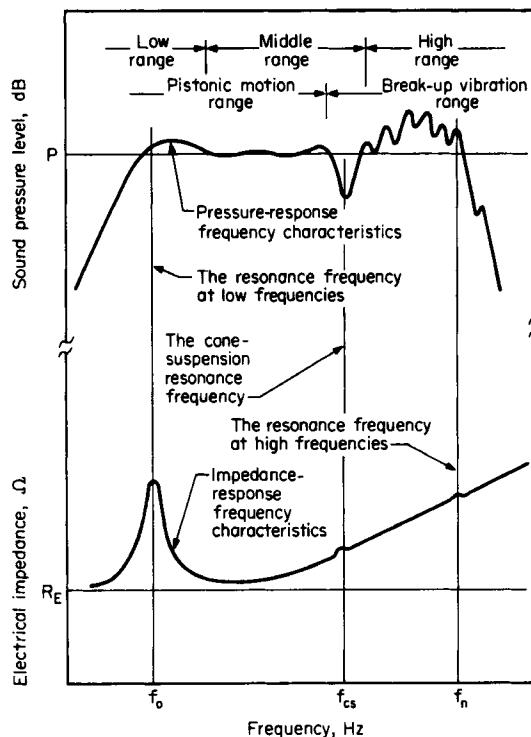


Figure 5.3.3 Frequency characteristics of the dynamic direct-radiator loudspeaker.

Q_0 can be obtained by calculation, but these constants must often actually be found by measurement. Voice-coil impedance near the resonance frequency is expressed as a sum of electrical impedance and motion impedance. That is,

$$Z_e = Z_c + \frac{A^2}{Z_M} \quad (5.3.1)$$

When $R_c \gg \omega L$,

$$Z_e = R_c + \frac{A^2}{Z_M} = R_c + \frac{1}{\frac{1}{r_m} + \frac{1}{j\omega n} + \frac{1}{j\omega C_m}} \quad (5.3.2)$$

5-30 Sound Reproduction Devices and Systems

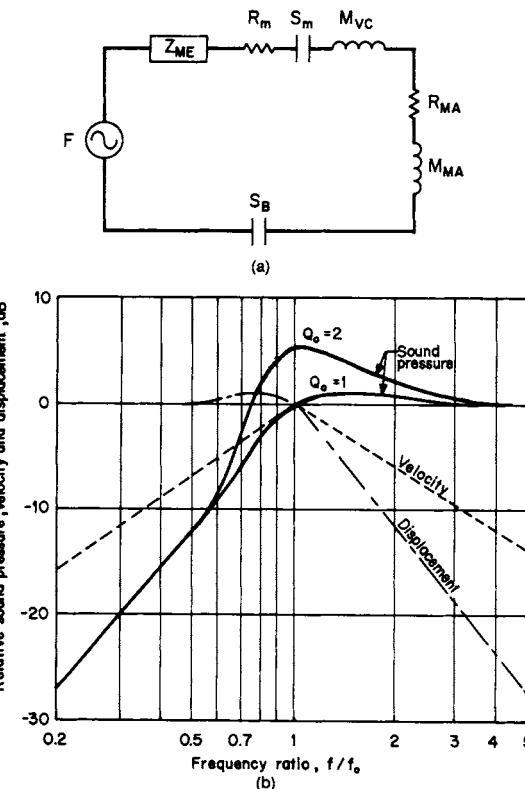


Figure 5.3.4 Loudspeaker characteristics: (a) mechanical equivalent circuit at a low-frequency range, (b) frequency characteristics of sound pressure, velocity, and displacement. Z_{ME} = motional impedance, mechanical ohms; R_m = resistance of vibrating system, mechanical ohms; S_m = stiffness of vibrating system, N/m; M_{VC} = mass of vibrating system, kg; R_{MA} = resistance of radiating system, mechanical ohms; M_{MA} = mass of radiating system, kg; S_B = stiffness of back cavity, N/m.

The vector impedance locus is shown in Figure 5.3.5. From these results, Figure 5.3.6 is obtained, and Q_0 can be found directly from electrical impedance. In midrange, cone suspension less rigid than the diaphragm produces a split vibration. This phenomenon appears typically near 1000 Hz with a speaker using a paper-cone diaphragm. The analytical results of this condition, using the finite-element method (FEM), are shown in Figure 5.3.7a. To eliminate this, damping material is coated and the shape is redesigned, thus controlling the resonance. For the diaphragm, specific resonance starts to appear, a peak and a dip in sound pressure response occur, and a strain may result. Regarding this shortcoming, the results of analysis by FEM are shown in Figure 5.3.7b. To control this specific resonance, materials with a larger internal loss are used, the shape of the diaphragm is changed from a simple cone to a paracurve, and corrugation is provided. Furthermore, when the frequency rises, elastic deformation concentrates at the junction between the drive coil and the diaphragm, and stiffness S_R appears there equivalently. Therefore,

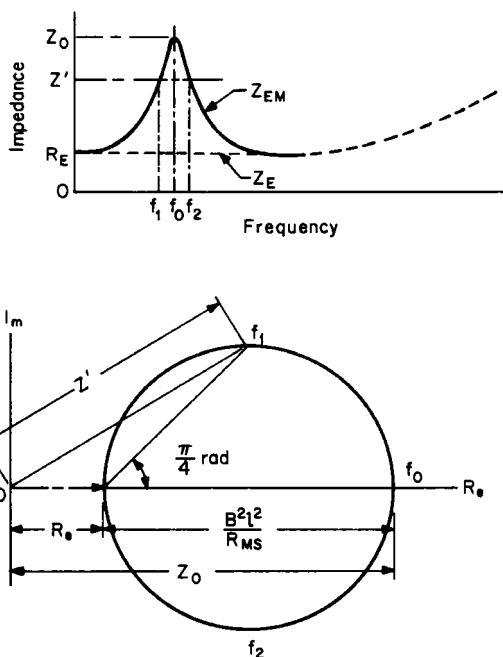


Figure 5.3.5 Loudspeaker voice-coil impedance and impedance locus. R_E = resistance of voice coil, Ω ; $Z = Z_0/\sqrt{2}$, Ω ; B = magnetic-flux density in the gap, Wb/m²; l = length of wire on voice-coil winding, m; R_{MS} = resistance of vibrating system, mechanical ohms; f_0 = resonance at low-frequency range, Hz; f = frequency at -3 dB, Hz.

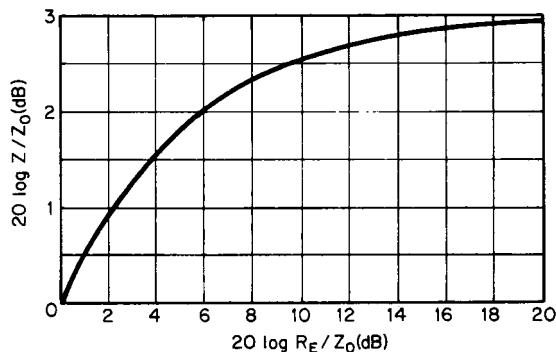


Figure 5.3.6 Relation between Z' and Z .

sound pressure is suddenly lowered at a higher level than the resonance frequency by S_R and the diaphragm mass, which actually presents the playback limit.

5-32 Sound Reproduction Devices and Systems

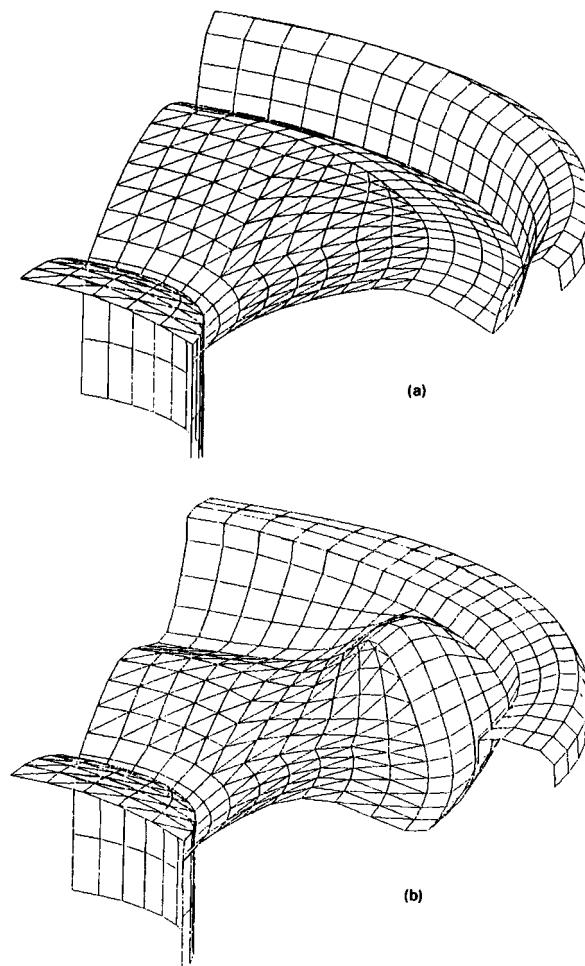


Figure 5.3.7 Breakup vibrating modes, estimated by FEM: (a) fundamental mode of the suspension, (b) axial mode of the cone.

5.3.2b Efficiency

Speaker efficiency is expressed in terms of the ratio of electrical input to acoustic output. The electrical input with due regard to only the real-number part in the equivalent circuit in Figure 5.3.2 is expressed by

$$W_e = R_c I^2 \quad (5.3.3)$$

Where:

W_e = electrical input, W

R_c = coil resistance, Ω

I = current flowing into the coil, A

Acoustic output W_a is determined by

$$W_a = r_R \left| \frac{F}{Z_m} \right| \quad (5.3.4)$$

Where:

W_a = acoustic output, W

r_R = acoustic radiation resistance, N/m

F = driving force, N

Z_m = vibration-system mechanical impedance, mechanical ohms

Consequently, efficiency η is found as follows

$$\begin{aligned} \eta &= \frac{W_a}{W_e + W_a} \\ &= \frac{1}{1 + W_e/W_a} \end{aligned} \quad (5.3.5)$$

With the diaphragm considered as a stiff disk, if it is an infinite baffle board, W_a can employ the approach shown previously in this section. If the acoustic output is constant, η is determined by

$$\eta = \frac{1}{1 + \frac{2cm^2R_c}{\pi pa^4B^2l^2}} \times 100 \quad (5.3.6)$$

Where:

η = conversion efficiency, percent

p = air density, kg/m³

c = sound velocity, m/s

m = vibration-system mass, kg

a = effective radius of diaphragm, m

B = flux density, We/m²

l = coil length, m

R_c = coil resistance, Ω

In Equation (5.3.6), the magnitude on the second term normally is approximately 50. The efficiency is only a few percentage points, which is very low.

5-34 Sound Reproduction Devices and Systems

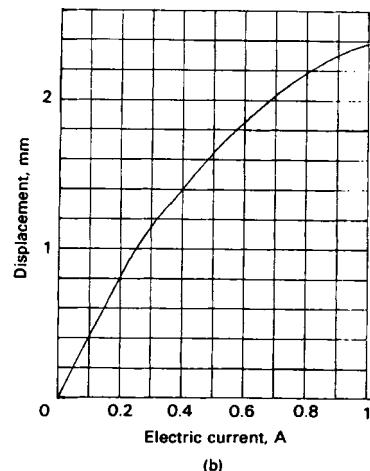
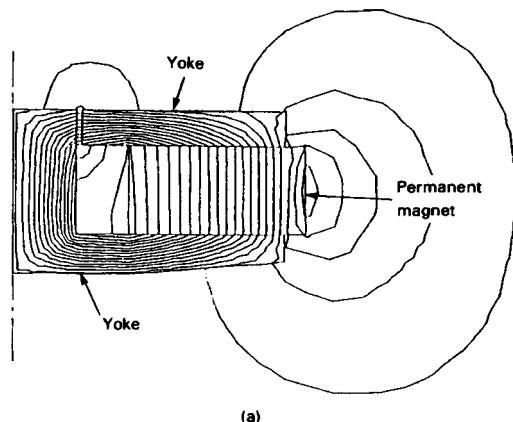


Figure 5.3.8 Flux distribution of the half-magnetic circuit: (a) typical flux lines in an air gap estimated by FEM, (b) relation between the electric current and the displacement.

5.3.2c Nonlinear Distortion

The strain that takes place in a dynamic speaker includes several types of distortion: 1) driving-force distortion, 2) support system distortion, 3) air distortion, and 4) frequency-modulation distortion.

Driving-Force Distortion

Driving-force distortion occurs mainly because a drive coil flows out from the uniform magnetic field as the amplitude varies, whereby the driving force ceases to be proportional to current. Figure 5.3.8a shows the magnetic-flux distribution and magnetic-flux-density distribution near the magnetic pole in a typical magnetic circuit. Figure 5.3.8b shows the relationship between the power coefficient generated by a coil located in such a magnetic circuit and the coil displacement. Consequently, the following nonlinear differential equation must be solved. Here, a study

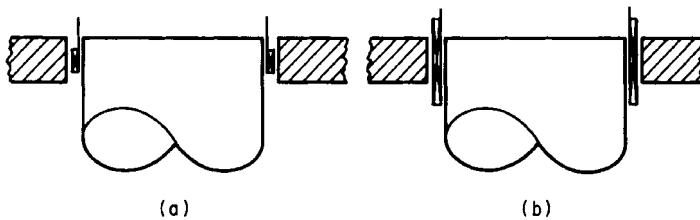


Figure 5.3.9 Relation between the voice coil and the magnetic circuit for reducing distortion: (a) voice coil shorter than the air gap, (b) voice coil longer than the air gap.

may be made at an ultralow frequency with a large amplitude. Therefore, radiation impedance can be approximated with radiation mass and vibration-system impedance with stiffness. The basic equations are as follows, assuming the stiffness to be linear:

$$M_{MA} \frac{d^2 \xi}{dt^2} = m \frac{d^2 \xi}{dt^2} + r \frac{d\xi}{dt} + s_n - A(\xi) I(\xi) \quad (5.3.7)$$

$$E_0 \sin \omega t = RI(\xi) + A(\xi) \frac{d\xi}{dt} \quad (5.3.8)$$

Where:

$A(\zeta) = \zeta$ function force factor, N/A

M_{MA} = radiation mass, kg

s_n = vibration-system stiffness, N/m

$E_0 \sin \omega t$ = applied voltage, V

R = coil resistance, Ω

$I(\zeta)$ = current of ζ function, A

ζ = displacement, m

To reduce this distortion, it is preferable to adopt a method of decreasing the coil-winding width as shown in Figure 5.3.9a so that it is not off the magnetic field or sufficiently increasing the coil-winding width as shown in Figure 5.3.9b. Because driving-force distortion develops as current distortion, this distortion can be reduced by detecting current flowing into a coil with a microresistance and feeding this current back into the input terminal of the amplifier. This is shown in Figure 5.3.10. Other driving-force distortions include a strain generated by hysteresis of the magnetic-circuit yoke. This can be substantially improved by using a silicon steel plate for a yoke and a magnetic material of small conductivity. This technique is shown in Figure 5.3.11.

Support-System Distortion

Support-system distortion is such that since elasticity of suspension is nonlinear, force and displacement cease to be proportional to each other. The force-versus-displacement characteristics

5-36 Sound Reproduction Devices and Systems

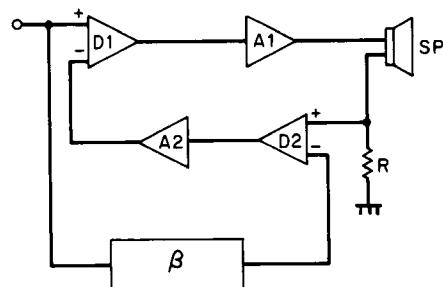


Figure 5.3.10 System for reducing current distortion. D_1, D_2 = differential amplifiers; A_1, A_2 = amplifiers; β = feedback circuit; SP = loudspeaker; R = resistor for detecting current distortion.

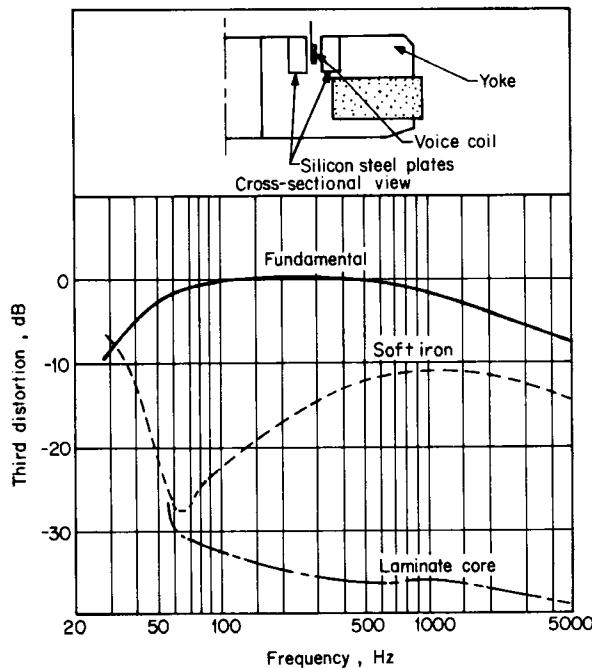


Figure 5.3.11 Comparison of the third-harmonic distortion between soft iron and silicon plates (solid line = fundamental current level, dashed line = soft-iron-type yoke, dash-dot line = laminate-core-type yoke).

of a general support system are shown in Figure 5.3.12. The function showing such a curve is expressed by the following equation.

$$F(\xi) = \alpha\xi + \beta\xi^2 + \gamma\xi^3 \quad (5.3.9)$$

Where:

$F(\zeta)$ = force at displacement V , N

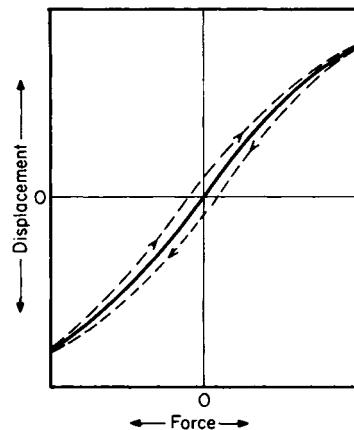


Figure 5.3.12 Relation between force and displacement in a typical suspension.

ζ = displacement, m

α, β, γ = constants

Consequently, stiffness ζ is found by

$$s(\xi) = \alpha + \beta\xi + \gamma\xi^2 \quad (5.3.10)$$

Assuming ω to be an ultralow frequency with this stiffness function substituted for the basic equation, Equations (5.3.7) and (5.3.8) are as follows.

$$M_{MA} \frac{d^2\xi}{dt^2} = m \frac{d^2\xi}{dt^2} + r \frac{d\xi}{dt} + s_n(\xi)\xi - A(\xi)I(\xi) \quad (5.3.11)$$

$$E_0 \sin \omega t = RI(\xi) + A(\xi) \frac{d\xi}{dt} \quad (5.3.12)$$

There are several methods of solving this equation. The calculation results on the assumption that the current is constant, using the indefinite-coefficient method and sample measurements, are shown in Figure 5.3.13. The point to be considered in the support system in particular is that, for a large amplitude, suspension elasticity is suddenly lost, forming a cropped wave and leading to rupture. Because the support system is nonlinear, not only does distortion occur, but a so-called jumping phenomenon is found. As shown in Figure 5.3.14, amplitude suddenly changes discontinuously for frequency and current. To prevent this, as large a suspension as possible is used, and such materials and construction are selected that the center-holding capacity is not lowered. Cone suspension uses a soft material wherever applicable, and corrugation and ribbing are provided to avoid edge resonance.

5-38 Sound Reproduction Devices and Systems

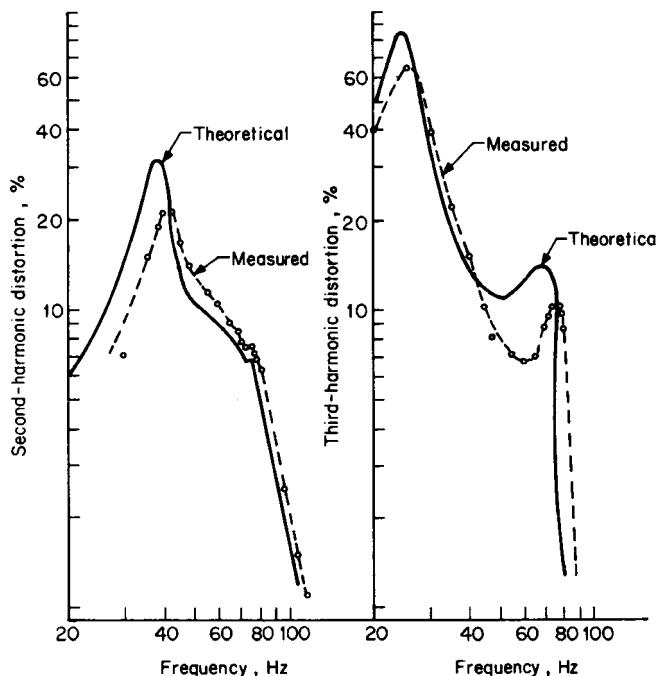


Figure 5.3.13 Distortion characteristics of a driving force, calculated from Equations (5.3.11) and (5.3.12).

Air Distortion

Generally, on the assumption that changes in volume are very small when the sound-surge equation is solved, the secondary or more terms are ignored. However, the smallest distortion cannot be ignored, and the high-order term cannot be ignored when sound pressure is large. Equation (5.3.12) shows the degree of second-harmonic wave due to nonlinearity on this high-order term:

$$p_2 = \frac{(\gamma+1)\omega^2}{2\sqrt{2\gamma}p_0c} p_r^2 r \quad (5.3.12)$$

Where:

p_2 = second-harmonic distorted sound pressure generated by plane waves at distance r , N/m²

p_r = fundamental wave sound pressure of plane wave at distance r , N/m²

p_0 = atmospheric pressure, N/m

γ = ratio between constant-pressure specific heat and constant-volume specific heat (air, 1:4)

ω = angular frequency, rad

c = sound velocity, m/s

r = distance, m

The calculation results of Equation (5.3.12) are shown in Figure 5.3.15.

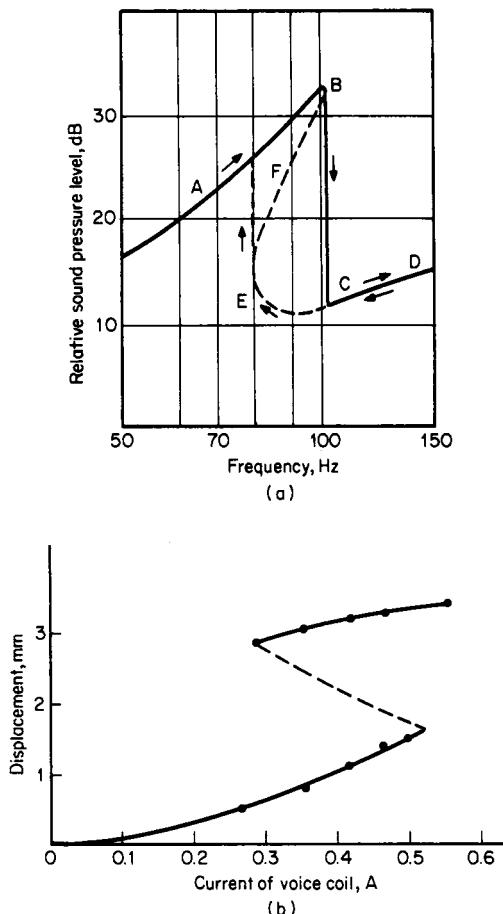


Figure 5.3.14 Nonlinear suspension system: (a) the unstable portion of the response frequency characteristic, indicated by the dashed line; (b) the unstable portion of the response current characteristic, indicated by the dashed line.

Frequency-Modulation Distortion

Signals input to a speaker have various frequency spectra. When low- and high-frequency sounds are radiated from a diaphragm at the same time, high-frequency sound is subjected to modulation because the diaphragm is moving forward and backward significantly according to low-frequency signals. The frequency-modulated wave generated thereby is expressed by a carrier and an unlimited number of sideband waves. The mean-square value of the ratio of sideband-wave energy to all energies of the sound wave is expressed in percentage as follows:

5-40 Sound Reproduction Devices and Systems

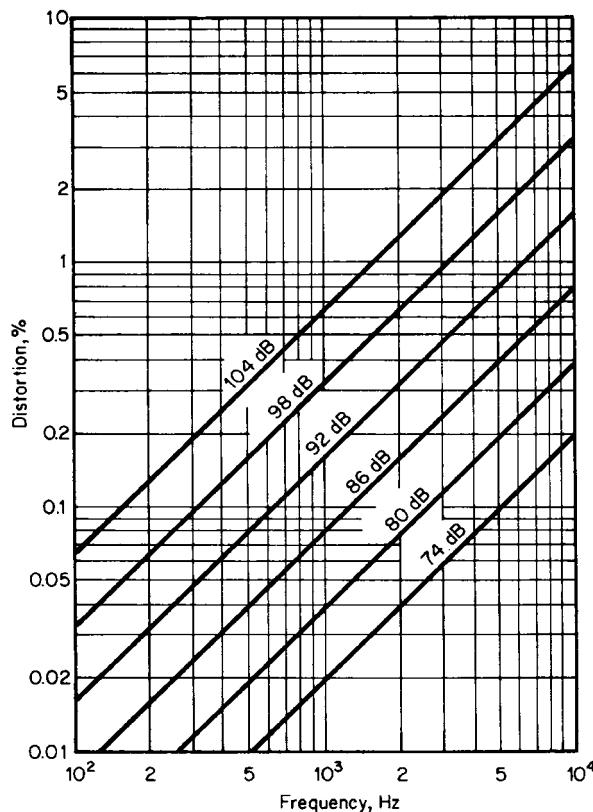


Figure 5.3.15 The distortion generated in the air gap between the cone and the listening-point distance in a direct-radiator speaker with a cone diameter of 20 cm, measured at a distance of 3 m.

$$D = 2900 \frac{f_2 \sqrt{p_1}}{f_1^2 d^2} \quad (5.3.13)$$

Where:

D = distortion, percent

f_2 = modulated wave (high-frequency), Hz

f_1 = modulated wave (low-frequency), Hz

p_1 = acoustic output of f_1 , W

d = cone diameter, m

One of the methods for reducing this distortion is to use a multiway speaker system.

5.3.2d Diaphragm and Support System

It is no exaggeration to say that the diaphragm and support system nearly determine speaker acoustic characteristics. A typical shape and features for the diaphragm and support system are described below.

Diaphragm

Diaphragms are classified by shape into cone, plane, and dome diaphragms. The cone diaphragm is one of the most frequently used types. Figure 5.3.16 shows some typical shapes. Any of these types is directed to widening the piston-motion area to enhance a high-range playback limit frequency and also to reduce distortion. For this purpose, it is important to know the vibrating conditions of the cone diaphragm, but it is very difficult to find them analytically.

In the *dome diaphragm*, a thin metallic foil, resin-impregnated cloth, or paper is formed into a sphere, and the periphery of the diagram is driven. A diaphragm with a smaller-aperture diameter is easy to realize because of circumferential drive, split vibration can be controlled up to a high frequency, and favorable directional characteristics are also obtained. Materials used in this diaphragm include the following:

- Sulfite cellulose
- Sulfate pulp
- Paper mixed with highly elastic fiber such as silicon carbide whiskers, carbon fiber, and alamido fiber
- Metal foil such as aluminum, titanium, and beryllium
- High-polymer film such as polyethylene terephthalate or highly elastic materials reinforced by deposition such as carbon, boron, and beryllium
- Composite materials using honeycomb and foamed urethane as a core

Support System

The support system is divided broadly into a cone suspension system and a center holder. The cone suspension system is required to absorb reflection from the frame as a termination of the diaphragm to control edge resonance and also to prevent an acoustic short circuit which would occur before and after the diaphragm along with a baffle board. This system must be constructed so that it is easy to move in the vibrating-axis direction of the diaphragm and difficult to move in the lateral direction along with the center holder. The principal construction features of the cone suspension system are shown in Figure 5.3.17. Materials having proper mechanical resistance are preferable.

Requirements for centering suspension include the following:

- Provide proper stiffness in order to maintain a restoration force
- Hold a voice coil in the center of the gap formed by the magnetic circuit in order to smooth movement in the axial direction
- Maintain favorable linearity of driving-force-to-displacement characteristics even when the diaphragm is given large amplitude

5-42 Sound Reproduction Devices and Systems

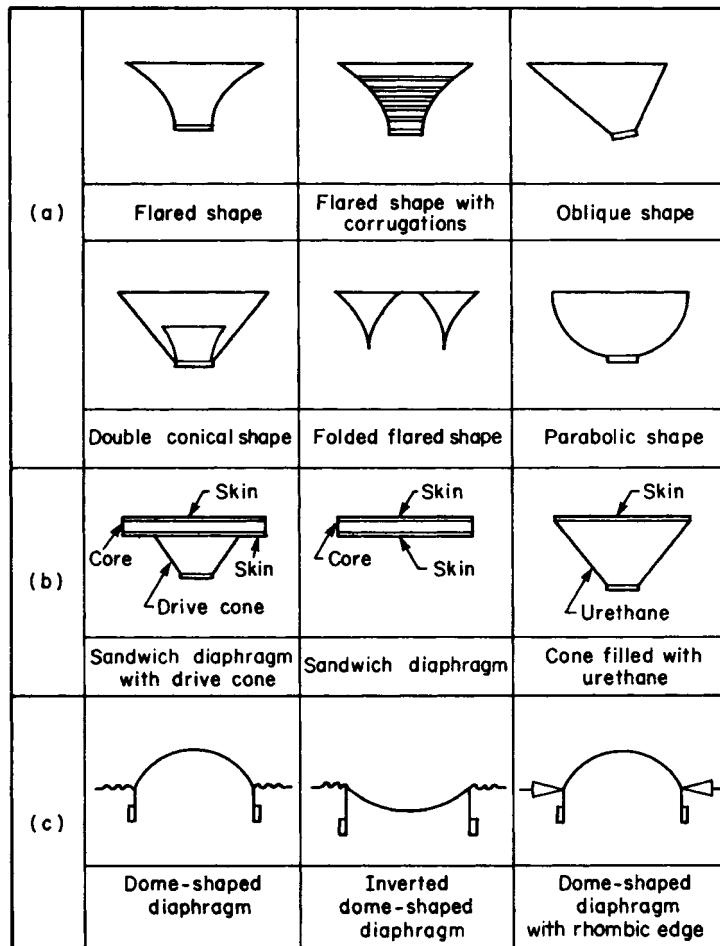


Figure 5.3.16 Sectional views of various diaphragm shapes: (a) diaphragm of the cone type extends to the reproducing band by changing the shape of the curved surface, (b) diaphragm of the plane type removes the cavity effect by using a flat radiation surface, (c) diaphragm of the dome type improves bending elasticity by forming thin plates into a domelike shape.

- Provide a light weight assembly

As shown in Figure 5.3.18, most shapes of the support system are corrugated, but linearity is improved and the maximum allowable range is increased.

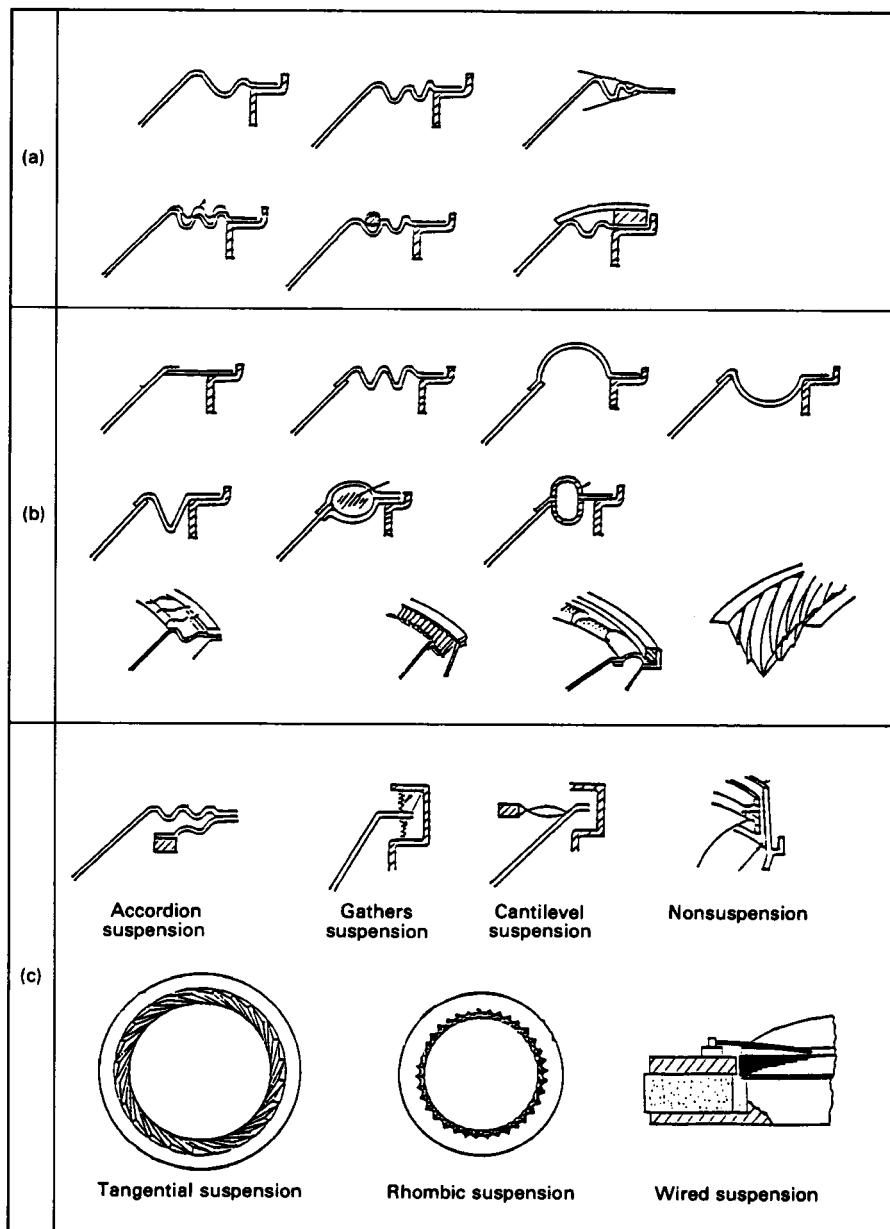
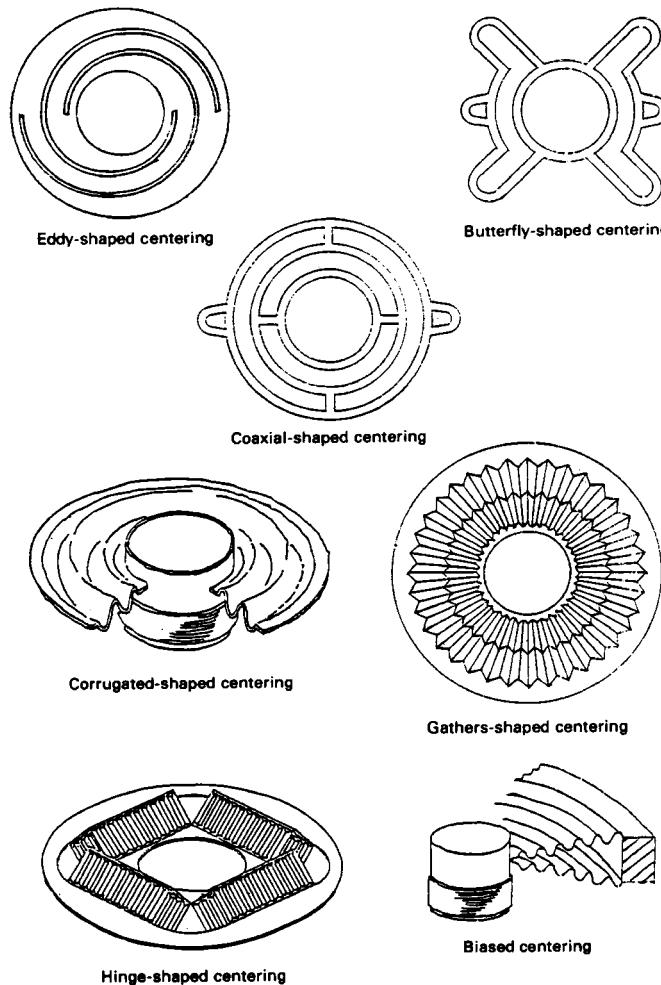


Figure 5.3.17 Sectional views of cone suspension systems: (a) the thinned edge of a diaphragm fulfills the function of the cone suspension, (b) material different from that of a diaphragm is used to fulfill the function of cone suspension, (c) exceptional cone suspensions.

5-44 Sound Reproduction Devices and Systems

**Figure 5.3.18** Various shapes of centering systems.**5.3.3 Bibliography**

- Allison, R., et al.: "On the Magnitude and Audibility of FM Distortion in Loudspeakers," *J. Audio Eng Soc.*, vol. 30, no. 10, pg. 694, 1982.
- Beranek, L. L.: *Acoustics*, McGraw-Hill, New York, N.Y., pg. 183, 1954.
- Hirata, Y.: "Study of Nonlinear Distortion in Audio Instruments," *J. Audio Eng. Soc.*, vol. 29, no. 9, pg. 607, 1981.
- Kinsler, L. E., et al: *Fundamentals of Acoustics*, Wiley, New York, N.Y., pg. 363, 1982.

Dynamic-Type Direct-Radiation Speakers 5-45

- Melillo, L., et al.: "Ferrolluids as a Means of Controlling Woofer Design Parameters," presented at the 63d Convention of the Audio Engineering Society, vol. 1, pg. 177, 1979.
- Niguchi, H., et al.: "Reinforced Olefin Polymer Diaphragm for Loudspeakers," *J. Audio Eng. Soc.*, vol. 29, no. 11, pg. 808, 1981.
- Okahara, M., et al: *Audio Handbook*, Ohm Sya, pg. 285, 1978 (in Japanese).
- Olson, H. F.: *Elements of Acoustical Engineering*, Van Nostrand, Princeton, N.J., pg. 123, 1957.
- Sakamoto, N.: *Loudspeaker and Loudspeaker Systems*, Nikkan Kogyo Shinbunshya, pg. 36, 1967 (in Japanese).
- Sakamotoet, N., et. al.: "Loudspeaker with Honeycomb Disk Diaphragm," *J. Audio Eng. Soc.*, vol. 29, no. 10, pg. 711, 1981.
- Shindo, T., et al: "Effect of Voice-Coil and Surround on Vibration and Sound Pressure Response of Loudspeaker Cones," *J. Audio Eng. Soc.*, vol. 28, no 7-8, pg. 490, 1980.
- Suwa, H., et al.: "Heat Pipe Cooling Enables Loudspeakers to Handle Higher Power," presented at the 63d Convention of the Audio Engineering Society, vol. 1, pg. 213, 1979.
- Suzuki, H., et al.: "Radiation and Diffraction Effects by Convex and Concave Domes," *J. Audio Eng Soc.*, vol. 29, no. 12, pg. 873, 1981.
- Takahashi, S., et al.: "Glass-Fiber and Graphite-Flake Reinforced Polyimide Composite Diaphragm for Loudspeakers," *J. Audio Eng. Soc.*, vol. 31, no. 10, pg. 723, 1983.
- Tsuchiya, H., et al.: "Reducing Harmonic Distortion in Loudspeakers," presented at the 63d Convention of the Audio Engineering Society, vol. 2, pg. 1, 1979.
- Yamamoto, T., et al.: "High-Fidelity Loudspeakers with Boronized Titanium Diaphragm," *J. Audio Eng. Soc.*, vol. 28, no. 12, pg. 868, 1980.
- Yoshihisa, N., et al.: "Nonlinear Distortion in Cone Loudspeakers," Chuyu-Ou Univ. Rep., vol. 23, pg. 271, 1980.

Section
6

Digital Coding of Audio Signals

Digital signal processing (DSP) techniques are being applied to the implementation of various stages of audio capture, processing, storage, and distribution systems for a number of reasons, including:

- Improved cost-performance considerations
- Future product-enhancement capabilities
- Greatly reduced alignment and testing requirements

A wide variety of video circuits and systems can be readily implemented using various degrees of embedded DSP. The most important parameters are signal bandwidth and S/N, which define, respectively, the required sampling rate and the effective number of bits required for the conversion. Additional design considerations include the stability of the sampling clock, quadrature channel matching, aperture uncertainty, and the cutoff frequency of the quantizer networks.

DSP devices differ from microprocessors in a number of ways. For one thing, microprocessors typically are built for a range of general-purpose functions and normally run large blocks of software. Also, microprocessors usually are not called upon to perform real-time computation. Typically, they are at liberty to shuffle workloads and to select an action branch, such as completing a printing job before responding to a new input command. The DSP, on the other hand, is dedicated to a single task or small group of related tasks. In a sophisticated video system, one or more DSPs may be employed as attached processors, assisting a general-purpose host microprocessor that manages the front-panel controls or other key functions of the unit.

One convenient way to classify DSP devices and applications is by their *dynamic range*. In this context, the dynamic range is the spread of numbers that must be processed in the course of an application. It takes a certain range of values, for example, to describe a particular signal, and that range often becomes even wider as calculations are performed on the input data. The DSP must have the capability to handle such data without overflow.

The processor capacity is a function of its data width, i. e., the number of bits it manipulates and the type of arithmetic that it performs (fixed or *floating point*). Floating point processing manipulates numbers in a form similar to scientific notation, enabling the device to accommodate an enormous breadth of data. Fixed arithmetic processing, as the name implies, restricts the processing capability of the device to a predefined value.

6-2 Section Six

Recent advancements in very large scale integration (VLSI) technologies in general, and DSP in particular, have permitted the integration of many video system functional blocks into a single device. Such designs typically offer excellent performance because of the elimination of the traditional interfaces required by discrete designs. This high level of integration also decreases the total parts count of the system, thereby increasing the overall reliability of the system.

The trend toward DSP operational blocks in video equipment of all types is perhaps the single most important driving force in video hardware today. It has reshaped products as diverse as cameras and displays. Thanks in no small part to research and development efforts in the computer industry, the impact is just now being felt in the television business.

In This Section:

Chapter 6.1: Analog/Digital Signal Conversion	6-5
Introduction	6-5
The Nyquist Limit and Aliasing	6-5
The A/D Conversion Process	6-6
Successive Approximation	6-9
Parallel/Flash	6-9
The D/A Conversion Process	6-11
Practical Implementation	6-12
Converter Performance Criteria	6-12
References	6-14
Chapter 6.2: Digital Filters	6-15
Introduction	6-15
FIR Filters	6-15
Design Techniques	6-18
Applications	6-18
Finite Wordlength Effects	6-19
Infinite Impulse Response Filters	6-20
Reference	6-22
Chapter 6.3: Digital Modulation	6-23
Introduction	6-23
Digital Modulation Techniques	6-23
QPSK	6-24
Signal Analysis	6-25
Digital Coding	6-26
Source Coding	6-26
Channel Coding	6-27
Error-Correction Coding	6-27
Reference	6-28
Chapter 6.4: DSP Devices and Systems	6-29
Introduction	6-29
Fundamentals of Digital Signal Processing	6-29

Discrete Systems	6-30
Impulse Response and Convolution	6-30
Complex Numbers	6-34
Mathematical Transforms	6-35
Unit Circle and Region of Convergence	6-39
Poles and Zeros	6-39
DSP Elements	6-41
Sources of Errors	6-42
DSP Integrated Circuits	6-43
DSP Applications	6-44
Digital Delay	6-44
Example DSP Device	6-46
Functional Overview	6-47
References	6-50

On the CD-ROM

- “Digital Audio” by P. Jeffrey Bloom, et. al., an archive chapter from the first edition of the *Audio Engineering Handbook*. This reference material explains the fundamental elements of digital audio coding, storage, and manipulation.

Reference Documents for this Section

Alkin, Oktay: “Digital Coding Schemes,” *The Electronics Handbook*, Jerry C. Whitaker (ed.), CRC Press, Boca Raton, Fla., pp. 1252–1258, 1996.

Benson, K. B., and D. G. Fink: “Digital Operations in Video Systems,” *HDTV: Advanced Television for the 1990s*, McGraw-Hill, New York, pp. 4.1–4.8, 1990.

Chambers, J. A., S. Tantaratana, and B. W. Bomar: “Digital Filters,” *The Electronics Handbook*, Jerry C. Whitaker (ed.), CRC Press, Boca Raton, Fla., pp. 749–772, 1996.

Garrod, Susan A. R.: “D/A and A/D Converters,” *The Electronics Handbook*, Jerry C. Whitaker (ed.), CRC Press, Boca Raton, Fla., pp. 723–730, 1996.

Garrod, Susan, and R. Borns: *Digital Logic: Analysis, Application, and Design*, Saunders College Publishing, Philadelphia, 1991.

Lee, E. A., and D. G. Messerschmitt: *Digital Communications*, 2nd ed., Kluwer, Norell, Mass., 1994.

Nyquist, H.: “Certain Factors Affecting Telegraph Speed,” *Bell System Tech. J.*, vol. 3, pp. 324–346, March 1924.

Parks, T. W., and J. H. McClellan: “A Program for the Design of Linear Phase Infinite Impulse Response Filters,” *IEEE Trans. Audio Electroacoustics*, AU-20(3), pp. 195–199, 1972.

Peterson, R., R. Ziemer, and D. Borth: *Introduction to Spread Spectrum Communications*, Prentice-Hall, Englewood Cliffs, N. J., 1995.

Pohlmann, Ken: *Principles of Digital Audio*, McGraw-Hill, New York, N.Y., 2000.

Sklar, B.: *Digital Communications: Fundamentals and Applications*, Prentice-Hall, Englewood Cliffs, N. J., 1988.

6-4 Section Six

TMS320C55x DSP Functional Overview, Texas Instruments, Dallas, TX, literature No. SRPU312, June 2000.

Ungerboeck, G.: "Trellis-Coded Modulation with Redundant Signal Sets," parts I and II, *IEEE Comm. Mag.*, vol. 25 (Feb.), pp. 5-11 and 12-21, 1987.

Ziemer, R., and W. Tranter: *Principles of Communications: Systems, Modulation, and Noise*, 4th ed., Wiley, New York, 1995.

Ziemer, Rodger E.: "Digital Modulation," *The Electronics Handbook*, Jerry C. Whitaker (ed.), CRC Press, Boca Raton, Fla., pp. 1213-1236, 1996.

Chapter
6.1
Analog/Digital Signal Conversion

Susan A. R. Garrod, K. Blair Benson, Donald G. Fink

Jerry C. Whitaker, Editor-in-Chief

6.1.1 Introduction

Analog-to-digital conversion (A/D) is the process of converting a continuous range of analog signals into specific digital codes. Such conversion is necessary to interface analog pickup elements and systems with digital devices and systems that process, store, interpret, transport, and manipulate the analog values. Analog-to-digital conversion is not an exact process; the comparison between the analog sample and a reference voltage is uncertain by the amount of the difference between one reference voltage and the next [1]. The uncertainty amounts to plus or minus one-half that difference. When words of 8 bits are used, this uncertainty occurs in essentially random fashion, so its effect is equivalent to the introduction of random noise (*quantization noise*). Fortunately, such noise is not prominent in the analog signal derived from the digital version. For example, in 8-bit digitization of the NTSC 4.2 MHz baseband at 13.5 megasamples per second (MS/s), the quantization noise is about 60 dB below the peak-to-peak signal level, far lower than the noise typically present in the analog signal from the camera.

6.1.2 The Nyquist Limit and Aliasing

A critical rule must be observed in sampling an analog signal if it is to be reproduced without spurious effects known as *aliasing*. The rule, first described by Nyquist in 1924 [2], states that the time between samples must be short compared with the rates of change of the analog waveform. In video terms, the sampling rate in megasamples per second must be at least twice the maximum frequency in megahertz of the analog signal. Thus, the 4.2 MHz maximum bandwidth in the luminance spectrum of the NTSC baseband requires that the NTSC signal be sampled at 8.4 MS/s or greater. Conversely, the 13.5 MS/s rate specified in the ITU-R studio digital standard

6-6 Digital Coding of Audio Signals

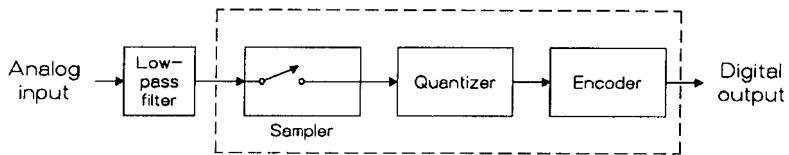


Figure 6.1.1 Basic elements of an analog-to-digital converter. (From [1]. Used with permission.)

can be applied to a signal having no higher frequency components than 6.75 MHz. If studio equipment exceeds this limit—and many cameras and associated amplifiers do—a low-pass filter must be inserted in the signal path before the conversion from analog to digital form takes place. A similar band limit must be met at 3.375 MHz in the chrominance channels before they are digitized in the NTSC system. If the sampling occurs at a rate lower than the Nyquist limit, the spectrum of the output analog signal contains spurious components, which are actually higher-frequency copies of the input spectrum that have been transposed so that they overlap the desired output spectrum. When this output analog signal is displayed, the spurious information shows up in a variety of forms, depending on the subject matter and its motions [1]. Moiré patterns are typical, as are distorted and randomly moving diagonal edges of objects. These aliasing effects often cover large areas and are visible at normal viewing distances.

Aliasing may occur, in fact, not only in digital sampling, but whenever any form of sampling of the image occurs. An example long familiar in motion pictures is that of vehicle wheels (usually wagon wheels) that appear to be rotating backward as the vehicle moves forward. This occurs because the image is sampled by the camera at 24 frames/s. If the rotation of the spokes of the wheel is not precisely synchronous with the film advance, another spoke takes the place of the adjacent one on the next frame, at an earlier time in its rotation. The two spokes are not separately identified by the viewer, so the spoke motion appears reversed. Many other examples of image sampling occur in television. The display similarly offers a series of samples in the vertical dimension, with results that depend not only on the time-vs.-light characteristics of the display device but also, and more important, on the time-vs.-sensation properties of the human eye.

6.1.3 The A/D Conversion Process

To convert a signal from the analog domain into a digital form, it is necessary to create a succession of digital words that comprise only two discrete values, 0 and 1 [1]. Figure 6.1.1 shows the essential elements of the analog-to-digital converter. The input analog signal must be confined to a limited spectrum to prevent spurious components in the reconverted analog output. A low-pass filter, therefore, is placed prior to the converter. The converter proper first samples the analog input, measuring its amplitude at regular, discrete intervals of time. These individual amplitudes then are matched, in the quantizer, against a large number of discrete levels of amplitude (256 levels to convert into 8-bit words). Each one of these discrete levels can be represented by a specific digital word. The process of matching each discrete amplitude with its unique word is carried out in the encoder, which, in effect, scans the list of words and picks out the one that matches

Table 6.1.1 Binary Values of Amplitude Levels for 8-Bit Words (From [1]. Used with permission.)

Amplitude	Binary Level	Amplitude	Binary Level	Amplitude	Binary Level
0	00000000	120	01111000	240	11110000
1	00000001	121	01111001	241	11110001
2	00000010	122	01111010	242	11110010
3	00000011	123	01111011	243	11110011
4	00000100	124	01111100	244	11110100
5	00000101	125	01111101	245	11110101
6	00000110	126	01111110	246	11110110
7	00000111	127	01111111	247	11110111
8	00001000	128	10000000	248	11111000
9	00001001	129	10000001	249	11111001
10	00001010	130	10000010	250	11111010
11	00001011	131	10000011	251	11111011
12	00001100	132	10000100	252	11111100
13	00001101	133	10000101	253	11111101
14	00001110	134	10000110	254	11111110
15	00001111	135	10000111	255	11111111

the amplitude then present. The encoder passes out the series of code words in a sequence corresponding to the sequence in which the analog signal was sampled. This *bit stream* is, consequently, the digital version of the analog input.

The list of digital words corresponding to the sampled amplitudes is known as a *code*. Table 6.1.1 represents a simple code showing amplitude levels and their 8-bit words in three ranges: 0 to 15, 120 to 135, and 240 to 255. Signals encoded in this way are said to be *pulse-code-modulated*. Although the basic pulse-code modulation (PCM) code sometimes is used, more elaborate codes—with many additional bits per word—generally are applied in circuits where errors may be introduced into the bit stream. Figure 6.2.2 shows a typical video waveform and several quantized amplitude levels based on the PCM coding scheme of Table 6.1.1.

The sampling rate, even in analog sampling systems, is crucial. Figure 6.1.3a shows the spectral consequence of a sampling rate that is too low for the input bandwidth; Figure 6.1.3b shows the result of a rate equal to the theoretical minimum value, which is impractical; and Figure 6.1.3c shows typical practice. The input spectrum must be limited by a low-pass filter to greatly attenuate frequencies near one-half the sampling rate and above. The higher the sampling rate, the easier and simpler the design of the input filter becomes. An excessively high sampling rate, however, is wasteful of transmission bandwidth and storage capacity, while a low but adequate rate complicates the design and increases the cost of input and output analog filters.

Analog signals can be converted to digital codes using a number of methods, including the following [3]:

6-8 Digital Coding of Audio Signals

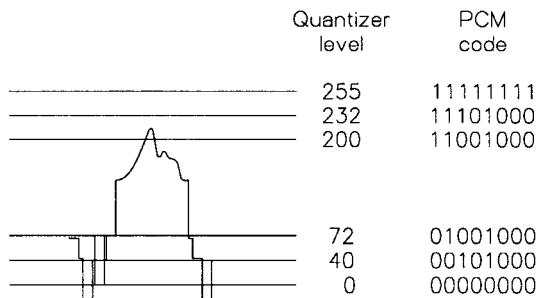


Figure 6.1.2 Video waveform quantized into 8-bit words.

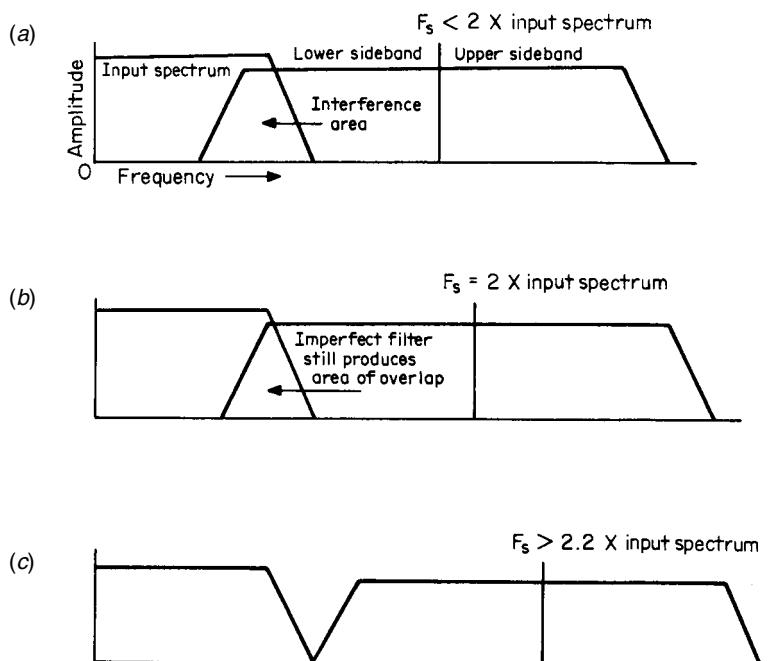


Figure 6.1.3 Relationship between sampling rate and bandwidth: (a) a sampling rate too low for the input spectrum, (b) the theoretical minimum sampling rate (F_s), which requires a theoretically perfect filter, (c) a practical sampling rate using a practical input filter.

- Integration
- Successive approximation
- Parallel (flash) conversion

- Delta modulation
- Pulse-code modulation
- Sigma-delta conversion

Two of the more common A/D conversion processes are successive approximation and parallel or flash. Very high-resolution digital video systems require specialized A/D techniques that often incorporate one of these general schemes in conjunction with proprietary technology.

6.1.3a Successive Approximation

Successive approximation A/D conversion is a technique commonly used in medium- to high-speed data-acquisition applications. One of the fastest A/D conversion techniques, it requires a minimum amount of circuitry to implement. The conversion times for successive approximation A/D conversion typically range from 10 to 300 μ s for 8-bit systems.

The successive approximation A/D converter can approximate the analog signal to form an n -bit digital code in n steps. The *successive approximation register* (SAR) individually compares an analog input voltage with the midpoint of one of n ranges to determine the value of 1 bit. This process is repeated a total of n times, using n ranges, to determine the n bits in the code. The comparison is accomplished as follows:

- The SAR determines whether the analog input is above or below the midpoint and sets the bit of the digital code accordingly.
- The SAR assigns the bits beginning with the most significant bit.
- The bit is set to a 1 if the analog input is greater than the midpoint voltage; it is set to a 0 if the input is less than the midpoint voltage.
- The SAR then moves to the next bit and sets it to a 1 or a 0 based on the results of comparing the analog input with the midpoint of the next allowed range.

Because the SAR must perform one approximation for each bit in the digital code, an n -bit code requires n approximations. A successive approximation A/D converter consists of four main functional blocks, as shown in Figure 6.1.4. These blocks are the SAR, the analog comparator, a D/A (digital-to-analog) converter, and a clock.

6.1.3b Parallel/Flash

Parallel or flash A/D conversion is used in high-speed applications such as video signal processing, medical imaging, and radar detection systems. A flash A/D converter simultaneously compares the input analog voltage with $2^n - 1$ threshold voltages to produce an n -bit digital code representing the analog voltage. Typical flash A/D converters with 8-bit resolution operate at 100 MHz to 1 GHz.

The functional blocks of a flash A/D converter are shown in Figure 6.1.5. The circuitry consists of a precision resistor ladder network, $2^n - 1$ analog comparators, and a digital priority encoder. The resistor network establishes threshold voltages for each allowed quantization level. The analog comparators indicate whether the input analog voltage is above or below the threshold at each level. The output of the analog comparators is input to the digital priority encoder. The priority encoder produces the final digital output code, which is stored in an output latch.

6-10 Digital Coding of Audio Signals

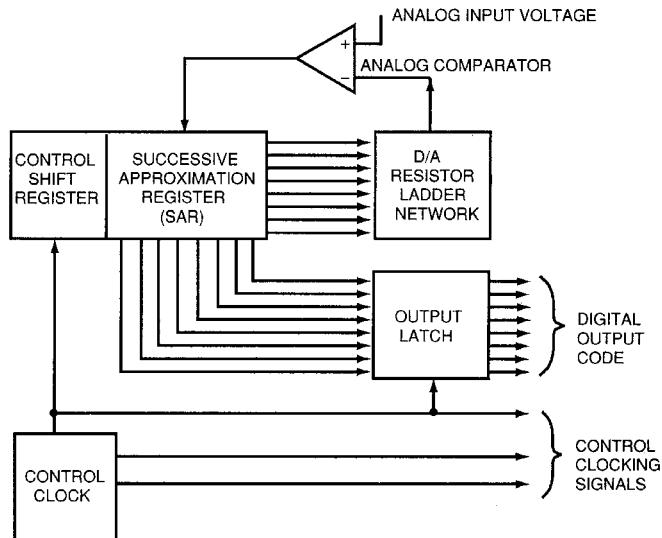


Figure 6.1.4 Successive approximation A/D converter block diagram. (After [4].)

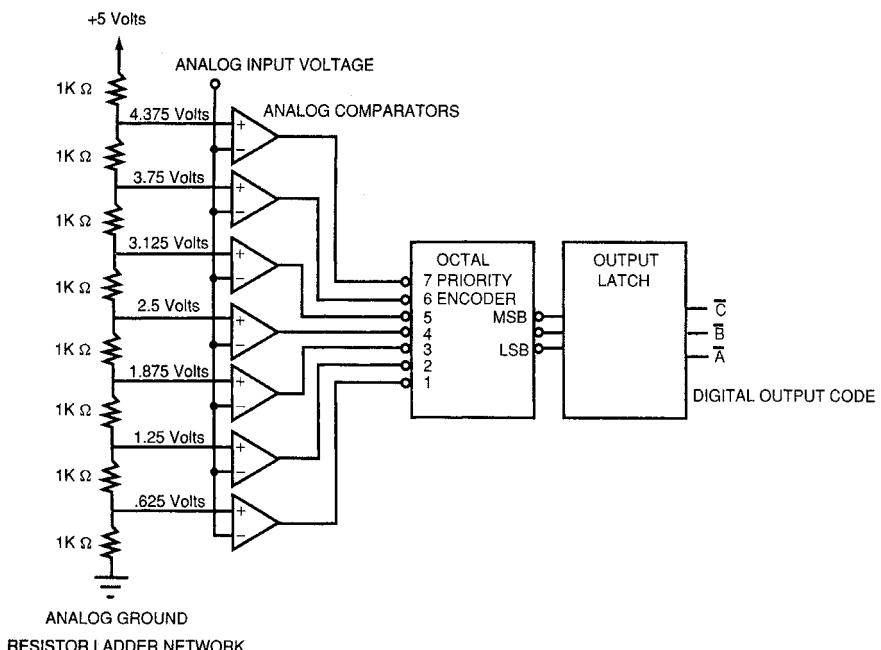


Figure 6.1.5 Block diagram of a flash A/D converter. (After [5].)

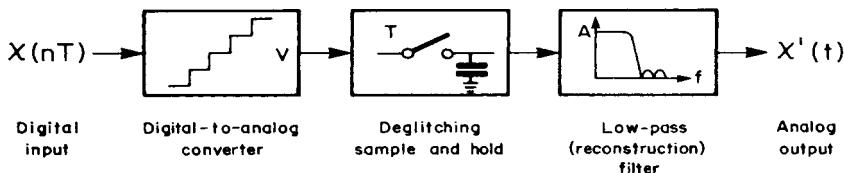


Figure 6.1.6 Digital-to-analog converter block diagram.

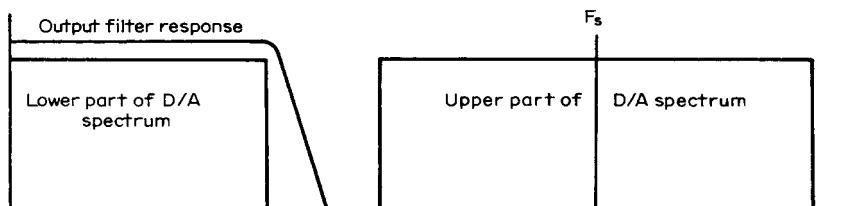


Figure 6.1.7 Output filter response requirements for a common D/A converter.

An 8-bit flash A/D converter requires 255 comparators. The cost of high-resolution A/D comparators escalates as the circuit complexity increases and the number of analog converters rises by $2^n - 1$. As a low-cost alternative, some manufacturers produce modified flash converters that perform the A/D conversion in two steps, to reduce the amount of circuitry required. These modified flash converters also are referred to as *half-flash* A/D converters because they perform only half of the conversion simultaneously.

6.1.4 The D/A Conversion Process

The digital-to-analog converter (DAC) is, in principle, quite simple. The digital stream of binary pulses is decoded into discrete, sequentially timed signals corresponding to the original sampling in the A/D. The output is an analog signal of varying levels. The time duration of each level is equal to the width of the sample taken in the A/D conversion process. The analog signal is separated from the sampling components by a low-pass filter. Figure 6.1.6 shows a simplified block diagram of a D/A. The deglitching sample-and-hold circuits in the center block set up the analog levels from the digital decoding and remove the unwanted high-frequency sampling components. Each digital number is converted to a corresponding voltage and stored until the next number is converted. Figure 6.1.7 shows the resulting spectrum. The energy surrounding the sampling frequency must be removed, and an output low-pass filter is used to accomplish that task. One cost-effective technique used in a variety of applications is called *oversampling*. A new sampling rate is selected that is a whole multiple of the input sampling rate. The new rate is typically two or four times the old rate. Every second or fourth sample is filled with the input value, while the others are set to zero. The result is passed through a digital filter that distributes the energy in the

6-12 Digital Coding of Audio Signals

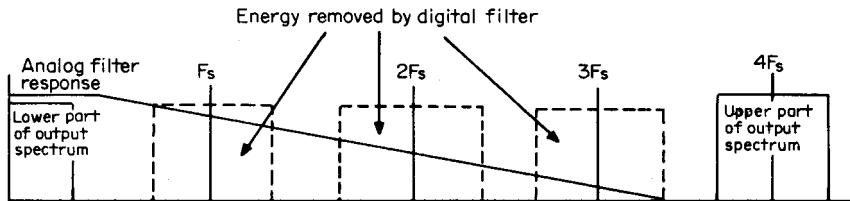


Figure 6.1.8 The filtering benefits of oversampling.

real samples among the empty ones and itself. The resulting spectrum (for a 4 \times oversampling system) is shown in Figure 6.1.8. The energy around the 4 \times sample frequency must be removed, which can be done simply because it is so distant from the upper band edge. The response of the output filter is chiefly determined by the digital processing and is therefore very stable with age, in contrast to a strictly analog filter, whose component values are susceptible to drift with age and other variables.

6.1.4a Practical Implementation

To convert digital codes to analog voltages, a voltage weight typically is assigned to each bit in the digital code, and the voltage weights of the entire code are summed [3]. A general-purpose D/A converter consists of a network of precision resistors, input switches, and level shifters to activate the switches to convert the input digital code to an analog current or voltage output. A D/A device that produces an analog current output usually has a faster settling time and better linearity than one that produces a voltage output.

D/A converters commonly have a fixed or variable reference level. The reference level determines the switching threshold of the precision switches that form a controlled impedance network, which in turn controls the value of the output signal. *Fixed-reference* D/A converters produce an output signal that is proportional to the digital input. In contrast, *multiplying* D/A converters produce an output signal that is proportional to the product of a varying reference level times a digital code.

D/A converters can produce bipolar, positive, or negative polarity signals. A *four-quadrant multiplying* D/A converter allows both the reference signal and the value of the binary code to have a positive or negative polarity.

6.1.5 Converter Performance Criteria

The major factors that determine the quality of performance of A/D and D/A converters are *resolution*, *sampling rate*, *speed*, and *linearity* [3]. The resolution of a D/A circuit is the smallest possible change in the output analog signal. In an A/D system, the resolution is the smallest change in voltage that can be detected by the system and produce a change in the digital code. The resolution determines the total number of digital codes, or quantization levels, that will be recognized or produced by the circuit.

Table 6.1.2 Representative Sampling of Converter Operating Parameters

Converter Type	Sampling Rate	Resolution	S/Nq¹	Max. Input Frequency	Power Consumption
A/D	400 Ms/s ²	8 bits	43 dB	1 GHz	3 W
	200 Ms/s	10 bits	58 dB	400 MHz	2 W
	120 Ms/s	12 bits	70 dB	350 MHz	1 W
	70 Ms/s	14 bits	75 dB	300 MHz	1.3 W
D/A	Sampling Rate	Resolution	Dynamic Range		Power Consumption
	500 Ms/s ³	10 bits	80 dB		250 mW
	300 Ms/s	12 bits	85 dB		300 mW
	200 Ms/s	14 bits	88 dB		350 mW

Notes:

¹ signal-to-quantization noise, ² megasamples per second, ³ settling time in megasamples per second

The resolution of a D/A or A/D device usually is specified in terms of the bits in the digital code, or in terms of the *least significant bit* (LSB) of the system. An n -bit code allows for 2^n quantization levels, or $2^n - 1$ steps between quantization levels. As the number of bits increases, the step size between quantization levels decreases, therefore increasing the accuracy of the system when a conversion is made between an analog and digital signal. The system resolution also can be specified as the voltage step size between quantization levels.

The speed of a D/A or A/D converter is determined by the amount of time it takes to perform the conversion process. For D/A converters, the speed is specified as the *settling time*. For A/D converters, the speed is specified as the *conversion time*. The settling time for a D/A converter varies with supply voltage and transition in the digital code; it is specified in the data sheet with the appropriate conditions stated.

A/D converters have a maximum sampling rate that limits the speed at which they can perform continuous conversions. The sampling rate is the number of times per second that the analog signal can be sampled and converted into a digital code. For proper A/D conversion, the minimum sampling rate must be at least 2 times the highest frequency of the analog signal being sampled to satisfy the Nyquist criterion. The conversion speed and other timing factors must be taken into consideration to determine the maximum sampling rate of an A/D converter. Nyquist A/D converters use a sampling rate that is slightly greater than twice the highest frequency in the analog signal. *Oversampling* A/D converters use sampling rates of N times rate, where N typically ranges from 2 to 64.

Both D/A and A/D converters require a voltage reference to achieve absolute conversion accuracy. Some conversion devices have internal voltage references, whereas others accept external voltage references. For high-performance systems, an external precision reference is required to ensure long-term stability, load regulation, and control over temperature fluctuations.

Measurement accuracy is specified by the converter's linearity. *Integral linearity* is a measure of linearity over the entire conversion range. It often is defined as the deviation from a straight line drawn between the endpoints and through zero (or the *offset value*) of the conversion range.

6-14 Digital Coding of Audio Signals

Integral linearity also is referred to as *relative accuracy*. The offset value is the reference level required to establish the zero or midpoint of the conversion range. *Differential linearity*, the linearity between code transitions, is a measure of the *monotonicity* of the converter. A converter is said to be monotonic if increasing input values result in increasing output values.

The accuracy and linearity values of a converter are specified in units of the LSB of the code. The linearity may vary with temperature, so the values often are specified at +25°C as well as over the entire temperature range of the device.

With each new generation of devices, A/D and D/A converter technology improves, yielding higher sampling rates with greater resolution. Table 6.1.2 shows some typical values as this book went to press.

6.1.6 References

1. Benson, K. B., and D. G. Fink: "Digital Operations in Video Systems," *HDTV: Advanced Television for the 1990s*, McGraw-Hill, New York, pp. 4.1–4.8, 1990.
2. Nyquist, H.: "Certain Factors Affecting Telegraph Speed," *Bell System Tech. J.*, vol. 3, pp. 324–346, March 1924.
3. Garrod, Susan A. R.: "D/A and A/D Converters," *The Electronics Handbook*, Jerry C. Whitaker (ed.), CRC Press, Boca Raton, Fla., pp. 723–730, 1996.
4. Garrod, Susan, and R. Borns: *Digital Logic: Analysis, Application, and Design*, Saunders College Publishing, Philadelphia, pg. 919, 1991.
5. Garrod, Susan, and R. Borns: *Digital Logic: Analysis, Application, and Design*, Saunders College Publishing, Philadelphia, pg. 928, 1991.

Chapter
6.2
Digital Filters

J. A. Chambers, S. Tantaratana, B. W. Bomar

Jerry C. Whitaker, Editor-in-Chief

6.2.1 Introduction

Digital filtering is concerned with the manipulation of discrete data sequences to remove noise, extract information, change the sample rate, and/or modify the input information in some form or context [1]. Although an infinite number of numerical manipulations can be applied to discrete data (e.g., finding the mean value, forming a histogram), the objective of digital filtering is to form a discrete output sequence $y(n)$ from a discrete input sequence $x(n)$. In some manner, each output sample is computed from the input sequence—not just from any one sample, but from many, possibly all, of the input samples. Those filters that compute their output from the present input and a finite number of past inputs are termed *finite impulse response* (FIR) filters; those that use all past inputs are termed *infinite impulse response* (IIR) filters.

6.2.2 FIR Filters

An FIR filter is a linear discrete-time system that forms its output as the weighted sum of the most recent, and a finite number of past, inputs [1]. A *time-invariant* FIR filter has finite memory, and its impulse response (its response to a discrete-time input that is unity at the first sample and otherwise zero) matches the fixed weighting coefficients of the filter. *Time-variant* FIR filters, on the other hand, may operate at various sampling rates and/or have weighting coefficients that adapt in sympathy with some statistical property of the environment in which they are applied.

Perhaps the simplest example of an FIR filter is the *moving average* operation described by the following linear constant-coefficient difference equation:

6-16 Digital Coding of Audio Signals

$$y[n] = \sum_{k=0}^{M} b_k x[n-k] \quad b_k = \frac{1}{M+1} \quad (6.2.1)$$

Where:

$y[n]$ = output of the filter at integer sample index n

$x[n]$ = input to the filter at integer sample index n

b_k = filter weighting coefficients, $k = 0, 1, \dots, M$

M = filter order

In a practical application, the input and output discrete-time signals will be sampled at some regular sampling time interval, T seconds, denoted $x[nT]$ and $y[nT]$, which is related to the sampling frequency by $f_s = 1/T$, samples per second. However, for generality, it is more convenient to assume that T is unity, so that the effective sampling frequency also is unity and the Nyquist frequency is one-half. It is, then, straightforward to scale, by multiplication, this normalized frequency range, i.e. $[0, 1/2]$, to any other sampling frequency.

The output of the simple moving average filter is the average of the $M + 1$ most recent values of $x[n]$. Intuitively, this corresponds to a smoothed version of the input, but its operation is more appropriately described by calculating the frequency response of the filter. First, however, the z -domain representation of the filter is introduced in analogy to the s - (or *Laplace*) domain representation of analog filters. The z -transform of a causal discrete-time signal $x[n]$ is defined by:

$$X(z) = \sum_{n=0}^{\infty} x[n] z^{-n} \quad (6.2.2)$$

Where:

$X(z)$ = z -transform of $x[n]$

z = complex variable

The z -transform of a delayed version of $x[n]$, namely $x[n - k]$ with k a positive integer, is found to be given by $z^{-k} X(z)$. This result can be used to relate the z -transform of the output, $y[n]$, of the simple moving average filter to its input:

$$Y(z) = \sum_{k=0}^{M} b_k z^{-k} X(z) \quad b_k = \frac{1}{M+1} \quad (6.2.3)$$

The z -domain transfer function, namely the ratio of the output to input transform, becomes:

$$H(z) = \frac{Y(z)}{X(z)} = \sum_{k=0}^{M} b_k z^{-k} \quad b_k = \frac{1}{M+1} \quad (6.2.4)$$

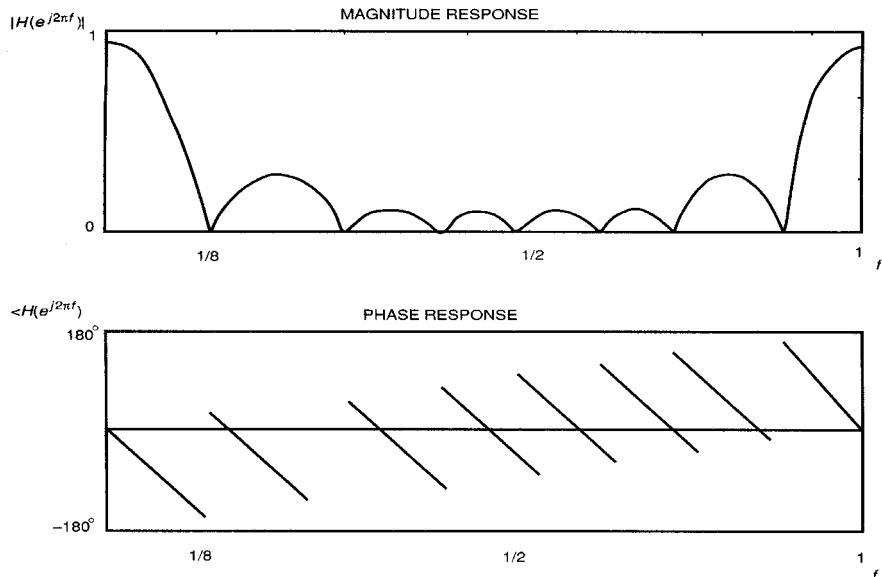


Figure 6.2.1 The magnitude and phase response of the simple moving average filter with $M = 7$. (From [1]. Used with permission.)

Notice the transfer function, $H(z)$, is entirely defined by the values of the weighting coefficients, b_k , $k = 0, 1, \dots, M$, which are identical to the discrete impulse response of the filter, and the complex variable z . The finite length of the discrete impulse response means that the transient response of the filter will last for only $M + 1$ samples, after which a steady state will be reached. The frequency-domain transfer function for the filter is found by setting

$$z = e^{j2\pi f} \quad (6.2.5)$$

Where $j = \sqrt{-1}$ and can be written as:

$$H(e^{j2\pi f}) = \frac{1}{M+1} \sum_{k=0}^M e^{-j2\pi fk} = \frac{1}{M+1} e^{-j\pi fM} \frac{\sin[\pi f(M+1)]}{\sin(\pi f)} \quad (6.2.6)$$

The magnitude and phase response of the simple moving average filter, with $M = 7$, are calculated from $H(e^{j2\pi f})$ and shown in Figure 6.2.1. The filter is seen clearly to act as a crude low-pass smoothing filter with a linear phase response. The sampling frequency periodicity in the magnitude and phase response is a property of discrete-time systems. The linear phase response is due to the $e^{-j\pi fM}$ term in $H(e^{j2\pi f})$ and corresponds to a constant $M/2$ group delay through

6-18 Digital Coding of Audio Signals

the filter. A phase discontinuity of $\pm 180^\circ$ is introduced each time the magnitude term changes sign. FIR filters that have center symmetry in their weighting coefficients have this constant frequency-independent group-delay property that is desirable in applications in which time dispersion is to be avoided, such as in pulse transmission, where it is important to preserve pulse shapes [2].

6.2.2a Design Techniques

Linear-phase FIR filters can be designed to meet various filter specifications, such as low-pass, high-pass, bandpass, and band-stop filtering [1]. For a low-pass filter, two frequencies are required. One is the maximum frequency of the passband below which the magnitude response of the filter is approximately unity, denoted the *passband corner frequency* f_p . The other is the minimum frequency of the stop-band above which the magnitude response of the filter must be less than some prescribed level, named the *stop-band corner frequency* f_s . The difference between the passband and stop-band corner frequencies is the *transition bandwidth*. Generally, the order of an FIR filter, M , required to meet some design specification will increase with a reduction in the width of the transition band. There are three established techniques for coefficient design:

- *Windowing*. A design method that calculates the weighting coefficients by sampling the ideal impulse response of an analog filter and multiplying these values by a smoothing window to improve the overall frequency-domain response of the filter.
- *Frequency sampling*. A technique that samples the ideal frequency-domain specification of the filter and calculates the weighting coefficients by inverse-transforming these values.
- Optimal approximations.

The best results generally can be obtained with the optimal approximations method. With the increasing availability of desktop and portable computers with fast microprocessors, large quantities of memory, and sophisticated software packages, optimal approximations is the preferred method for weighting coefficient design. The impulse response and magnitude response for a 40th-order optimal half-band FIR low-pass filter designed with the *Parks-McClellan* algorithm [3] are shown in Figure 6.2.2, together with the ideal frequency-domain design specification. Notice the zeros in the impulse response. This algorithm minimizes the peak deviation of the magnitude response of the design filter from the ideal magnitude response. The magnitude response of the design filter alternates about the desired specification within the passband and above the specification in the stop-band. The maximum deviation from the desired specification is equalized across the passband and stop-band; this is characteristic of an *optimal solution*.

6.2.2b Applications

In general, digitally implemented FIR filters exhibit the following attributes [1]:

- Absence of drift
- Reproducibility
- Multirate realizations
- Ability to adapt to time-varying environments

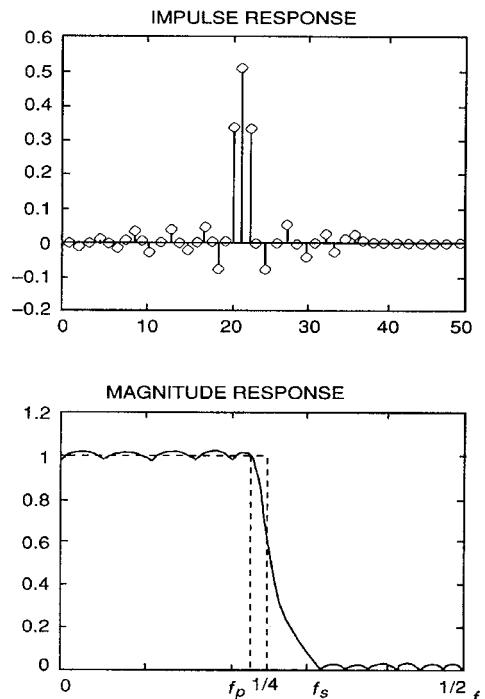


Figure 6.2.2 The impulse and magnitude response of an optimal 40th-order half-band FIR filter. (From [1]. Used with permission.)

These features have led to the widespread use of FIR filters in a variety of applications, particularly in telecommunications. The primary advantage of the fixed-coefficient FIR filter is its unconditional stability because of the lack of feedback within its structure and its exact linear phase characteristics. Nonetheless, for applications that require sharp, selective, filtering—in standard form—they do require relatively large orders. For some applications, this may be prohibitive; therefore, recursive IIR filters are a valuable alternative.

6.2.2c Finite Wordlength Effects

Practical digital filters must be implemented with finite precision numbers and arithmetic [1]. As a result, both the filter coefficients and the filter input and output signals are in discrete form. This leads to four types of finite wordlength effects:

- *Discretization* (quantization) of the filter coefficients has the effect of perturbing the location of the filter poles and zeroes. As a result, the actual filter response differs slightly from the ideal response. This deterministic frequency response error is referred to as *coefficient quantization error*.
- The use of finite precision arithmetic makes it necessary to quantize filter calculations by rounding or truncation. *Roundoff noise* is that error in the filter output that results from rounding or truncating calculations within the filter. As the name implies, this error looks like low-level noise at the filter output.

6-20 Digital Coding of Audio Signals

- Quantization of the filter calculations also renders the filter slightly nonlinear. For large signals this nonlinearity is negligible, and roundoff noise is the major concern. However, for recursive filters with a zero or constant input, this nonlinearity can cause spurious oscillations called *limit cycles*.
- With fixed-point arithmetic it is possible for filter calculations to overflow. The term *overflow oscillation* refers to a high-level oscillation that can exist in an otherwise stable filter because of the nonlinearity associated with the overflow of internal filter calculations. Another term for this is *adder overflow limit cycle*.

6.2.3 Infinite Impulse Response Filters

A digital filter with impulse response having infinite length is known as an *infinite impulse response* filter [1]. Compared to an FIR filter, an IIR filter requires a much lower order to achieve the same requirement of the magnitude response. However, whereas an FIR filter is always stable, an IIR filter may be unstable if the coefficients are not chosen properly. Because the phase of a stable causal IIR filter cannot be made linear, FIR filters are preferable to IIR filters in applications for which linear phase is essential.

Practical *direct form* realizations of IIR filters are shown in Figure 6.2.3. The realization shown in Figure 6.2.3a is known as *direct form I*. Rearranging the structure results in *direct form II*, as shown in Figure 6.2.3b. The results of transposition are *transposed direct form I* and *transposed direct form II*, as shown in Figures 6.2.3c and 6.2.3d, respectively. Other realizations for IIR filters are *state-space structure*, *wave structure*, and *lattice structure*. In some situations, it is more convenient or suitable to use software realizations that are implemented by programming a general-purpose microprocessor or a digital signal processor. (See [1] for details on IIR filter implementations.)

Designing an IIR filter involves choosing the coefficients to satisfy a given specification, usually a magnitude response parameter. There are various IIR filter design techniques, including:

- Design using an analog prototype filter, in which an analog filter is designed to meet the (analog) specification and the analog filter transfer function is transformed to a digital system function.
- Design using digital frequency transformation, which assumes that a given digital low-pass filter is available, and the desired digital filter is then obtained from the digital low-pass filter by a digital frequency transformation.
- Computer-aided design (CAD), which involves the execution of algorithms that choose the coefficients so that the response is as close as possible to the desired filter.

The first two methods are easily accomplished; they are suitable for designing standard filters (low-pass, high-pass, bandpass, and band-stop). The CAD approach, however, can be used to design both standard and nonstandard filters.

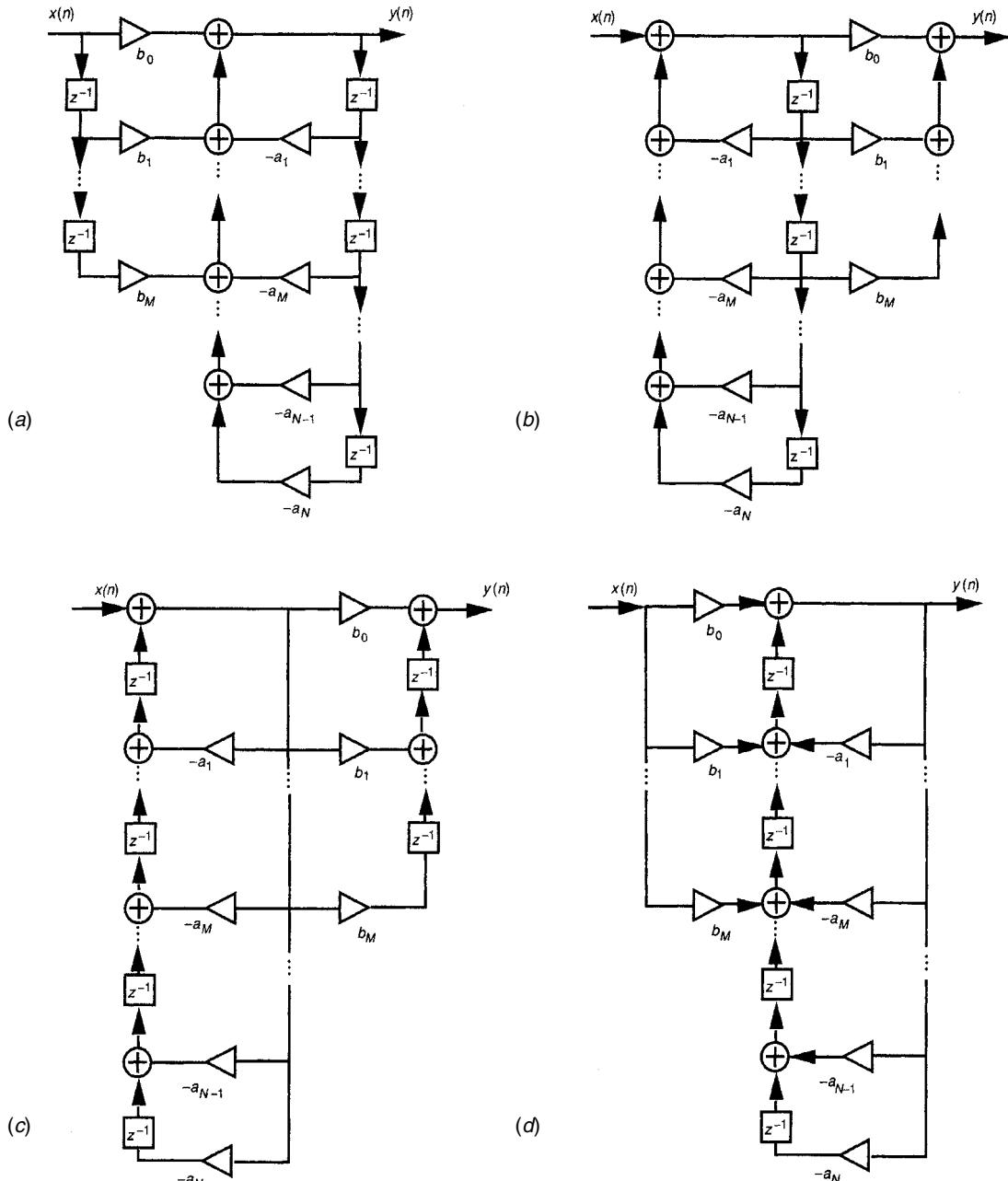


Figure 6.2.3 Direct form realizations of IIR filters: (a) direct form I, (b) direct form II, (c) transposed direct form I, (d) transposed direct form II. (From [1]. Used with permission.)

6-22 Digital Coding of Audio Signals**6.2.4 Reference**

1. Chambers, J. A., S. Tararatana, and B. W. Bomar: "Digital Filters," *The Electronics Handbook*, Jerry C. Whitaker (ed.), CRC Press, Boca Raton, Fla., pp. 749–772, 1996.
2. Lee, E. A., and D. G. Messerschmitt: *Digital Communications*, 2nd ed., Kluwer, Norell, Mass., 1994.
3. Parks, T. W., and J. H. McClellan: "A Program for the Design of Linear Phase Infinite Impulse Response Filters," *IEEE Trans. Audio Electroacoustics*, AU-20(3), pp. 195–199, 1972.

Chapter
6.3
Digital Modulation

Rodger E. Ziemer, Oktay Alkin

Jerry C. Whitaker, Editor-in-Chief

6.3.1 Introduction

Digital modulation is necessary before digital data can be transmitted through a channel, be it a satellite link or HDTV. *Modulation* is the process of varying some attribute of a carrier waveform as a function of the input intelligence to be transmitted. Attributes that can be varied include amplitude, frequency, and phase.

6.3.2 Digital Modulation Techniques

With digital modulation, the message sequence is a stream of digits, typically of binary value [1]. In the simplest case, parameter variation is on a symbol-by-symbol basis; no memory is involved. Carrier parameters that can be varied under this scenario include the following:

- Amplitude, resulting in *amplitude-shift keying* (ASK)
- Frequency, resulting in *frequency-shift keying* (FSK)
- Phase, resulting in *phase-shift keying* (PSK)

So-called higher-order modulation schemes impose memory over several symbol periods. Such modulation techniques can be classified as *binary* or *M-ary*, depending on whether one of two possible signals or $M > 2$ signals per signaling interval can be sent. (Binary signaling may be defined as any signaling scheme in which the number of possible signals sent during any given signaling interval is two. *M-ary* signaling, on the other hand, is a signaling system in which the number of possible signals sent during any given signaling interval is M .) For the case of *M-ary* modulation when the source digits are binary, it is clear that several bits must be grouped to make up an *M-ary* word.

Another classification for digital modulation is *coherent* vs. *noncoherent*, depending upon whether a reference carrier at the receiver coherent with the received carrier is required for

6-24 Digital Coding of Audio Signals

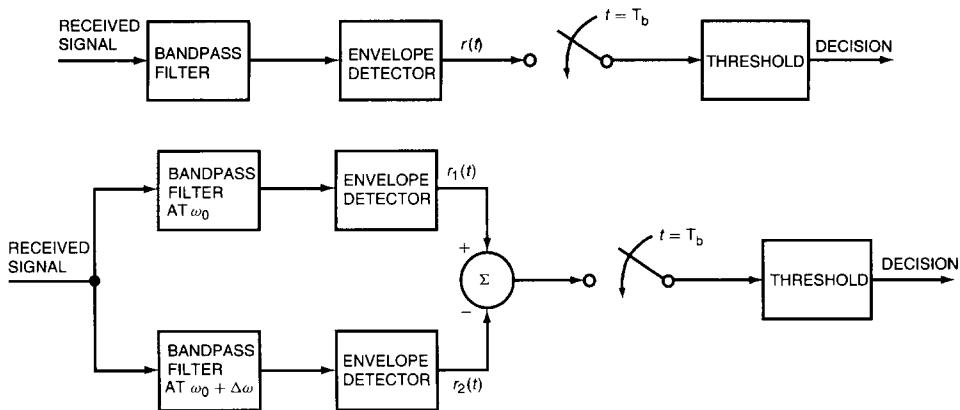


Figure 6.3.1 Receiver systems for noncoherent detection of binary signals: (a) ASK, (b) FSK. (From [1]. Used with permission.)

demodulation (the coherent case) or not (the noncoherent case). For situations in which it is difficult to maintain phase stability—for example, in channels subject to fading—it is useful to employ a modulation technique that does not require the acquisition of a reference signal at the receiver that is phase-coherent with the received carrier. ASK and FSK are two modulation techniques that lend themselves well to noncoherent detection. Receivers for detection of ASK and FSK noncoherently are shown in Figure 6.3.1.

One other binary modulation technique is, in a sense, noncoherent: *differentially coherent PSK* (DPSK). With DPSK, the phase of the preceding bit interval is used as a reference for the current bit interval. This technique depends on the channel being sufficiently stable so that phase changes resulting from channel perturbations from a given bit interval to the succeeding one are inconsequential. It also depends on there being a known phase relationship from one bit interval to the next. This requirement is ensured by differentially encoding the bits before phase modulation at the transmitter. Differential encoding is illustrated in Table 6.3.1. An arbitrary reference bit is chosen to start the process. In the table a 1 has been chosen. For each bit of the encoded sequence, the present bit is used as the reference for the following bit in the sequence. A 0 in the message sequence is encoded as a transition from the state of the reference bit to the opposite state in the encoded message sequence. A 1 is encoded as no change of state. Using these rules, the result is the encoded sequence shown in the table.

6.3.2a QPSK

Consider the case of an MPSK signal where $M = 4$, commonly referred to as *quadrature-phase-shift keying* (QPSK) [1]. This common modulation technique utilizes four signals in the signal set distinguished by four phases 90° apart. For the case of an MASK signal where $M = 4$, a *quadrature-amplitude-shift keying* (QASK) condition results. With QASK, both the phase and amplitude of

Table 6.3.1 Example of the differential encoding process (After [1].)

Message Sequence		1	0	0	1	1	1	0
Encoded Sequence	1	1	0	1	1	1	1	0
Transmitted Phase Radians	0	0	p	0	0	0	0	p

the carrier take on a set of values in one-to-one correspondence with the digital data to be transmitted.

Several variations of QPSK have been developed to meet specific operational requirements. One such scheme is referred to as *offset QPSK* (OQPSK) [2]. This format is produced by allowing only $\pm 90^\circ$ phase changes in a QPSK format. Furthermore, the phase changes can take place at multiples of a half-symbol interval, or a bit period. The reason for limiting phase changes to $\pm 90^\circ$ is to prevent the large envelope deviations that occur when QPSK is filtered to restrict side-lobe power, and regrowth of the sidelobes after amplitude limiting is used to produce a constant-envelope signal. This condition often is encountered in satellite systems in which, because of power-efficiency considerations, hard limiting repeaters are used in the communications system.

Another modulation technique closely related to QPSK and OQPSK is *minimum shift keying* (MSK) [2]. MSK is produced from OQPSK by weighting the in-phase and quadrature components of the baseband OQPSK signal with half-sinusoids. The phase changes linearly over a bit interval. As with OQPSK, the goal of MSK is to produce a modulated signal with a spectrum of reduced sidelobe power, one that behaves well when filtered and limited. Many different forms of MSK have been proposed and investigated over the years.

A modulation scheme related to 8-PSK is $\pi/4$ -differential QPSK ($\pi/4$ -DQPSK) [3]. This technique is essentially an 8-PSK format with differential encoding where, from a given phase state, only specified phase shifts of $\pm\pi/4$ or $\pm 3\pi/4$ are allowed.

Continuous phase modulation (CPM) [4] comprises a host of modulation schemes. These formats employ continuous phase trajectories over one or more symbols to get from one phase to the next in response to input changes. CPM schemes are employed in an attempt to simultaneously improve power and bandwidth efficiency.

6.3.2b Signal Analysis

The ideal choice of a modulation technique depends on many factors. Two of the most basic are the *bandwidth efficiency* and *power efficiency*. These parameters are defined as follows:

- Bandwidth efficiency is the ratio of the bit rate to the bandwidth occupied for a digital modulation scheme. Technically, it is dimensionless, but for convenience it is usually given the dimensions of bits/second/hertz.
- Power efficiency is the energy per bit over the noise power spectral density (E_b/N_o) required to provide a given probability of bit error for a digital modulation scheme.

Computation of these parameters is beyond the scope of this chapter. Interested readers are directed to [1] for a detailed discussion of performance parameters.

6-26 Digital Coding of Audio Signals

6.3.3 Digital Coding

Two issues are fundamental in assessing the performance of a digital communication system [5]:

- The reliability of the system in terms of accurately transmitting information from one point to another
- The rate at which information can be transmitted with an acceptable level of reliability

In an ideal communication system, information would be transmitted at an infinite rate with an infinite level of reliability. In reality, however, fundamental limitations affect the performance of any communication system. No physical system is capable of instantaneous response to changes, and the range of frequencies that a system can reliably handle is limited. These real-world considerations lead to the concept of *bandwidth*. In addition, random noise affects any signal being transmitted through any communication medium. Finite bandwidth and additive random noise are two fundamental limitations that prevent designers from achieving an infinite rate of transmission with infinite reliability. Clearly, a compromise is necessary. What makes the situation even more challenging is that the reliability and the rate of information transmission usually work against each other. For a given system, a higher rate of transmission normally means a lower degree of reliability, and vice versa. To favorably affect this balance, it is necessary to improve the efficiency and the robustness of the communication system. *Source coding* and *channel coding* are the means for accomplishing this task.

6.3.3a Source Coding

Most information sources generate signals that contain redundancies [5]. For example, consider a picture that is made up of pixels, each of which represents one of 256 grayness levels. If a fixed coding scheme is used that assigns 8 binary digits to each pixel, a 100×100 picture of random patterns and a 100×100 picture that consists of only white pixels would both be coded into the same number of binary digits, although the white-pixel version would have significantly less information than the random-pattern version.

One simple method of source encoding is the *Huffman* coding technique, which is based on the idea of assigning a code word to each symbol of the source alphabet such that the length of each code word is approximately equal to the amount of information conveyed by that symbol. As a result, symbols with lower probabilities get longer code words. Huffman coding is achieved through the following process:

- List the source symbols in descending order of probabilities.
- Assign a binary 0 and a binary 1, respectively, to the last two symbols in the list.
- Combine the last two symbols in the list into a new symbol with its probability equal to the sum of two symbol probabilities.
- Reorder the list, and continue in this manner until only one symbol is left.
- Trace the binary assignments in reverse order to obtain the code word for each symbol.

A tree diagram for decoding a coded sequence of symbols is shown in Figure 6.3.2. It can easily be verified that the *entropy* of the source under consideration is 2.3382 bits/symbol, and the average code-word length using Huffman coding is 2.37 bits/symbol.

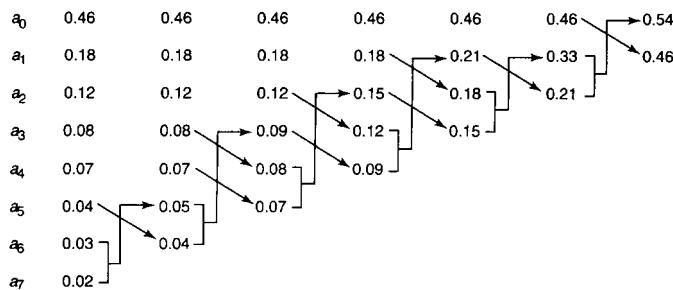


Figure 6.3.2 The Huffman coding algorithm. (From [5]. Used with permission.)

At this point it is appropriate to define *entropy*. In a general sense, entropy is a measure of the disorder or randomness in a closed system. With regard to digital communications, it is defined as a measure of the number of bits necessary to transmit a message as a function of the probability that the message will consist of a specific set of symbols.

6.3.3b Channel Coding

The previous section identified the need for removing redundancies from the message signal to increase efficiency in transmission [5]. From an efficiency point of view, the ideal scenario would be to obtain an average word length that is numerically equal to the entropy of the source. From a practical perspective, however, this would make it impossible to detect or correct errors that may occur during transmission. Some redundancy must be added to the signal in a controlled manner to facilitate detection and correction of transmission errors. This process is referred to as channel coding.

A variety of techniques exist for detection and correction of errors. For the purposes of this chapter, however, it is sufficient to understand that error-correction coding is important to reliable digital transmission and that it adds to the total bit rate of a given information stream. For closed systems, where retransmission of garbled data is possible, a minimum of error-correction overhead is practical. The error-checking *parity* system is a familiar technique. However, for transmission channels where 2-way communication is not possible, or the channel restrictions do not permit retransmission of specific packets of data, robust error correction is a requirement. More information on the basic principles of error correction can be found in [5].

6.3.3c Error-Correction Coding

Digital modulation schemes in their basic form have dependency between signaling elements over only one signaling division [1]. There are advantages, however, to providing memory over several signaling elements from the standpoint of error correction. Historically, this has been accomplished by adding redundant symbols for error correction to the encoded data, and then

6-28 Digital Coding of Audio Signals

using the encoded symbol stream to modulate the carrier. The ratio of information symbols to total encoded symbols is referred to as the *code rate*. At the receiver, demodulation is accomplished followed by decoding.

The drawback to this approach is that redundant symbols are added, requiring a larger transmission bandwidth, assuming the same data throughput. However, the resulting signal is more immune to channel-induced errors resulting from, among other things, a marginal S/N for the channel. The end result for the system is a *coding gain*, defined as the ratio of the signal-to-noise ratios without and with coding.

There are two widely used coding methods:

- *Block coding*, a scheme that encodes the information symbols block-by-block by adding a fixed number of error-correction symbols to a fixed block length of information symbols.
- *Convolutional coding*, a scheme that encodes a sliding window of information symbols by means of a shift register and two or more modulo-2 adders for the bits in the shift register that are sampled to produce the encoded output.

Although an examination of these coding methods is beyond the scope of this chapter, note that coding used in conjunction with modulation always expands the required transmission bandwidth by the inverse of the code rate, assuming the overall bit rate is held constant. In other words, the power efficiency goes up, but the bandwidth efficiency goes down with the use of a well-designed code. Certain techniques have been developed to overcome this limitation, including *trellis-coded modulation* (TCM), which is designed to simultaneously conserve power and bandwidth [6].

6.3.4 Reference

1. Ziemer, Rodger E.: "Digital Modulation," *The Electronics Handbook*, Jerry C. Whitaker (ed.), CRC Press, Boca Raton, Fla., pp. 1213–1236, 1996.
2. Ziemer, R., and W. Tranter: *Principles of Communications: Systems, Modulation, and Noise*, 4th ed., Wiley, New York, 1995.
3. Peterson, R., R. Ziemer, and D. Borth: *Introduction to Spread Spectrum Communications*, Prentice-Hall, Englewood Cliffs, N. J., 1995.
4. Sklar, B.: *Digital Communications: Fundamentals and Applications*, Prentice-Hall, Englewood Cliffs, N. J., 1988.
5. Alkin, Oktay: "Digital Coding Schemes," *The Electronics Handbook*, Jerry C. Whitaker (ed.), CRC Press, Boca Raton, Fla., pp. 1252–1258, 1996.
6. Ungerboeck, G.: "Trellis-Coded Modulation with Redundant Signal Sets," parts I and II, *IEEE Comm. Mag.*, vol. 25 (Feb.), pp. 5-11 and 12-21, 1987.

Chapter

6.4

DSP Devices and Systems

Ken Polmann

6.4.1 Introduction

To efficiently process digital signals, considerable computational power is required. The impressive advancements in the performance of microprocessors intended for personal computer applications have enabled a host of new devices intended for communications systems. For receivers, the most important of these is the *digital signal processor* (DSP), which is a class of processor intended for a specific application or range of applications. The DSP is, in essence, a microprocessor that sacrifices flexibility (or *instruction set*) for speed. There are a number of tradeoffs in DSP design, however, with each new generation of devices, those constraints are minimized while performance is improved.

6.4.2 Fundamentals of Digital Signal Processing¹

Digital signal processing is used to generate, analyze, or otherwise manipulate signals in the digital domain [1]. Digital processing of acquired waveforms offers several advantages over processing of continuous-time signals. Fundamentally, the use of unambiguous discrete samples promotes:

- Use of components with lower tolerances.
- Predetermined system accuracy.
- Identically reproducible circuits.
- Theoretically unlimited number of successive operations on a sample.
- Reduced sensitivity to external effects such as noise, temperature, and aging.

1. Portions of this chapter were adapted from: Pohlmann, Ken: *Principles of Digital Audio*, McGraw-Hill, New York, N.Y., 2000. Used with permission.

6-30 Digital Coding of Audio Signals

The programmable nature of discrete-time signals permits changes in function without changes in hardware. Some operations implemented with digital processing are difficult or impossible using analog means. On the other hand, DSP also has certain disadvantages, including:

- The technology always requires power; there is no passive form of DSP circuitry.
- Digital representation of a signal requires a larger bandwidth than the corresponding analog signal.
- DSP technology can be expensive to develop.
- Circuits capable of performing fast computation are required.
- When used for analog applications, A/D and D/A conversion are required.

DSP presents rich possibilities for professional video and audio applications. Error correction, multiplexing, sample rate conversion, speech and music synthesis, data reduction and data compression, filtering, adaptive equalization, dynamic compression and expansion, reverberation, ambience processing, time alignment, mixing and editing, encryption and watermarking, and acoustical analysis can all be performed with digital signal processing.

6.4.2a Discrete Systems

A discrete system is any system that accepts one or more discrete input signals $x(n)$ and produces one or more discrete output signals $y(n)$ in accordance with a set of operating rules [1]. The input and output discrete time signals are represented by a sequence of numbers. If an analog signal $x(t)$ is sampled every T seconds, the discrete time signal is $x(nT)$, where n is an integer. Time can be *normalized* so that the signal is written as $x(n)$.

Linearity and time-invariance are two important criteria for discrete systems. A *linear system* exhibits the property of *superposition*: the response of a linear system to a sum of signals is the sum of the responses to each individual input. That is, the input $x_1(n) + x_2(n)$ yields the output $y_1(n) + y_2(n)$. A linear system exhibits the property of *homogeneity*: the amplitude of the output of a linear system is proportional to that of the input. That is, an input $ax(n)$ yields the output $ay(n)$. Combining these properties, a linear discrete system with the input signal $ax_1(n) + bx_2(n)$ produces an output signal $ay_1(n) + by_2(n)$ where a and b are constants. The input signals are treated independently, output amplitude is proportional to that of the input, and no new signal components are introduced. As described in the following paragraphs, all z -transforms and Fourier transforms are linear.

A discrete time system is time-invariant if the input signal $x(n - k)$ produces an output signal $y(n - k)$ where k is an integer. In other words, a *linear time-invariant discrete* (LTD) system behaves the same way at all times. For example, an input delayed by k samples generates an output delayed by k samples.

A discrete system is *causal* if at any instant the output signal corresponding to any input signal is independent of the values of the input signal after that instant. In other words, there are no output values before there has been an input signal. The output does not depend on future inputs.

6.4.2b Impulse Response and Convolution

The impulse response $h(t)$ gives a full description of a linear time-invariant discrete system in the time domain [1]. A LTD system, like any discrete system, converts an input signal into an output

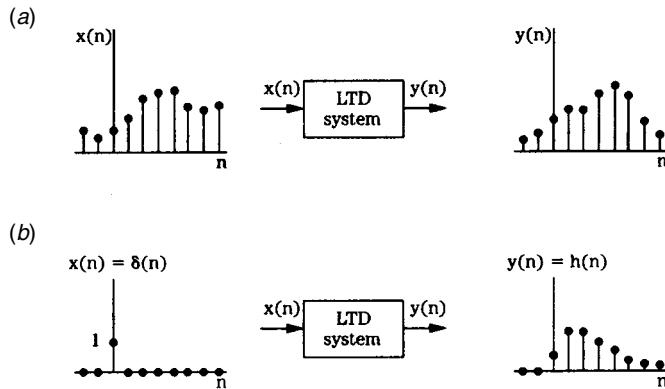


Figure 6.4.1 Two properties of linear-time-invariant discrete (LTD) systems: (a) LTD systems produce an output signal based on the input, (b) an LTD system can be characterized by its impulse response. (*van den Enden and Verhoeckx. From [1]. Used with permission.*)

signal, as shown in Figure 6.4.1a. However, an LTD system has a special property such that when an impulse (a *delta function*) is applied to an LTD system, the output is the system's impulse response, as shown in Figure 6.4.1b. The impulse response describes the system in the time domain, and can be used to reveal the frequency response of the system in the frequency domain. Practically speaking, most digital filters are LTD systems, and yield this property. A system is stable if any input signal of finite amplitude produces an output signal of finite amplitude. In other words, the sum of the absolute value of every input and the impulse response must yield a finite number. Useful discrete systems are stable.

Furthermore, the impulse response can be sampled and used to filter a signal. Audio and video samples themselves are impulses, represented as numbers. The signal could be filtered, for example, by using the samples as *scaling values*: all of the values of a filter's impulse response are multiplied by each signal value. This yields a series of filter impulse responses scaled to each signal sample. To obtain the result, each scaled filter impulse response is substituted for its multiplying signal sample. The filter response can extend over many samples; thus, several scaled values might overlap. When these are added together, the series of sums forms the new filtered signal values.

This is the process of *convolution*. The output of a linear system is the convolution of the input and the system's impulse response. Convolution is a time-domain process that is equivalent to the multiplication of the frequency responses of two networks. Convolution in one domain (such as the time domain) is equivalent to multiplication in the conjugate domain (such as frequency). Furthermore, the duality exists such that multiplication in the time domain is equivalent to convolution in the frequency domain.

Because convolution is not an intuitive phenomenon, a graphical illustration of its nature is useful. Consider the waveform in Figure 6.4.2a. It can be divided into discrete pieces such that $x(t) = x_1(t) + x_2(t) + x_3(t) + ..$

In other words

6-32 Digital Coding of Audio Signals

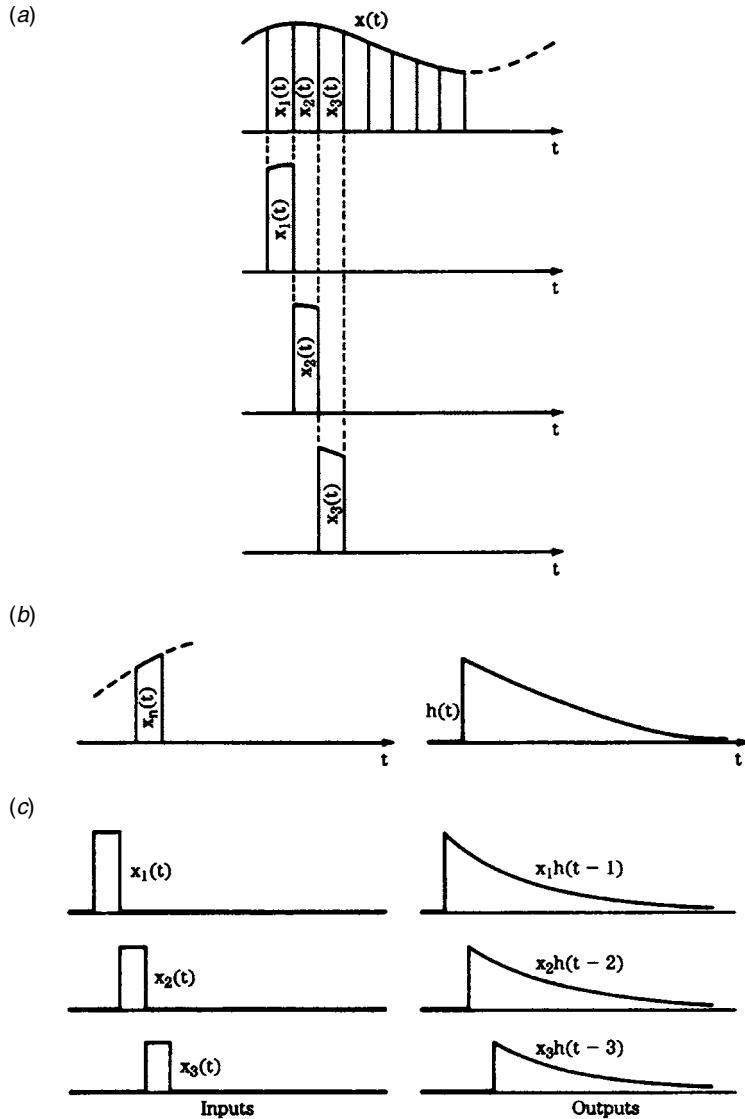


Figure 6.4.2 A graphical representation of convolution: (a) the samples comprising a discrete signal may be considered singly, (b) when applied to a discrete processing system such as a digital filter, each sample produces an output response, (c) the overall response is the summation of the individual responses. (*Blesser. From [1]. Used with permission.*)

$$x(t) = \sum_{k=-\infty}^{\infty} x_k(t) \quad (6.4.1)$$

where $k = 1, 2, 3, \dots$

Consider a network that produces an output $h(t)$ when a single piece of the waveform is input, as shown in Figure 6.4.2b. The output $h(t)$ defines the network; from this single response we can find the network's response to any input. Similarly, the inputs that follow produce outputs that are scaled and delayed by the delay of the input, as shown in Figure 6.4.3c. The sum of the individual responses is the full response to the input waveform:

$$y(n) = \sum_{k=1}^{\infty} h(k)x(n-k) \quad (6.4.2)$$

Equivalently,

$$y(n) = \sum_{k=1}^{\infty} x(k)h(n-k) \quad (6.4.3)$$

This is *convolution*, mathematically expressed as:

$$y(n) = h(n)*x(n) \text{ where } * \text{ denotes convolution} \quad (6.4.4)$$

To view convolution in action, consider a series of snapshots of the terms present at five consecutive sample times

$$\begin{bmatrix} t = 0T & t = 1T & t = 2T & t = 3T & t = 4T \\ x_0h_0 & x_0h_1 & x_0h_2 & x_0h_3 & x_0h_4 \\ x_1h_0 & x_1h_1 & x_1h_2 & x_1h_3 & \\ x_2h_0 & x_2h_1 & x_2h_2 & & \\ x_3h_0 & x_3h_1 & & & \\ x_4h_0 & & & & \end{bmatrix} \quad (6.4.5)$$

The response is the sum of the terms in each column:

6-34 Digital Coding of Audio Signals

$$\left[\begin{array}{l} y_0 = x_0 h_0 \\ y_1 = x_0 h_1 + x_1 h_0 \\ y_2 = x_0 h_2 + x_1 h_1 + x_2 h_0 \\ y_3 = x_0 h_3 + x_1 h_2 + x_2 h_1 + x_3 h_0 \\ y_4 = x_0 h_4 + x_1 h_3 + x_2 h_2 + x_3 h_1 + x_4 h_0 \end{array} \right] \quad (6.4.6)$$

The convolved response is found by reversing the impulse response, and aligning h_0 with the current x_k sample to generate the ordered weighted product. The rest of the sequence is obtained by moving the reversed impulse response until it has passed through the duration of the samples of interest, be it finite or infinite in length.

More generally, when two waveforms are multiplied together, their spectra are convolved, and if two spectra are multiplied, their determining waveforms are multiplied. The response to any input waveform can be determined from the impulse response of the network, and its response to any part of the input waveform. As noted, the convolution of two signals in the time domain corresponds to multiplication of their Fourier transforms in the frequency domain (as well as the dual correspondence). It is apparent, then, that any signal can be considered to be a sum of impulses.

6.4.2c Complex Numbers

Analog and digital networks share a common mathematical basis [1]. Fundamentally, whether the discussion is one of resistors, capacitors, and inductors, or scaling, delay, and addition (all linear, time-invariant elements), processors can be understood through complex numbers. A complex number z is any number that can be written in the form $z = x + jy$ where x and y are real numbers, and where x is the real part, and jy is the imaginary part of the complex number. An imaginary number is any real number multiplied by j , where j is the square root of -1 . The form $x + jy$ is the *rectangular form* of a complex number, and represents the two-dimensional aspects of numbers. For example, the real part can denote distance, and the imaginary part can denote direction. A vector can thus be constructed, showing the indicated location.

A waveform can also be described by a complex number. This is often expressed in polar form, with two parameters: r and θ . The form $re^{j\theta}$ also can be used. If a dot is placed on a circle and rotated, perhaps representing a waveform changing over time, the dot's location can be expressed by a complex number. A location of 450 would be expressed as $0.707 + 0.707j$. A location of 900 would be $0 + 1j$, 135° would be $-0.707 + 0.707j$, and 180° would be $-1 + 0j$. The size of the circle could be used to indicate the magnitude of the number.

The j operator can also be used to convert between imaginary and real numbers. A real number multiplied by an imaginary number becomes complex, and an imaginary number multiplied by an imaginary number becomes real. Multiplication by a complex number is analogous to phase shifting. For example, multiplication by j represents a 90° phase shift, and multiplication by $0.707 + 0.707j$ represents a 45° phase shift. In the digital domain, phase shift is performed by time delay. A digital network comprised of delays can be analyzed by changing each delay to a phase shift. For example, a delay of 10° corresponds to the complex number $0.984 - 0.174j$. If the input signal is multiplied by this complex number, the output result would be a signal of the same magnitude, but delayed by 10° .

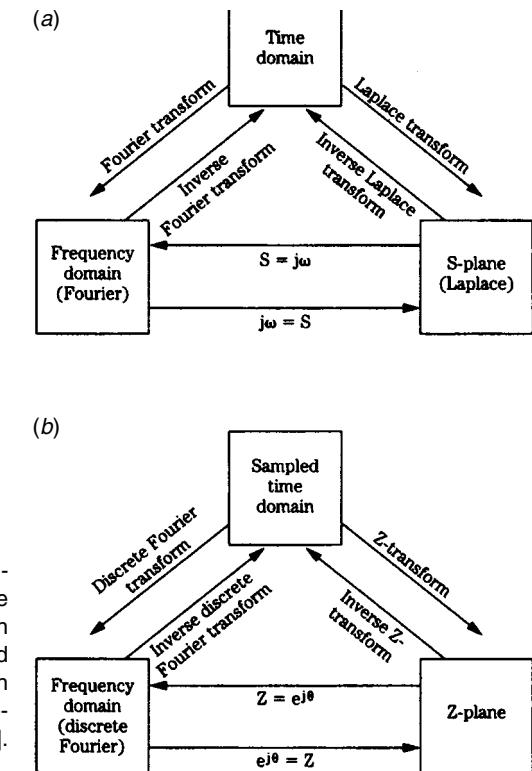


Figure 6.4.3 Transforms are used to mathematically convert a signal from one domain to another: (a) analog signals can be expressed in the time, frequency, and s-plane domains, (b) discrete signals can be expressed in the sampled-time, frequency, and z-plane domains. (From [1]. Used with permission.)

6.4.2d Mathematical Transforms

Signal processing, either analog or digital, can be considered in either of two domains [1]. Together, they offer two perspectives on a unified theory. For analog signals, the domains are time and frequency. For sampled signals, they are discrete time and discrete frequency. A transform is a mathematical tool used to move between the time and frequency domains. Continuous transforms are used with signals continuous in time and frequency; series transforms are applied to continuous time, discrete frequency signals; and discrete transforms are applied to discrete time and frequency signals.

The analog relationships between a continuous signal, its Fourier transform, and Laplace transform are shown in Figure 6.4.3a. The discrete-time relationships between a discrete signal, its discrete Fourier transform, and z-transform are shown in Figure 6.4.3b.

The Laplace transform is used to analyze continuous time and frequency signals; it maps a time domain function $x(t)$ into a frequency domain, complex frequency function $X(s)$. The Laplace transform takes the form

$$X(s) = \int_{-x}^x x(t)e^{-st} dt \quad (6.4.7)$$

6-36 Digital Coding of Audio Signals

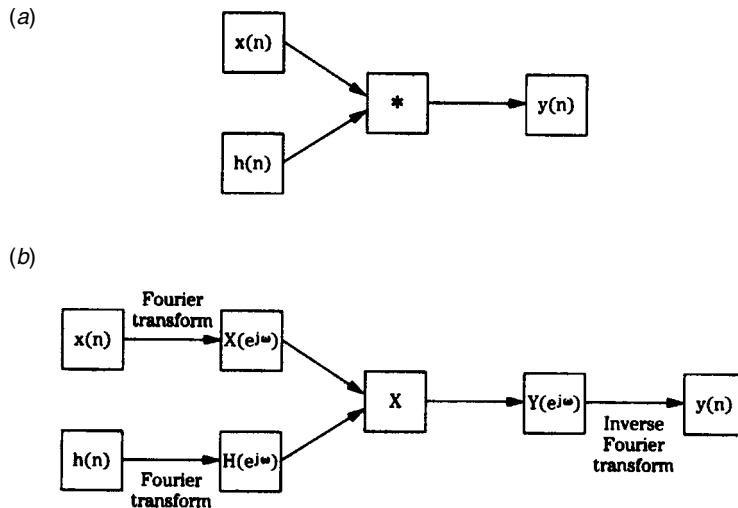


Figure 6.4.4 Given an input signal $x(n)$ and impulse response $h(n)$ the output signal $y(n)$ can be calculated through: (a) direct convolution, or (b) Fourier transformation, multiplication, and inverse Fourier transformation. In practice, the latter method is often an easier calculation. (From [1]. Used with permission.)

The inverse Laplace transform performs the reverse mapping. Laplace transforms are useful for analog design.

The Fourier transform is a special kind of Laplace transform; it maps a time domain function $x(t)$ into a frequency domain function $X(j\omega)$, where $X(j\omega)$ describes the spectrum (frequency response) of the signal $x(t)$. The Fourier transform takes the form

$$X(j\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t} dt \quad (6.4.8)$$

This equation (and the inverse Fourier transform), are identical to the Laplace transforms when $s = j\omega$; the Laplace transform equals the Fourier transform when the real part of s is zero. The Fourier series is a special case of the Fourier transform and results when a signal contains only discrete frequencies, and the signal is periodic in the time domain.

Figure 6.4.4 shows how transforms are used. Specifically, two methods can be used to compute an output signal: convolution in the time domain, and multiplication in the frequency domain. Although convolution is conceptually concise, in practice, the second method using transforms and multiplication in the frequency domain is usually preferable. Transforms also are invaluable in analyzing a signal, to determine its spectral characteristics. In either case, the effect of filtering a discrete signal can be predictably known.

The Fourier transform for discrete signals generates a continuous spectrum but is difficult to compute. Thus, a sampled spectrum for discrete time signals of finite duration is implemented as

the discrete Fourier transform (DFT). Just as the Fourier transform generates the spectrum of a continuous signal, the DFT generates the spectrum of a discrete signal, expressed as a set of harmonically related sinusoids with unique amplitude and phase. The DFT takes samples of a waveform and operates on them as if they were an infinitely long waveform comprised of sinusoids, harmonically related to a fundamental frequency corresponding to the original sample period. An inverse DFT can recover the original sampled signal.

The DFT is the Fourier transform of a sampled signal. When a finite number of samples (N) are considered, the N -point DFT transform is expressed as

$$X(m) = \sum_{n=0}^{N-1} x(n)e^{(-j(2\pi/N))mn} \quad (6.4.9)$$

The $X(m)$ term is often called bin m and describes the amplitude of the frequencies in signal $x(n)$, computed at N equally spaced frequencies. The $m = 0$, or bin 0 term describes the dc content of the signal, and all other frequencies are all harmonically related to the fundamental frequency corresponding to $m = 1$, or bin 1. Bin numbers thus specify the harmonics that comprise the signal, and the amplitude in each bin describes the power spectrum (square of the amplitude). The DFT thus describes all the frequencies contained in signal $x(n)$. There are identical positive and negative frequencies; usually only the positive half is shown, and multiplied by 2 to obtain the actual amplitudes.

An example of DFT operations is shown in Figure 6.4.5. The input signal to be analyzed is a simple periodic function $x(n) = \cos(2\pi n/6)$. The function is periodic over six samples because $x(n) = x(n + 6)$. Three N -point DFTs are used, with $N = 6, 12$, and 16 . In the first two cases, N is equal to 6 or is an integer multiple of 6; a larger N yields greater spectral resolution. In the third case, $N = 16$, the discrete spectrum positions cannot exactly represent the input signal; spectral leakage occurs in all bins. In all cases, the spectrum is symmetrical.

The DFT is computation-intensive, requiring N^2 complex multiplications and $N(N - 1)$ complex additions. The DFT is often generated with the fast Fourier transform (FFT), a collection of fast and efficient algorithms for spectral computation that takes advantage of computational symmetries and redundancies in the DFT; it requires $N \log_2 N$ computations, 100 times fewer than DFT. The FFT can only be used when N is an integral power of 2; zero samples can be padded to satisfy this requirement. The FFT is not another type of transformation, but rather an efficient method of calculating the DFT. In general, a number of short length DFTs are calculated, then the results are combined. The FFT can be applied to various calculation methods and strategies, including analysis of signals and filter design.

The FFT will transform a time series, such as the impulse response of a network, into the real and imaginary parts of the impulse response in the frequency domain. In this way, the magnitude and phase of the network's transfer function can be obtained. An inverse FFT can produce a time domain signal. FFT filtering is accomplished through multiplication of spectra. The impulse response of the filter is transformed to the frequency domain. Real and imaginary arrays, obtained by FFT transformation of overlapping segments of the signal, are multiplied by filter arrays, and an inverse FFT produces a filtered signal. Because the FFT can be efficiently computed, it can be used as an alternative to time domain convolution if the overall number of multiplications is fewer.

6-38 Digital Coding of Audio Signals

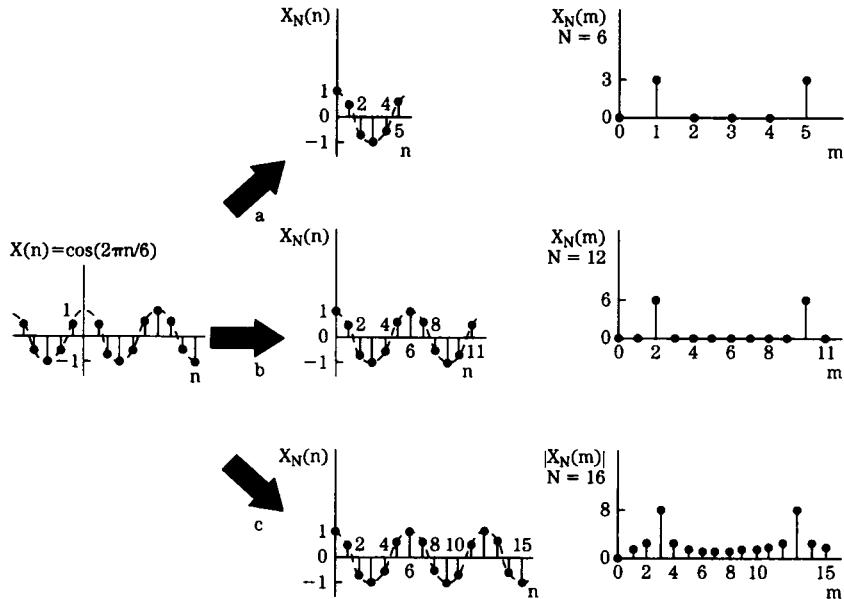


Figure 6.4.5 Examples of a periodic signal applied to an N -point DFT for three different values of N . Greater resolution is obtained as N is increased. When N is not equal to an integral number of waveform periods, spectral leakage occurs. (van den Enden and Verhoeckx. From [1]. Used with permission.)

The z -transform operates on discrete signals in the same way that the Laplace transform operates on continuous signals. In the same way that the Laplace transform is a generalization of the Fourier transform, the z -transform is a generalization of the DFT. Whereas the Fourier transform operates on a particular complex value e^{-jw} the z -transform operates with any complex value. When $z = e^{jw}$, the z -transform is identical to the Fourier transform. The DFT is thus a special case of the z -transform. The z -transform of a sequence $x(n)$ is defined as

$$X(z) = \sum_{n=-\infty}^{\infty} x(n)z^{-n} \quad (6.4.10)$$

where z is a complex variable and z^{-1} represents a unit delay element. The z -transform has an inverse transform, often obtained through partial fraction expansion.

Whereas the DFT is used for literal operations, the z -transform is a mathematical tool used in digital signal processing theory. Convolution in the time domain is equivalent to multiplication in the z -domain. For example, we could take the z -transform of the convolution equation, such that the z -transform of an input multiplied by the z -transform of a filter's impulse response is equal to the z -transform of the filter output. In other words, the ratio of the filter output transform to the filter input transform—that is, the *transfer function* $H(z)$ —is the z -transform of the impulse

response. Furthermore, this ratio, the transfer function $H(z)$, is a fixed function determined by the filter. In the z -domain, given an impulse input, the transfer function equals the output.

6.4.2e Unit Circle and Region of Convergence

The Fourier transform of a discrete signal corresponds to the z -transform on the unit circle in the z -plane. The equation $z = e^{j\omega}$ defines the *unit circle* in the complex plane. The evaluation of the z -transform along the unit circle yields the frequency response of the function.

The variable z is complex, and $X(z)$ is the function of the complex variable. The set of z in the complex plane for which the magnitude of $X(z)$ is finite is said to be in the *region of convergence*. The set of z in the complex plane for which the magnitude of $X(z)$ is infinite is said to diverge, and is outside the region of convergence. The function $X(z)$ is defined over the entire z -plane but is only valid in the region of convergence. The complex variable s is used to describe complex frequency; this is a function of the Laplace transform. The s variable lies on the complex s -plane. The s -plane can be mapped to the z -plane; vertical lines on the s -plane map as circles in the z -plane.

Because there is a finite number of samples, practical systems must be designed within the region of convergence. The unit circle is the smallest region in the z -plane that falls within the region of convergence for all finite stable sequences. Poles must be placed inside the unit circle on the z -plane for proper stability. Improper placement of the poles constitutes an instability.

Mapping from the s -plane to the z -plane is an important process. Theoretically, this function allows the designer to choose an analog transfer function and find the z -transform of that function. Unfortunately, the s -plane generally does not map into the unit circle of the z -plane. Stable analog filters, for example, do not always map into stable digital filters. This is avoided by multiplying by a transform constant, used to match analog and digital frequency response. There also is a nonlinear relationship between analog and digital break frequencies that must be accounted for. The nonlinearities are known as *warping effects* and the use of the constant is known as *pre-warping* the transfer function.

Often, a digital implementation can be derived from an existing analog representation. For example, a stable analog filter can be described by the system function $H(s)$. Its frequency response is found by evaluating $H(s)$ at points on the imaginary axis of the s -plane. In the function $H(s)$, s can be replaced by a rational function of z , which will map the imaginary axis of the s -plane onto the unit circle of the z -plane. The resulting system function $H(z)$ is evaluated along the unit circle and will take on the same values of $H(s)$ evaluated along its imaginary axis.

6.4.2f Poles and Zeros

Summarizing, the transfer function $H(z)$ of a linear, time-invariant discrete-time filter is defined to be the z -transform of the impulse response $h(n)$. The spectrum of a function is equal to the z -transform evaluated on the unit circle. The transfer function of a digital filter can be written in terms of its z transform; this permits analysis in terms of the filter's poles and zeros. The zeros are the roots of the numerator's polynomial of the transfer function of the filter, and the poles are the denominator's roots. Mathematically, zeros make $H(z) = 0$, and poles make $H(z)$ nonanalytic. When the magnitude of $H(z)$ is plotted as a function of z , poles appear at a distance above the z -plane and zeros touch the z -plane. One might imagine the flat z -plane and above it a flexible contour—the *magnitude transfer function*—passing through the poles and zeros, with peaks on top

6-40 Digital Coding of Audio Signals

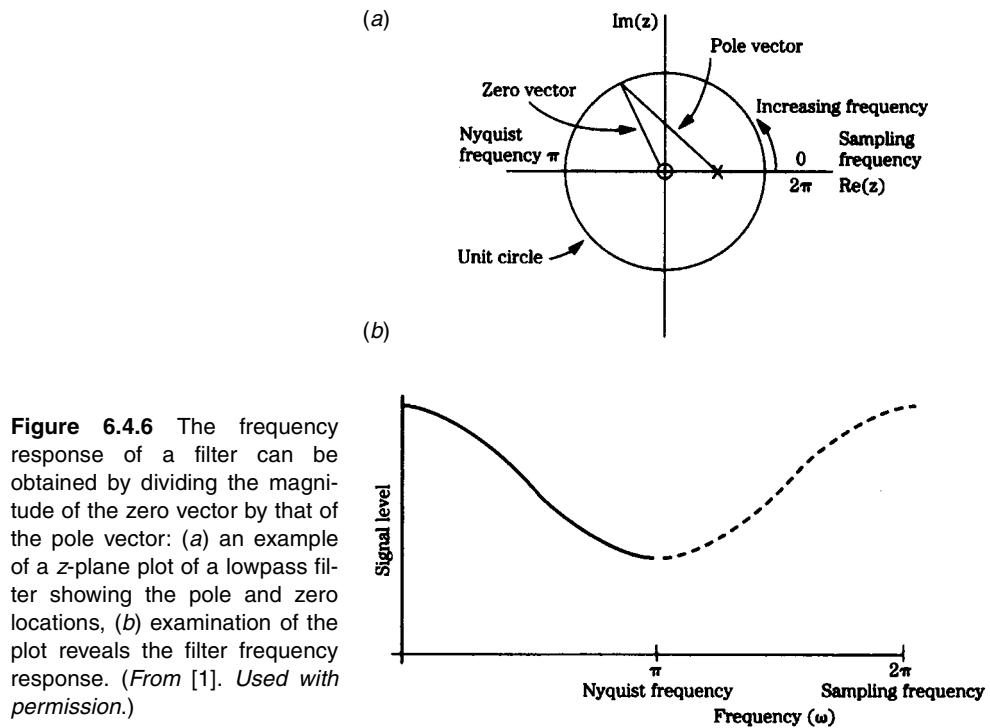


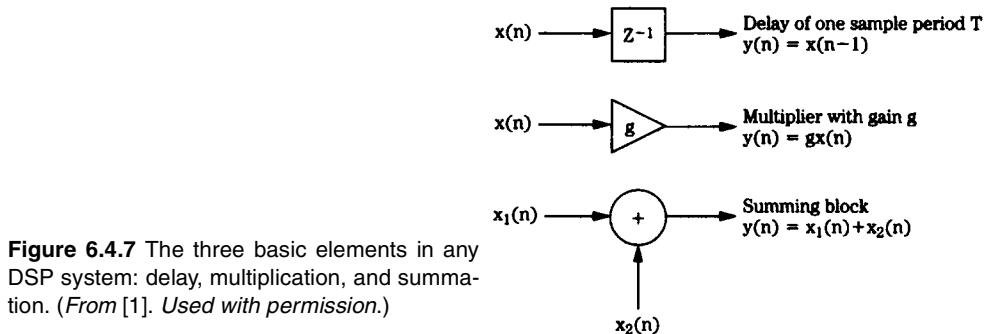
Figure 6.4.6 The frequency response of a filter can be obtained by dividing the magnitude of the zero vector by that of the pole vector: (a) an example of a z -plane plot of a lowpass filter showing the pole and zero locations, (b) examination of the plot reveals the filter frequency response. (From [1]. Used with permission.)

of poles, and valleys centered on zeros. Tracing the rising and falling of the contour around the unit circle yields the frequency response. For example, the gain of a filter at any frequency can be measured by the magnitude of the contour. The phase shift at any frequency is the angle of the complex number that represents the system response at that frequency.

If we plot $|z| = 1$ on the complex plane, we get the unit circle; $|z| > 1$ specifies all points on the complex plane that lie outside the unit circle; and $|z| < 1$ specifies all points inside it. The z -transform of a sequence can be represented by plotting the locations of the poles and zeros on the complex plane.

Figure 6.4.6a shows an example of a z -plane plot. Among other approaches, the response can be analyzed by examining the relationships between the pole and zero vectors. In the z -plane, angular frequency is represented as an angle, with a rotation of 360° corresponding to the sampling frequency. The Nyquist frequency is thus located at π in the figure. The example shows a single pole (X) and zero (0). The amplitude of the frequency response can be determined by dividing the magnitude of the zero vector by that of the pole vector. The frequency response from 0 to the Nyquist frequency is seen to be that of a lowpass filter, as shown in Figure 6.4.6b. Similarly, the phase response can be determined by subtracting the argument of the pole vector from that of the zero vector. As the positions of the pole and zero are varied, the response of the filter changes. For example, if the pole is moved along the negative real axis, the filter's response changes to that of a highpass filter.

Zeros are created by summing input samples, and poles are created by feedback. The order of a filter is equal to the number of poles or zeros it exhibits, whichever is greater. A filter is stable



only if all its poles are inside the unit circle of the z -plane. Zeros can lie anywhere. When all zeros lie inside the unit circle, the system is called a *minimum-phase network*. If all poles are inside the unit circle and all zeros are outside, and if poles and zeros are always reflections of one another in the unit circle, the system is a *constant-amplitude*, or *all-pass network*. If a system has zeros only, except for the origin, and they are reflected in pairs in the unit circle, the system is phase linear. No real function can have more zeros than poles. When the coefficients are real, poles and zeros occur in complex conjugate pairs; their plot is symmetrical across the real z -axis. The closer its location to the unit circle, the greater the effect of each pole and zero on frequency response.

6.4.3 DSP Elements

Successful DSP applications require sophisticated hardware and software [1]. However, all DSP processing can be considered in three basic processing operations (Figure 6.4.7):

- **Summing**, where multiple digital values are added to produce a single result.
- **Multiplication**, where a gain change is accomplished by multiplying the sample value by a coefficient.
- **Time delay**, where a digital value is stored for one sample period ($n - 1$).

The delay element (realized with shift registers or memory locations) is alternatively notated as z^{-1} because a delay of one sampling period in the time domain corresponds to multiplication by z^{-1} in the z -domain; thus $z^{-1}x(n) = x(n - 1)$. Delays can be cascaded, for example, a z^{-2} term describes a two-sample ($n - 2$) delay. Although it is usually most convenient to operate with sample numbers, the time of a delay can be obtained by taking nT , where T is the sampling interval. Figure 6.4.8 shows two examples of simple networks and their impulse responses. LTD systems such as these are completely described by the impulse response.

In practice, these elemental operations are performed many times for each sample, in specific configurations depending on the desired result. In this way, algorithms can be devised to perform operations useful to signal processing, such as reverberation, equalization, data compression, limiting, and noise removal. Of course, for real-time operation, all processing for each sample must be completed within one sampling period.

6-42 Digital Coding of Audio Signals

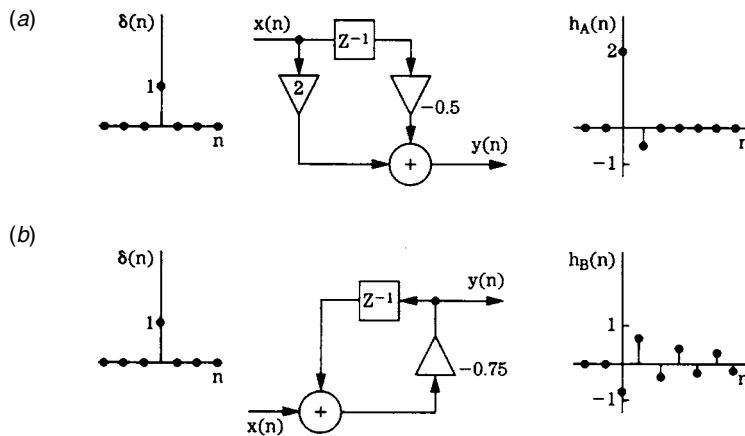


Figure 6.4.8 LTD systems can be characterized by their impulse responses: (a) simple nonrecursive system and its impulse response, (b) simple recursive system and its impulse response. (*van den Enden and Verhoeckx*. From [1]. Used with permission.)

6.4.3a Sources of Errors

In general, errors in digital processors can be classified as coefficient errors, limit cycle errors, overflow, truncation, and round-off errors [1]. Coefficient errors occur when a coefficient is not specified with sufficient accuracy. Limit cycle error might occur when a signal is removed from a filter, leaving a decaying sum. This decay might become zero or might oscillate at a constant amplitude, known as *limit cycle oscillation*. This effect can be eliminated, for example, by offsetting the filter output so that truncation always produces a zero output. Overflow occurs when a register length is exceeded, resulting in a computational error. In the case of wrap-around, when a 1 is added to the maximum value positive two's complement number, the result is the maximum value negative number. In short, the information has overflowed into a nonexistent bit. To prevent this, *saturating arithmetic* can be used so that when the addition of two positive numbers would result in a negative number, the maximum positive sum is substituted instead.

Truncation and round-off errors occur whenever the word length of a computed result is limited. Errors accumulate both inside the processor during calculation, and when word length is reduced for output through a D/A converter. However, A/D conversion always results in quantization error, and computation error can appear in different guises. For example, when two n -bit numbers are multiplied, the number of output bits will be $2n - 1$. Thus, multiplication almost doubles the number of bits required to represent the output. Although many hardware multipliers can perform double precision computation, a finite word length must be maintained following multiplication, thus limiting precision. Discarded data results in an error that is analogous to A/D quantization. To be properly modeled, multiplication must be followed by quantization. Multiplication does not introduce error, but inability to keep the extra bits does.

Rather than truncate a word—for example, following multiplication—the value can be *rounded*; that is, the word is taken to the nearest available value. This results in a peak error of $1/2$ LSB, and a RMS value of $1/(12)^{1/2}$, or 0.288 LSB. This round-off error will accumulate over successive calculations. In general, the number of calculations must be large for significant error.

However, in addition, dither information can be lost during computation. For example, when a properly dithered 16-bit word is input to a 32-bit processor, even though computation is of high precision, the output signal can be truncated to 16 bits for conversion through the output D/A converter. For example, a 16-bit signal that is delayed and scaled by a 12-dB attenuation would result in a 12-bit undithered signal. To overcome this, digital dithering to the resolution of the next processing step should be used in a computation.

6.4.3b DSP Integrated Circuits

A DSP chip is a specialized hardware device that performs digital signal processing under the control of software algorithms [1]. DSP chips are stand-alone processors, often independent of host CPUs (central processing units), and are specially designed for operations used in spectral and numerical applications. For example, large numbers of multiplications are possible, as well as special addressing modes such as bit-reverse and circular addressing. When the memory and input/output circuits are added, the result is an integrated digital signal processor. Such a general-purpose DSP chip is software programmable, and thus can be used for a variety of signal-processing applications. Alternatively, a custom signal processor can be designed to accomplish a specific task.

DSP chips are designed according to two arithmetic types, fixed integer and floating point, which define the format of the data they operate on. A fixed integer chip uses two's complement, binary integer data. Floating point chips use integer and floating point numbers (represented as a mantissa and exponent). The dynamic range of a fixed integer chip is based on its word length; data must be scaled to prevent overflow; this can increase programming complexity. The scientific notation used in a floating point chip allows larger dynamic range, without overflow problems. However, the resolution of floating point representation is limited by the word length of the exponent.

DSP chips often use a pipelining architecture so that several instructions can be paralleled. For example, a fetch (fetch instruction from memory and update program counter), decode (decode instruction and generate operand address), read (read operand from memory), and execute (perform necessary operations) can be effectively executed in one clock cycle with pipelining. A pipeline manager, aided by proficient user programming, helps ensure rapid processing.

DSP chips, like all computers, are comprised of input and output devices, an arithmetic unit, a control unit, and memory, interconnected by buses. All computers originally used a single sequential bus (*von Neumann architecture*), shared by data, memory addresses, and instructions. However, in a DSP chip a particularly large number of operations must be performed quickly for real-time operation. Thus, parallel bus structures are used (such as the *Harvard architecture*) that store data and instructions in separate memories and transfers them via separate buses. For example, a chip can have separate buses for program, data, and DMA, providing parallel program fetches, data reads, as well as DMA operations with slower peripherals.

General purpose DSP chips tend to follow a similar architecture; typical elements include the following:

- Multiply-accumulate unit
- Data address generator
- Data RAM
- Coefficient RAM

6-44 Digital Coding of Audio Signals

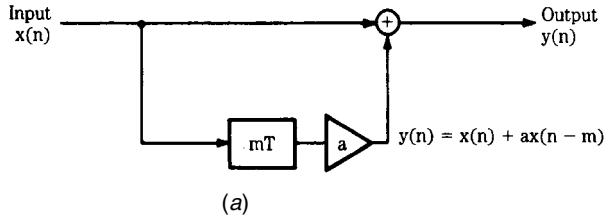
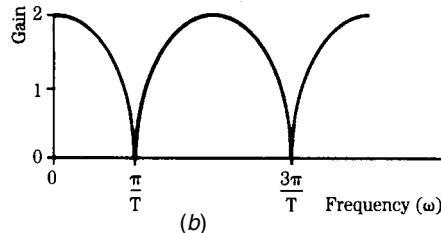


Figure 6.4.9 A delay block can be used to create an echo circuit: (a) the circuit contains an mT delay and gain stage, (b) with shorter delay times, a comb filter response will result. (*Bloom, Berkhout, and Eggermont*. From [1]. Used with permission.)



- Coefficient address generator
- Program control unit
- Program ROM

These components are connected by the following buses:

- The data bus
- Coefficient bus
- Control bus

6.4.3c DSP Applications

In addition to digital filtering, some of the most powerful and creative applications of digital signal processing come in the form of specialized processing of audio and video signals. Building on the basic operations of multiplication and delay, sophisticated operations can be developed. Reverberation perhaps epitomizes the degree of time manipulation possible in the digital domain; it is possible to synthesize reverberation to both simulate natural acoustical environments and to create acoustical environments that could not physically exist.

Digital Delay

A delay block is a simple storage unit, such as a memory location [1]. A sample is placed in memory, stored, then recalled some time later and output. A delay unit can be described by the equation $y(n) = x(n - m)$ where m is the delay in samples. Generally, when the delay is small, the frequency response of the signal is altered; when the delay is longer, an echo results. Just as in filtering, a simple delay can be used to create sophisticated effects. For example, Figure 6.4.9a

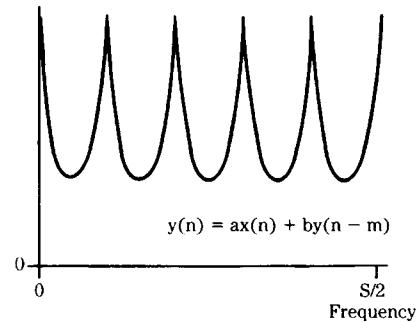


Figure 6.4.10 A recursive comb filter creates a delay with feedback, yielding a toothed frequency response. (From [1]. Used with permission.)

shows an echo circuit using a delay block. Delay mT is of duration m samples, and samples are multiplied by a gain coefficient (a scaling factor) less than unity. If the delay time is set between 10 and 50 ms, an echo results; with shorter fixed delays, a comb filter response results, as shown in Figure 6.4.9b. Peaks and dips are equally spaced through the frequency response from 0 Hz to the Nyquist frequency. The number of peaks depends on the delay time; the longer the delay, the more peaks.

A comb filter can be either recursive or nonrecursive. It cascades a series of delay elements, creating a new response. Mathematically, a nonrecursive comb filter can be designed by adding the input sample to the same sample but delayed

$$y(n) = x(n) + ax(n - m) \quad (6.4.11)$$

where m is the delay time in samples. A recursive comb filter creates a delay with feedback. The delayed signal is attenuated and fed back into the delay

$$y(n) = ax(n) + by(n - m) \quad (6.4.12)$$

This yields a response as shown in Figure 6.4.10. The number of peaks depends on the duration of the delay; the longer the delay, the greater the number of peaks.

An all-pass filter is one that has a flat frequency response from 0 Hz to the Nyquist frequency. However, its phase response causes different frequencies to be delayed by different amounts. An all-pass filter can be described as

$$y(n) = -ax(n) + x(n - 1) + by(n - 1) \quad (6.4.13)$$

If the delay in the foregoing circuits is replaced by a digital all-pass filter or a cascade of all-pass filters, a *phasing effect* is achieved

$$y(n) = -ax(n) + x(n - m) + by(n - m) \quad (6.4.14)$$

The effect becomes more pronounced as the delay increases. The system exhibits nonuniformly spaced notches in its frequency response, varying independently in time.

6-46 Digital Coding of Audio Signals

Table 6.4.1 Characteristics of the TMS320C55x Processors (After [2]. Courtesy of Texas Instruments.)

Parameter	VC5510
Memory	
On-chip SARAM	32 K words (64 K bytes)
On-chip DARAM	128 K words (256 K bytes)
On-chip ROM	16 K words (32 K bytes)
Total addressable memory (internal/external)	8M words (16 M bytes)
On-chip bootloader (in ROM)	Yes
Peripherals	
McBSPs	3
DMA controller	Yes
EHPI (16-bit)	Yes
Configurable instruction cache	24K bytes
Timers	2
Programmable DLL clock generator	Yes
General Purpose I/O Pins	
Dedicated input/output	Yes
XF—dedicated output	1
Multiplexed with McBSP (input/output)	21
Multiplexed with timer (output only)	2
CPU Cycle Time/Speed	
160 MHz (6.25 ns)	Yes
200 MHz (5 ns)	Yes
Package Type	240-pin BGA

6.4.3d Example DSP Device

The functional description of a representative DSP device will help to illustrate the concepts outline previously in this chapter. Table 6.4.1 lists the overall characteristics for members of the TMS320C55x generation of fixed-point digital signal processors (Texas Instruments). Features for the high performance, low-power C55x CPU include the following [2]:

- Advanced multiple-bus architecture with one internal program memory bus and five internal data buses (three dedicated to *reads* and two dedicated to *writes*)
- Unified program/data memory architecture

- Dual 17-bit \times 17-bit multipliers coupled to 40-bit dedicated adders for non-pipelined single-cycle *multiply accumulate* (MAC) operations
- *Compare, select, and store unit* (CSSU) for the add/compare section of the Viterbi operator
- Exponent encoder to compute an exponent value of a 40-bit accumulator value in a single cycle
- Two address generators with eight auxiliary registers and two auxiliary register arithmetic units
- Data buses with bus holders
- $8 \text{ M} \times 16\text{-bit}$ (16 Mbyte) total addressable memory space
- Single-instruction repeat or block repeat operations for program code
- Conditional execution
- Seven-stage pipeline for high instruction throughput
- Instruction buffer unit that loads, parses, queues, and decodes instructions to decouple the program fetch function from the pipeline
- Program flow unit that coordinates program actions among multiple parallel CPU functional units
- Address data flow unit that provides data address generation and includes a 16-bit arithmetic unit capable of performing arithmetic, logical, shift, and saturation operations
- Data computation unit containing the primary computation units of the CPU, including a 40-bit arithmetic logic unit, two multiply-accumulate units, and a shifter

Because many DSP chips are used in portable systems, power consumption is a key operating parameter. As a result, power control features are an important DSP specification. Hardware and software functions designed to conserve power include:

- Software-programmable *idle domains* that provide configurable low-power modes
- Automatic power management
- Advanced low-power CMOS process

Functional Overview

The C55x architecture achieves power-efficient performance through increased parallelism and a focus on reduction in power dissipation. The CPU supports an internal bus structure composed of the following elements [2]:

- One program bus
- Three data read buses
- Two data write buses
- Additional buses dedicated to peripheral and DMA activity

These buses provide the ability to perform up to three data reads and two data writes in a single cycle. In parallel, the DMA controller can perform up to two data transfers per cycle independent

6-48 Digital Coding of Audio Signals

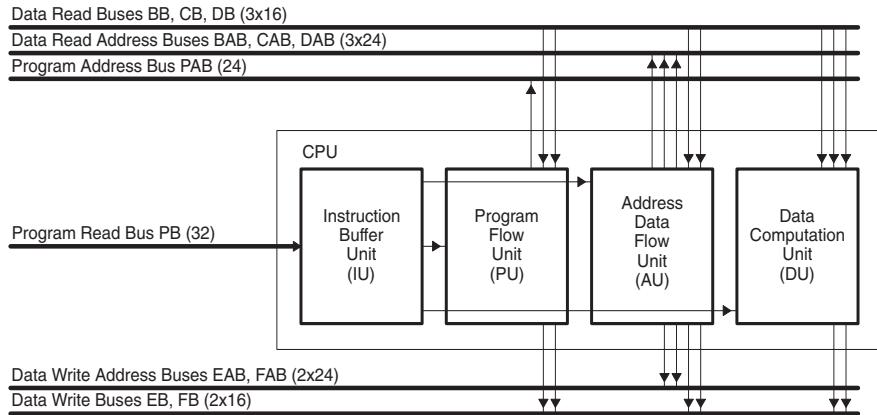


Figure 6.4.11 Functional Block Diagram of the TMS320C55x DSP series. (From [2]. Courtesy of Texas Instruments.)

of CPU activity. The C55x CPU provides two multiply-accumulate units each capable of 17-bit \times 17-bit multiplication in a single cycle. A central 40-bit arithmetic/logic unit (ALU) is supported by an additional 16-bit ALU. Use of ALUs is subject to instruction set control. This programmability provides the capacity to optimize parallel activity and power consumption. These resources are managed in the address data flow unit (AU) and data computation unit (DU) of the C55x CPU.

The C55x architecture supports a variable byte width instruction set for improved code density. The instruction buffer unit (IU) performs 32-bit program fetches from internal or external memory and queues instructions for the program unit (PU). The program unit decodes the instructions, directs tasks to AU and DU resources, and manages the fully-protected pipeline.

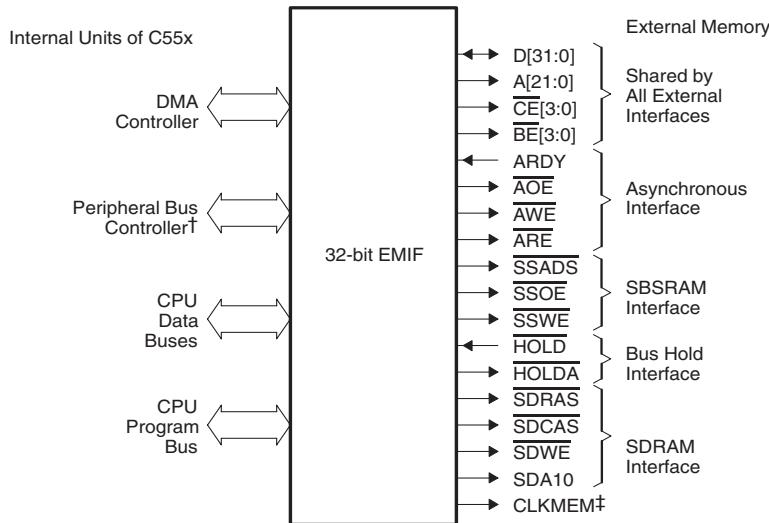
A configurable instruction cache is also available to minimize external memory accesses, improving data throughput and conserving system power.

The C55x architecture is built around four primary blocks:

- The instruction buffer unit
- Program flow unit
- Address data flow unit
- Data computation unit

These functional units exchange program and data information with each other and with memory through multiple dedicated internal buses. Figure 6.4.11 shows the principal blocks and bus structure in the C55x devices.

Program fetches are performed using the 24-bit program address bus (PAB) and the 32-bit program read bus (PB). The functional units read data from memory via three 16-bit data read buses named B-bus (BB), C-bus (CB), and D-bus (DB). Each data read bus also has an associ-



† This connection allows the CPU to access the EMIF registers.

‡ The CLKMEM signal is shared by the SDRAM and SBSRAM interfaces.

Figure 6.4.12 Block diagram of EMIF for the TMS320C55x DSP. (From [2]. Courtesy of Texas Instruments.)

ated 24-bit data read address bus (BAB, CAB, and DAB). Single operand reads are performed on the D-bus. Dual-operand reads use C-bus and D-bus. B-bus provides a third read path and can be used to provide coefficients for dual-multiply operations.

Program and data writes are performed on two 16-bit data write buses called E-bus (EB) and F-bus (FB). The write buses also have associated 24-bit data write address buses (EAB and FAB). Additional buses are present on the C55x devices to provide dedicated service to the DMA controller and the peripheral controller.

All C55x DSP devices use the same CPU structure but are capable of supporting different on-chip peripherals and memory configurations. The on-chip peripherals include:

- Digital phase-locked loop (DPLL) clock generation
- Instruction cache
- External memory interface (EMIF)
- Direct memory access (DMA) controller
- 16-bit enhanced host port interface (EHPI)
- Multichannel serial ports (McBSPs)
- 16-bit timers with 4-bit prescalers
- General-purpose I/O (GPIO) pins
- Trace FIFO (for emulation purposes only)

6-50 Digital Coding of Audio Signals

Peripheral control registers are mapped to an I/O space separate from the main memory space. The peripheral bus controller handles exchange of data between peripherals and the CPU via dedicated peripheral buses.

The EMIF supports a “glueless” interface from the C55x to a variety of external memory devices. For each memory type, the EMIF supports 8-bit, 16-bit, and 32-bit accesses for both reads and writes. For writes, the EMIF controls the byte enable signals and the data bus to perform 8-bit transfers or 16-bit transfers. For reads, the entire 32-bit bus is read. Then, it is internally parsed by the EMIF. The EMIF block diagram, Figure 6.4.12, shows the interface between external memory and the internal resources of the C55x.

6.4.4 References

1. Pohlmann, Ken: *Principles of Digital Audio*, McGraw-Hill, New York, N.Y., 2000.
2. *TMS320C55x DSP Functional Overview*, Texas Instruments, Dallas, TX, literature No. SRPU312, June 2000.

Section
7

Compression Technologies for Audio

A number of existing and proposed audio compression systems employ a combination of processing techniques. Any scheme that becomes widely adopted can enjoy economies of scale and reduced market confusion. Timing, however, is critical to market acceptance of any standard. If a standard is selected well ahead of market demand, more cost-effective or higher-performance approaches may become available before the market takes off. On the other hand, a standard may be merely academic if it is established after alternative schemes already have become well entrenched in the marketplace.

These factors have placed a great deal of importance on the standards-setting activities of leading organizations around the world. It has been through hard work, inspiration, and even a little compromise that the various standards have developed and evolved to the levels of utility and acceptance that they enjoy today.

With these important benchmarks in place, audio industry manufacturers have been able to focus on implementation of the technology and offering specific user-centered features. Fortunately, the days of the video (and audio) tape “format wars” appear to have passed as the standards-setting bodies take the lead in product direction and interface.

The function of any audio compression device or system is to provide for efficient storage and/or transmission of information from one location or device to another. The encoding process, naturally, is the beginning point of this chain. Like any chain, audio encoding represents not just a single link but many interconnected and interdependent links. The bottom line in audio encoding is to ensure that the compressed signal or data stream represents the information required for recording and/or transmission, and *only* that information. If there is additional information of any nature remaining in the data stream, it will take bits to store and/or transmit, which will result in fewer bits being available for the required data. Surplus information is irrelevant because the intended recipient(s) do not require it and can make no use of it.

Surplus information can take many forms. For example, it can be information in the original signal or data stream that exceeds the capabilities of the receiving device to process and reproduce. There is little point in transmitting more information than the receiving device can use.

Noise is another form of surplus information. Noise is—by nature—random or nearly so, and this makes it essentially incompressible. Many other types of artifacts exist, ranging from filter ringing to disc scratches. Some may seem trivial, but in the field of compression they can be very important. Compression relies on order and consistency for best performance, and such artifacts

7-2 Section Seven

can compromise the final reproduction or at least lower the achievable bit rate reduction. Generally speaking, compression systems are designed for particular tasks, and make use of certain basic assumptions about the nature of the data being compressed.

Such requirements have brought about a true “systems approach” to compression. From the algorithm to the input audio, every step must be taken with care and precision for the overall product to be of high quality.

These forces are shaping the audio and video technologies of tomorrow. Any number of scenarios have been postulated as to the hardware and software that will drive the digital facility of the future. One thing is certain, however: It will revolve around compressed audio and video signals.

In This Section:

Chapter 7.1: Audio Compression Systems	7-7
Introduction	7-7
PCM Versus Compression	7-7
Audio Bit Rate Reduction	7-8
Redundancy and Irrelevancy	7-8
Human Auditory System	7-8
Quantization	7-9
Sampling Frequency and Bit Rate	7-10
Prediction and Transform Algorithms	7-11
Subband Coding	7-11
Subband Gain	7-11
APCM Coding	7-12
Processing and Propagation Delay	7-13
Bit Rate and Compression Ratio	7-14
Bit Rate Mismatch	7-14
Editing Compressed Data	7-14
Common Audio Compression Techniques	7-14
apt-X100	7-15
ISO/MPEG-1 Layer 2	7-16
MPEG-2 AAC	7-20
MPEG-4	7-21
Dolby E Coding System	7-23
Architectural Overview	7-25
Coded Frame Format	7-25
Objective Quality Measurements	7-26
Perspective on Audio Compression	7-28
References	7-29
Bibliography	7-30
Chapter 7.2: ATSC DTV System Compression Issues	7-31
Introduction	7-31
MPEG-2 Layer Structure	7-31
Slices	7-32
Pictures, Groups of Pictures, and Sequences	7-33

Compression Technologies for Audio 7-3

I-Frames	7-34
P-Frames	7-34
B-Frames	7-34
Motion Estimation	7-35
Vector Search Algorithm	7-35
Motion-Vector Precision	7-35
Motion-Vector Coding	7-35
Encoder Prediction Loop	7-35
Spatial Transform Block—DCT	7-37
Quantizer	7-37
Entropy Coder	7-38
Inverse Quantizer	7-39
Inverse Spatial Transform Block—IDCT	7-39
Motion Compensator	7-39
Dual Prime Prediction Mode	7-39
Adaptive Field/Frame Prediction Mode	7-40
Image Refresh	7-40
Periodic Transmission of I-Frames	7-40
Progressive Refresh	7-41
Discrete Cosine Transform	7-41
Blocks of 8×8 Pixels	7-41
Adaptive Field/Frame DCT	7-42
Adaptive Quantization	7-42
Perceptual Weighting	7-42
Entropy Coding of Video Data	7-43
Huffman Coding	7-43
Run Length Coding	7-44
Channel Buffer	7-45
Decoder Block Diagram	7-45
Spatial and S/N Scalability	7-45
References	7-46

Chapter 7.3: DTV Audio Encoding and Decoding 7-47

Introduction	7-47
AES Audio	7-47
AES3 Data Format	7-49
SMPTE 324M	7-50
Audio Compression	7-50
Encoding	7-51
Decoding	7-53
Implementation of the AC-3 System	7-53
Audio-Encoder Interface	7-54
Sampling Parameters	7-55
Output Signal Specification	7-55
Operational Details of the AC-3 Standard	7-56
Transform Filterbank	7-57
Window Function	7-57
Time-Division Aliasing Cancellation Transform	7-57

7-4 Section Seven

Transient Handling	7-57
Coded Audio Representation	7-57
Exponent Coding	7-58
Mantissas	7-58
Bit Allocation	7-58
Forward Adaptive	7-59
Rematrixing	7-59
Coupling	7-59
Bit Stream Elements and Syntax	7-60
Splicing and Insertion	7-60
Error-Detection Codes	7-61
Loudness and Dynamic Range	7-61
Dynamic Range Compression	7-61
Encoding the AC-3 Bit Stream	7-61
Input Word Length/Sample Rate	7-62
AC-3/MPEG Bit Stream	7-62
Decoding the AC-3 Bit Stream	7-64
Continuous or Burst Input	7-64
Synchronization and Error Detection	7-64
Decoding Components	7-66
Algorithmic Details	7-67
Bit Allocation	7-67
Audio System Level Control	7-68
Dialogue Normalization	7-68
Example Situation	7-69
Dynamic Range Compression	7-69
Example Situation	7-70
Heavy Compression	7-71
Audio System Features	7-72
Complete Main Audio Service (CM)	7-72
Main Audio Service, Music and Effects (ME)	7-72
Visually Impaired (VI)	7-73
Hearing Impaired (HI)	7-73
Dialogue (D)	7-73
Commentary (C)	7-74
Emergency (E)	7-74
Voice-Over (V0)	7-74
Multilingual Services	7-75
Channel Assignments and Levels	7-75
References	7-77
Bibliography	7-78

On the CD-ROM:

- ATSC, "Digital Audio Compression Standard (AC-3)," Advanced Television Systems Committee, Washington, D.C., Doc. A/52, Dec. 20, 1995.
- ATSC, "Digital Television Standard," Advanced Television Systems Committee, Washington, D.C., Doc. A/53, Sept. 16, 1995.
- ATSC, "Guide to the Use of the Digital Television Standard," Advanced Television Systems Committee, Washington, D.C., Doc. A/54, Oct. 4, 1995.
- The Tektronix publication *A Guide to MPEG Fundamentals and Protocol Analysis* as an Acrobat (PDF) file. This document, copyright 1997 by Tektronix, provides a detailed discussion of MPEG as applied to DTV and DVB, and quality analysis requirements and measurement techniques for MPEG-based systems.

Reference Documents for this Section

- Brandenburg, K., and Gerhard Stoll: "ISO-MPEG-1 Audio: A Generic Standard for Coding of High Quality Digital Audio," *92nd AES Convention Proceedings*, Audio Engineering Society, New York, N.Y., 1992, revised 1994.
- Ehmer, R. H.: "Masking Patterns of Tones," *J. Acoust. Soc. Am.*, vol. 31, pp. 1115–1120, August 1959.
- Fibush, David K.: "Testing MPEG-Compressed Signals," *Broadcast Engineering*, Overland Park, Kan., pp. 76–86, February 1996.
- Herre, J., and B. Grill: "MPEG-4 Audio—Overview and Perspectives for the Broadcaster," *IBC 2000 Proceedings*, International Broadcast Convention, Amsterdam, September 2000.
- IEEE Standard Dictionary of Electrical and Electronics Terms*, ANSI/IEEE Standard 100-1984, Institute of Electrical and Electronics Engineers, New York, 1984.
- ITU-R Recommendation BS-775, "Multi-channel Stereophonic Sound System with and Without Accompanying Picture."
- Lyman, Stephen, "A Multichannel Audio Infrastructure Based on Dolby E Coding," *Proceedings of the NAB Broadcast Engineering Conference*, National Association of Broadcasters, Washington, D.C., 1999.
- Moore, B. C. J., and B. R. Glasberg: "Formulae Describing Frequency Selectivity as a Function of Frequency and Level, and Their Use in Calculating Excitation Patterns," *Hearing Research*, vol. 28, pp. 209–225, 1987.
- Robin, Michael, and Michel Poulin: *Digital Television Fundamentals*, McGraw-Hill, New York, N.Y., 1998.
- SMPTE Standard for Television: "12-Channel Serial Interface for Digital Audio and Auxiliary Data," SMPTE 324M (Proposed), SMPTE, White Plains, N.Y., 1999.
- SMPTE Standard for Television: "Channel Assignments and Levels on Multichannel Audio Media," SMPTE 320M-1999, SMPTE, White Plains, N.Y., 1999.

7-6 Section Seven

- Smyth, Stephen: "Digital Audio Data Compression," *Broadcast Engineering*, Intertec Publishing, Overland Park, Kan., February 1992.
- Terry, K. B., and S. B. Lyman: "Dolby E—A New Audio Distribution Format for Digital Broadcast Applications," *International Broadcasting Convention Proceedings*, IBC, London, England, pp. 204–209, September 1999.
- Todd, C., et. al.: "AC-3: Flexible Perceptual Coding for Audio Transmission and Storage," AES 96th Convention, Preprint 3796, Audio Engineering Society, New York, February 1994.
- Vernon, S., and T. Spath: "Carrying Multichannel Audio in a Stereo Production and Distribution Infrastructure," *Proceedings of IBC 2000*, International Broadcasting Convention, Amsterdam, September 2000.
- Wylie, Fred: "Audio Compression Techniques," *The Electronics Handbook*, Jerry C. Whitaker (ed.), CRC Press, Boca Raton, Fla., pp. 1260–1272, 1996.
- Wylie, Fred: "Audio Compression Technologies," *NAB Engineering Handbook*, 9th ed., Jerry C. Whitaker (ed.), National Association of Broadcasters, Washington, D.C., 1998.
- Zwicker, E.: "Subdivision of the Audible Frequency Range Into Critical Bands (Frequenzgruppen)," *J. Acoust. Soc. of Am.*, vol. 33, p. 248, February 1961.

Chapter
7.1
Audio Compression Systems

Fred Wylie

Jerry C. Whitaker, Editor-in-Chief

7.1.1 Introduction

As with video, high on the list of priorities for the professional audio industry is to refine and extend the range of digital equipment capable of the capture, storage, post production, exchange, distribution, and transmission of high-quality audio, be it mono, stereo, or 5.1 channel AC-3 [1]. This demand being driven by end-users, broadcasters, film makers, and the recording industry alike, who are moving rapidly towards a “tapeless” environment. Over the last two decades, there have been continuing advances in DSP technology, which have supported research engineers in their endeavors to produce the necessary hardware, particularly in the field of digital audio data compression or—as it is often referred to—*bit-rate reduction*. There exist a number of real-time or—in reality—near instantaneous compression coding algorithms. These can significantly lower the circuit bandwidth and storage requirements for the transmission, distribution, and exchange of high-quality audio.

The introduction in 1983 of the compact disc (CD) digital audio format set a quality benchmark that the manufacturers of subsequent professional audio equipment strive to match or improve. The discerning consumer now expects the same quality from radio and television receivers. This leaves the broadcaster with an enormous challenge.

7.1.1a PCM Versus Compression

It can be an expensive and complex technical exercise to fully implement a linear *pulse code modulation* (PCM) infrastructure, except over very short distances and within studio areas [1]. To demonstrate the advantages of distributing compressed digital audio over wireless or wired systems and networks, consider again the CD format as a reference. The CD is a 16 bit linear PCM process, but has one major handicap: the amount of circuit bandwidth the digital signal occupies in a transmission system. A stereo CD transfers information (data) at 1.411 Mbits/s, which would require a circuit with a bandwidth of approximately 700 kHz to avoid distortion of the digital signal. In practice, additional bits are added to the signal for channel coding, synchro-

7-8 Compression Technologies for Audio

nization, and error correction; this increases the bandwidth demands yet again. 1.5 MHz is the commonly quoted bandwidth figure for a circuit capable of carrying a CD or similarly coded linear PCM digital stereo signal. This can be compared with the 20 kHz needed for each of two circuits to distribute the same stereo audio in the analog format, a 75-fold increase in bandwidth requirements.

7.1.1b Audio Bit Rate Reduction

In general, analog audio transmission requires fixed input and output bandwidths [2]. This condition implies that in a real-time compression system, the quality, bandwidth, and distortion/noise level of both the original and the decoded output sound should not be *subjectively* different, thus giving the appearance of a lossless and real-time process.

In a technical sense, all practical real-time bit-rate-reduction systems can be referred to as “lossy.” In other words, the digital audio signal at the output is not identical to the input signal data stream. However, some compression algorithms are, for all intents and purposes, lossless; they lose as little as 2 percent of the original signal. Others remove approximately 80 percent of the original signal.

Redundancy and Irrelevancy

A complex audio signal contains a great deal of information, some of which, because the human ear cannot hear it, is deemed irrelevant. [2]. The same signal, depending on its complexity, also contains information that is highly predictable and, therefore, can be made redundant.

Redundancy, measurable and quantifiable, can be removed in the coder and replaced in the decoder; this process often is referred to as *statistical compression*. *Irrelevancy*, on the other hand, referred to as *perceptual coding*, once removed from the signal cannot be replaced and is lost, irretrievably. This is entirely a subjective process, with each proprietary algorithm using a different psychoacoustic model.

Critically perceived signals, such as pure tones, are high in redundancy and low in irrelevancy. They compress quite easily, almost totally a statistical compression process. Conversely, noncritically perceived signals, such as complex audio or noisy signals, are low in redundancy and high in irrelevancy. These compress easily in the perceptual coder, but with the total loss of all the irrelevancy content.

Human Auditory System

The sensitivity of the human ear is biased toward the lower end of the audible frequency spectrum, around 3 kHz [2]. At 50 Hz, the bottom end of the spectrum, and 17 kHz at the top end, the sensitivity of the ear is down by approximately 50 dB relative to its sensitivity at 3 kHz (Figure 7.1.1). Additionally, very few audio signals—music- or speech-based—carry fundamental frequencies above 4 kHz. Taking advantage of these characteristics of the ear, the structure of audible sounds, and the redundancy content of the PCM signal is the basis used by the designers of the *predictive* range of compression algorithms.

Another well-known feature of the hearing process is that loud sounds mask out quieter sounds at a similar or nearby frequency. This compares with the action of an automatic gain control, turning the gain down when subjected to loud sounds, thus making quieter sounds less likely to be heard. For example, as illustrated in Figure 7.1.2, if we assume a 1 kHz tone at a level of 70

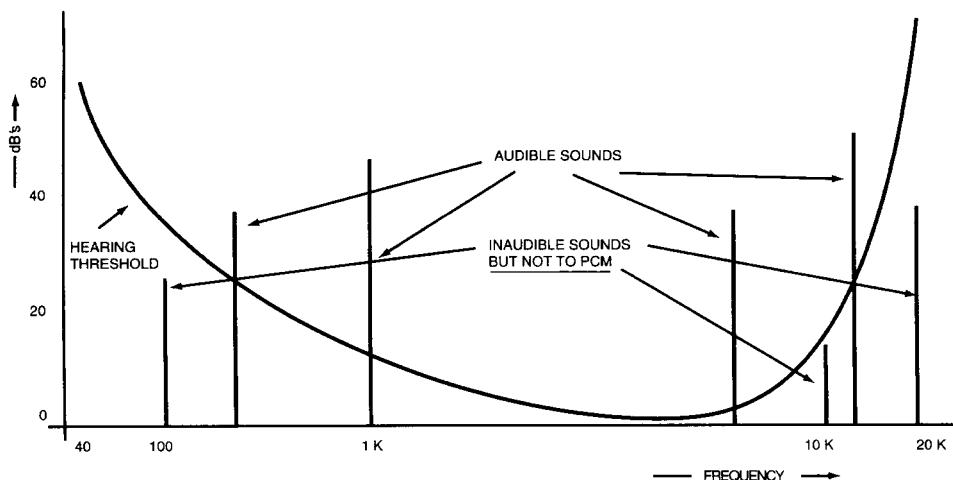


Figure 7.1.1 Generalized frequency response of the human ear. Note how the PCM process captures signals that the ear cannot distinguish. (From [2]. Used with permission.)

dB_u, levels of greater than 40 dB_u at 750 Hz and 2 kHz would be required for those frequencies to be heard. The ear also exercises a degree of temporal masking, being exceptionally tolerant of sharp transient sounds.

It is by mimicking these additional psychoacoustic features of the human ear and identifying the irrelevancy content of the input signal that the *transform* range of low bit-rate algorithms operate, adopting the principle that if the ear is unable to hear the sound then there is no point in transmitting it in the first place.

Quantization

Quantization is the process of converting an analog signal to its representative digital format or, as in the case with compression, the requantizing of an already converted signal [2]. This process is the limiting of a finite level measurement of a signal sample to a specific preset integer value. This means that the *actual* level of the sample may be greater or smaller than the preset *reference* level it is being compared with. The difference between these two levels, called the *quantization error*, is compounded in the decoded signal as *quantization noise*.

Quantization noise, therefore, will be injected into the audio signal after each A/D and D/A conversion, the level of that noise being governed by the bit allocation associated with the coding process (i.e., the number of bits allocated to represent the level of each sample taken of the analog signal). For linear PCM, the bit allocation is commonly 16. The level of each audio sample, therefore, will be compared with one of 2^{16} or 65,536 discrete levels or steps.

Compression or bit-rate reduction of the PCM signal leads to the requantizing of an already quantized signal, which will unavoidably inject further quantization noise. It always has been good operating practice to restrict the number of A/D and D/A conversions in an audio chain.

7-10 Compression Technologies for Audio

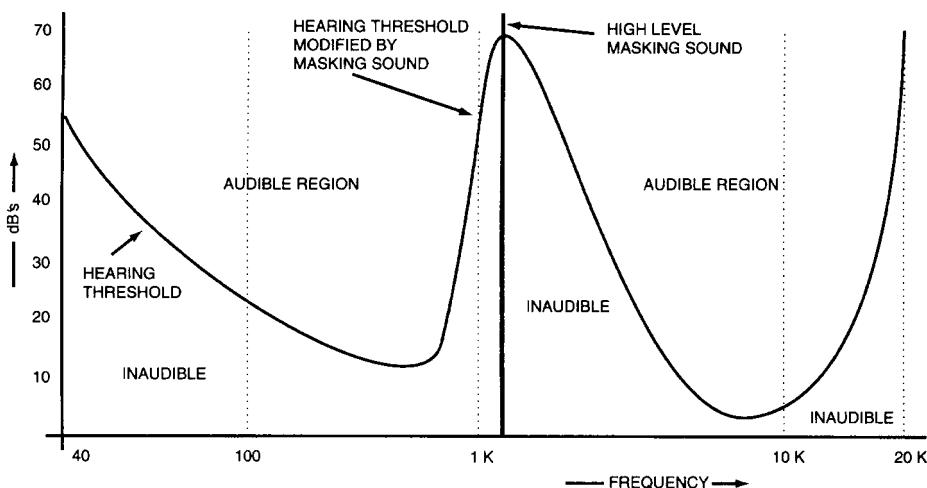


Figure 7.1.2 Example of the masking effect of a high-level sound. (From [2]. Used with permission.)

Nothing has changed in this regard, and now the number of compression stages also should be kept to a minimum. Additionally, the bit rates of these stages should be set as high as practical; put another way, the compression ratio should be as low as possible.

Sooner or later—after a finite number of A/D, D/A conversions and passes of compression coding, of whatever type—the accumulation of quantization noise and other unpredictable signal degradations eventually will break through the noise/signal threshold, be interpreted as part of the audio signal, be processed as such, and be heard by the listener.

Sampling Frequency and Bit Rate

The bit rate of a digital signal is defined by:

$$\text{sampling frequency} \times \text{bit resolution} \times \text{number of audio channels}$$

The rules regarding the selection of a sampling frequency are based on Nyquist's theorem [2]. This ensures that, in particular, the lower sideband of the sampling frequency does not encroach into the baseband audio. Objectionable and audible aliasing effects would occur if the two bands were to overlap. In practice, the sampling rate is set slightly above twice the highest audible frequency, which makes the filter designs less complex and less expensive.

In the case of a stereo CD with the audio signal having been sampled at 44.1 kHz, this sampling rate produces audio bandwidths of approximately 20 kHz for each channel. The resulting audio bit rate = $44.1 \text{ kHz} \times 16 \times 2 = 1.411 \text{ Mbits/s}$, as discussed previously.

7.1.1c Prediction and Transform Algorithms

Most audio-compression systems are based upon one of two basic technologies [2]:

- Predictive or *adaptive differential* PCM (ADPCM) time-domain coding
- Transform or *adaptive* PCM (APCM) frequency-domain coding

It is in their approaches to dealing with the redundancy and irrelevancy of the PCM signal that these techniques differ.

The time domain or *prediction* approach includes G.722, which has been a universal standard since the mid-70s, and was joined in 1989 by a proprietary algorithm, apt-X100. Both these algorithms deal mainly with redundancy.

The frequency domain or *transform* method adopted by a number of algorithms deal in irrelevancy, adopting psychoacoustic masking techniques to identify and remove those unwanted sounds. This range of algorithms include the industry standards ISO/MPEG-1 Layers 1, 2, and 3; apt-Q; MUSICAM; Dolby AC-2 and AC3; and others.

Subband Coding

Without exception, all of the algorithms mentioned in the previous section process the PCM signal by splitting it into a number of frequency subbands, in one case as few as two (G.722) or as many as 1024 (apt-Q) [1]. MPEG-1 Layer 1, with 4:1 compression, has 32 frequency subbands and is the system found in the Digital Compact Cassette (DCC). The MiniDisc ATRAC proprietary algorithm at 5:1 has a more flexible multisubband approach, which is dependent on the complexity of the audio signal.

Subband coding enables the frequency domain redundancies within the audio signals to be exploited. This permits a reduction in the coded bit rate, compared to PCM, for a given signal fidelity. Spectral redundancies are also present as a result of the signal energies in the various frequency bands being unequal at any instant in time. By altering the bit allocation for each subband, either by dynamically adapting it according to the energy of the contained signal or by fixing it for each subband, the quantization noise can be reduced across all bands. This process compares favorably with the noise characteristics of a PCM coder performing at the same overall bit rate.

Subband Gain

On its own, subband coding, incorporating PCM in each band, is capable of providing a performance improvement or *gain* compared with that of full band PCM coding, both being fed with the same complex, constant level input signal [1]. The improvement is defined as *subband gain* and is the ratio of the variations in quantization errors generated in each case while both are operating at the same transmission rate. The gain increases as the number of subbands increase, and with the complexity of the input signal. However, the implementation of the algorithm also becomes more difficult and complex.

Quantization noise generated during the coding process is constrained within each subband and cannot interfere with any other band. The advantage of this approach is that the masking by each of the subband dominant signals is much more effective because of the reduction in the noise bandwidth. Figure 7.1.3 charts subband gain as a function of the number of subbands for four essentially stationary, but differing, complex audio signals.

7-12 Compression Technologies for Audio

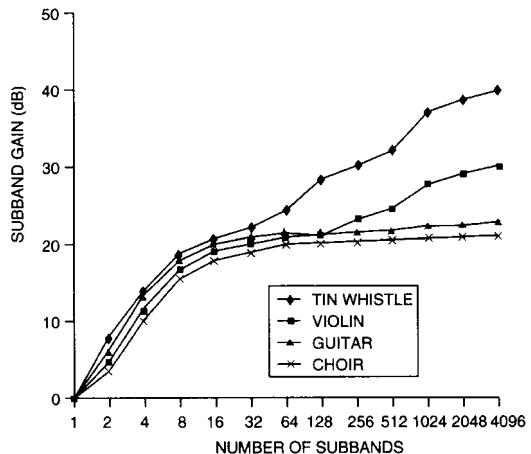


Figure 7.1.3 Variation of subband gain as a function of the number of subbands. (From [2]. Used with permission.)

In practical implementations of compression codecs, several factors tend to limit the number of subbands employed. The primary considerations include:

- The level variation of normal audio signals leading to an averaging of the energy across bands and a subsequent reduction in the coding gain
- The coding or processing delay introduced by additional subbands
- The overall computational complexity of the system

The two key issues in the analysis of a subband framework are:

- Determining the likely improvement associated with additional subbands
- Determining the relationships between subband gain, the number of subbands, and the response of the filter bank used to create those subbands

APCM Coding

The APCM processor acts in a similar fashion to an automatic gain control system, continually making adjustments in response to the dynamics—at all frequencies—of the incoming audio signal [1]. Transform coding takes a time block of signal, analyzes it for frequency and energy, and identifies irrelevant content. Again, to exploit the spectral response of the ear, the frequency spectrum of the signal is divided into a number of subbands, and the most important criteria are coded with a bias toward the more sensitive low frequencies. At the same time, through the use of psychoacoustic masking techniques, those frequencies which it is assumed will be masked by the ear are also identified and removed. The data generated, therefore, describes the frequency content and the energy level at those frequencies, with more bits being allocated to the higher-energy frequencies than those with lower energy.

The larger the time block of signal being analyzed, the better the frequency resolution and the greater the amount of irrelevancy identified. The penalty, however, is an increase in coding delay and a decrease in temporal resolution. A balance has been struck with advances in perceptual

Table 7.1.1 Operational Parameters of Subband APCM Algorithm (After [2].)

Coding System	Compression Ratio	Subbands	Bit Rate, kbits/s	A to A Delay, ms ¹	Audio Bandwidth, kHz
Dolby AC-2	6:1	256	256	45	20
ISO Layer 1	4:1	32	384	19	20
ISO Layer 2	Variable	32	192–256	>40	20
IOS Layer 3	12:1	576	128	>80	20
MUSICAM	Variable	32	128–384	>35	20

¹ The total system delay (encoder-to-decoder) of the coding system.

coding techniques and psychoacoustic modeling leading to increased efficiency. It is reported in [2] that, with this approach to compression, some 80 percent of the input audio can be removed with acceptable results.

This hybrid arrangement of working with time-domain subbands and simultaneously carrying out a spectral analysis can be achieved by using a *dynamic bit allocation* process for each subband. This subband APCM approach is found in the popular range of software-based MUSICAM, Dolby AC-2, and ISO/MPEG-1 Layers 1 and 2 algorithms. Layer 3—a more complex method of coding and operating at much lower bit rates—is, in essence, a combination of the best functions of MUSICAM and ASPEC, another adaptive transform algorithm. Table 7.1.1 lists the primary operational parameters for these systems.

Additionally, some of these systems exploit the significant redundancy between stereo channels by using a technique known as *joint stereo coding*. After the common information between left and right channels of a stereo signal has been identified, it is coded only once, thus reducing the bit-rate demands yet again.

Each of the subbands has its own defined *masking threshold*. The output data from each of the filtered subbands is requantized with just enough bit resolution to maintain adequate headroom between the quantization noise and the masking threshold for each band. In more complex coders (e.g., ISO/MPEG-1 Layer 3), any spare bit capacity is utilized by those subbands with the greater need for increased masking threshold separation. The maintenance of these signal-to-masking threshold ratios is crucial if further compression is contemplated for any postproduction or transmission process.

7.1.1d Processing and Propagation Delay

As noted previously, the current range of popular compression algorithms operate—for all intents and purposes—in real time [1]. However, this process does of necessity introduce some measurable delay into the audio chain. All algorithms take a finite time to analyze the incoming signal, which can range from a few milliseconds to tens and even hundreds of milliseconds. The amount of processing delay will be crucial if the equipment is to be used in any interactive or two-way application. As a rule of thumb, any more than 20 ms of delay in a two-way audio exchange is problematic. Propagation delay in satellite and long terrestrial circuits is a fact of life. A two-way hook up over a 1000 km, full duplex, telecom digital link has a propagation delay

7-14 Compression Technologies for Audio

of 3 ms in each direction. This is comparable to having a conversation with someone standing 1 m away. It is obvious that even over a very short distance, the use of a codec with a long processing delay characteristic will have a dramatic effect on operation.

7.1.1e Bit Rate and Compression Ratio

The ITU has recommend the following bit rates when incorporating data compression in an audio chain [1]:

- 128 kbytes/s per mono channel (256 kbytes/s for stereo) as the minimum bit rate for any stage if further compression is anticipated or required.
- 192 kbytes/s per mono channel (384 kbytes/s for stereo) as the minimum bit rate for the first stage of compression in a complex audio chain.

These markers place a 4:1 compression ratio at the “safe” end in the scale. However, more aggressive compression ratios, currently up to a nominal 20:1, are available. Keep in mind, though, that low bit rate, high-level compression can lead to problems if any further stages of compression are required or anticipated.

With successive stages of compression, either or both the noise floor and the audio bandwidth will be set by the stage operating at the lowest bit rate. It is, therefore, worth emphasizing that after these platforms have been set by a low bit rate stage, they cannot be subsequently improved by using a following stage operating at a higher bit rate.

Bit Rate Mismatch

A stage of compression may well be followed in the audio chain by another digital stage, either of compression or linear, but—more importantly—operating at a different sampling frequency [1]. If a D/A conversion is to be avoided, a sample rate converter must be used. This can be a stand alone unit or it may already be installed as a module in existing equipment. Where a following stage of compression is operating at the same sampling frequency but a different compression ratio, the bit resolution will change by default.

If the stages have the same sampling frequencies, a direct PCM or AES/EBU digital link can be made, thus avoiding the conversion to the analog domain.

7.1.1f Editing Compressed Data

The linear PCM waveform associated with standard audio workstations is only useful if decoded [1]. The resolution of the compressed data may or may not be adequate to allow direct editing of the audio signal. The minimum audio sample that can be removed or edited from a transform-coded signal will be determined by the size of the time block of the PCM signal being analyzed. The larger the time block, the more difficult the editing of the compressed data becomes.

7.1.2 Common Audio Compression Techniques

Subband APCM coding has found numerous applications in the professional audio industry, including [2]:

- The digital compact cassette (DCC)—uses the simplest implementation of subband APCM with the PASC/ISO/MPEG-1 Layer 1 algorithm incorporating 32 subbands offering 4:1 compression and producing a bit rate of 384 kbytes/s.
- The MiniDisc with the proprietary ATRAC algorithm—produces 5:1 compression and 292 kbytes/s bit rate. This algorithm uses a *modified discrete cosine transform* (MDCT) technique ensuring greater signal analysis by processing time blocks of the signal in nonuniform frequency divisions, with fewer divisions being allocated to the least sensitive higher frequencies.
- ISO/MPEG-1 Layer 2 (MUSICAM by another name)—a software-based algorithm that can be implemented to produce a range of bit rates and compression ratios commencing at 4:1.
- The ATSC DTV system—uses the subband APCM algorithm in Dolby AC-3 for the audio surround system associated with the ATSC DTV standard. AC-3 delivers five audio channels plus a bass-only effects channel in less bandwidth than that required for one stereo CD channel. This configuration is referred to as 5.1 channels.

For the purposes of illustration, two commonly used audio compression systems will be examined in some detail:

- apt-X100
- ISO/MPEG-1 Layer 2

7.1.2a apt-X100

apt-X100 is a four subband prediction (ADPCM) algorithm [1]. Differential coding reduces the bit rate by coding and transmitting or storing only the difference between a predicted level for a PCM audio sample and the absolute level of that sample, thus exploiting the redundancy contained in the PCM signal.

Audio exhibits relatively slowly varying energy fluctuations with respect to time. Adaptive differential coding, which is dependent on the energy of the input signal, dynamically alters the step size for each quantizing interval to reflect these fluctuations. In apt-X100, this equates to the *backwards adaptation process* and involves the analysis of 122 previous samples. Being a continuous process, this provides an almost constant and optimal signal-to-quantization noise ratio across the operating range of the quantizer.

Time domain subband algorithms implicitly model the hearing process and indirectly exploit a degree of irrelevancy by accepting that the human ear is more sensitive at lower frequencies. This is achieved in the four subband derivative by allocating more bits to the lower frequency bands. This is the only application of psychoacoustics exercised in apt-X100. All the information contained in the PCM signal is processed, audible or not (i.e., no attempt is made to remove irrelevant information). It is the unique fixed allocation of bits to each of the four subbands, coupled with the filtering characteristics of each individual listeners' hearing system, that achieves the satisfactory audible end result.

The user-defined output bit rates range from 56 to 384 kbytes/s, achieved by using various sampling frequencies from 16 kHz to 48 kHz, which produce audio bandwidths from 7.5 kHz mono to 22 kHz stereo.

7-16 Compression Technologies for Audio

Auxiliary data up to 9.6 kbytes/s can also be imbedded into the data stream without incurring a bit overhead penalty. When this function is enabled, an audio bit in one of the higher frequency subbands is replaced by an auxiliary data bit, again with no audible effect.

An important feature of this algorithm is its inherent robustness to random bit errors. No audible distortion is apparent for normal program material at a *bit error rate* (BER) of 1:10,000, while speech is still intelligible down to a BER of 1:10.

Distortions introduced by bit errors are constrained within each subband and their impact on the decoder subband predictors and quantizers is proportional to the magnitude of the differential signal being decoded at that instant. Thus, if the signal is small—which will be the case for a low level input signal or for a resonant, highly predictable input signal—any bit error will have minimal effect on either the predictor or quantizer.

The 16 bit linear PCM signal is processed in time blocks of four samples at a time. These are filtered into four equal-width frequency subbands; for 20 kHz, this would be 0–5 kHz, 5–10 kHz, and so on. The four outputs from the *quadrature mirror filter* (QMF) tree are still in the 16 bit linear PCM format, but are now frequency-limited.

As shown in Figure 7.1.4, the compression process can be mapped by taking, for example, the first and lowest frequency subband. The first step is to create the difference signal. After the system has settled down on initiation, there will be a reconstructed 16 bit difference signal at the output of the inverse quantizer. This passes into a prediction loop that, having analyzed 122 previous samples, will make a prediction for the level of the next full level sample arriving from the filter tree. This prediction is then compared with the actual level.

The output of the comparator is the resulting 16-bit difference signal. This is requantized to a new 7-bit format, which in turn is inverse quantized back to 16 bits again to enable the prediction loop.

The output from the inverse quantizer is also analyzed for energy content, again for the same 122 previous samples. This information is compared with on-board look up tables and a decision is made to dynamically adjust, up or down as required, the level of each step of the 1024 intervals in the 7-bit quantizer. This ensures that the quantizer will always have adequate range to deal with the varying energy levels of the audio signal. Therefore, the input to the multiplexer will be a 7 bit word but the range of those bits will be varying in relation to the signal energy.

The three other subbands will go through the same process, but the number of bits allocated to the quantizers are much less than for the first subband.

The output of the multiplexer or bit stream formatter is a new 16-bit word that represents four input PCM samples and is, therefore, one quarter of the input rate; a reduction of 4:1.

The decoding process is the complete opposite of the coding procedure. The incoming 16-bit compressed data word is demultiplexed and used to control the operation of four subband decoder sections, each with similar predictor and quantizer step adjusters. A QMF filter tree finally reconstructs a linear PCM signal and separates any auxiliary data that may be present.

7.1.2b ISO/MPEG-1 Layer 2

This algorithm differs from Layer 1 by adopting more accurate quantizing procedures and by additionally removing redundancy and irrelevancy on the generated scale factors [1]. The ISO/MPEG-1 Layer 2 scheme operates on a block of 1152 PCM samples, which at 48 kHz sampling represents a 24 ms time block of the input audio signal. Simplified block diagrams of the encoding/decoding systems are given in Figure 7.1.5.

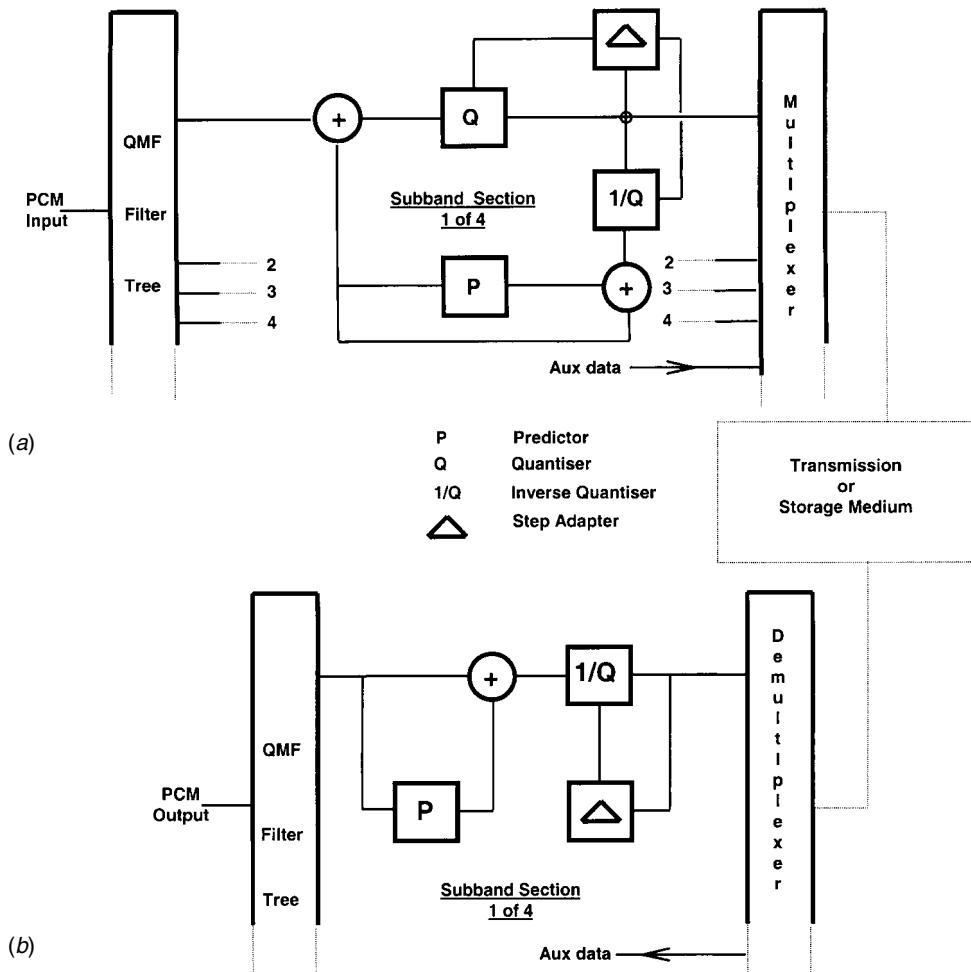


Figure 7.1.4 apt-X100 audio coding system: (a) encoder block diagram, (b) decoder block diagram. (Courtesy of Audio Processing Technology.)

The incoming linear PCM signal block is divided into 32 equally spaced subbands using a polyphase analysis filter bank (Figure 7.1.5a). At 48 kHz sampling, this equates to the bandwidth of each subband being 750 Hz. The bit allocation for the requantizing of these subband samples is then dynamically controlled by information derived from analyzing the audio signal, measured against a preset psychoacoustic model.

The filter bank, which displays manageable delay and minimal complexity, optimally adapts each block of audio to achieve a balance between the effects of temporal masking and inaudible pre-echoes.

7-18 Compression Technologies for Audio

The PCM signal is also fed to a *fast Fourier transform* (FFT) running in parallel with the filter bank. The aural sensitivities of the human auditory system are exploited by using this FFT process to detect the differences between the wanted and unwanted sounds and the quantization noise already present in the signal, and then to adjust the signal-to-mask thresholds, conforming to a preset perceptual model.

This psychoacoustic model is only found in the coder, thus making the decoder less complex and permitting the freedom to exploit future improvements in coder design. The actual number of levels for each quantizer is determined by the bit allocation. This is arrived at by setting the *signal-to-mask ratio* (SMR) parameter, defined as the difference between the minimum masking threshold and the maximum signal level. This minimum masking threshold is calculated using the psychoacoustic model and provides a reference noise level of “just noticeable” noise for each subband.

In the decoder, after demultiplexing and decyphering of the audio and side information data, a dual synthesis filter bank reconstructs the linear PCM signal in blocks of 32 output samples (Figure 7.1.5b).

A scale factor is determined for each 12 subband sample block. The maximum of the absolute values of these 12 samples generates a *scale factor* word consisting of 6 bits, a range of 63 different levels. Because each frame of audio data in Layer 2 corresponds to 36 subband samples, this process will generate 3 scale factors per frame. However, the transmitted data rate for these scale factors can be reduced by exploiting some redundancy in the data. Three successive subband scale factors are analyzed and a pattern is determined. This pattern, which is obviously related to the nature of the audio signal, will decide whether one, two or all three scale factors are required. The decision will be communicated by the insertion of an additional *scale factor select information* data word of 2 bits (SCFSI).

In the case of a fairly stationary tonal-type sound, there will be very little change in the scale factors and only the largest one of the three is transmitted; the corresponding data rate will be (6 + 2) or 8 bits. However, in a complex sound with rapid changes in content, the transmission of two or even three scale factors may be required, producing a maximum bit rate demand of (6 + 6 + 6 + 2) or 20 bits. Compared with Layer 1, this method of coding the scale factors reduces the allocation of data bits required for them by half.

The number of data bits allocated to the overall bit pool is limited or fixed by the data rate parameters. These parameters are set out by a combination of sampling frequency, compression ratio, and—where applicable—the transmission medium. In the case of 20 kHz stereo being transmitted over ISDN, for example, the maximum data rate is 384 kbits/s, sampling at 48kHz, with a compression ratio of 4:1.

After the number of side information bits required for scale factors, bit allocation codes, CRC, and other functions have been determined, the remaining bits left in the pool are used in the re-coding of the audio subband samples. The allocation of bits for the audio is determined by calculating the SMR, via the FFT, for each of the 12 subband sample blocks. The bit allocation algorithm then selects one of 15 available quantizers with a range such that the overall bit rate limitations are met and the quantization noise is masked as far as possible. If the composition of the audio signal is such that there are not enough bits in the pool to adequately code the subband samples, then the quantizers are adjusted down to a best-fit solution with (hopefully) minimum damage to the decoded audio at the output.

If the signal block being processed lies in the lower one third of the 32 frequency subbands, a 4-bit code word is simultaneously generated to identify the selected quantizer; this word is, again, carried as side information in the main data frame. A 3-bit word would be generated for

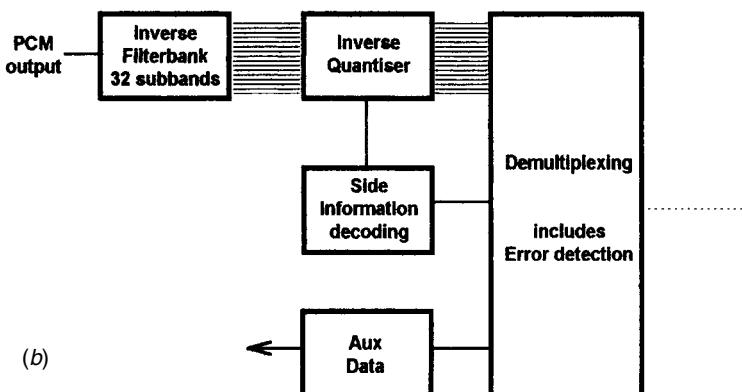
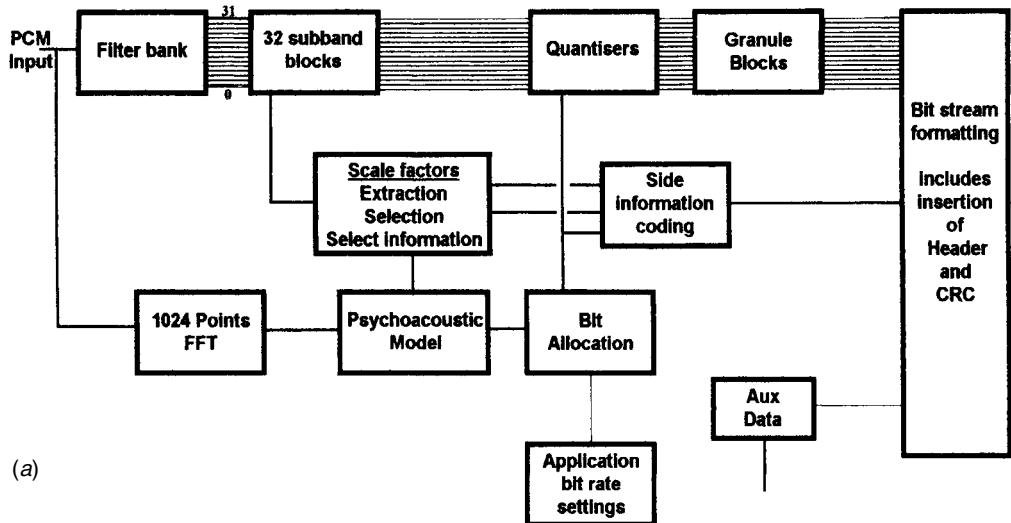


Figure 7.1.5 ISO/MPEG-1 Layer 2 system: (a) encoder block diagram, (b) decoder block diagram. (After [1].)

processing in the mid frequency subbands and a 2-bit word for the higher frequency subbands. When the audio analysis demands it, this allows for *at least* 15, 7, and 3 quantization levels, respectively, in each of the three spectrum groupings. However, each quantizer can, if required,

7-20 Compression Technologies for Audio

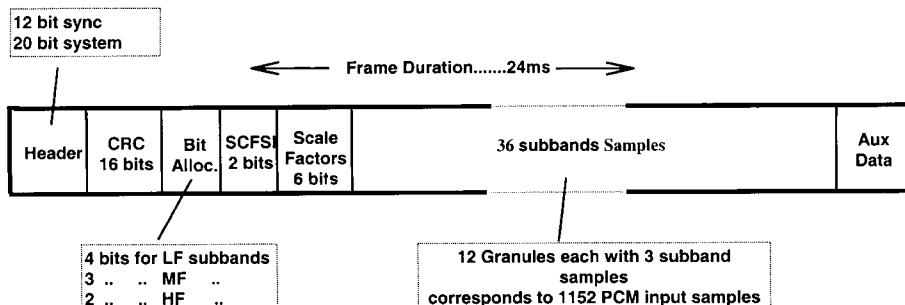


Figure 7.1.6 ISO/MPEG-1 Layer 2 data frame structure. (After [1].)

cover from 3 to 65,535 levels and additionally, if no signal is detected then no quantization takes place.

As with the scale factor data, some further redundancy can be exploited, which increases the efficiency of the quantising process. For the lowest quantizer ranges (i.e., 3, 5, and 9 levels), three successive subband sample blocks are grouped into a “granule” and this—in turn—is defined by only one code word. This is particularly effective in the higher frequency subbands where the quantizer ranges are invariably set at the lower end of the scale.

Error detection information can be relayed to the decoder by inserting a 16 bit CRC word in each data frame. This parity check word allows for the detection of up to three single bit errors or a group of errors up to 16 bits in length. A codec incorporating an error concealment regime can either mute the signal in the presence of errors or replace the impaired data with a previous, error free, data frame. The typical data frame structure for ISO/MPEG-1 Layer 2 audio is given in Figure 7.1.6.

7.1.2c MPEG-2 AAC

Also of note is MPEG-2 *advanced audio coding* (AAC), a highly advanced perceptual code, used initially for digital radio applications. The AAC code improves on previous techniques to increase coding efficiency. For example, an AAC system operating at 96 kbit/s produces the same sound quality as ISO/MPEG-1 Layer 2 operating at 192 kbit/s—a 2:1 reduction in bit rate. There are three modes (Profiles) in the AAC standard:

- *Main*—used when processing power, and especially memory, are readily available.
- *Low complexity* (LC)—used when processing cycles and memory use are constrained.
- *Scalable sampling rate* (SSR)—appropriate when a *scalable decoder* is required. A scalable decoder can be designed to support different levels of audio quality from a common bit stream; for example, having both high- and low-cost implementations to support higher and lower audio qualities, respectively.

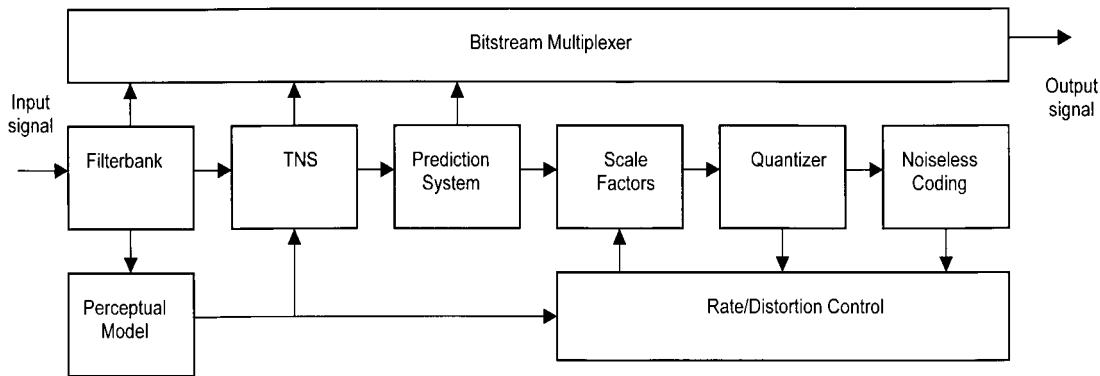


Figure 7.1.7 Functional block diagram of the MPEG-2 AAC coding system.

Different Profiles trade off encoding complexity for audio quality at a given bit rate. For example, at 128 kbit/s, the Main Profile AAC code has a more complex encoder structure than the LC AAC code at the same bit rate, but provides better audio quality as a result.

A block diagram of the AAC system general structure is given in Figure 7.1.7. The blocks in the drawing are referred to as “tools” that the coding algorithm uses to compress the digital audio signal. While many of these tools exist in most audio perceptual codes, two are unique to AAC—the *temporal noise shaper* (TNS) and the *filterbank* tool. The TNS uses a backward adaptive prediction process to remove redundancy between the frequency channels that are created by the filterbank tool.

MPEG-2 AAC provides the capability of up to 48 main audio channels, 16 low frequency effects channels, 16 overdub/multilingual channels, and 10 data streams. By comparison, ISO/MPEG-1 Layer 1 provides two channels and Layer 2 provides 5.1 channels (maximum). AAC is not backward compatible with the Layer 1 and Layer 2 codes.

7.1.2d MPEG-4

MPEG-4, as with the MPEG-1 and MPEG-2 efforts, is not concerned solely with the development of audio coding standards, but also encompasses video coding and data transmission elements. In addition to building upon the audio coding standards developed for MPEG-2, MPEG-4 includes a revolutionary new element—synthesized sound. Tools are provided within MPEG-4 for coding of both natural sounds (speech and music) and for synthesizing sounds based on structured descriptions. The representations used for synthesizing sounds can be formed by text or by instrument descriptions, and by coding other parameters to provide for effects, such as reverberation and spatialization.

Natural audio coding is supported within MPEG-4 at bit rates ranging from 2–64 kbit/s, and includes the MPEG-2 AAC standard (among others) to provide for general compression of audio

7-22 Compression Technologies for Audio

in the upper bit rate range (8–64 kbit/s), the range of most interest to broadcasters. Other types of coders, primarily voice coders (or *vocoders*) are used to support coding down to the 2 kbit/s rate.

For synthesized sounds, decoders are available that operate based on so-called *structured inputs*, that is, input signals based on descriptions of sounds and not the sounds themselves. Text files are one example of a structured input. In MPEG-4, text can be converted to speech in a *text-to-speech* (TTS) decoder. Synthetic music is another example, and may be delivered at extremely low bit rates while still describing an exact sound signal. The standard's *structured audio decoder* uses a language to define an orchestra made up of instruments, which can be downloaded in the bit stream, not fixed in the decoder.

TTS support is provided in MPEG-4 for unembellished text, or text with prosodic (pitch contour, phoneme duration, etc.) parameters, as an input to generate intelligible synthetic speech. It includes the following functionalities:

- Speech synthesis using the prosody of the original speech
- Facial animation control with phoneme information (important for multimedia applications)
- *Trick mode* functionality: pause, resume, jump forward, jump backward
- International language support for text
- International symbol support for phonemes
- Support for specifying the age, gender, language, and dialect of the speaker

MPEG-4 does not standardize a method of synthesis, but rather specifies a method of describing synthesis.

Compared to previous MPEG coding standards, the goals of MPEG-4 go far beyond just achieving higher coding efficiency [3]. Specifically, MPEG-4 is conceived as a set of interoperable technologies implementing the following concepts:

- **Universality:** Rather than serving specific application areas, MPEG-4 is an attempt to provide solutions for almost any conceivable scenario using audiovisual compression, ranging from very low bit rates to studio-quality applications. Because there is currently no single coding technology serving all these cases equally well, MPEG-4 Audio provides both a so-called General Audio coder and a number of coders specifically targeting certain types of signals (e.g., speech or music) or bit rates.
- **Scalability:** The MPEG-4 concept of scalable coding enables the transmission and decoding of a scalable bitstream with a bit rate that can be adapted to dynamically varying requirements, such as the instantaneous transmission channel capacity. This capability offers significant advantages for transmitting content over channels with a variable channel capacity (e.g., the Internet and wireless links) or connections for which the available channel capacity is unknown at the time of encoding.
- **Object-based representation and composition:** As suggested by the standard's name ("Generic Coding of Audiovisual Objects"), MPEG-4 represents audiovisual content as a set of *objects* rather than a flat representation of the entire audiovisual scene. The relation of the coded objects with each other and the way to construct the scene from these objects ("composition") is described by a *scene description*.

- **Content-based interactivity:** The combination of object-based representation and scene description/composition allows exciting new capabilities. For example, during a presentation, the user can interact with coded objects and control the way they are rendered by the composition unit. Examples of this so-called *content-based interactivity* include omitting reproduction for certain objects or controlling their scene composition parameters, such as spatial coordinates and reproduction level.
- **Natural and synthetic representations:** Synthetic content, such as computer graphics or synthesized audio, has gained increasing importance and is widely deployed, fuelled by the success of the personal computer. Converging both worlds, MPEG-4 defines representations for both natural and synthetic objects and allows arbitrary combinations of these object types within a scene.

7.1.2e Dolby E Coding System

Dolby E coding was developed to expand the capacity of existing two channel AES/EBU digital audio infrastructures to make them capable of carrying up to eight channels of audio plus the metadata required by the Dolby Digital coders used in the ATSC DTV transmission system [4]. This allows existing digital videotape recorders, routing switchers, and other video plant equipment, as well as satellite and telco facilities, to be used in program contribution and distribution systems that handle multichannel audio. The coding system was designed to provide broadcast quality output even when decoded and re-encoded many times, and to provide clean transitions when switching between programs.

Dolby E encodes up to eight audio channels plus the necessary metadata and inserts this information into the payload space of a single AES digital audio pair. Because the AES protocol is used as the transport mechanism for the Dolby E encoded signal, digital VTRs, routing switchers, DAs, and all other existing digital audio equipment in a typical video facility can handle multichannel programming. It is possible to do insert or assemble edits on tape or to make audio-follow-video cuts between programs because the Dolby E data is synchronized with the accompanying video. The metadata is multiplexed into the compressed audio, so it is switched with and stays in sync with the audio.

The main challenge in designing a bit-rate reduction system for multiple generations is to prevent coding artifacts from appearing in the recovered audio after several generations. The coding artifacts are caused by a buildup of noise during successive encoding and decoding cycles, so the key to good multigeneration performance is to manage the noise optimally.

This noise is caused by the rate reduction process itself. Digitizing (quantizing) a signal leads to an error that appears in the recovered signal as a broadband noise. The smaller the quantizer steps (i.e., the more resolution or bits used to quantize the signal), the lower the noise will be. This quantizing noise is related to the signal, but becomes “whiter” as the quantizer resolution rises. With resolutions less than about 5 or 6 bits and no dither, the quantizing noise is clearly related to the program material.

Bit rate reduction systems try to squeeze the data rates down to the equivalent of a few bits (or less) per sample and, thus, tend to create quantizing noise in quite prodigious quantities. The key to recovering signals that are subjectively indistinguishable from the original signals, or in which the quantizing noise is inaudible, is in allocating the available bits to the program signal components in a way that takes advantage of the ear's natural ability to mask low level signals with higher level ones.

7-24 Compression Technologies for Audio

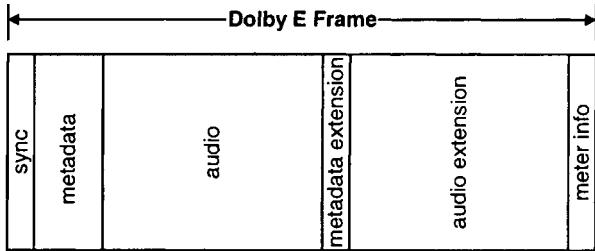


Figure 7.1.8 Basic frame structure of the Dolby E coding system. (After [4].)

The rate reduction encoder sends information about the frequency spectrum of the program signal to the decoder. A set of reconstruction filters in the decoder confines the quantizing noise produced by the bit allocation process in the encoder to the bandwidth of those filters. This allows the system designer to keep the noise (ideally) below the masking thresholds produced by the program signal. The whole process of allocating different numbers of bits to different program signal components (or of quantizing them at different resolutions) creates a noise floor that is related to the program signal and to the rate reduction algorithm used. The key to doing this is to have an accurate model of the masking characteristics of the ear, and in allocating the available bits to each signal component so that the masking threshold is not exceeded.

When a program is decoded and then re-encoded, the re-encoding process (and any subsequent ones) adds its noise to the noise already present. Eventually, the noise present in some part of the spectrum will build up to the point where it becomes audible, or exceeds the allowable *coding margin*. A codec designed for minimum data rate has to use lower coding margins (or more aggressive bit allocation strategies) than one intended to produce high quality signals after many generations.

The design strategy for a multigeneration rate reduction system, such as one used for Dolby E, is therefore quite different than that of a minimum data rate codec intended for program transmission applications.

Dolby E signals are carried in the AES3 interface using a packetized structure [5]. The packets are based on the coded Dolby E frame, which is illustrated in Figure 7.1.8. Each Dolby E frame consists of a *synchronization field*, *metadata field*, *coded audio field*, and a *meter field*. The metadata field contains a complete set of parameters so that each Dolby E frame can be decoded independently. The Dolby E frames are embedded into the AES3 interface by mapping the Dolby E data into the audio sample word bits of the AES3 frames utilizing both channels within the signal. (See Figure 7.1.9.) The data can be packed to utilize 16, 20, or 24 bits in each AES3 sub-frame. The advantage of utilizing more bits per sub-frame is that a higher data rate is available for carrying the coded information. With a 48 kHz AES3 signal, the 16 bit mode allows a data rate of up to 1.536 Mbits/s for the Dolby E signal, while the 20 bit mode allows 1.92 Mbits/s. Higher data rate allows more generations and/or more channels of audio to be supported. However, some AES3 data paths may be restricted in data rate (e.g., some storage devices will only record 16 or 20 bits). Dolby E therefore allows the user to choose the optimal data rate for a given application.

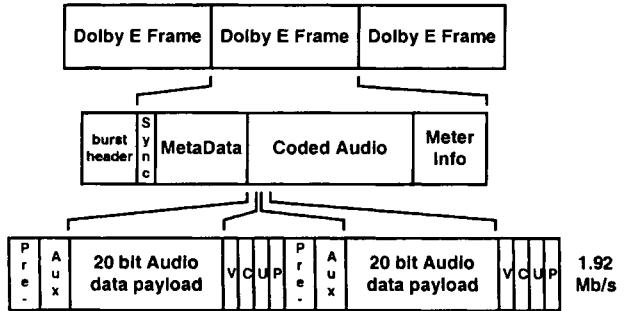


Figure 7.1.9 Overall coding scheme of Dolby E. (After [5].)

Architectural Overview

The basic architecture of the Dolby E encoder/decoder is shown in Figure 7.1.10 [6]. In the encoder, the incoming multichannel PCM audio is first passed through a set of *sample rate converters* (SRC), which convert the audio sample rate to a simple multiple of the video frame rate. The sample rate converter clock is derived from the video reference signal via a phase-locked loop (PLL).

The output of the sample rate converter is then fed to the encoder core, which includes the audio data compression engine. In addition, incoming metadata parameters are also passed to the encoder core. The output of this core is a series of coded data frames, each containing a combination of compressed multichannel audio and metadata, delivered at a rate that is synchronous with the video signal.

The decoder architecture is a straightforward reversal of the encoder. The coded bitstream is passed into the decoder core, which reconstructs the multichannel audio samples and the metadata. Because the reconstructed audio sample rate is a function of the video frame rate, a second set of sample rate converters is used to convert the output audio to a standard 48 kHz rate.

The sample rate converters used in this design are a consequence of the need to support a wide variety of video frame rates. Table 7.1.2 lists several of the most common video frame rates used in current broadcast practice, as well as the number of 48 kHz PCM samples per frame associated with each rate. Not only does the number of samples vary depending on the frame rate, but for one case the number is not even an integer.

In order to simplify the design of the audio compression engine, sample rate converters were introduced to ensure that the number of samples per frame was a constant. For this system, each frame duration is the equivalent of 1792 audio samples. As a result, the sample rate seen at the input of the encoder core is no longer 48 kHz, but rather varies with the video frame rate.

Table 7.1.3 lists the internal sample rates used by this system during audio data compression encoding and decoding, as a function of the associated video frame rate.

Coded Frame Format

In order to meet certain compatibility requirements, the coded output frame is structured to look in many ways like stereo PCM [6]. Specifically, the output bitstream sample rate is set to 48 kHz, and the coded data is aligned in 20-bit words. This format is shown in Figure 7.1.11. Note that system options also allow for 16-bit or 24-bit output formats, however 20-bit is the most common standard.

7-26 Compression Technologies for Audio

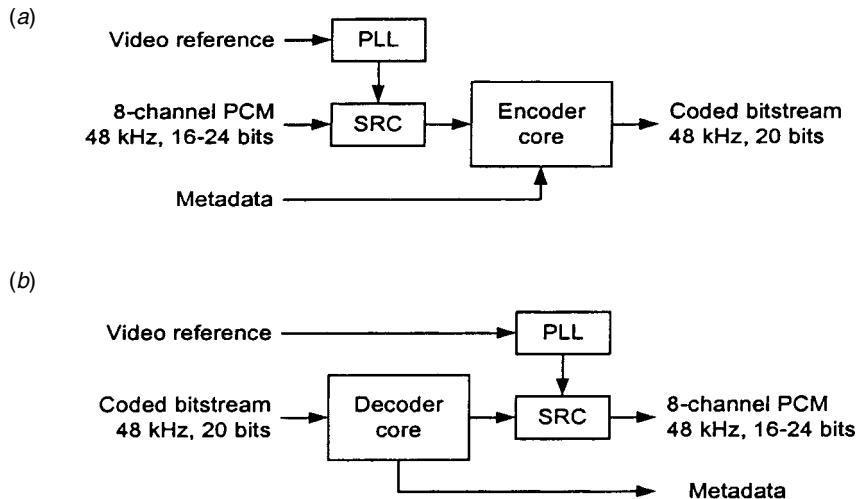


Figure 7.1.10 Dolby E architecture: (a) encoder, (b) decoder. (After [6].)

Table 7.1.2 PCM Sample Count as a Function of Video Frame Rate (After [7].)

Frame Rate (f/s)	Samples per Frame
23.98	2002
24	2000
25	1920
29.97	1601.6
30	1600

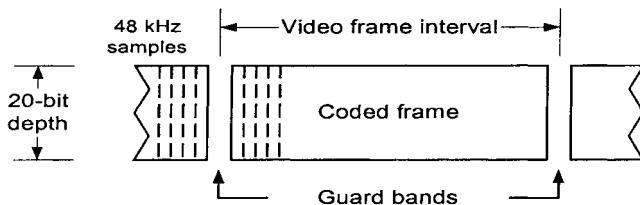
As shown in Figure 7.1.11, not all of the output channel data rate is used for carrying coded audio. Instead, gaps of about 5 percent of the total frame duration are introduced between successive audio frames. These gaps act as *switching guard bands*, providing a measure of tolerance for real-time splicing and editing of coded bitstreams without concern for damaging adjacent frame data.

7.1.3 Objective Quality Measurements

Perceptual audio coding has revolutionized the processing and distribution of digital audio signals. One aspect of this technology, not often emphasized, is the difficulty of determining, *objectively*, the quality of perceptually coded signals. Audio professionals could greatly benefit from

Table 7.1.3: Internal Sample Rate as a Function of Video Frame Rate (After [7].)

Frame Rate (f/s)	Internal Sample Rate (kHz)
23.98	42.965
24	43.008
25	44.800
29.97	53.706
30	53.760

**Figure 7.1.11** Dolby E coded frame format. (After [6].)

an objective approach to signal characterization because it would offer a simple but accurate approach for verification of good audio quality within a given facility.

Most of the discussions regarding this topic involve reference to the results of subjective evaluations of audio quality, where for example, groups of listeners compare reference audio material to coded audio material and then judge the *level of impairment* caused by the coding process. A procedure for this process has been standardized in ITU-R Rec. BS.1116, and makes use of the ITU-R five grade impairment scale:

- 5.0—Imperceptible
- 4.0—Perceptible but not annoying
- 3.0—Slightly annoying
- 2.0—Annoying
- 1.0—Very annoying

Quality measurements made with properly executed subjective evaluations are widely accepted and have been used for a variety of purposes, from determining which of a group of perceptual coders performs best, to assessing the overall performance of an audio broadcasting system.

The problem with subjective evaluations is that, while accurate, they are time consuming and expensive to undertake. Traditional objective benchmarks of audio performance, such as signal-to-noise ratio or total harmonic distortion, are not reliable measures of perceived audio quality, especially when perceptually coded signals are being considered.

To remedy this situation, ITU-R established Task Group 10-4 to develop a method of objectively assessing perceived audio quality. Conceptually, the result of this effort would be a device

7-28 Compression Technologies for Audio

Table 7.1.4 Target Applications for ITU-R Rec. BS.1116 PEAQ

Category	Application	Version
Diagnostic	Assessment of implementations	Both
	Equipment or connection status	Advanced
	Codec identification	Both
Operational	Perceptual quality line-up	Basic
	On-line monitoring	Basic
Development	Codec development	Both
	Network planning	Both
	Aid to subjective assessment	Advanced

having two inputs—a reference and the audio signal to be evaluated—and would generate an audio quality estimate based on these sources.

Six organizations proposed models for accomplishing this objective, and over the course of several years these models were evaluated for effectiveness, in part by using source material from previously documented subjective evaluations. Ultimately, the task group decided that none of the models by themselves fully met the stated requirements. The group decided, instead, to use the best parts of the different models to create another model that would meet the sought-after requirements.

This approach resulted in an objective measurement method known as *Perceptual Evaluation of Audio Quality* (PEAQ). The method contains two versions—a basic version designed to support real-time implementations, and an advanced version optimized for the highest accuracy but not necessarily implementable in real-time. The primary applications for PEAQ are summarized in Table 7.1.4.

7.1.3a Perspective on Audio Compression

A balance must be struck between the degree of compression available and the level of distortion that can be tolerated, whether the result of a single coding pass or the result of a number of passes, as would be experienced in a complex audio chain or network [1]. There have been many outstanding successes for digital audio data compression in communications and storage, and as long as the limitations of the various compression systems are fully understood, successful implementations will continue to grow in number. Table 7.1.5 compares several common audio coding systems.

Compression is a tradeoff and in the end you get what you pay for. Quality must be measured against the coding algorithm being used, the compression ratio, bit rate, and coding delay resulting from the process.

There is continued progress in expanding the arithmetical capabilities of digital signal processors, and the supporting hardware developments would seem to be following a parallel course. It is possible to obtain a single chip containing both encoder and decoder elements, including ste-

Table 7.1.5 Comparison of Audio Compression Coding Systems (After [7].)

Audio System	Total Bit Rates (kbits/s)	Filter Bank	Frequency Resolution @ 48 kHz	Temporal Resolution @ 48 kHz (ms)	Frame Length @ 48 kHz (ms)	Bit Rate Target (kbits/s per channel)
MPEG Layer I	32–448	PQMF	750 Hz	0.66	8	128
MPEG Layer II	32–384	PQMF	750 Hz	0.66	24	128
MPEG Layer III	32–320	PQMF/MDCT	41.66 Hz	4	24	64
apt-X	Fixed 4:1 compression ¹	PQMF	12 kHz	<0.1	2.54	192
apt-Q	Fixed 12:1 and 18:1 compression ²	MDCT	23.44 Hz	5.34	42.66	64
AC-3	32–640	MDCT	93.75 Hz	2.66	32	64

1 Bit rates are from 56 to 384 kbits/s when the sampling frequency varies from 16 to 48 kHz.
 2 Bit rates can vary from 56 kbits/s for a mono signal sampled at 32 kHz to 128 kbits/s for a stereo signal sampled at 48 kHz.

re capabilities. In every five year period, it is not unreasonable to expect a tenfold increase in the processing capabilities of a single DSP chip, thus, increasing flexibility and processing power. Speculation could point to an eventual position when a completely lossless algorithm with an extremely high compression ratio would become available. In any event, the art of compressing audio data streams into narrower and narrower digital pipes will undoubtedly continue.

7.1.4 References

1. Wylie, Fred: "Audio Compression Technologies," *NAB Engineering Handbook*, 9th ed., Jerry C. Whitaker (ed.), National Association of Broadcasters, Washington, D.C., 1998.
2. Wylie, Fred: "Audio Compression Techniques," *The Electronics Handbook*, Jerry C. Whitaker (ed.), CRC Press, Boca Raton, Fla., pp. 1260–1272, 1996.
3. Herre, J., and B. Grill: "MPEG-4 Audio—Overview and Perspectives for the Broadcaster," *IBC 2000 Proceedings*, International Broadcast Convention, Amsterdam, September 2000.
4. Lyman, Stephen, "A Multichannel Audio Infrastructure Based on Dolby E Coding," *Proceedings of the NAB Broadcast Engineering Conference*, National Association of Broadcasters, Washington, D.C., 1999.
5. Terry, K. B., and S. B. Lyman: "Dolby E—A New Audio Distribution Format for Digital Broadcast Applications," *International Broadcasting Convention Proceedings*, IBC, London, England, pp. 204–209, September 1999.

7-30 Compression Technologies for Audio

6. Vernon, S., and T. Spath: "Carrying Multichannel Audio in a Stereo Production and Distribution Infrastructure," *Proceedings of IBC 2000*, International Broadcasting Convention, Amsterdam, September 2000.
7. Robin, Michael, and Michel Poulin: *Digital Television Fundamentals*, McGraw-Hill, New York, N.Y., 1998.

7.1.5 Bibliography

Brandenburg, K., and Gerhard Stoll: "ISO-MPEG-1 Audio: A Generic Standard for Coding of High Quality Digital Audio," *92nd AES Convention Proceedings*, Audio Engineering Society, New York, N.Y., 1992, revised 1994.

Smyth, Stephen: "Digital Audio Data Compression," *Broadcast Engineering*, Intertec Publishing, Overland Park, Kan., February 1992.

Chapter
7.2

ATSC DTV System Compression Issues

Jerry C. Whitaker, Editor-in-Chief

7.2.1 Introduction

The primary application of interest when the MPEG-2 standard was first defined was “true” television broadcast resolution, as specified by ITU-R Rec. 601. This is roughly four times more picture information than the MPEG-1 standard provides. MPEG-2 is a superset, or extension, of MPEG-1. As such, an MPEG-2 decoder also should be able to decode an MPEG-1 stream. This broadcast version adds to the MPEG-1 toolbox provisions for dealing with interlace, graceful degradation, and hierarchical coding.

Although MPEG-1 and MPEG-2 each were specified with a particular range of applications and resolutions in mind, the committee’s specifications form a set of techniques that support multiple coding options, including picture types and macroblock types. Many variations exist with regard to picture size and bit rates. Also, although MPEG-1 can run at high bit rates and at full ITU-R Rec. 601 resolution, it processes frames, not fields. This fact limits the attainable quality, even at data rates approaching 5 Mbits/s.

The MPEG specifications apply only to decoding, not encoding. The ramifications of this approach are:

- Owners of existing decoding software can benefit from future breakthroughs in encoding processing. Furthermore, the suppliers of encoding equipment can differentiate their products by cost, features, encoding quality, and other factors.
- Different schemes can be used in different situations. For example, although *Monday Night Football* must be encoded in real time, a film can be encoded in non-real time, allowing for fine-tuning of the parameters via computer or even a human operator.

7.2.2 MPEG-2 Layer Structure

To allow for a simple yet upgradable system, MPEG-2 defines only the functional elements—syntax and semantics—of coded streams. Using the same system of *I*-, *P*-, and *B*-frames developed for MPEG-1, MPEG-2 employs a 6-layer hierarchical structure that breaks the data into simplified units of information, as given in Table 7.2.1.

7-32 Compression Technologies for Audio

Table 7.2.1 Layers of the MPEG-2 Video Bit-Stream Syntax

Syntax layer	Functionality
Video sequence layer	Context unit
Group of pictures (GOP) layer	Random access unit: video coding
Picture layer	Primary coding unit
Slice layer	Resynchronization unit
Macroblock layer	Motion-compensation unit
Block layer	DCT unit

The top *sequence layer* defines the decoder constraints by specifying the context of the video sequence. The sequence-layer data header contains information on picture format and application-specific details. The second level allows for random access to the decoding process by having a periodic series of pictures; it is fundamentally this GOP layer that provides the bidirectional frame prediction. Intraframe-coded (*I*) frames are the entry-point frames, which require no data from other frames in order to reconstruct. Between the *I*-frames lie the predictive (*P*) frames, which are derived from analyzing previous frames and performing motion estimation. These *P*-frames require about one-third as many bits per frame as *I*-frames. *B*-frames, which lie between two *I*-frames or *P*-frames, are bidirectionally encoded, making use of past and future frames. The *B*-frames require only about one-ninth of the data per frame, compared with *I*-frames.

These different compression ratios for the frames lead to different data rates, so that buffers are required at both the encoder output and the decoder input to ensure that the sustained data rate is constant. One difference between MPEG-1 and MPEG-2 is that MPEG-2 allows for a variety of data-buffer sizes, to accommodate different picture dimensions and to prevent buffer under- and overflows.

The data required to decode a single picture is embedded in the *picture layer*, which consists of a number of horizontal *slice layers*, each containing several macroblocks. Each *macroblock layer*, in turn, is made up of a number of individual blocks. The picture undergoes DCT processing, with the slice layer providing a means of synchronization, holding the precise position of the slice within the image frame.

MPEG-2 places the motion vectors into the coded macroblocks for *P*- and *B*- frames; these are used to improve the reconstruction of predicted pictures. MPEG-2 supports both field- and frame-based prediction, thus accommodating interlaced signals.

The last layer of MPEG-2's video structure is the *block layer*, which provides the DCT coefficients of either the transformed image information for *I*-frames or the residual prediction error of *B*- and *P*- frames.

7.2.2a Slices

Two or more contiguous macroblocks within the same row are grouped together to form *slices* [1]. The order of the macroblocks within a slice is the same as the conventional television raster scan, being from left to right.

Slices provide a convenient mechanism for limiting the propagation of errors. Because the coded bit stream consists mostly of variable-length code words, any uncorrected transmission

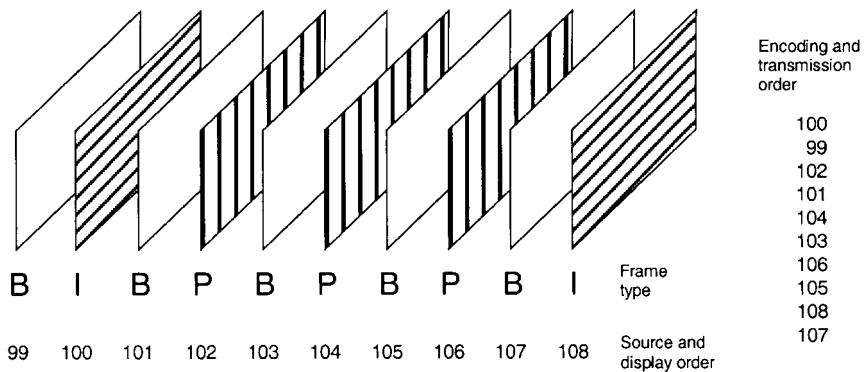


Figure 7.2.1 Sequence of video frames for the MPEG-2/ATSC DTV system. (From [1]. Used with permission.)

errors will cause a decoder to lose its sense of code word alignment. Each slice begins with a slice start code. Because the MPEG code word assignment guarantees that no legal combination of code words can emulate a start code, the slice start code can be used to regain the sense of code-word alignment after an error. Therefore, when an error occurs in the data stream, the decoder can skip to the start of the next slice and resume correct decoding.

The number of slices affects the compression efficiency; partitioning the data stream to have more slices provides for better error recovery, but claims bits that could otherwise be used to improve picture quality.

In the DTV system, the initial macroblock of every horizontal row of macroblocks is also the beginning of a slice, with a possibility of several slices across the row.

7.2.2b Pictures, Groups of Pictures, and Sequences

The primary coding unit of a video sequence is the individual video frame or picture [1]. A video picture consists of the collection of slices, constituting the *active picture area*.

A *video sequence* consists of a collection of two or more consecutive pictures. A video sequence commences with a sequence header and is terminated by an end-of-sequence code in the data stream. A video sequence may contain additional sequence headers. Any video-sequence header can serve as an *entry point*. An entry point is a point in the coded video bit stream after which a decoder can become properly initialized and correctly parse the bit-stream syntax.

Two or more pictures (frames) in sequence may be combined into a GOP to provide boundaries for interframe picture coding and registration of time code. GOPs are optional within both MPEG-2 and the ATSC DTV system. Figure 7.2.1 illustrates a typical time sequence of video frames.

7-34 Compression Technologies for Audio

I-Frames

Some elements of the compression process exploit only the spatial redundancy within a single picture (frame or field) [1]. These processes constitute intraframe coding, and do not take advantage of the temporal correlation addressed by temporal prediction (interframe) coding. Frames that do not use any interframe coding are referred to as *I*-frames (where “I” denotes *intraframe-coded*). The ATSC video-compression system utilizes both intraframe and interframe coding.

The use of periodic *I*-frames facilitates receiver initializations and channel acquisition (for example, when the receiver is turned on or the channel is changed). The decoder also can take advantage of the intraframe coding mode when noncorrectable channel errors occur. With motion-compensated prediction, an initial frame must be available at the decoder to start the prediction loop. Therefore, a mechanism must be built into the system so that if the decoder loses synchronization for any reason, it can rapidly reacquire tracking.

The frequency of occurrence of *I*-pictures may vary and is selected at the encoder. This allows consideration to be given to the need for random access and the location of scene cuts in the video sequence.

P-Frames

P-frames, where the temporal prediction is in the forward direction only, allow the exploitation of interframe coding techniques to improve the overall compression efficiency and picture quality [1]. *P*-frames may include portions that are only intraframe-coded. Each macroblock within a *P*-frame can be either forward-predicted or intraframe-coded.

B-Frames

The *B*-frame is a picture type within the coded video sequence that includes prediction from a future frame as well as from a previous frame [1]. The referenced future or previous frames, sometimes called *anchor frames*, are in all cases either *I*- or *P*-frames.

The basis of the *B*-frame prediction is that a video frame is correlated with frames that occur in the past as well as those that occur in the future. Consequently, if a future frame is available to the decoder, a superior prediction can be formed, thus saving bits and improving performance. Some of the consequences of using future frames in the prediction are:

- The *B*-frame cannot be used for predicting future frames.
- The transmission order of frames is different from the displayed order of frames.
- The encoder and decoder must reorder the video frames, thereby increasing the total latency.

In the example illustrated in Figure 7.2.1, there is one *B*-frame between each pair of *I*- and *P*-frames. Each frame is labeled with both its display order and transmission order. The *I* and *P* frames are transmitted out of sequence, so the video decoder has both anchor frames decoded and available for prediction.

B-frames are used for increasing the compression efficiency and perceived picture quality when encoding latency is not an important factor. The use of *B*-frames increases coding efficiency for both interlaced- and progressive-scanned material. *B*-frames are included in the DTV system because the increase in compression efficiency is significant, especially with progressive scanning. The choice of the number of bidirectional pictures between any pair of reference (*I* or *P*) frames can be determined at the encoder.

Motion Estimation

The efficiency of the compression algorithm depends on, first, the creation of an estimate of the image being compressed and, second, subtraction of the pixel values of the estimate or prediction from the image to be compressed [1]. If the estimate is good, the subtraction will leave a very small residue to be transmitted. In fact, if the estimate or prediction were perfect, the difference would be zero for all the pixels in the frame of differences, and no new information would need to be sent; this condition can be approached for still images.

If the estimate is not close to zero for some pixels or many pixels, those differences represent information that needs to be transmitted so that the decoder can reconstruct a correct image. The kinds of image sequences that cause large prediction differences include severe motion and/or sharp details.

7.2.2c Vector Search Algorithm

The video-coding system uses motion-compensated prediction as part of the data-compression process [1]. Thus, macroblocks in the current frame of interest are predicted by macroblock-sized regions in previously transmitted frames. Motion compensation refers to the fact that the locations of the macroblock-sized regions in the reference frame can be offset to account for local motions. The macroblock offsets are known as *motion vectors*.

The DTV standard does not specify how encoders should determine motion vectors. One possible approach is to perform an exhaustive search to identify the vertical and horizontal offsets that minimize the total difference between the offset region in the reference frame and the macroblock in the frame to be coded.

7.2.2d Motion-Vector Precision

The estimation of interframe displacement is calculated with half-pixel precision, in both vertical and horizontal dimensions [1]. As a result, the displaced macroblock from the previous frame can be displaced by noninteger displacements and will require interpolation to compute the values of displaced picture elements at locations not in the original array of samples. Estimates for half-pixel locations are computed by averages of adjacent sample values.

7.2.2e Motion-Vector Coding

Motion vectors within a slice are differenced, so that the first value for a motion vector is transmitted directly, and the following sequence of motion-vector differences is sent using *variable-length codes* (VLC) [1]. Motion vectors are constrained so that all pixels from the motion-compensated prediction region in the reference picture fall within the picture boundaries.

7.2.2f Encoder Prediction Loop

The encoder prediction loop, shown in the simplified block diagram of Figure 7.2.2, is the heart of the video-compression system for DTV [1]. The prediction loop contains a prediction function that estimates the picture values of the next picture to be encoded in the sequence of successive pictures that constitute the TV program. This prediction is based on previous information that is

7-36 Compression Technologies for Audio

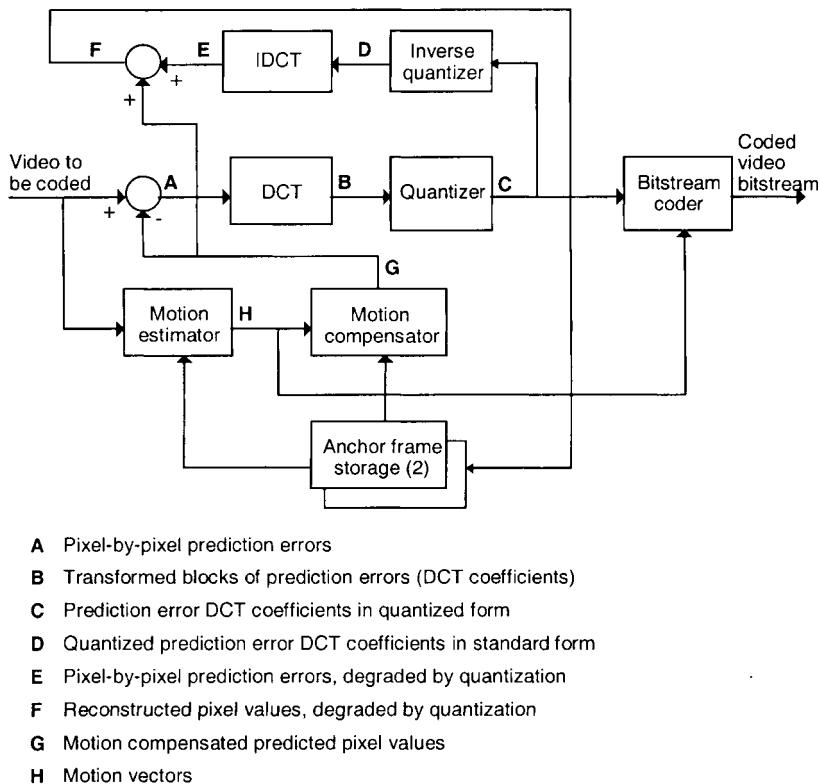


Figure 7.2.2 Simplified encoder prediction loop. (From [1]. Used with permission.)

available within the loop, derived from earlier pictures. The transmission of the predicted compressed information works because the same information used to make the prediction also is available at the receiving decoder (barring transmission errors, which are usually infrequent within the primary coverage area).

The subtraction of the predicted picture values from the new picture to be coded is at the core of predictive coding. The goal is to do such a good job of predicting the new values that the result of the subtraction function at the beginning of the prediction loop is zero or close to zero most of the time.

The prediction differences are computed separately for the luminance and two chrominance components before further processing. As explained in previous discussion of *I*-frames, there are times when prediction is not used, for part of a frame or for an entire frame.

Spatial Transform Block—DCT

The image prediction differences (sometimes referred to as *prediction errors*) are organized into 8×8 blocks, and a spatial transform is applied to the blocks of difference values [1]. In the intraframe case, the spatial transform is applied to the raw, undifferenced picture data. The luminance and two chrominance components are transformed separately. Because the chrominance data is subsampled vertically and horizontally, each 8×8 block of chrominance (C_b or C_r) data corresponds to a 16×16 macroblock of luminance data, which is not subsampled.

The spatial transform used is the discrete cosine transform. The formula for transforming the data is given by:

$$F(u, v) = \frac{1}{4} C(u) C(v) \sum_{x=0}^{7} \sum_{y=0}^{7} f(x, y) \cos\left[\frac{(2x+1)u\pi}{16}\right] \cos\left[\frac{(2y+1)v\pi}{16}\right] \quad (7.2.1)$$

where x and y are pixel indices within an 8×8 block, u and v are DCT coefficient indices within an 8×8 block, and:

$$C(w) = \frac{1}{\sqrt{2}} \quad \text{for } w = 0 \quad (7.2.2)$$

$$C(w) = 1 \quad \text{for } w = 1, 2, \dots, 7 \quad (7.2.3)$$

Thus, an 8×8 array of numbers $f(x, y)$ is the input to a mathematical formula, and the output is an 8×8 array of different numbers, $F(u, v)$. The inverse transform is given by:

$$f(x, y) = \frac{1}{4} \sum_{u=0}^{7} \sum_{v=0}^{7} C(u) C(v) F(u, v) \cos\left[\frac{(2x+1)u\pi}{16}\right] \cos\left[\frac{(2y+1)v\pi}{16}\right] \quad (7.2.4)$$

It should be noted that for the DTV implementation, the *inverse discrete cosine transform* (IDCT) must conform to the specifications noted in [2].

In principle, applying the IDCT to the transformed array would yield exactly the same array as the original. In that sense, transforming the data does not modify the data, but merely represents it in a different form.

The decoder uses the inverse transformation to approximately reconstruct the arrays that were transformed at the encoder, as part of the process of decoding the received compressed data. The approximation in that reconstruction is controlled in advance during the encoding process for the purpose of minimizing the visual effects of coefficient inaccuracies while reducing the quantity of data that needs to be transmitted.

Quantizer

The process of transforming the original data organizes the information in a way that exposes the spatial frequency components of the images or image differences [1]. Using information about

7-38 Compression Technologies for Audio

the response of the human visual system to different spatial frequencies, the encoder can selectively adjust the precision of transform coefficient representation. The goal is to include as much information about a particular spatial frequency as necessary—and as possible, given the constraints on data transmission—while not using more precision than is needed, based upon visual perception criteria.

For example, in a portion of a picture that is “busy” with a great deal of detail, imprecision in reconstructing spatial high-frequency components in a small region might be masked by the picture’s local “busyness.” On the other hand, highly precise representation and reconstruction of the average value or dc term of the DCT block would be important in a smooth area of sky. The dc $F(0,0)$ term of the transformed coefficients represents the average of the original 64 coefficients.

As stated previously, the DCT of each 8×8 block of pixel values produces an 8×8 array of DCT coefficients. The relative precision accorded to each of the 64 DCT coefficients can be selected according to its relative importance in human visual perception. The relative coefficient precision information is represented by a *quantizer matrix*, which is an 8×8 array of values. Each value in the quantizer matrix represents the coarseness of quantization of the related DCT coefficient.

Two types of quantizer matrices are supported:

- A matrix used for macroblocks that are intraframe-coded
- A matrix used for non-intraframe-coded macroblocks

The video-coding system defines default values for both the intraframe-quantizer and the non-intraframe-quantizer matrices. Either or both of the quantizer matrices can be overridden at the picture level by transmission of appropriate arrays of 64 values. Any quantizer matrix overrides stay in effect until the following sequence start code.

The transform coefficients, which represent the bulk of the actual coded video information, are quantized to various degrees of coarseness. As indicated previously, some portions of the picture will be more affected in appearance than others by the loss of precision through coefficient quantization. This phenomenon is exploited by the availability of the quantizer scale factor, which allows the overall level of quantization to vary for each macroblock. Consequently, entire macroblocks that are deemed to be visually less important can be quantized more coarsely, resulting in fewer bits being needed to represent the picture.

For each coefficient other than the dc coefficient of intraframe-coded blocks, the quantizer scale factor is multiplied by the corresponding value in the appropriate quantizer matrix to form the quantizer step size. Quantization of the dc coefficients of intraframe-coded blocks is unaffected by the quantizer scale factor and is governed only by the (0, 0) element of the intraframe-quantizer matrix, which always is set to be 8 (ISO/IEC 13818-2).

Entropy Coder

An important effect of the quantization of transform coefficients is that many coefficients will be rounded to zero after quantization [1]. In fact, a primary method of controlling the encoded data rate is the control of quantization coarseness, because a coarser quantization leads to an increase in the number of zero-value quantized coefficients.

Inverse Quantizer

At the decoder, the coded coefficients are decoded, and an 8×8 block of quantized coefficients is reconstructed [1]. Each of these 64 coefficients is *inverse-quantized* according to the prevailing quantizer matrix, quantizer scale, and frame type. The result of inverse quantization is a block of 64 DCT coefficients.

Inverse Spatial Transform Block—IDCT

The decoded and inverse-quantized coefficients are organized as 8×8 blocks of DCT coefficients, and the inverse discrete cosine transform is applied to each block [1]. This results in a new array of pixel values, or pixel difference values that correspond to the output of the subtraction at the beginning of the prediction loop. If the prediction loop was in the interframe mode, the values will be pixel differences. If the loop was in the intraframe mode, the inverse transform will produce pixel values directly.

Motion Compensator

If a portion of the image has not moved, then it is easy to see that a subtraction of the old portion from the new portion of the image will produce zero or nearly zero pixel differences, which is the goal of the prediction [1]. If there has been movement in the portion of the image under consideration, however, the direct pixel-by-pixel differences generally will not be zero, and might be statistically very large. The motion in most natural scenes is organized, however, and can be approximately represented locally as a translation in most cases. For this reason, the video-coding system allows for *motion-compensated* prediction, whereby macroblock-sized regions in the reference frame may be translated vertically and horizontally with respect to the macroblock being predicted, to compensate for local motion.

The pixel-by-pixel differences between the current macroblock and the motion-compensated prediction are transformed by the DCT and quantized using the composition of the non-intraframe-quantizer matrix and the quantizer scale factor. The quantized coefficients then are coded.

7.2.2g Dual Prime Prediction Mode

The dual prime prediction mode is an alternative “special” prediction mode that is built on field-based motion prediction but requires fewer transmitted motion vectors than conventional field-based prediction [1]. This mode of prediction is available only for interlaced material and only when the encoder configuration does not use *B*-frames. This mode of prediction can be particularly useful for improving encoder efficiency for low-delay applications.

The basis of dual prime prediction is that field-based predictions of both fields in a macroblock are obtained by averaging two separate predictions, which are predicted from the two nearest decoded fields in time. Each of the macroblock fields is predicted separately, although the four vectors (one pair per field) used for prediction all are derived from a single transmitted field-based motion vector. In addition to the single field-based motion vector, a small *differential vector* (limited to vertical and horizontal component values of +1, 0, and -1) also is transmitted for each macroblock. Together, these vectors are used to calculate the pairs of motion vectors for each macroblock. The first prediction in the pair is simply the transmitted field-based motion vector. The second prediction vector is obtained by combining the differential vector with a

7-40 Compression Technologies for Audio

scaled version of the first vector. After both predictions are obtained, a single prediction for each macroblock field is calculated by averaging each pixel in the two original predictions. The final averaged prediction then is subtracted from the macroblock field being encoded.

7.2.2h Adaptive Field/Frame Prediction Mode

Interlaced pictures may be coded in one of two ways: either as two separate fields or as a single frame [1]. When the picture is coded as separate fields, all of the codes for the first field are transmitted as a unit before the codes for the second field. When the picture is coded as a frame, information for both fields is coded for each macroblock.

When frame-based coding is used with interlaced pictures, each macroblock may be selectively coded using either field prediction or frame prediction. When frame prediction is used, a motion vector is applied to a picture region that is made up of both parity fields interleaved together. When field prediction is used, a motion vector is applied to a region made up of scan lines from a single field. Field prediction allows the selection of either parity field to be used as a reference for the field being predicted.

7.2.2i Image Refresh

As discussed previously, a given picture may be sent by describing the differences between it and one or two previously transmitted pictures [1]. For this scheme to work, there must be some way for decoders to become initialized with a valid picture upon tuning into a new channel, or to become reinitialized with a valid picture after experiencing transmission errors. Additionally, it is necessary to limit the number of consecutive predictions that can be performed in a decoder to control the buildup of errors resulting from *IDCT mismatch*.

IDCT mismatch occurs because the video-coding system, by design, does not completely specify the results of the IDCT operation. MPEG did not fully specify the results of the IDCT to allow for evolutionary improvements in implementations of this computationally intensive operation. As a result, it is possible for the reconstructed pictures in a decoder to “drift” from those in the encoder if many successive predictions are used, even in the absence of transmission errors. To control the amount of drift, each macroblock is required to be coded without prediction (intraframe-coded) at least once in any 132 consecutive frames.

The process whereby a decoder becomes initialized or reinitialized with valid picture data—without reference to previously transmitted picture information—is termed *image refresh*. Image refresh is accomplished by the use of intraframe-coded macroblocks. The two general classes of image refresh, which can be used either independently or jointly, are:

- Periodic transmission of *I*-frames
- Progressive refresh

Periodic Transmission of I-Frames

One simple approach to image refresh is to periodically code an entire frame using only intraframe coding [1]. In this case, the intra-coded frame is typically an *I*-frame. Although prediction is used within the frame, no reference is made to previously transmitted frames. The period between successive intracoded frames may be constant, or it may vary. When a receiver

tunes into a new channel where *I*-frame coding is used for image refresh, it may perform the following steps:

- Ignore all data until receipt of the first sequence header
- Decode the sequence header, and configure circuits based on sequence parameters
- Ignore all data until the next received *I*-frame
- Commence picture decoding and presentation

When a receiver processes data that contains uncorrectable errors in an *I*- or *P*-frame, there typically will be a propagation of picture errors as a result of predictive coding. Pictures received after the error may be decoded incorrectly until an error-free *I*-frame is received.

Progressive Refresh

An alternative method for accomplishing image refresh is to encode only a portion of each picture using the intraframe mode [1]. In this case, the intraframe-coded regions of each picture should be chosen in such a way that, over the course of a reasonable number of frames, all macroblocks are coded intraframe at least once. In addition, constraints might be placed on motion-vector values to avoid possible contamination of refreshed regions through predictions using unrefreshed regions in an uninitialized decoder.

7.2.2j Discrete Cosine Transform

Predictive coding in the MPEG-2 compression algorithm exploits the temporal correlation in the sequence of image frames [1]. Motion compensation is a refinement of that temporal prediction, which allows the coder to account for apparent motions in the image that can be estimated. Aside from temporal prediction, another source of correlation that represents redundancy in the image data is the spatial correlation within an image frame or field. This spatial correlation of images, including parts of images that contain apparent motion, can be accounted for by a spatial transform of the prediction differences. In the intraframe-coding case, where there is by definition no attempt at prediction, the spatial transform applies to the actual picture data. The effect of the spatial transform is to concentrate a large fraction of the signal energy in a few transform coefficients.

To exploit spatial correlation in intraframe and predicted portions of the image, the image-prediction residual pixels are represented by their DCT coefficients. For typical images, a large fraction of the energy is concentrated in a few of these coefficients. This makes it possible to code only a few coefficients without seriously affecting the picture quality. The DCT is used because it has good energy-compaction properties and results in real coefficients. Furthermore, numerous fast computational algorithms exist for implementation of DCT.

Blocks of 8×8 Pixels

Theoretically, a large DCT will outperform a small DCT in terms of coefficient decorrelation and block energy compaction [1]. Better overall performance can be achieved, however, by subdividing the frame into many smaller regions, each of which is individually processed.

If the DCT of the entire frame is computed, the whole frame is treated equally. For a typical image, some regions contain a large amount of detail, and other regions contain very little.

7-42 Compression Technologies for Audio

Exploiting the changing characteristics of different images and different portions of the same image can result in significant improvements in performance. To take advantage of the varying characteristics of the frame over its spatial extent, the frame is partitioned into blocks of 8×8 pixels. The blocks then are independently transformed and adaptively processed based on their local characteristics. Partitioning the frame into small blocks before taking the transform not only allows spatially adaptive processing, but also reduces the computational and memory requirements. The partitioning of the signal into small blocks before computing the DCT is referred to as the *block DCT*.

An additional advantage of using the DCT domain representation is that the DCT coefficients contain information about the spatial frequency content of the block. By utilizing the spatial frequency characteristics of the human visual system, the precision with which the DCT coefficients are transmitted can be in accordance with their perceptual importance. This is achieved through the quantization of these coefficients, as explained in the following sections.

Adaptive Field/Frame DCT

As noted previously, the DCT makes it possible to take advantage of the typically high degree of spatial correlation in natural scenes [1]. When interlaced pictures are coded on a frame basis, however, it is possible that significant amounts of motion result in relatively low spatial correlation in some regions. This situation is accommodated by allowing the DCTs to be computed either on a field basis or on a frame basis. The decision to use field- or frame-based DCT is made individually for each macroblock.

Adaptive Quantization

The goal of video compression is to maximize the video quality at a given bit rate, and this requires a careful distribution of the limited number of available bits [1]. By exploiting the perceptual irrelevancy and statistical redundancy within the DCT domain representation, an appropriate bit allocation can yield significant improvements in performance. Quantization is performed to reduce the precision of the DCT coefficient values, and through quantization and code word assignment, the actual bit-rate compression is achieved. The quantization process is the source of virtually all the loss of information in the compression algorithm. This is an important point, as it simplifies the design process and facilitates fine-tuning of the system.

The degree of subjective picture degradation caused by coefficient quantization tends to depend on the nature of the scenery being coded. Within a given picture, distortions of some regions may be less apparent than in others. The video-coding system allows for the level of quantization to be adjusted for each macroblock in order to save bits, where possible, through coarse quantization.

Perceptual Weighting

The human visual system is not uniformly sensitive to coefficient quantization error [1]. Perceptual weighting of each source of coefficient quantization error is used to increase quantization coarseness, thereby lowering the bit rate. The amount of visible distortion resulting from quantization error for a given coefficient depends on the coefficient number, or frequency, the local brightness in the original image, and the duration of the temporal characteristic of the error. The dc coefficient error results in mean value distortion for the corresponding block of pixels, which

can expose block boundaries. This is more visible than higher-frequency coefficient error, which appears as noise or texture.

Displays and the HVS exhibit nonuniform sensitivity to detail as a function of local average brightness. Loss of detail in dark areas of the picture is not as visible as it is in brighter areas. Another opportunity for bit savings is presented in textured areas of the picture, where high-frequency coefficient error is much less visible than in relatively flat areas. Brightness and texture weighting require analysis of the original image because these areas may be well predicted. Additionally, distortion can be easily masked by limiting its duration to one or two frames. This effect is most profitably used after scene changes, where the first frame or two can be greatly distorted without perceptible artifacts at normal speed.

When transform coefficients are being quantized, the differing levels of perceptual importance of the various coefficients can be exploited by “allocating the bits” to shape the quantization noise into the perceptually less important areas. This can be accomplished by varying the relative step sizes of the quantizers for the different coefficients. The perceptually important coefficients may be quantized with a finer step size than the others. For example, low spatial frequency coefficients may be quantized finely, and the less important high-frequency coefficients may be quantized more coarsely. A simple method to achieve different step sizes is to normalize or weight each coefficient based on its visual importance. All of the normalized coefficients may then be quantized in the same manner, such as rounding to the nearest integer (uniform quantization). Normalization or weighting effectively scales the quantizer from one coefficient to another. The MPEG-2 video-compression system utilizes perceptual weighting, where the different DCT coefficients are weighted according to a perceptual criterion prior to uniform quantization. The perceptual weighting is determined by quantizer matrices. The compression system allows for modifying the quantizer matrices before each picture.

7.2.2k Entropy Coding of Video Data

Quantization creates an efficient, discrete representation for the data to be transmitted [1]. Code word assignment takes the quantized values and produces a digital bit stream for transmission. Hypothetically, the quantized values could be simply represented using uniform- or fixed-length code words. Under this approach, every quantized value would be represented with the same number of bits. Greater efficiency—in terms of bit rate—can be achieved with entropy coding.

Entropy coding attempts to exploit the statistical properties of the signal to be encoded. A signal, whether it is a pixel value or a transform coefficient, has a certain amount of information, or entropy, based on the probability of the different possible values or events occurring. For example, an event that occurs infrequently conveys much more new information than one that occurs often. The fact that some events occur more frequently than others can be used to reduce the average bit rate.

Huffman Coding

Huffman coding, which is utilized in the ATSC DTV video-compression system, is one of the most common entropy-coding schemes [1]. In Huffman coding, a code book is generated that can approach the minimum average description length (in bits) of events, given the probability distribution of all the events. Events that are more likely to occur are assigned shorter-length code words, and those less likely to occur are assigned longer-length code words.

7-44 Compression Technologies for Audio

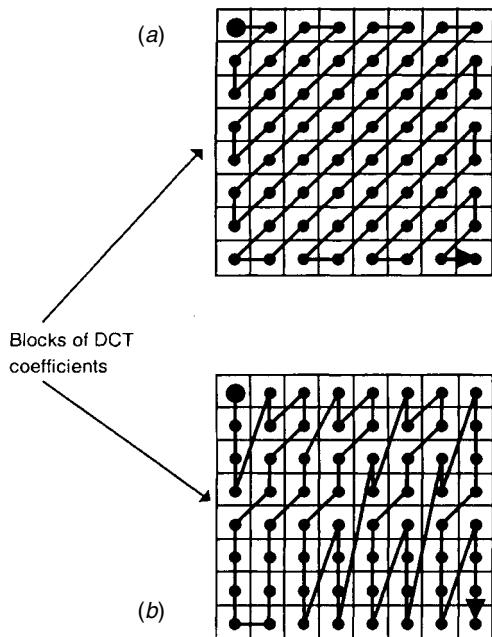


Figure 7.2.3 Scanning of coefficient blocks: (a) alternate scanning of coefficients, (b) zigzag scanning of coefficients. (From [1]. Used with permission.)

Run Length Coding

In video compression, most of the transform coefficients frequently are quantized to zero [1]. There may be a few non-zero low-frequency coefficients and a sparse scattering of non-zero high-frequency coefficients, but most of the coefficients typically have been quantized to zero. To exploit this phenomenon, the 2-dimensional array of transform coefficients is reformatted and prioritized into a 1-dimensional sequence through either a zigzag- or alternate-scanning process. This results in most of the important non-zero coefficients (in terms of energy and visual perception) being grouped together early in the sequence. They will be followed by long runs of coefficients that are quantized to zero. These zero-value coefficients can be efficiently represented through *run length encoding*.

In run length encoding, the number (run) of consecutive zero coefficients before a non-zero coefficient is encoded, followed by the non-zero coefficient value. The run length and the coefficient value can be entropy-coded, either separately or jointly. The scanning separates most of the zero and the non-zero coefficients into groups, thereby enhancing the efficiency of the run length encoding process. Also, a special *end-of-block* (EOB) marker is used to signify when all of the remaining coefficients in the sequence are equal to zero. This approach can be extremely efficient, yielding a significant degree of compression.

In the alternate-/zigzag-scan technique, the array of 64 DCT coefficients is arranged in a 1-dimensional vector before run length/amplitude code word assignment. Two different 1-dimensional arrangements, or *scan types*, are allowed, generally referred to as *zigzag scan* (shown in Figure 7.2.3a) and *alternate scan* (shown in Figure 7.2.3b). The scan type is specified before coding each picture and is permitted to vary from picture to picture.

Channel Buffer

Whenever entropy coding is employed, the bit rate produced by the encoder is variable and is a function of the video statistics [1]. Because the bit rate permitted by the transmission system is less than the peak bit rate that may be produced by the variable-length coder, a *channel buffer* is necessary at the decoder. This buffering system must be carefully designed. The buffer controller must allow efficient allocation of bits to encode the video and also ensure that no overflow or underflow occurs.

Buffer control typically involves a feedback mechanism to the compression algorithm whereby the amplitude resolution (quantization) and/or spatial, temporal, and color resolution may be varied in accordance with the instantaneous bit-rate requirements. If the bit rate decreases significantly, a finer quantization can be performed to increase it.

The ATSC DTV standard specifies a channel buffer size of 8 Mbits. The *model buffer* is defined in the DTV video-coding system as a reference for manufacturers of both encoders and decoders to ensure interoperability. To prevent overflow or underflow of the model buffer, an encoder may maintain measures of buffer occupancy and scene complexity. When the encoder needs to reduce the number of bits produced, it can do so by increasing the general value of the quantizer scale, which will increase picture degradation. When it is able to produce more bits, it can decrease the quantizer scale, thereby decreasing picture degradation.

Decoder Block Diagram

As shown in Figure 7.2.4, the ATSC DTV video decoder contains elements that invert, or undo, the processing performed in the encoder [1]. The incoming coded video bit stream is placed in the channel buffer, and bits are removed by a *variable length decoder* (VLD).

The VLD reconstructs 8×8 arrays of quantized DCT coefficients by decoding run length/amplitude codes and appropriately distributing the coefficients according to the scan type used. These coefficients are dequantized and transformed by the IDCT to obtain pixel values or prediction errors.

In the case of interframe prediction, the decoder uses the received motion vectors to perform the same prediction operation that took place in the encoder. The prediction errors are summed with the results of motion-compensated prediction to produce pixel values.

7.2.2I Spatial and S/N Scalability

Because MPEG-2 was designed in anticipation of the need for handling different picture sizes and resolutions, including standard-definition television and high-definition television, provisions were made for a hierarchical split of the picture information into a base layer and two enhancement layers [1]. In this way, SDTV decoders would not be burdened with the cost of decoding an HDTV signal.

An encoder for this scenario could work as follows. The HDTV signal would be used as the starting point. It would be spatially filtered and subsampled to create a standard resolution image, which then would be MPEG-encoded. The higher-definition information could be included in an enhancement layer.

Another use of a hierarchical split would be to provide different picture quality without changing the spatial resolution. An encoder quantizer block could realize both coarse and fine filtering levels. Better error correction could be provided for the more coarse data, so that as sig-

7-46 Compression Technologies for Audio

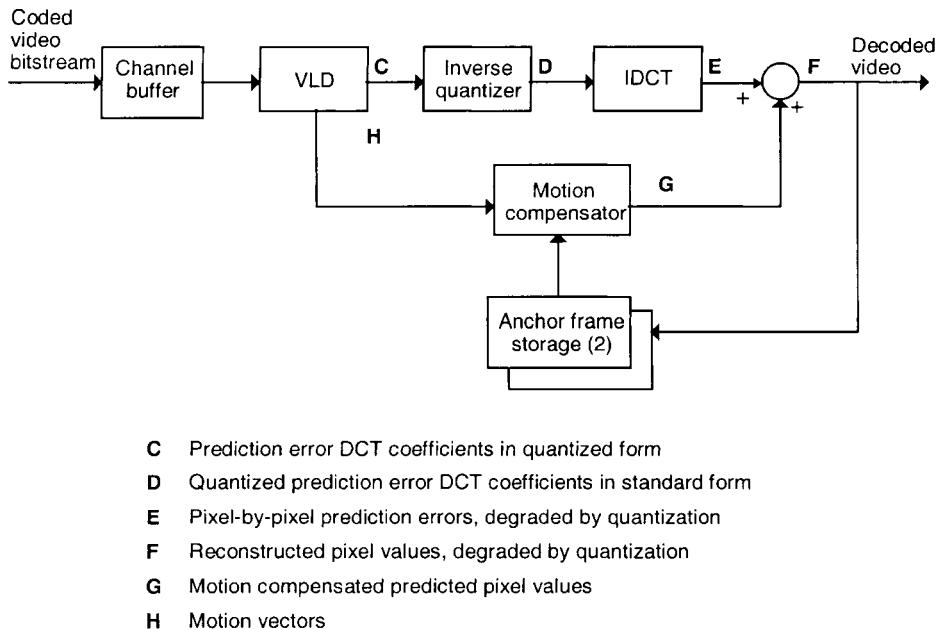


Figure 7.2.4 ATSC DTV video system decoder functional block diagram. (*From [1]. Used with permission.*)

nal strength weakened, a step-by-step reduction in the picture signal-to-noise ratio would occur in a way similar to that experienced in broadcast analog signals today. Viewers with poor reception, therefore, would experience a more graceful degradation in picture quality instead of a sudden dropout.

7.2.3 References



1. ATSC, "Guide to the Use of the ATSC Digital Television Standard," Advanced Television Systems Committee, Washington, D.C., doc. A/54, Oct. 4, 1995.
2. "IEEE Standard Specifications for the Implementation of 8×8 Inverse Discrete Cosine Transform," std. 1180-1990, Dec. 6, 1990.

Chapter
7.3

DTV Audio Encoding and Decoding

Jerry C. Whitaker, Editor-in-Chief

7.3.1 Introduction

Monophonic sound is the simplest form of aural communication. A wide range of acceptable listening positions are practical, although it is obvious from most positions that the sound is originating from one source rather than occurring in the presence of the listener. Consumers have accepted this limitation without much thought in the past because it was all that was available. However, monophonic sound creates a poor illusion of the sound field that the program producer might want to create.

Two channel stereo improves the illusion that the sound is originating in the immediate area of the reproducing system. Still, there is a smaller acceptable listening area. It is difficult to keep the sound image centered between the left and right speakers, so that the sound and the action stay together as the listener moves in the room.

The AC-3 surround sound system is said to have 5.1 channels because there is a left, right, center, left surround, and right surround, which make up the 5 channels. A sixth channel is reserved for the lower frequencies and consumes only 120 Hz of the bandwidth; it is referred to as the 0.1 or *low-frequency enhancement* (LFE) channel. The center channel restores the variety of listening positions possible with monophonic sound.

The AC-3 system is effective in providing either an enveloping (ambient) sound field or allowing precise placement and movement of special effects because of the channel separation afforded by the multiple speakers in the system.

For efficient and reliable interconnection of audio devices, standardization of the interface parameters is of critical importance. The primary interconnection scheme for professional digital audio systems is AES Audio.

7.3.2 AES Audio

AES audio is a standard defined by the Audio Engineering Society and the European Broadcasting Union. Each AES stream carries two audio channels, which can be either a stereo pair or two independent feeds. The signals are pulse code modulated (PCM) data streams carrying digitized audio. Each sample is quantized to 20 or 24 bits, creating an audio *sample word*. Each word is then formatted to form a *subframe*, which is multiplexed with other subframes to form the AES

7-48 Compression Technologies for Audio

Table 7.3.1 Theoretical S/N as a Function of the Number of Sampling Bits

Number of Sampling Bits	Resolution (number of quantizing steps)	Maximum Theoretical S/N
18	262,144	110 dB
20	1,048,576	122 dB
24	16,777,216	146 dB

digital audio stream. The AES stream can then be serialized and transmitted over coaxial or twisted-pair cable. The sampling rates supported range from 32 to 50 kHz. Common rates and applications include the following:

- 32 kHz—used for radio broadcast links
- 44.1 kHz—used for CD players
- 48 kHz—used for professional recording and production

Although 18-bit sampling was commonly used in the past, 20 bits has become prevalent today.

At 24 bits/sample, the S/N is 146 dB. This level of performance is generally reserved for high-end applications such as film recording and CD mastering. Table 7.3.1 lists the theoretical S/N ratios as a function of sampling bits for audio A/D conversion.

Of particular importance is that the AES format is designed to be independent of the audio conversion sample rate. The net data rate is exactly 64 times the sample rate, which is generally 48 kHz for professional applications. Thus, the most frequently encountered bit rate for AES3 data is 3.072 Mbits/s.

The AES3-1992 standard document precisely defines the AES3 twisted pair interconnection scheme. The signal, which is transmitted on twisted pair copper cable in a balanced format, is *biphasic coded*. Primary signal parameters include the following:

- Output level can range from 2–10 V p-p
- Source impedance 110Ω
- Receiver sensitivity 200 mV minimum
- Input impedance is recommended to be 110Ω
- Interconnecting cable characteristic impedance 110Ω

Electrical interface guidelines also have been set by the SMPTE and AES3 committees to permit transmission of AES3 data on coaxial cable. This single-ended interface is known as AES3-ID. The signal level, when terminated with 75Ω is 1 V p-p, ± 20 percent. The source impedance is 75Ω .

AES3 is inherently synchronous. A master local digital audio reference is normally used so that all audio equipment will be frequency- and phase-locked. The master reference can originate from the digital audio equipment in a single room or an external master system providing a reference signal for larger facilities.

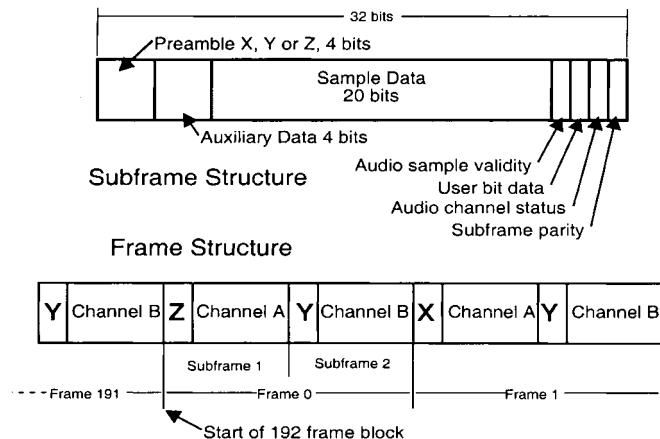


Figure 7.3.1 AES audio data format structure. (From [1]. Used with permission.)

7.3.2a AES3 Data Format

The basic format structure of the AES data frames is shown in Figure 7.3.1. Each sample is carried by a subframe containing the following elements [1]:

- 20 bits of sample data
- 4 bits of auxiliary data, which may be used to extend the sample to 24 bits
- 4 additional bits of data
- A preamble

Two subframes make up a frame, which contains one sample from each of the two channels. Frames are further grouped into 192-frame blocks, which define the limits of *user data* and *channel status data* blocks. A special preamble indicates the channel identity for each sample (X or Y preamble) and the start of a 192-frame block (Z preamble). To minimize the direct current (dc) component on the transmission line, facilitate clock recovery, and make the interface polarity insensitive, the data is channel coded in the biphasemark mode.

The preambles specifically violate the biphasemark rules for easy recognition and to ensure synchronization. When digital audio is embedded in the serial digital video data stream, the start of the 192-frame block is indicated by the Z bit, which corresponds to the occurrence of the Z-type preamble.

The *validity bit* indicates whether the audio sample bits in the subframe are suitable for conversion to an analog audio signal. User data is provided to carry other information, such as time code. Channel status data contains information associated with each audio channel.

There are three levels of implementation of the channel status data: minimum, standard, and enhanced. The standard implementation is recommended for use in professional video applications; the channel status data typically contains information about signal emphasis, sampling frequency, channel mode (stereo, mono, etc.), use of auxiliary bits (extend audio data to 24 bits or other use), and a CRC for error checking of the total channel status block.

7-50 Compression Technologies for Audio

7.3.2b SMPTE 324M

SMPTE 324M (proposed at this writing) defines a synchronous, self-clocking serial interface for up to 12 channels of linearly encoded audio and auxiliary data [2]. The interface is designed to allow multiplexing of six two-channel streams compliant with AES3. Audio sampled at 48 kHz and clock-locked to video is the preferred implementation for studio applications. However, the 324M interlace supports any frequency of operation supported by AES3, provided that all the audio channels are sampled by a common clock. Ideally, all the channels should be *audio synchronous* for guaranteed audio phase coherence. An audio channel is defined as being synchronous with another when the two channels are running from the same clock and the analog inputs are concurrently sampled.

The 324M standard is intended to provide a reliable method of distributing multiple cophased channels of digital audio around the studio without losing the initial relative sample-phase relationship. A mechanism is provided to allow more than one 12-channel stream to be realigned after a relative misalignment of up to ± 8 samples.

The interface, intended to be compatible with the complete range of digital television scanning standards and standard film rates, can be used for distribution of multiple channels of audio in either a pre-mix or post-mix situation. In the post-mix case, channel assignment is defined in SMPTE 320M.

7.3.2c Audio Compression

Efficient recording and/or transmission of digital audio signals demands a reduction in the amount of information required to represent the aural signal [3]. The amount of digital information needed to accurately reproduce the original PCM samples taken of an analog input may be reduced by applying a digital compression algorithm, resulting in a digitally compressed representation of the original signal. (In this context, the term *compression* applies to the digital information that must be stored or recorded, not to the dynamic range of the audio signal.) The goal of any digital compression algorithm is to produce a digital representation of an audio signal which, when decoded and reproduced, sounds the same as the original signal, while using a minimum amount of digital information (bit rate) for the compressed (or encoded) representation. The AC-3 digital compression algorithm specified in the ATSC DTV system can encode from 1 to 5.1 channels of source audio from a PCM representation into a serial bit stream at data rates ranging from 32 to 640 kbytes/s.

A typical application of the bit-reduction algorithm is shown in Figure 7.3.2. In this example, a 5.1 channel audio program is converted from a PCM representation requiring more than 5 Mbytes/s (6 channels \times 48 kHz \times 18 bits = 5.184 Mbytes/s) into a 384 kbytes/s serial bit stream by the AC-3 encoder. Radio frequency (RF) transmission equipment converts this bit stream into a modulated waveform that is applied to a satellite transponder. The amount of bandwidth and power thus required by the transmission has been reduced by more than a factor of 13 by the AC-3 digital compression system. The received signal is demodulated back into the 384 kbytes/s serial bit stream, and decoded by the AC-3 decoder. The result is the original 5.1 channel audio program.

Digital compression of audio is useful wherever there is an economic benefit to be obtained by reducing the amount of digital information required to represent the audio signal. Typical applications include the following:

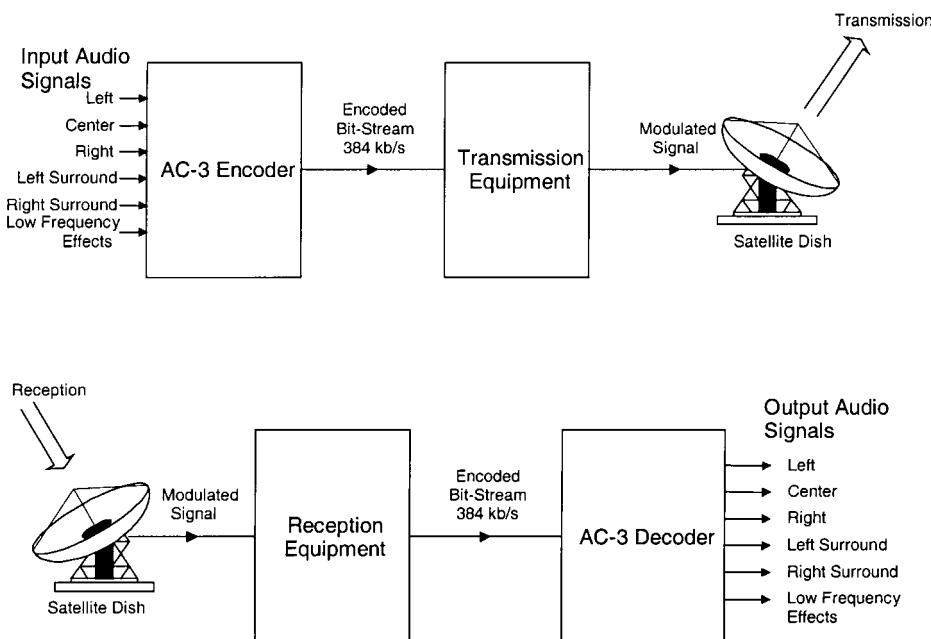


Figure 7.3.2 Example application of the AC-3 audio subsystem for satellite audio transmission. (From [3]. Used with permission.)

- Terrestrial audio broadcasting
- Delivery of audio over metallic or optical cables, or over RF links
- Storage of audio on magnetic, optical, semiconductor, or other storage media

7.3.2d Encoding

The AC-3 encoder accepts PCM audio and produces the encoded bit stream for the ATSC DTV standard [3]. The AC-3 algorithm achieves high *coding gain* (the ratio of the input bit rate to the output bit rate) by coarsely quantizing a frequency-domain representation of the audio signal. A block diagram of this process is given in Figure 7.3.3. The first step in the encoding chain is to transform the representation of audio from a sequence of PCM time samples into a sequence of blocks of frequency coefficients. This is done in the *analysis filterbank*. Overlapping blocks of 512 time samples are multiplied by a time window and transformed into the frequency domain. Because of the overlapping blocks, each PCM input sample is represented in two sequential transformed blocks. The frequency-domain representation then may be decimated by a factor of 2, so that each block contains 256 frequency coefficients. The individual frequency coefficients are represented in binary exponential notation as a *binary exponent* and a *mantissa*. The set of exponents is encoded into a coarse representation of the signal spectrum, referred to as the *spec-*

7-52 Compression Technologies for Audio

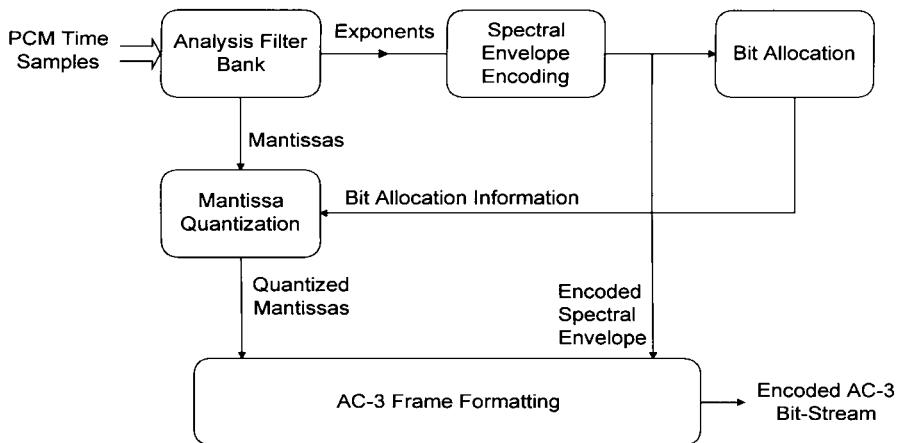


Figure 7.3.3 Overview of the AC-3 audio-compression system encoder. (From [3]. Used with permission.)

This spectral envelope is used by the core bit-allocation routine, which determines how many bits should be used to encode each individual mantissa. The spectral envelope and the coarsely quantized mantissas for six audio blocks (1536 audio samples) are formatted into an AC-3 frame. The AC-3 bit stream is a sequence of AC-3 frames.

The actual AC-3 encoder is more complex than shown in the simplified system of Figure 7.3.3. The following functions also are included:

- A frame header is attached, containing information (bit rate, sample rate, number of encoded channels, and other data) required to synchronize to and decode the encoded bit stream.
- Error-detection codes are inserted to allow the decoder to verify that a received frame of data is error-free.
- The analysis filterbank spectral resolution may be dynamically altered to better match the time/frequency characteristic of each audio block.
- The spectral envelope may be encoded with variable time/frequency resolution.
- A more complex bit-allocation may be performed, and parameters of the core bit-allocation routine may be modified to produce a more optimum bit allocation.
- The channels may be coupled at high frequencies to achieve higher coding gain for operation at lower bit rates.
- In the 2-channel mode, a rematrixing process may be selectively performed to provide additional coding gain, and to allow improved results to be obtained in the event that the 2-channel signal is decoded with a matrix surround decoder.

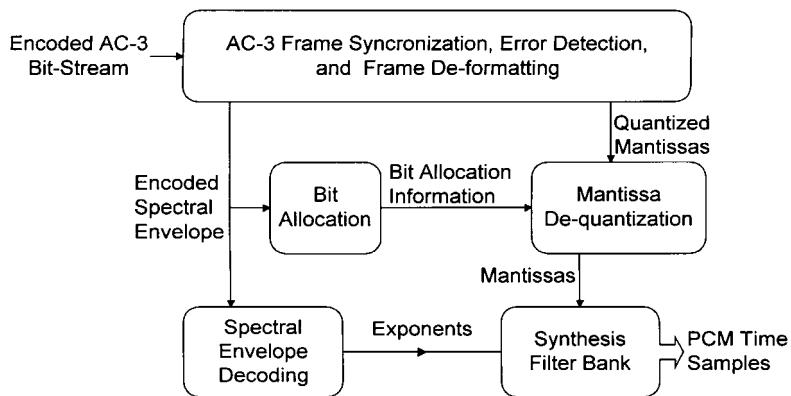


Figure 7.3.4 Overview of the AC-3 audio-compression system decoder. (From [3]. Used with permission.)

7.3.2e Decoding

The decoding process is, essentially, the inverse of the encoding process [3]. The basic decoder, shown in Figure 7.3.4, must synchronize to the encoded bit stream, check for errors, and deformat the various types of data (i.e., the encoded spectral envelope and the quantized mantissas). The bit-allocation routine is run, and the results are used to unpack and dequantize the mantissas. The spectral envelope is decoded to produce the exponents. The exponents and mantissas are transformed back into the time domain to produce the decoded PCM time samples. Additional steps in the audio decoding process include the following:

- Error concealment or muting may be applied in the event a data error is detected.
- Channels that have had their high-frequency content coupled must be decoupled.
- Dematrixing must be applied (in the 2-channel mode) whenever the channels have been *rematrixed*.
- The synthesis filterbank resolution must be dynamically altered in the same manner as the encoder analysis filterbank was altered during the encoding process.

7.3.3 Implementation of the AC-3 System

As illustrated in Figure 7.3.5, the audio subsystem of the ATSC DTV standard comprises the audio-encoding/decoding function and resides between the audio inputs/outputs and the transport subsystem [4]. The audio encoder is responsible for generating the *audio elementary stream*, which is an encoded representation of the baseband audio input signals. (Note that more than one audio encoder may be used in a system.) The flexibility of the transport system allows multiple audio elementary streams to be delivered to the receiver. At the receiver, the transport subsystem is responsible for selecting which audio streams to deliver to the audio subsystem. The audio

7-54 Compression Technologies for Audio

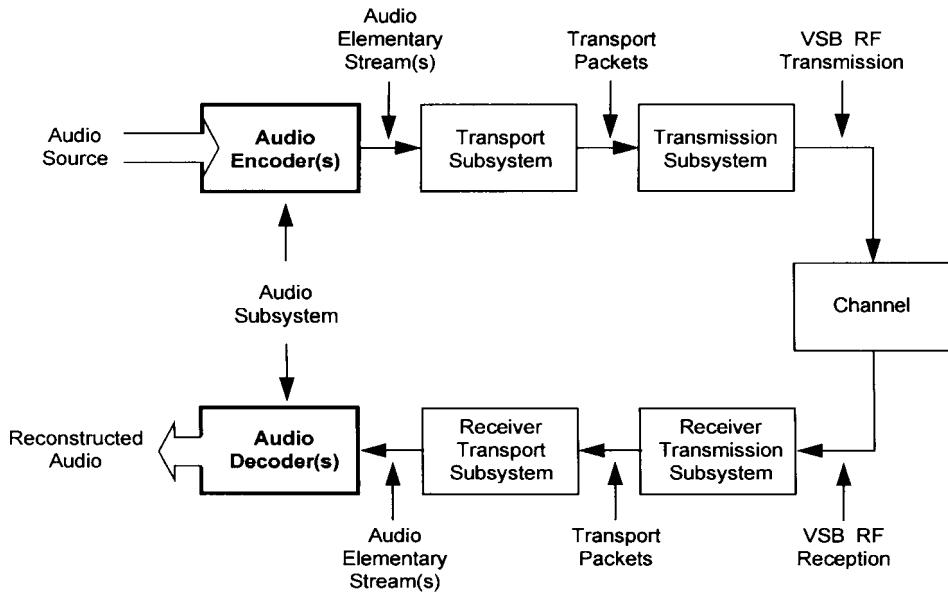


Figure 7.3.5 The audio subsystem in the DTV standard. (From [4]. Used with permission.)

subsystem is then responsible for decoding the audio elementary stream back into baseband audio.

An audio program source is encoded by a *digital television audio encoder*. The output of the audio encoder is a string of bits that represent the audio source (the audio elementary stream). The transport subsystem packetizes the audio data into PES (*packetized elementary system*) packets, which are then further packetized into *transport packets*. The transmission subsystem converts the transport packets into a modulated RF signal for transmission to the receiver. At the receiver, the signal is demodulated by the receiver transmission subsystem. The receiver transport subsystem converts the received audio packets back into an audio elementary stream, which is decoded by the digital television audio decoder.

The partitioning shown in Figure 7.3.5 is conceptual, and practical implementations may differ. For example, the transport processing may be broken into two blocks; the first would perform PES packetization, and the second would perform transport packetization. Or, some of the transport functionality may be included in either the audio coder or the transmission subsystem.

7.3.3a Audio-Encoder Interface

The audio system accepts baseband inputs with up to six channels per audio program bit stream in a channelization scheme consistent with ITU-R Rec. BS-775 [5]. The six audio channels are:

- Left
- Center

- Right
- Left surround
- Right surround
- Low-frequency enhancement (LFE)

Multiple audio elementary bit streams may be conveyed by the transport system.

The bandwidth of the LFE channel is limited to 120 Hz. The bandwidth of the other (main) channels is limited to 20 kHz. Low-frequency response may extend to dc, but it is more typically limited to approximately 3 Hz (-3 dB) by a dc-blocking high-pass filter. Audio-coding efficiency (and thus audio quality) is improved by removing dc offset from audio signals before they are encoded. The input audio signals may be in analog or digital form.

For analog input signals, the input connector and signal level are not specified [4]. Conventional broadcast practice may be followed. One commonly used input connector is the 3-pin XLR female (the incoming audio cable uses the male connector) with pin 1 ground, pin 2 hot or positive, and pin 3 neutral or negative.

Likewise, for digital input signals, the input connector and signal format are not specified. Commonly used formats such as the AES3-1992 2-channel interface are suggested. When multiple 2-channel inputs are used, the preferred channel assignment is:

- Pair 1: Left, Right
- Pair 2: Center, LFE
- Pair 3: Left surround, Right surround

Sampling Parameters

The AC-3 system conveys digital audio sampled at a frequency of 48 kHz, locked to the 27 MHz system clock [4]. If analog signal inputs are employed, the A/D converters should sample at 48 kHz. If digital inputs are employed, the input sampling rate should be 48 kHz, or the audio encoder should contain sampling rate converters that translate the sampling rate to 48 kHz. The sampling rate at the input to the audio encoder must be locked to the video clock for proper operation of the audio subsystem.

In general, input signals should be quantized to at least 16-bit resolution. The audio-compression system can convey audio signals with up to 24-bit resolution.

7.3.3b Output Signal Specification

Conceptually, the output of the audio encoder is an elementary stream that is formed into PES packets within the transport subsystem [4]. It is possible that digital television systems will be implemented wherein the formation of audio PES packets takes place within the audio encoder. In this case, the output of the audio encoder would be PES packets. Physical interfaces for these outputs (elementary streams and/or PES packets) may be defined as voluntary industry standards by SMPTE or other organizations; they are not, however, specified in the core ATSC standard.

7-56 Compression Technologies for Audio

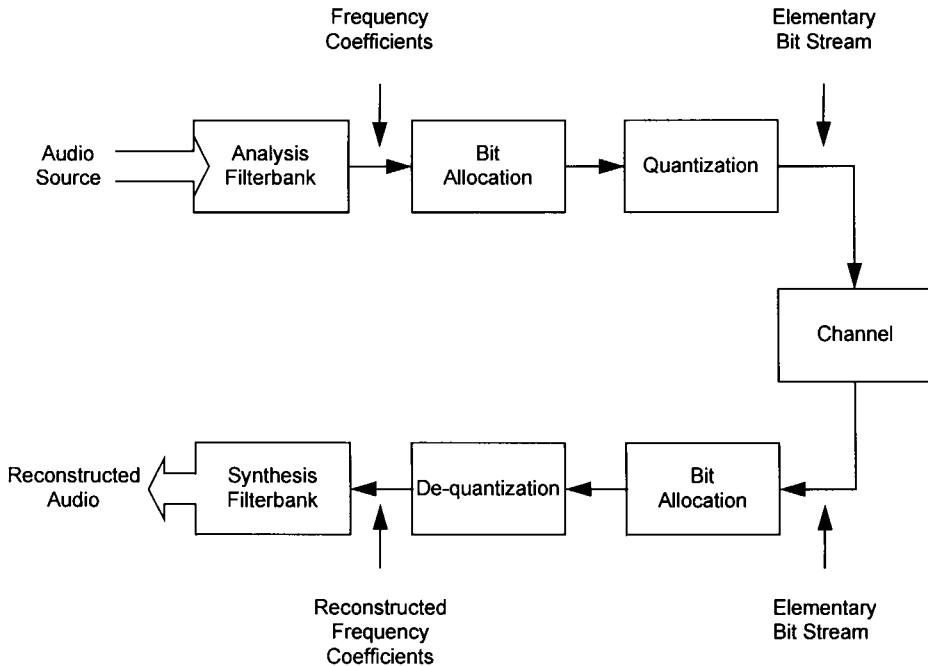


Figure 7.3.6 Overview of the AC-3 audio-compression system. (From [6]. Used with permission.)

7.3.4 Operational Details of the AC-3 Standard

The AC-3 audio-compression system consists of three basic operations, as illustrated in Figure 7.3.6 [6]. In the first stage, the representation of the audio signal is changed from the time domain to the frequency domain, which is a more efficient domain in which to perform psychoacoustically based audio compression. The resulting frequency-domain coefficients are then encoded. The frequency-domain coefficients may be coarsely quantized because the resulting quantizing noise will be at the same frequency as the audio signal, and relatively low S/N ratios are acceptable because of the phenomenon of psychoacoustic masking. Based on a psychoacoustic model of human hearing, a bit-allocation operation determines the actual S/N acceptable for each individual frequency coefficient. Finally, the frequency coefficients are coarsely quantized to the necessary precision and formatted into the audio elementary stream.

The basic unit of encoded audio is the AC-3 *sync frame*, which represents 1536 audio samples. Each sync frame of audio is a completely independent encoded entity. The elementary bit stream contains the information necessary to allow the audio decoder to perform the identical (to the encoder) bit allocation. This permits the decoder to unpack and dequantize the elementary bit-stream frequency coefficients, resulting in the reconstructed frequency coefficients. The synthesis filterbank is the inverse of the analysis filterbank, and it converts the reconstructed frequency coefficients back into a time-domain signal.

7.3.4a Transform Filterbank

The process of converting the audio from the time domain to the frequency domain requires that the audio be blocked into overlapping blocks of 512 samples [6]. For every 256 new audio samples, a 512-sample block is formed from the 256 new samples and the 256 previous samples. Each audio sample is represented in two audio blocks, so the number of samples to be processed initially is doubled. The overlapping of blocks is necessary to prevent audible blocking artifacts. New audio blocks are formed every 5.33 ms. A group of six blocks is coded into one AC-3 sync frame.

Window Function

Prior to being transformed into the frequency domain, the block of 512 time samples is *windowed* [6]. The windowing operation involves a vector multiplication of the 512-point block with a 512-point window function. The window function has a value of 1.0 in its center, tapering down to almost zero at the ends. The shape of the window function is such that the overlap/add processing at the decoder will result in a reconstruction free of blocking artifacts. The window function shape also determines the shape of each individual filterbank filter.

Time-Division Aliasing Cancellation Transform

The analysis filterbank is based on the fast Fourier transform [6]. The particular transformation employed is the oddly stacked *time-domain aliasing cancellation* (TDAC) transform. This particular transformation is advantageous because it allows removal of the 100 percent redundancy that was introduced in the blocking process. The input to the TDAC transform is 512 windowed time-domain points, and the output is 256 frequency-domain coefficients.

Transient Handling

When extreme time-domain transients exist (an impulse, such as a castanets click), there is a possibility that quantization error—incurred by coarsely quantizing the frequency coefficients of the transient—will become audible as a result of *time smearing* [6]. The quantization error within a coded audio block is reproduced throughout the block. It is possible for the portion of the quantization error that is reproduced prior to the impulse to be audible. Time smearing of quantization noise may be reduced by altering the length of the transform that is performed. Instead of a single 512-point transform, a pair of 256-point transforms may be performed—one on the first 256 windowed samples, and one on the last 256 windowed samples. A transient detector in the encoder determines when to alter the transform length. The reduction in transform length prevents quantization error from spreading more than a few milliseconds in time, which is adequate to prevent audibility.

7.3.4b Coded Audio Representation

The frequency coefficients that result from the transformation are converted to a binary floating point notation [6]. The scaling of the transform is such that all values are smaller than 1.0. An example value in binary notation (base 2) with 16-bit precision would be:

0.0000 0000 1010 11002

7-58 Compression Technologies for Audio

The number of leading zeros in the coefficient, 8 in this example, becomes the *raw exponent*. The value is left-shifted by the exponent, and the value to the right of the decimal point (1010 1100) becomes the *normalized mantissa* to be coarsely quantized. The exponents and the coarsely quantized mantissas are encoded into the bit stream.

Exponent Coding

A certain amount of processing is applied to the raw exponents to reduce the amount of data required to encode them [6]. First, the raw exponents of the six blocks to be included in a single AC-3 sync frame are examined for block-to-block differences. If the differences are small, a single exponent set is generated that is usable by all six blocks, thus reducing the amount of data to be encoded by a factor of 6. If the exponents undergo significant changes within the frame, exponent sets are formed over blocks where the changes are not significant. Because of the frequency response of the individual filters in the analysis filterbank, exponents for adjacent frequencies rarely differ by more than ± 2 . To take advantage of this fact, exponents are encoded differentially in frequency. The first exponent is encoded as an absolute, and the difference between the current exponent and the following exponent then is encoded. This reduces the exponent data rate by a factor of 2. Finally, where the spectrum is relatively flat, or an exponent set only covers 1 or 2 blocks, differential exponents may be shared across 2 or 4 frequency coefficients, for an additional savings of a factor of 2 or 4.

The final coding efficiency for AC-3 exponents is typically 0.39 bits/exponent (or 0.39 bits/sample, because there is an exponent for each audio sample). Exponents are coded only up to the frequency needed for the perception of full frequency response. Typically, the highest audio frequency component in the signal that is audible is at a frequency lower than 20 kHz. In the case that signal components above 15 kHz are inaudible, only the first 75 percent of the exponent values are encoded, reducing the exponent data rate to less than 0.3 bits/sample.

The exponent processing changes the exponent values from their original values. The encoder generates a local representation of the exponents that is identical to the decoded representation that will be used by the decoder. The decoded representation then is used to shift the original frequency coefficients to generate the normalized mantissas that are subsequently quantized.

Mantissas

The frequency coefficients produced by the analysis filterbank have a useful precision that is dependent upon the word length of the input PCM audio samples as well as the precision of the transform computation [6]. Typically, this precision is on the order of 16 to 18 bits, but may be as high as 24 bits. Each normalized mantissa is quantized to a precision from 0 to 16 bits. Because the goal of audio compression is to maximize the audio quality at a given bit rate, an optimum (or near-optimum) allocation of the available bits to the individual mantissas is required.

7.3.4c Bit Allocation

The number of bits allocated to each individual mantissa value is determined by the bit-allocation routine [6]. The identical core routine is run in both the encoder and the decoder, so that each generates an identical bit allocation.

The core bit-allocation algorithm is considered *backward adaptive*, in that some of the encoded audio information within the bit stream (fed back into the encoder) is used to compute

the final bit allocation. The primary input to the core allocation routine is the decoded exponent values, which give a general picture of the signal spectrum. From this version of the signal spectrum, a *masking curve* is calculated. The calculation of the masking model is based on a model of the human auditory system. The masking curve indicates, as a function of frequency, the level of quantizing error that may be tolerated. Subtraction (in the log power domain) of the masking curve from the signal spectrum yields the required S/N as a function of frequency. The required S/N values are mapped into a set of *bit-allocation pointers* (BAPs) that indicate which quantizer to apply to each mantissa.

Forward Adaptive

The AC-3 encoder may employ a more sophisticated psychoacoustic model than that used by the decoder [6]. The core allocation routine used by both the encoder and the decoder makes use of a number of adjustable parameters. If the encoder employs a more sophisticated psychoacoustic model than that of the core routine, the encoder may adjust these parameters so that the core routine produces a better result. The parameters are subsequently inserted into the bit stream by the encoder and fed forward to the decoder.

In the event that the available bit-allocation parameters do not allow the ideal allocation to be generated, the encoder can insert explicit codes into the bit stream to alter the computed masking curve, hence the final bit allocation. The inserted codes indicate changes to the base allocation and are referred to as *delta bit-allocation codes*.

7.3.4d Rematrixing

When the AC-3 encoder is operating in a 2-channel stereo mode, an additional processing step is inserted to enhance interoperability with Dolby Surround 4-2-4 matrix encoded programs [6]. This extra step is referred to as *rematrixing*.

The signal spectrum is broken into four distinct rematrixing frequency bands. Within each band, the energy of the left, right, sum, and difference signals are determined. If the largest signal energy is in the left and right channels, the band is encoded normally. If the dominant signal energy is in the sum and difference channels, then those channels are encoded instead of the left and right channels. The decision as to whether to encode left and right or sum and difference is made on a band-by-band basis and is signaled to the decoder in the encoded bit stream.

7.3.4e Coupling

In the event that the number of bits required to transparently encode the audio signals exceeds the number of bits that are available, the encoder may invoke *coupling* [6]. Coupling involves combining the high-frequency content of individual channels and sending the individual channel signal envelopes along with the combined coupling channel. The psychoacoustic basis for coupling is that within narrow frequency bands, the human ear detects high-frequency localization based on the signal envelope rather than on the detailed signal waveform.

The frequency above which coupling is invoked, and the channels that participate in the process, are determined by the AC-3 encoder. The encoder also determines the frequency banding structure used by the coupling process. For each coupled channel and each coupling band, the encoder creates a sequence of *coupling coordinates*. The coupling coordinates for a particular channel indicate what fraction of the common coupling channel should be reproduced out of that

7-60 Compression Technologies for Audio

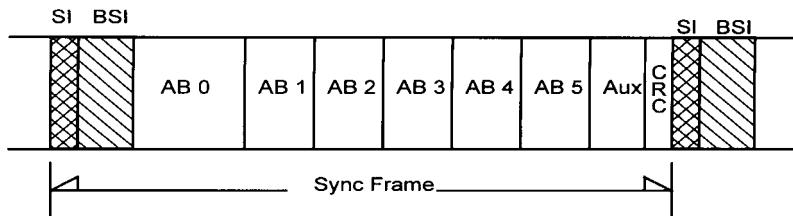


Figure 7.3.7 The AC-3 synchronization frame. (From [6]. Used with permission.)

particular channel output. The coupling coordinates represent the individual signal envelopes for the channels. The encoder determines the frequency with which coupling coordinates are transmitted. If the signal envelope is steady, the coupling coordinates do not need to be sent every block, but can be reused by the decoder until new coordinates are sent. The encoder determines how often to send new coordinates, and it can send them as often as each block (every 5.3 ms).

7.3.4f Bit Stream Elements and Syntax

An AC-3 serial-coded audio bit stream is made up of a sequence of *synchronization frames*, as illustrated in Figure 7.3.7 [6]. Each synchronization frame contains six coded audio blocks, each of which represent 256 new audio samples. A *synchronization information* (SI) header at the beginning of each frame contains information needed to acquire and maintain synchronization. A *bit-stream information* (BSI) header follows each SI, containing parameters describing the coded audio service. The coded audio blocks may be followed by an auxiliary data (Aux) field. At the end of each frame is an error-check field that includes a CRC word for error detection. An additional CRC word, the use of which is optional, is located in the SI header.

A number of bit-stream elements have values that may be transmitted, but whose meaning has been reserved. If a decoder receives a bit stream that contains reserved values, the decoder may or may not be able to decode and produce audio.

Splicing and Insertion

The ideal place to splice encoded audio bit streams is at the boundary of a sync frame [6]. If a bit stream splice is performed at the boundary of the sync frame, the audio decoding will proceed without interruption. If a bit stream splice is performed randomly, there will be an audio interruption. The frame that is incomplete will not pass the decoder's error-detection test, and this will cause the decoder to mute. The decoder will not find sync in its proper place in the next frame, and it will enter a sync search mode. After the sync code of the new bit stream is found, synchronization will be achieved, and audio reproduction will resume. This type of outage will be on the order of two frames, or about 64 ms. Because of the windowing process of the filter-bank, when the audio goes to mute, there will be a gentle fadedown over a period of 2.6 ms. When the audio is recovered, it will fade up over a period of 2.6 ms. Except for the approximately 64 ms of time during which the audio is muted, the effect of a random splice of an AC-3 elementary stream is relatively benign.

Error-Detection Codes

Each AC-3 sync frame ends with a 16-bit CRC error-check code [6]. The decoder may use this code to determine whether a frame of audio has been damaged or is incomplete. Additionally, the decoder may make use of error flags provided by the transport system. In the case of detected errors, the decoder may try to perform error concealment, or it may simply mute.

7.3.4g Loudness and Dynamic Range

It is important for the digital television system to provide uniform subjective loudness for all audio programs [6]. Consumers often find it annoying when audio levels fluctuate between broadcast channels (observed when channel hopping) or between program segments on a particular channel (such as commercials being much louder than entertainment programs). One element found in most audio programming is the human voice. Achieving an approximate level match for dialogue (spoken in a normal voice, without shouting or whispering) in all audio programming is a desirable goal. The AC-3 audio system provides syntactical elements that make this goal achievable.

Because the digital audio-coding system can provide more than 100 dB of dynamic range, there is no technical reason for dialogue to be encoded anywhere near 100 percent, as it commonly is in NTSC television. However, there is no assurance that all program channels, or all programs or program segments on a given channel, will have dialogue encoded at the same (or even a similar) level. Without a uniform coding level for dialogue (which would imply a uniform headroom available for all programs), there would be inevitable audio-level fluctuations between program channels or even between program segments.

Dynamic Range Compression

It is common practice for high-quality programming to be produced with wide dynamic range audio, suitable for the highest-quality audio reproduction environment [6]. Because they serve audiences with a wide range of receiver capabilities, however, broadcasters typically process audio to reduce its dynamic range. This processed audio is more suitable for most of the audience, which does not have an audio reproduction environment that matches the original audio production studio. In the case of NTSC, all viewers receive the same audio with the same dynamic range; it is impossible for any viewer to enjoy the original wide dynamic range of the audio production.

For DTV, the audio-coding system provides an embedded dynamic range control scheme that allows a common encoded bit stream to deliver programming with a dynamic range appropriate for each individual listener. A *dynamic range control value* (DynRng) is provided in each audio block (every 5 ms). These values are used by the audio decoder to alter the level of the reproduced sound for each audio block. Level variations of up to ± 24 dB can be indicated.

7.3.4h Encoding the AC-3 Bit Stream

Because the ATSC DTV standard AC-3 audio system is specified by the syntax and decoder processing, the encoder itself is not precisely specified [3]. The only normative requirement on the encoder is that the output elementary bit stream follow the AC-3 syntax. Therefore, encoders of varying levels of sophistication may be produced. More sophisticated encoders may offer supe-

7-62 Compression Technologies for Audio

rior audio performance, and they may make operation at lower bit rates acceptable. Encoders are expected to improve over time, and all decoders will benefit from encoder improvements. The encoder described in this section, although basic in operation, provides good performance and offers a starting point for future designs. A flow chart diagram of the encoding process is given in Figure 7.3.8.

Input Word Length/Sample Rate

The AC-3 encoder accepts audio in the form of PCM words [3]. The internal dynamic range of AC-3 allows input word lengths of up to 24 bits to be useful.

The input sample rate must be locked to the output bit rate so that each AC-3 sync frame contains 1536 samples of audio. If the input audio is available in a PCM format at a different sample rate than that required, sample rate conversion must be performed to conform the sample rate.

Individual input channels may be high-pass filtered. Removal of dc components of the input signals can allow more efficient coding because the available data rate then is not used to encode dc. However, there is the risk that signals that do not reach 100 percent PCM level before high-pass filtering will exceed the 100 percent level after filtering, and thus be clipped. A typical encoder would high-pass filter the input signals with a single pole filter at 3 Hz.

The LFE channel normally is low-pass-filtered at 120 Hz. A typical encoder would filter the LFE channel with an 8th-order elliptic filter whose cutoff frequency is 120 Hz.

Transients are detected in the full-bandwidth channels to decide when to switch to short-length audio blocks to improve pre-echo performance. High-pass filtered versions of the signals are examined for an increase in energy from one subblock time segment to the next. Subblocks are examined at different time scales. If a transient is detected in the second half of an audio block in a channel, that channel switches to a short block.

The transient detector is used to determine when to switch from a *long transform block* (length 512) to a *short transform block* (length 256). It operates on 512 samples for every audio block. This is done in two passes, with each pass processing 256 samples. Transient detection is broken down into four steps:

- High-pass filtering
- Segmentation of the block into submultiples
- Peak amplitude detection within each subblock segment
- Threshold comparison

7.3.4i AC-3/MPEG Bit Stream

The AC-3 elementary bit stream is included in an MPEG-2 multiplex bit stream in much the same way an MPEG-1 audio stream would be included, with the AC-3 bit stream packetized into PES packets [7]. An MPEG-2 multiplex bit stream containing AC-3 elementary streams must meet all audio constraints described in the MPEG model. It is necessary to unambiguously indicate that an AC-3 stream is, in fact, an AC-3 stream, and not an MPEG audio stream. The MPEG-2 standard does not explicitly state codes to be used to indicate an AC-3 stream. Also, the MPEG-2 standard does not have an audio descriptor adequate to describe the contents of the AC-3 bit stream in its internal tables. The solution to this problem is beyond the scope of this chapter; interested readers should consult [7] for additional information on the subject.

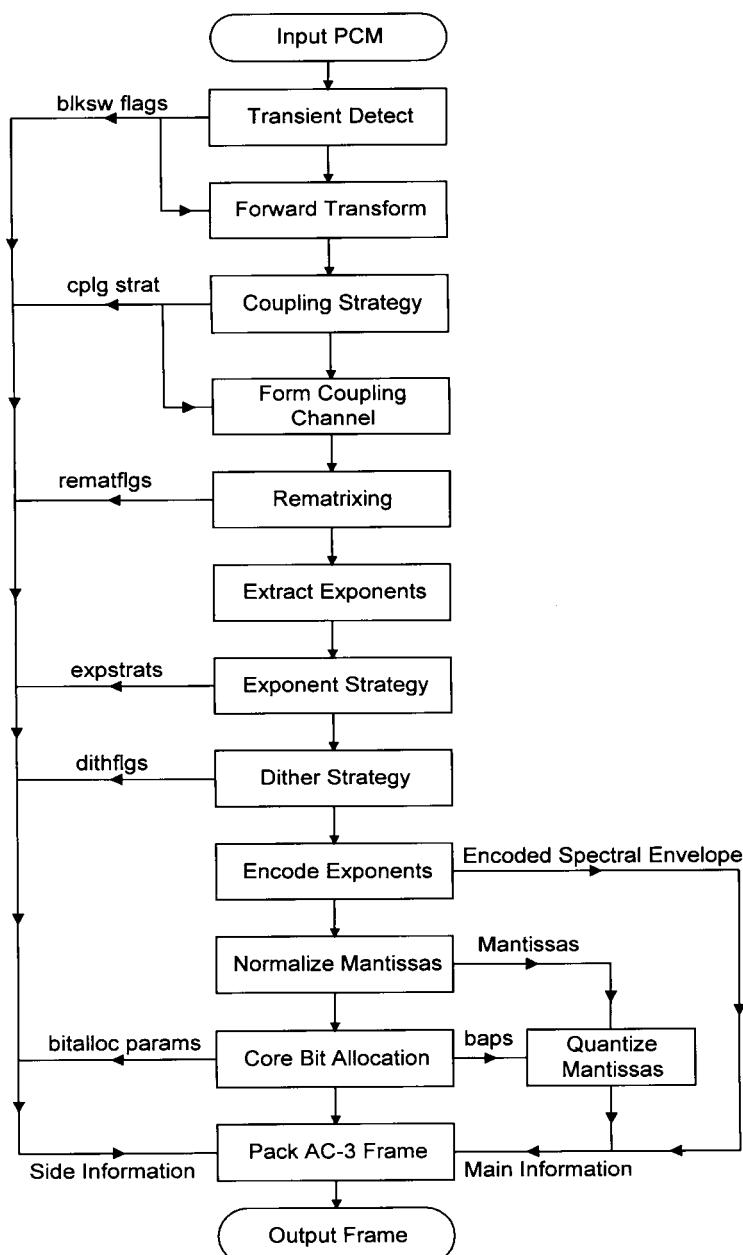


Figure 7.3.8 Generalized flow diagram of the AC-3 encoding process. (From [7]. Used with permission.)

7-64 Compression Technologies for Audio

7.3.4j Decoding the AC-3 Bit Stream

An overview of AC-3 decoding is diagrammed in Figure 7.3.9, where the decoding process flow is shown as a sequence of blocks down the center of the illustration, and some of the key information flow is indicated by arrowed lines at the sides [3]. This decoder should be considered only as an example; other methods certainly exist to implement decoders, and those other methods may have advantages in certain areas (such as instruction count, memory requirements, number of transforms required, and other parameters). The input bit stream typically will come from a transmission or storage system. The interface between the source of AC-3 data and the AC-3 decoder is not specified in the ATSC DTV standard.

Continuous or Burst Input

The encoded AC-3 data may be input to the decoder as a continuous data stream at the nominal bit rate, or chunks of data may be burst into the decoder at a high rate with a low duty cycle [3]. For burst-mode operation, either the data source or the decoder may be the master controlling the burst timing. The AC-3 decoder input buffer may be smaller if the decoder can request bursts of data on an as-needed basis, but the external buffer memory may need to be larger.

Most applications of the standard will convey the elementary AC-3 bit stream with byte or (16-bit) word alignment. The *sync frame* is always an integral number of words in length. The decoder may receive data as a continuous serial stream of bits without any alignment, or the data may be input to the decoder with either byte or word alignment. Byte or word alignment of the input data may allow some simplification of the decoder. Alignment does reduce the probability of false detection of the sync word.

Synchronization and Error Detection

The AC-3 bit steam format allows for rapid synchronization [3]. The 16-bit sync word has a low probability of false detection. With no input stream alignment, the probability of false detection of the sync word is 0.0015 percent per input stream bit position. For a bit rate of 384 kbits/s, the probability of false sync word detection is 19 percent per frame. Byte alignment of the input stream drops this probability to 2.5 percent, and word alignment drops it to 1.2 percent.

When a sync pattern is detected, the decoder may be estimated to be in sync, and one of the CRC words (CRC1 or CRC2) may be checked. Because CRC1 comes first and covers the first five-eighths of the frame, the result of a CRC1 check may be available after only five-eighths of the frame has been received. Or, the entire frame size can be received and CRC2 checked. If either CRC word checks, the decoder may safely be presumed to be in sync, and decoding and reproduction of audio may proceed. The chance of false sync in this case would be the concatenation of the probabilities of a false sync word detection and a CRC misdetection of error. The CRC check is reliable to 0.0015 percent. This probability, concatenated with the probability of a false sync detection in a byte-aligned input bit stream, yields a probability of false synchronization of 0.000035 percent (or about once in 3 million synchronization attempts).

If this small probability of false sync is too large for a specific application, several methods may be used to reduce it. The decoder may only presume correct sync in the case that both CRC words check properly. The decoder also may require multiple sync words to be received with the proper alignment. If the data transmission or storage system is aware that data is in error, this information may be made known to the decoder.

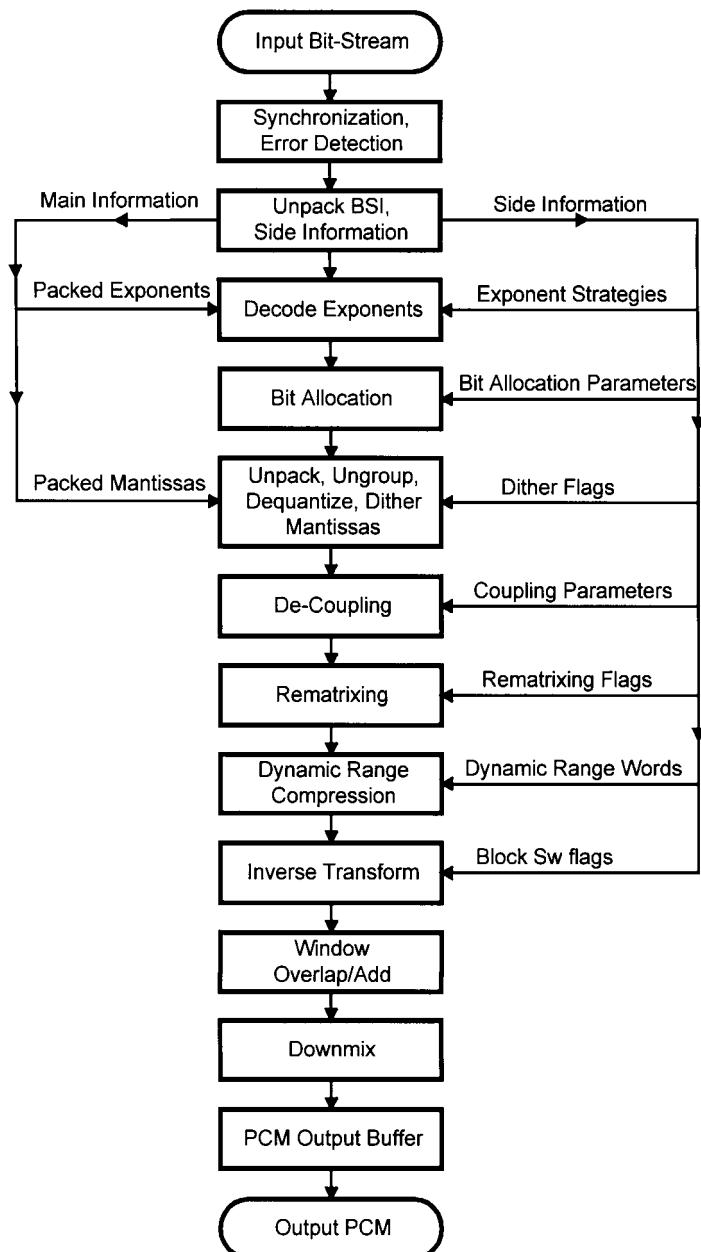


Figure 7.3.9 Generalized flow diagram of the AC-3 decoding process. (From [3]. Used with permission.)

7-66 Compression Technologies for Audio

Inherent to the decoding process is the *unpacking* (demultiplexing) of the various types of information included in the bit stream. Among the options for distribution of this bit stream information are:

- Selected data may be copied from the input buffer to dedicated registers.
- Data from the input buffer may be copied to specific working memory locations.
- The data may simply be located in the input buffer, with pointers to the data saved to another location for use when the information is required.

Decoding Components

The audio-compression system exponents are delivered in the bit stream in an encoded form [3]. To unpack and decode the exponents, two types of “side information” are required:

- The number of exponents must be known.
- The exponent “strategy” in use by each channel must be known.

The *bit-allocation computation* reveals how many bits are used for each mantissa. The inputs to the bit-allocation computation are the decoded exponents and the bit-allocation side information. The outputs of the bit-allocation computation are a set of *bit-allocation pointers* (BAPs), one BAP for each coded mantissa. The BAP indicates the quantizer used for the mantissa, and how many bits in the bit stream were used for each mantissa.

The coarsely quantized mantissas make up the bulk of the AC-3 data stream. Each mantissa is quantized to a level of precision indicated by the corresponding BAP. To pack the mantissa data more efficiently, some mantissas are grouped together into a single transmitted value. For instance, two 11-level quantized values are conveyed in a single 7-bit code (3.5 bits/value) in the bit stream.

The mantissa data is unpacked by peeling off groups of bits as indicated by the BAPs. Grouped mantissas must be ungrouped. The individual coded mantissa values are converted into a dequantized value. Mantissas that are indicated as having zero bits may be reproduced as either zero or by a random dither value (under control of a dither flag).

Other steps in the decoding process include the following:

- *Decoupling*. When *coupling* is in use, the channels that are coupled must be decoupled. Decoupling involves reconstructing the high-frequency section (exponents and mantissas) of each coupled channel, from the common coupling channel and the coupling coordinates for the individual channel. Within each coupling band, the coupling-channel coefficients (exponent and mantissa) are multiplied by the individual channel coupling coordinates.
- *Rematrixing*. In the 2/0 audio-coding mode, rematrixing may be employed as indicated by a *rematrix flag*. When the flag indicates that a band is rematrixed, the coefficients encoded in the bit stream are sum and difference values, instead of left and right values.
- *Dynamic range compression*. For each block of audio, a dynamic range control value may be included in the bit stream. The decoder, by default, will use this value to alter the magnitude of the coefficient (exponent and mantissa) as required to properly process the data.
- *Inverse transform*. The decoding steps described in this section will result in a set of frequency coefficients for each encoded channel. The inverse transform converts these blocks of frequency coefficients into blocks of time samples.

- *Window, overlap/add.* The individual blocks of time samples must be windowed, and adjacent blocks are overlapped and added together to reconstruct the final continuous-time-output PCM audio signal.
- *Downmixing.* If the number of channels required at the decoder output is smaller than the number of channels that are encoded in the bit stream, then downmixing is required. Downmixing in the time domain is shown in the example decoder of Figure 7.3.9. Because the inverse transform is a linear operation, it also is possible to downmix in the frequency domain prior to transformation.
- *PCM output buffer.* Typical decoders will provide PCM output samples at the PCM sampling rate. Because blocks of samples result from the decoding process, an output buffer typically is required.
- *Output PCM.* The output PCM samples are delivered in a form suitable for interconnection to a digital-to-analog converter (D/A), or in some other form required by the receiver.

7.3.4k Algorithmic Details

The actual audio information conveyed by the AC-3 bit stream consists of the quantized frequency coefficients [3]. The coefficients, delivered in floating point form, are 5-bit values that indicate the number of leading zeros in the binary representation of a frequency coefficient. The exponent acts as a scale factor for each mantissa, equal to 2^{exp} . Exponent values are allowed to range from 0 (for the largest-value coefficients with no leading zeros) to 24. Exponents for coefficients that have more than 24 leading zeros are fixed at 24, and the corresponding mantissas are allowed to have leading zeros. Exponents require 5 bits to represent all allowed values.

AC-3 bit streams contain coded exponents for all independent channels, all coupled channels, and for the coupling and low-frequency effects channels (when they are enabled). Because audio information is not shared across frames, block 0 of every frame will include new exponents for every channel. Exponent information may be shared across blocks within a frame, so blocks 1 through 5 may reuse exponents from previous blocks.

AC-3 exponent transmission employs *differential coding*, in which the exponents for a channel are differentially coded across frequency. These differential exponents are combined into groups in the audio block. This grouping is done by one of three methods, which are referred to as *exponent strategies*. The number of grouped differential exponents placed in the audio block for a particular channel depends on the exponent strategy and on the frequency bandwidth information for that channel. The number of exponents in each group depends only on the exponent strategy.

An AC-3 audio block contains two types of fields with exponent information. The first type defines the exponent coding strategy for each channel, and the second type contains the actual coded exponents for channels requiring new exponents. For independent channels, frequency bandwidth information is included along with the exponent strategy fields. For coupled channels, and the coupling channel, the frequency information is found in the coupling strategy fields.

7.3.4l Bit Allocation

The bit allocation routine analyzes the spectral envelope of the audio signal being coded with respect to masking effects to determine the number of bits to assign to each transform coefficient

7-68 Compression Technologies for Audio

mantissa [3]. In the encoder, the bit allocation is performed globally on the ensemble of channels as an entity, from a common bit pool. Because there are no preassigned exponent or mantissa bits, the routine is allowed to flexibly allocate bits across channels, frequencies, and audio blocks in accordance with signal demand.

The bit allocation contains a parametric model of human hearing for estimating a noise-level threshold, expressed as a function of frequency, which separates audible from inaudible spectral components. Various parameters of the hearing model can be adjusted by the encoder depending upon signal characteristics. For example, a prototype masking curve is defined in terms of two piecewise continuous line segments, each with its own slope and *y*-axis intercept. One of several possible slopes and intercepts is selected by the encoder for each line segment. The encoder may iterate on one or more such parameters until an optimal result is obtained. When all parameters used to estimate the noise-level threshold have been selected by the encoder, the final bit allocation is computed. The model parameters are conveyed to the decoder with other side information. The decoder then executes the routine in a single pass.

The estimated noise-level threshold is computed over 50 bands of nonuniform bandwidth (an approximate 1/6-octave scale). The defined banding structure is independent of sampling frequency. The required bit allocation for each mantissa is established by performing a table lookup based upon the difference between the input signal *power spectral density* (PSD), evaluated on a fine-grain uniform frequency scale, and the estimated noise-level threshold, evaluated on the coarse-grain (*banded*) frequency scale. Therefore, the bit allocation result for a particular channel has spectral granularity corresponding to the exponent strategy employed.

7.3.5 Audio System Level Control

The AC-3 system provides elements that allow the encoded bit stream to satisfy listeners in many different situations. Two principal techniques are used to control the subjective loudness of the reproduced audio signals:

- Dialogue normalization
- Dynamic range compression

7.3.5a Dialogue Normalization

The *dialogue normalization* (DialNorm) element permits uniform reproduction of spoken dialogue when decoding any AC-3 bit stream [3]. When audio from different sources is reproduced, the apparent loudness often varies from source to source. Examples include the following:

- Audio elements from different program segments during a broadcast (for example, a movie vs. a commercial message)
- Different broadcast channels
- Different types of media (for example, disc vs. tape)

The AC-3 coding technology solves this problem by explicitly coding an indication of loudness into the AC-3 bit stream.

The subjective level of normal spoken dialogue is used as a reference. The 5-bit dialogue normalization word that is contained in the bit stream, DialNorm, is an indication of the subjective

loudness of normal spoken dialogue compared with digital 100 percent. The 5-bit value is interpreted as an unsigned integer (most significant bit transmitted first) with a range of possible values from 1 to 31. The unsigned integer indicates the headroom in decibels above the subjective dialogue level. This value also may be interpreted as an indication of how many decibels the subjective dialogue level is below digital 100 percent.

The DialNorm value is not directly used by the AC-3 decoder. Rather, the value is used by the section of the sound reproduction system responsible for setting the reproduction volume, such as the system volume control. The system volume control generally is set based on listener input as to the desired loudness, or *sound-pressure level* (SPL). The listener adjusts a volume control that directly adjusts the reproduction system gain. With AC-3 and the DialNorm value, the reproduction system gain becomes a function of both the listener's desired reproduction sound-pressure level for dialogue, and the DialNorm value that indicates the level of dialogue in the audio signal. In this way, the listener is able to reliably set the volume level of dialogue, and the subjective level of dialogue will remain uniform no matter which AC-3 program is decoded.

Example Situation

An example will help to illustrate the DialNorm concept [3]. The listener adjusts the volume control to 67 dB. (With AC-3 dialogue normalization, it is possible to calibrate a system volume control directly in sound-pressure level, and the indication will be accurate for any AC-3 encoded audio source). A high quality entertainment program is being received, and the AC-3 bit stream indicates that the dialogue level is 25 dB below the 100 percent digital level. The reproduction system automatically sets the reproduction system gain so that full-scale digital signals reproduce at a sound-pressure level of 92 dB. Therefore, the spoken dialogue (down 25 dB) will reproduce at 67 dB SPL.

The broadcast program cuts to a commercial message, which has dialogue level at -15 dB with respect to 100 percent digital level. The system level gain automatically drops, so that digital 100 percent is now reproduced at 82 dB SPL. The dialogue of the commercial (down 15 dB) reproduces at a 67 dB SPL, as desired.

For the dialogue normalization system to work, the DialNorm value must be communicated from the AC-3 decoder to the system gain controller so that DialNorm can interact with the listener-adjusted volume control. If the volume-control function for a system is performed as a digital multiplier inside the AC-3 decoder, then the listener-selected volume setting must be communicated into the AC-3 decoder. The listener-selected volume setting and the DialNorm value must be combined to adjust the final reproduction system gain.

Adjustment of the system volume control is not an AC-3 function. The AC-3 bit stream simply conveys useful information that allows the system volume control to be implemented in a way that automatically removes undesirable level variations between program sources.

7.3.5b Dynamic Range Compression

The *dynamic range compression* (DynRng) element allows the program provider to implement subjectively pleasing dynamic range reduction for most of the intended audience, while allowing individual members of the audience the option to experience more (or all) of the original dynamic range [3].

A consistent problem in the delivery of audio programming is that members of the audience may prefer differing amounts of dynamic range. Original high-quality programs (such as feature

7-70 Compression Technologies for Audio

films) typically are mixed with quite a wide dynamic range. Using dialogue as a reference, loud sounds, such as explosions, often are at least 20 dB louder; faint sounds, such as rustling leaves, may be 50 dB quieter. In many listening situations, it is objectionable to allow the sound to become very loud, so the loudest sounds must be compressed downward in level. Similarly, in many listening situations, the very quiet sounds would be inaudible, and they must be brought upward in level to be heard. Because most of the television audience will benefit from a limited program dynamic range, motion picture soundtracks that have been mixed with a wide dynamic range generally are compressed. The dynamic range is reduced by bringing down the level of the loud sounds and bringing up the level of the quiet sounds. Although this satisfies the needs of much of the audience, some audience members may prefer to experience the original sound program in its intended form. The AC-3 audio-coding technology solves this conflict by allowing dynamic range control values to be placed into the AC-3 bit stream.

The dynamic range control values, DynRng, indicate a gain change to be applied in the decoder to implement dynamic range compression. Each DynRng value can indicate a gain change of ± 24 dB. The sequence of DynRng values constitute a compression control signal. An AC-3 encoder (or a bit stream processor) will generate the sequence of DynRng values. Each value is used by the AC-3 decoder to alter the gain of one or more audio blocks. The DynRng values typically indicate gain reductions during the loudest signal passages and gain increases during the quiet passages. For the listener, it is often desirable to bring the loudest sounds down in level, toward dialogue level, and bring the quiet sounds up in level, again toward dialogue level. Sounds that are at the same loudness as normal spoken dialogue typically will not have their gain changed.

The compression actually is applied to the audio in the AC-3 decoder. The encoded audio has full dynamic range. It is permissible for the AC-3 decoder to (optionally, under listener control) ignore the DynRng values in the bit stream. This will result in reproduction of the full dynamic range of the audio. It also is permissible (again under listener control) for the decoder to use some fraction of the DynRng control value and to use a different fraction of positive or negative values. Therefore, the AC-3 decoder can reproduce sounds according to one of the following parameters:

- Fully compressed audio (as intended by the compression control circuit in the AC-3 encoder)
- Full dynamic range audio
- Audio with partially compressed dynamic range, with different amounts of compression for high-level and low-level signals.

Example Situation

A feature film soundtrack is encoded into AC-3 [3]. The original program mix has dialogue level at -25 dB. Explosions reach a full-scale peak level of 0 dB. Some quiet sounds that are intended to be heard by all listeners are 50 dB below dialogue level (-75 dB). A compression control signal (a sequence of DynRng values) is generated by the AC-3 encoder. During those portions of the audio program when the audio level is higher than dialogue level, the DynRng values indicate negative gain, or gain reduction. For full-scale 0 dB signals (the loudest explosions), a gain reduction of -15 dB is encoded into DynRng. For very quiet signals, a gain increase of 20 dB is encoded into DynRng.

A listener wishes to reproduce this soundtrack quietly so as not to disturb anyone, but wishes to hear all of the intended program content. The AC-3 decoder is allowed to reproduce the

default, which is full compression. The listener adjusts dialogue level to 60 dB SPL. The explosions will go only as loud as 70 dB (they are 25 dB louder than dialogue but receive -15 dB applied gain), and the quiet sounds will reproduce at 30 dB SPL (20 dB of gain is applied to their original level of 50 dB below dialogue level). The reproduced dynamic range, therefore, will be $70\text{ dB} - 30\text{ dB} = 40\text{ dB}$.

The listening situation changes, and the listener now wishes to raise the reproduction level of dialogue to 70 dB SPL, but still wishes to limit the loudness of the program. Quiet sounds may be allowed to play as quietly as before. The listener instructs the AC-3 decoder to continue to use the DynRng values that indicate gain reduction, but to attenuate the values that indicate gain increases by a factor of 1/2. The explosions still will reproduce 10 dB above dialogue level, which is now 80 dB SPL. The quiet sounds now are increased in level by $20\text{ dB}/2 = 10\text{ dB}$. They now will be reproduced 40 dB below dialogue level, at 30 dB SPL. The reproduced dynamic range is now $80\text{ dB} - 30\text{ dB} = 50\text{ dB}$.

Another listener prefers the full original dynamic range of the audio. This listener adjusts the reproduced dialogue level to 75 dB SPL and instructs the AC-3 decoder to ignore the dynamic range control signal. For this listener, the quiet sounds reproduce at 25 dB SPL, and the explosions hit 100 dB SPL. The reproduced dynamic range is $100\text{ dB} - 25\text{ dB} = 75\text{ dB}$. This reproduction is exactly as intended by the original program producer.

For this dynamic range control method to be effective, it must be used by all program providers. Because all broadcasters wish to supply programming in the form that is most usable by their audiences, nearly all will apply dynamic range compression to any audio program that has a wide dynamic range. This compression is not reversible unless it is implemented by the technique embedded in AC-3. If broadcasters make use of the embedded AC-3 dynamic range control system, listeners can have significant control over the reproduced dynamic range at their receivers. Broadcasters must be confident that the compression characteristic that they introduce into AC-3 will, by default, be heard by the listeners. Therefore, the AC-3 decoder must, by default, implement the compression characteristic indicated by the DynRng values in the data stream. AC-3 decoders may optionally allow listener control over the use of the DynRng values, so that the listener may select full or partial dynamic range reproduction.

7.3.5c Heavy Compression

The *compression* (COMPR) element allows the program provider (or broadcaster) to implement a large dynamic range reduction (heavy compression) in a way that ensures that a monophonic downmix will not exceed a certain peak level [3]. The heavily compressed audio program may be desirable for certain listening situations, such as movie delivery to a hotel room or to an airline seat. The peak level limitation is useful when, for example, a monophonic downmix will feed an RF modulator, and overmodulation must be avoided.

Some products that decode the AC-3 bit stream will need to deliver the resulting audio via a link with very restricted dynamic range. One example is the case of a television signal decoder that must modulate the received picture and sound onto an RF channel to deliver a signal usable by a low-cost television receiver. In this situation, it is necessary to restrict the maximum peak output level to a known value—with respect to dialogue level—to prevent overmodulation. Most of the time, the dynamic range control signal, DynRng, will produce adequate gain reduction so that the absolute peak level will be constrained. However, because the dynamic range control system is intended to implement a subjectively pleasing reduction in the range of perceived loud-

7-72 Compression Technologies for Audio

ness, there is no assurance that it will control instantaneous signal peaks adequately to prevent overmodulation.

To allow the decoded AC-3 signal to be constrained in peak level, a second control signal, COMPR, (COMPR2 for channel 2 in 1+1 mode) may be included in the AC-3 data stream. This control signal should be present in all bit streams that are intended to be received by, for example, a television set-top decoder. The COMPR control signal is similar to the DynRng control signal in that it is used by the decoder to alter the reproduced audio level. The COMPR control signal has twice the control range as DynRng (± 48 dB compared with ± 24 dB) with half the resolution (0.5 vs. 0.25 dB).

7.3.6 Audio System Features

The audio subsystem offers a host of services and features to meet varied applications and audiences [4]. An AC-3 elementary stream contains the encoded representation of a single audio service. Multiple audio services are provided by multiple elementary streams. Each elementary stream is conveyed by the transport multiplex with a unique *program ID* (PID). A number of audio service types may be coded (individually) into each elementary stream; each AC-3 elementary stream is tagged as to its service type. There are two types of *main service* and six types of *associated service*. Each associated service may be tagged (in the AC-3 audio descriptor) as being associated with one or more main audio services. Each AC-3 elementary stream also may be tagged with a language code.

Associated services may contain complete program mixes or only a single program element. Associated services that are complete mixes may be decoded and used “as is.” Associated services that contain only a single program element are intended to be combined with the program elements from a main audio service.

In general, a complete audio program (what is presented to the listener over the set of loudspeakers) may consist of a main audio service, an associated audio service that is a complete mix, or a main audio service combined with an associated audio service. The capability to simultaneously decode one main service and one associated service is required in order to form a complete audio program in certain service combinations. This capability may not exist in some receivers.

7.3.6a Complete Main Audio Service (CM)

The CM type of main audio service contains a complete audio program (complete with dialogue, music, and effects) [4]. This is the type of audio service normally provided. The CM service may contain from 1 to 5.1 audio channels, and it may be further enhanced by means of the VI, HI, C, E, or VO associated services described in the following sections. Audio in multiple languages may be provided by supplying multiple CM services, each in a different language.

7.3.6b Main Audio Service, Music and Effects (ME)

The ME type of main audio service contains the music and effects of an audio program, but not the dialogue for the program [4]. The ME service may contain from 1 to 5.1 audio channels. The primary program dialogue is missing and (if any exists) is supplied by simultaneously encoding a

D associated service. Multiple D associated services in different languages may be associated with a single ME service.

7.3.6c Visually Impaired (VI)

The VI associated service typically contains a narrative description of the visual program content [4]. In this case, the VI service is a single audio channel. The simultaneous reproduction of both the VI associated service and the CM main audio service allows the visually impaired user to enjoy the main multichannel audio program, as well as to follow (by ear) the on-screen activity.

The dynamic range control signal in this type of service is intended to be used by the audio decoder to modify the level of the main audio program. Thus, the level of the main audio service will be under the control of the VI service provider, and the provider may signal the decoder (by altering the dynamic range control words embedded in the VI audio elementary stream) to reduce the level of the main audio service by up to 24 dB to ensure that the narrative description is intelligible.

Besides being provided as a single narrative channel, the VI service may be provided as a complete program mix containing music, effects, dialogue, and the narration. In this case, the service may be coded using any number of channels (up to 5.1), and the dynamic range control signal would apply only to this service.

7.3.6d Hearing Impaired (HI)

The HI associated service typically contains only dialogue that is intended to be reproduced simultaneously with the CM service [4]. In this case, the HI service is a single audio channel. This dialogue may have been processed for improved intelligibility by hearing-impaired users. Simultaneous reproduction of both the CM and HI services allows the hearing-impaired users to hear a mix of the CM and HI services in order to emphasize the dialogue while still providing some music and effects.

Besides being available as a single dialogue channel, the HI service may be provided as a complete program mix containing music, effects, and dialogue with enhanced intelligibility. In this case, the service may be coded using any number of channels (up to 5.1).

7.3.6e Dialogue (D)

The D associated service contains program dialogue intended for use with an ME main audio service [4]. The language of the D service is indicated in the AC-3 bit stream and in the audio descriptor. A complete audio program is formed by simultaneously decoding the D service and the ME service, then mixing the D service into the center channel of the ME main service (with which it is associated).

If the ME main audio service contains more than two audio channels, the D service is monophonic (1/0 mode). If the main audio service contains two channels, the D service may also contain two channels (2/0 mode). In this case, a complete audio program is formed by simultaneously decoding the D service and the ME service, mixing the left channel of the ME service with the left channel of the D service, and mixing the right channel of the ME service with the right channel of the D service. The result will be a 2-channel stereo signal containing music, effects, and dialogue.

7-74 Compression Technologies for Audio

Audio in multiple languages may be provided by supplying multiple D services (each in a different language) along with a single ME service. This is more efficient than providing multiple CM services, but, in the case of more than two audio channels in the ME service, requires that dialogue be restricted to the center channel.

Some receivers may not have the capability to simultaneously decode an ME and a D service.

7.3.6f Commentary (C)

The commentary associated service is similar to the D service, except that instead of conveying essential program dialogue, the C service conveys optional program commentary [4]. The C service may be a single audio channel containing only the commentary content. In this case, simultaneous reproduction of a C service and a CM service will allow the listener to hear the added program commentary.

The dynamic range control signal in the single-channel C service is intended to be used by the audio decoder to modify the level of the main audio program. Thus, the level of the main audio service will be under the control of the C service provider; the provider may signal the decoder (by altering the dynamic range control words embedded in the C audio elementary stream) to reduce the level of the main audio service by up to 24 dB to ensure that the commentary is intelligible.

Besides providing the C service as a single commentary channel, the C service may be provided as a complete program mix containing music, effects, dialogue, and the commentary. In this case, the service may be provided using any number of channels (up to 5.1).

7.3.6g Emergency (E)

The E associated service is intended to allow the insertion of emergency or high priority announcements [4]. The E service is always a single audio channel. An E service is given priority in transport and in audio decoding. Whenever the E service is present, it will be delivered to the audio decoder. Whenever the audio decoder receives an E-type associated service, it will stop reproducing any main service being received and reproduce only the E service out of the center channel (or left and right channels if a center loudspeaker does not exist). The E service also may be used for nonemergency applications. It may be used whenever the broadcaster wishes to force all decoders to quit reproducing the main audio program and reproduce a higher priority single audio channel.

7.3.6h Voice-Over (VO)

The VO associated service is a single-channel service intended to be reproduced along with the main audio service in the receiver [4]. It allows typical voice-overs to be added to an already encoded audio elementary stream without requiring the audio to be decoded back to baseband and then reencoded. The VO service is always a single audio channel and has second priority; only the E service has higher priority. It is intended to be simultaneously decoded and mixed into the center channel of the main audio service. The dynamic range control signal in the VO service is intended to be used by the audio decoder to modify the level of the main audio program. Thus, the level of the main audio service may be controlled by the broadcaster, and the broadcaster may signal the decoder (by altering the dynamic range control words embedded in the VO audio ele-

mentary stream) to reduce the level of the main audio service by up to 24 dB during the voice-over.

Some receivers may not have the capability to simultaneously decode and reproduce a voice-over service along with a program audio service.

7.3.6i Multilingual Services

Each audio bit stream may be in any language [4]. Table 7.3.2 lists the language codes for the ATSC DTV system. To provide audio services in multiple languages, a number of main audio services may be provided, each in a different language. This is the (artistically) preferred method, because it allows unrestricted placement of dialogue along with the dialogue reverberation. The disadvantage of this method is that as much as 384 kbytes/s is needed to provide a full 5.1-channel service for each language. One way to reduce the required bit rate is to reduce the number of audio channels provided for languages with a limited audience. For instance, alternate language versions could be provided in 2-channel stereo with a bit rate of 128 kbytes/s. Or, a mono version could be supplied at a bit rate of approximately 64 to 96 kbytes/s.

Another way to offer service in multiple languages is to provide a main multichannel audio service (ME) that does not contain dialogue. Multiple single-channel dialogue associated services (D) can then be provided, each at a bit rate in the range of 64 to 96 kbytes/s. Formation of a complete audio program requires that the appropriate language D service be simultaneously decoded and mixed into the ME service. This method allows a large number of languages to be efficiently provided, but at the expense of artistic limitations. The single channel of dialogue would be mixed into the center reproduction channel, and could not be panned. Also, reverberation would be confined to the center channel, which is not optimum. Nevertheless, for some types of programming (sports and news, for example), this method is very attractive because of the savings in bit rate that it offers. Some receivers may not have the capability to simultaneously decode an ME and a D service.

Stereo (2-channel) service without artistic limitation can be provided in multiple languages with added efficiency by transmitting a stereo ME main service along with stereo D services. The D and appropriate-language ME services are combined in the receiver into a complete stereo program. Dialogue may be panned, and reverberation may be included in both channels. A stereo ME service can be sent with high quality at 192 kbytes/s, and the stereo D services (voice only) can make use of lower bit rates, such as 128 or 96 kbytes/s per language. Some receivers may not have the capability to simultaneously decode an ME and a D service.

Note that during those times when dialogue is not present, the D services can be momentarily removed, and the data capacity can be used for other purposes. Table 7.3.3 lists the typical bit rates for various types of service.

7.3.6j Channel Assignments and Levels

To facilitate the reliable exchange of programs, the SMPTE produced a standard for channel assignments and levels on multichannel audio media. The standard, SMPTE 320M, provides specifications for the placement of a 5.1 channel audio program onto multitrack audio media [8]. As specified in ITU-R BS.775-1, the internationally recognized multichannel sound system consists of left, center, right, left surround, right surround, and low-frequency effects (LFE) channels. SMPTE RP 173 specifies the locations and relative level calibration of the loudspeakers

7-76 Compression Technologies for Audio

Table 7.3.2 Language Code Table for AC-3 (After [4].)

Code	Language	Code	Language	Code	Language	Code	Language
0x00	unknown/not applicable	0x20	Polish	0x40	background sound/clean feed	0x60	Moldavian
0x01	Albanian	0x21	Portuguese	0x41		0x61	Malaysian
0x02	Breton	0x22	Romanian	0x42		0x62	Malagasy
0x03	Catalan	0x23	Romansh	0x43		0x63	Macedonian
0x04	Croatian	0x24	Serbian	0x44		0x64	Laotian
0x05	Welsh	0x25	Slovak	0x45	Zulu	0x65	Korean
0x06	Czech	0x26	Slovene	0x46	Vietnamese	0x66	Khmer
0x07	Danish	0x27	Finnish	0x47	Uzbek	0x67	Kazakh
0x08	German	0x28	Swedish	0x48	Urdu	0x68	Kannada
0x09	English	0x29	Turkish	0x49	Ukrainian	0x69	Japanese
0x0A	Spanish	0x2A	Flemish	0x4A	Thai	0x6A	Indonesian
0x0B	Esperanto	0x2B	Walloon	0x4B	Telugu	0x6B	Hindi
0x0C	Estonian	0x2C		0x4C	Tatar	0x6C	Hebrew
0x0D	Basque	0x2D		0x4D	Tamil	0x6D	Hausa
0x0E	Faroese	0x2E		0x4E	Tadzhik	0x6E	Gurani
0x0F	French	0x2F		0x4F	Swahili	0x6F	Gujarati
0x10	Frisian	0x30	reserved	0x50	Sranan Tongo	0x70	Greek
0x11	Irish	0x31	"	0x51	Somali	0x71	Georgian
0x12	Gaelic	0x32	"	0x52	Sinhalese	0x72	Fulani
0x13	Galician	0x33	"	0x53	Shona	0x73	Dari
0x14	Icelandic	0x34	"	0x54	Serbo-Croat	0x74	Churash
0x15	Italian	0x35	"	0x55	Ruthenian	0x75	Chinese
0x16	Lappish	0x36	"	0x56	Russian	0x76	Burmese
0x17	Latin	0x37	"	0x57	Quechua	0x77	Bulgarian
0x18	Latvian	0x38	"	0x58	Pustu	0x78	Bengali
0x19	Luxembourgian	0x39	"	0x59	Punjabi	0x79	Belorussian
0x1A	Lithuanian	0x3A	"	0x5A	Persian	0x7A	Bambora
0x1B	Hungarian	0x3B	"	0x5B	Papamiento	0x7B	Azerbaijani
0x1C	Maltese	0x3C	"	0x5C	Oriya	0x7C	Assamese
0x1D	Dutch	0x3D	"	0x5D	Nepali	0x7D	Armenian
0x1E	Norwegian	0x3E	"	0x5E	Ndebele	0x7E	Arabic
0x1F	Occitan	0x3F	"	0x5F	Marathi	0x7F	Amharic

intended to reproduce these channels. SMPTE 320M specifies a mapping between the audio signals intended to feed loudspeakers, and a sequence of audio tracks on multitrack audio storage

Table 7.3.3 Typical Bit Rates for Various Services (After [4].)

Type of Service	Number of Channels	Typical Bit Rates
CM, ME, or associated audio service containing all necessary program elements	5	320–384 kbits/s
CM, ME, or associated audio service containing all necessary program elements	4	256–384 kbits/s
CM, ME, or associated audio service containing all necessary program elements	3	192–320 kbits/s
CM, ME, or associated audio service containing all necessary program elements	2	128–256 kbits/s
VI, narrative only	1	48–128 kbits/s
HI, narrative only	1	48–96 kbits/s
D	1	64–128 kbits/s
D	2	96–192 kbits/s
C, commentary only	1	32–128 kbits/s
E	1	32–128 kbits/s
VO	1	64–128 kbits/s

media. The standard also specifies the relative levels of the audio signals. Media prepared according to the standard should play properly on a loudspeaker system calibrated according to RP 173.

In consumer audio systems, the LFE channel is considered optional in reproduction. Media that conform to SMPTE 320M should be prepared so that they sound satisfactory even if the LFE channel is not reproduced. When an audio program originally produced as a feature film for theatrical release is transferred to consumer media, the LFE channel is often derived from the dedicated theatrical subwoofer channel. In the cinema, the dedicated subwoofer channel is always reproduced, and thus film mixes may use the subwoofer channel to convey important low frequency program content. Therefore, when transferring programs originally produced for the cinema over to television media, it may be necessary to remix some of the content of the subwoofer channel into the main full bandwidth channels.

7.3.7 References

1. Fibush, David K.: *A Guide to Digital Television Systems and Measurements*, Tektronix, Beaverton, Ore., 1997.
2. SMPTE Standard for Television: “12-Channel Serial Interface for Digital Audio and Auxiliary Data,” SMPTE 324M (Proposed), SMPTE, White Plains, N.Y., 1999.
3. ATSC, “Digital Audio Compression Standard (AC-3),” Advanced Television Systems Committee, Washington, D.C., Doc. A/52, Dec. 20, 1995. This document is included on the CD-ROM.



7-78 Compression Technologies for Audio



4. ATSC, "Digital Television Standard," Advanced Television Systems Committee, Washington, D.C., Doc. A/53, Sept.16, 1995. This document is included on the CD-ROM.
5. ITU-R Recommendation BS-775, "Multi-channel Stereophonic Sound System with and Without Accompanying Picture."
6. ATSC, "Guide to the Use of the Digital Television Standard," Advanced Television Systems Committee, Washington, D.C., Doc. A/54, Oct. 4, 1995. This document is included on the CD-ROM.
7. ATSC, "Digital Audio Compression Standard (AC-3), Annex A: AC-3 Elementary Streams in an MPEG-2 Multiplex," Advanced Television Systems Committee, Washington, D.C., Doc. A/52, Dec. 20, 1995. This document is included on the CD-ROM.
8. SMPTE Standard for Television: "Channel Assignments and Levels on Multichannel Audio Media," SMPTE 320M-1999, SMPTE, White Plains, N.Y., 1999.

7.3.8 Bibliography

- Ehmer, R. H.: "Masking of Tones Vs. Noise Bands," *J. Acoust. Soc. Am.*, vol. 31, pp. 1253–1256, September 1959.
- Ehmer, R. H.: "Masking Patterns of Tones," *J. Acoust. Soc. Am.*, vol. 31, pp. 1115–1120, August 1959.
- Moore, B. C. J., and B. R. Glasberg: "Formulae Describing Frequency Selectivity as a Function of Frequency and Level, and Their Use in Calculating Excitation Patterns," *Hearing Research*, vol. 28, pp. 209–225, 1987.
- Todd, C., et. al.: "AC-3: Flexible Perceptual Coding for Audio Transmission and Storage," AES 96th Convention, Preprint 3796, Audio Engineering Society, New York, February 1994.
- Zwicker, E.: "Subdivision of the Audible Frequency Range Into Critical Bands (Frequenzgruppen)," *J. Acoust. Soc. of Am.*, vol. 33, p. 248, February 1961.

Section **8**

Audio Networking

Network communication is the transport of data, voice, video, image, or facsimile (fax) from one location to another achieved by compatibly combining elements of hardware, software, and media. From a business perspective, network communications is delivering the right information to the right person at the right place and time for the right cost. Because there are so many variables involved in the analysis, design, and implementation of such networks, a structured methodology must be followed in order to assure that the implemented network meets the communications needs of the intended business, organization, or individual.

For an audio facility, the starting point is the business-level objectives. What is the company trying to accomplish by installing this network? Without a clear understanding of business level objectives, it is nearly impossible to configure and implement a successful network.

After business objectives are understood, the *applications* that will be running on the computer systems attached to these networks must be considered. After all, it is the applications that will be generating the traffic that will travel over the implemented network.

After applications are understood and have been documented, the *data* that those applications generate must be examined. In this case, the term “data” is used in a general sense as today’s networks are likely to transport a variety of payloads including voice, video, image, and fax in addition to “true” data. Data traffic analysis will determine not only the amount of data to be transported, but also important characteristics about the nature of that data.

Given these fundamental requirements, the task is to determine the specifications of a network that will possess the capability to deliver the expected data in a timely, cost-effective manner. These network performance criteria could be referred to as *what* the implemented network must do in order to meet the business objectives outlined at the outset of the process. Such requirements are known as the *logical network design*.

In This Section:

Chapter 8.1: Network Concepts	8-5
Introduction	8-5
OSI Model	8-5
Physical Layer	8-6
Data Link Layer	8-7

8-2 Section Eight

Network Layer	8-8
Transport Layer	8-8
Session Layer	8-8
Presentation Layer	8-9
Application Layer	8-9
Network Classifications	8-9
References	8-10
Bibliography	8-10
 Chapter 8.2: Serial Digital Video/Audio Systems	 8-11
Introduction	8-11
Serial Digital Interface	8-11
Embedded Audio	8-12
Error Detection and Handling	8-13
Packetized Data Transport	8-14
Optical Interconnect	8-15
High-Definition Serial Digital Interface	8-15
A Practical Implementation	8-17
Audio Interface Provisions	8-20
MPEG-2 Audio Transport	8-23
Data Services	8-23
Time Division Multiplexing on SMPTE 292M	8-24
Packet Transport	8-25
540 Mbits/s Interface	8-26
Serial Data Transport Interface	8-27
SDTI Data Structure	8-27
SDTI in Computer-Based Systems	8-29
SDTI Content Package Format	8-30
SMPTE 326M	8-30
Data Packet Encapsulation	8-32
Timing Issues	8-33
MPEG Decoder Templates	8-33
References	8-33
Bibliography	8-34
 Chapter 8.3: Video/Audio Networking Systems	 8-35
Introduction	8-35
Architecture of ATM	8-35
ATM Cell Structure	8-37
IEEE 1394	8-39
Operating Modes	8-39
Data Rates and Cable Lengths	8-39
Isochronous and Asynchronous Transmissions	8-40
Fibre Channel	8-42
Gigabit Ethernet	8-42
Network Bottlenecks	8-44
A Network Solution	8-44
Quality of Service	8-45

Reference Documents for this Section

- ATSC, "Guide to the Use of the Digital Television Standard," Advanced Television Systems Committee, Washington, D.C., Doc. A/54, Oct. 4, 1995.
- Craig, Donald: "Network Architectures: What does Isochronous Mean?," *IBC Daily News*, IBC, Amsterdam, September 1999.
- Dahlgren, Michael W.: "Servicing Local Area Networks," *Broadcast Engineering*, Intertec Publishing, Overland Park, Kan., November 1989.
- Fibush, David: *A Guide to Digital Television Systems and Measurement*, Tektronix, Beaverton, OR, 1994.
- Gaggioni, H., M. Ueda, F. Saga, K. Tomita, and N. Kobayashi, "The Development of a High-Definition Serial Digital Interface," Sony Technical Paper, Sony Broadcast Group, San Jose, Calif., 1998.
- Gallo and Hancock: *Networking Explained*, Digital Press, pp. 191–235, 1999.
- Goldman, J: *Applied Data Communications: A Business Oriented Approach*, 2nd ed., Wiley, New York, N.Y., 1998.
- Goldman, J: *Local Area Networks: A Business Oriented Approach*, 2nd ed., Wiley, New York, N.Y., 2000.
- Goldman, James E.: "Network Communication," in *The Electronics Handbook*, Jerry C. Whitaker (ed.), CRC Press, Boca Raton, Fla., 1996.
- Held, G.: *Ethernet Networks: Design Implementation, Operation and Management*, Wiley, New York, N.Y., 1994.
- Held, G.: *Internetworking LANs and WANs*, Wiley, New York, N.Y., 1993.
- Held, G.: *Local Area Network Performance Issues and Answers*, Wiley, New York, N.Y., 1994.
- Held, G.: *The Complete Modem Reference*, Wiley, New York, N.Y., 1994.
- International Organization for Standardization: "Information Processing Systems—Open Systems Interconnection—Basic Reference Model," ISO 7498, 1984.
- Legault, Alain, and Janet Matey: "Interconnectivity in the DTV Era—The Emergence of SDTI," *Proceedings of Digital Television '98*, Intertec Publishing, Overland Park, Kan., 1998.
- Miller, Mark A.: "Servicing Local Area Networks," *Microservice Management*, Intertec Publishing, Overland Park, Kan., February 1990.
- Miller, Mark A.: *LAN Troubleshooting Handbook*, M&T Books, Redwood City, Calif., 1990.
- "Networking and Internet Broadcasting," Omneon Video Networks, Campbell, Calif., 1999.
- "Networking and Production," Omneon Video Networks, Campbell, Calif., 1999.
- Owen, Peter: "Gigabit Ethernet for Broadcast and Beyond," *Proceedings of DTV99*, Intertec Publishing, Overland Park, Kan., November 1999.

8-4 Section Eight

- Piercy, John: "ATM Networked Video: Moving From Leased-Lines to Packetized Transmission," *Proceedings of the Transition to Digital Conference*, Intertec Publishing, Overland Park, Kan., 1996.
- "SMPTE Recommended Practice—Error Detection Checkwords and Status Flags for Use in Bit-Serial Digital Interfaces for Television," RP 165-1994, SMPTE, White Plains, N.Y., 1994.
- "SMPTE Recommended Practice—SDTI-CP MPEG Decoder Templates," RP 204 (Proposed), SMPTE, White Plains, N.Y., 1999.
- "SMPTE Standard for Television—24-Bit Digital Audio Format for HDTV Bit-Serial Interface," SMPTE 299M-1997, SMPTE, White Plains, N.Y., 1997.
- "SMPTE Standard for Television—Ancillary Data Packet and Space Formatting," SMPTE 291M-1998, SMPTE, White Plains, N.Y., 1998.
- "SMPTE Standard for Television—Bit-Serial Digital Interface for High-Definition Television Systems," SMPTE 292M-1998, SMPTE, White Plains, N.Y., 1998.
- "SMPTE Standard for Television—Element and Metadata Definitions for the SDTI-CP," SMPTE 331M-2000, SMPTE, White Plains, N.Y., 2000.
- "SMPTE Standard for Television—Encapsulation of Data Packet Streams over SDTI (SDTI-PF)," SMPTE 332M-2000, SMPTE, White Plains, N.Y., 2000.
- "SMPTE Standard for Television—Mapping of AES3 Data into MPEG-2 Transport Stream," SMPTE 302M-1998, SMPTE, White Plains, N.Y., 1998.
- "SMPTE Standard for Television—SDTI Content Package Format (SDTI-CP)," SMPTE 326M-2000, SMPTE, White Plains, N.Y., 2000.
- "SMPTE Standard for Television—Serial Data Transport Interface," SMPTE 305M-1998, SMPTE, White Plains, N.Y., 1998.
- SMPTE 344M (Proposed), "540 Mb/s Serial Digital Interface," SMPTE, White Plains, N.Y., 2000.
- "SMPTE Standard for Television (Proposed)—High Data-Rate Serial Data Transport Interface (HD-SDTI)," SMPTE 348M, SMPTE, White Plains, N.Y., 2000.
- "SMPTE Standard for Television (Proposed)—Signals and Generic Data over High-Definition Interfaces," SMPTE 346M, SMPTE, White Plains, N.Y., 2000.
- "SMPTE Standard for Television (Proposed)—Vertical Ancillary Data Mapping for Bit-Serial Interface," SMPTE 334M, SMPTE, White Plains, N.Y., 2000.
- "Technology Brief—Networking and Storage Strategies," Omneon Video Networks, Campbell, Calif., 1999.
- Turow, Dan: "SDTI and the Evolution of Studio Interconnect," *International Broadcasting Convention Proceedings*, IBC, Amsterdam, September 1998.
- Wilkinson, J. H., H. Sakamoto, and P. Horne: "SDDI as a Video Data Network Solution," *International Broadcasting Convention Proceedings*, IBC, Amsterdam, September 1997.
- Wu, Tsong-Ho: "Network Switching Concepts," *The Electronics Handbook*, Jerry C. Whitaker (ed.), CRC Press, Boca Raton, Fla., p. 1513, 1996.

Chapter
8.1
Network Concepts

James E. Goldman, Michael W. Dahlgren

Jerry C. Whitaker, Editor-in-Chief

8.1.1 Introduction

The open system interconnections (OSI) model is the most broadly accepted explanation of LAN transmissions in an open system. The reference model was developed by the International Organization for Standardization (ISO) to define a framework for computer communication. The OSI model divides the process of data transmission into the following steps:

- Physical layer
- Data link layer
- Network layer
- Transport layer
- Session layer
- Presentation layer
- Application layer

8.1.2 OSI Model

The OSI model allows data communications technology developers as well as standards developers to talk about the interconnection of two networks or computers in common terms without dealing in proprietary vendor jargon [1]. These common terms are the result of the layered architecture of the seven-layer OSI model. The architecture breaks the task of two computers communicating with each other into separate but interrelated tasks, each represented by its own layer. The top layer (layer 7) represents network services provided to the application program running on each computer and is therefore aptly named the *application layer*. The bottom layer (layer 1) is concerned with the actual physical connection of the two computers or networks and is there-

8-6 Audio Networking

fore named the *physical layer*. The remaining layers (2–6) may not be as obvious but, nonetheless, represent a sufficiently distinct logical group of functions required to connect two computers as to justify a separate layer.

To use the OSI model, a network analyst lists the known protocols for each computing device or network node in the proper layer of its own seven-layer OSI model. The collection of these known protocols in their proper layers is known as the *protocol stack* of the network node. For example, the physical media employed, such as unshielded twisted pair, coaxial cable, or fiber optic cable, would be entered as a layer 1 protocol, whereas Ethernet or token ring network architectures might be entered as a layer 2 protocol.

The OSI model allows network analysts to produce an accurate inventory of the protocols present on any given network node. This protocol profile represents a unique personality of each network node and gives the network analyst some insight into what *protocol conversion*, if any, may be necessary in order to allow any two network nodes to communicate successfully. Ultimately, the OSI model provides a structured methodology for determining what hardware and software technologies will be required in the physical network design in order to meet the requirements of the logical network design.

The basic elements and parameters of each layer are detailed in the following sections.

8.1.2a Physical Layer

Layer 1 of the OSI model is responsible for carrying an electrical current through the computer hardware to perform an exchange of information [2]. The physical layer is defined by the following parameters:

- Bit transmission rate
- Type of transmission medium (twisted pair, coaxial cable, or fiber optic cable), sometimes referred to as Layer 0
- Electrical specifications, including voltage- or current-based, and balanced or unbalanced
- Type of connectors used (for example, RJ-45 or DB-9)

Many different implementations exist at the physical layer

Layer 1 can exhibit error messages as a result of over usage. For example, if a file server is being burdened with requests from workstations, the results may show up in error statistic that reflect the server's inability to handle all incoming requests. An overabundance of *response timeouts* may also be noted in this situation. A response timeout (in this context) is a message sent back to the workstation stating that the waiting period allotted for a response from the file server has passed without action from the server.

Error messages of this sort, which can be gathered by any number of commercially available software diagnostic utilities, can indicate an overburdened file server or a hardware flaw within the system. Intermittent response timeout errors can also be caused by a corrupted *network interface card* (NIC) in the server. A steady flow of timeout errors throughout all nodes on the network may indicate the need for another server or bridge.

Hardware problems are among the easiest to locate in a networked system. In a simple configuration where something has suddenly gone wrong, the physical layer and the data-link layer are usually the first suspects.

8.1.2b Data Link Layer

Layer 2 of the OSI model, the data-link layer, describes hardware that enables data transmission (NICs and cabling systems) [2]. This layer integrates data packets into messages for transmission and checks them for integrity. Sometimes layer 2 will also send an “arrived safely” or “did not arrive correctly” message back to the transport layer (layer 4), which monitors this communications layer. The data link layer must define the frame (or package) of bits that is transmitted down the network cable. Incorporated within the frame are several important fields:

- Addresses of source and destination workstations
- Data to be transmitted between workstations
- Error control information, such as a *cyclic redundancy check* (CRC), which assures the integrity of the data

The data link layer must also define the method by which the network cable is accessed, because only one workstation can transmit at a time on a baseband LAN. The two predominant schemes are:

- *Token passing*, used with token ring and related networks
- *Carrier sense multiple access with collision detection* (CSMA/CD), used with Ethernet and and related networks

At the data link layer, the true identity of the LAN begins to emerge.

Because most functions of the data-link layer (in a PC-based system¹) take place in integrated circuits on NICs, software analysis is generally not required in the event of a failure. As mentioned previously, when something happens on the network, the data-link layer is among the first to suspect. Because of the complexities of linking multiple topologies, cabling systems, and operating systems, the following failure modes may be experienced:

- RF disturbance. Transmitters, ac power controllers, and other computers can all generate energy that may interfere with data transmitted on the cable. RF interference (RFI) is usually the single biggest problem in a broadband network. This problem can manifest itself through excessive checksum errors and/or garbled data.
- Excessive cable runs. Problems related to the data-link layer can result from long cable runs. Ethernet runs can stretch 1,000 ft. or more, depending on the cable and the Ethernet implementation. A basic token ring system can stretch 600 ft. or so with the same qualification. The need for additional distance can be accommodated by placing a *bridge*, *gateway*, *active hub*, equalizer, or amplifier on the line.

The data-link layer usually includes some type of routing hardware, including one or more of the following:

- Active hub
- Passive hub
- Multiple access units (for token ring-type networks)

1. In this context, the term “PC” is used to describe any computer, workstation, or laptop device.

8-8 Audio Networking

8.1.2c Network Layer

Layer 3 of the OSI model guarantees the delivery of transmissions as requested by the upper layers of the OSI [2]. The network layer establishes the physical path between the two communicating endpoints through the *communications subnet*, the common name for the physical, data link, and network layers taken collectively. As such, layer 3 functions (routing, switching, and network congestion control) are critical. From the viewpoint of a single LAN, the network layer is not required. Only one route—the cable—exists. Inter-network connections are a different story, however, because multiple routes are possible. The Internet Protocol (IP) and Internet Packet Exchange (IPX) are two examples of layer 3 protocols.

The network layer confirms that signals get to their designated targets, and then translates logical addresses into physical addresses. The physical address determines where the incoming transmission is stored. Lost data or similar errors can usually be traced back to the network layer, in most cases incriminating the network operating system. The network layer is also responsible for statistical tracking, and communications with other environments, including gateways. Layer 3 decides which route is the best to take, given the needs of the transmission. If router tables are being corrupted or excessive time is required to route from one network to another, an operating system error on the network layer may be involved.

8.1.2d Transport Layer

Layer 4, the transport layer, acts as an interface between the bottom three and the upper three layers, ensuring that the proper connections are established and maintained [2]. It does the same work as the network layer, only on a local level. The network operating system driver performs transport layer tasks.

Connection difficulties between computers on a network can sometimes be attributed to the shell driver. The transport layer may have the ability to save transmissions that were en route in the case of a system crash, or re-route a transmission to its destination in case of primary route failure. The transport layer also monitors transmissions, checking to make sure that packets arriving at the destination node are consistent with the *build specifications* given to the sending node in layer 2. The data-link layer in the sending node builds a series of packets according to specifications sent down from higher levels, then transmits the packets to a *destination node*. The transport layer monitors these packets to ensure they arrive according to specifications indicated in the original build order. If they do not, the transport layer calls for a retransmission. Some operating systems refer to this technique as a *sequenced packet exchange* (SPX) transmission, meaning that the operating system guarantees delivery of the packet. Transmission control protocol (TCP) is the transport layer protocol associated with the IP network layer protocol.

8.1.2e Session Layer

Layer 5 is responsible for turning communications on and off between communicating parties [2]. Unlike other levels, the session layer can receive instructions from the application layer through the network basic input/output operation system (netBIOS), skipping the layer directly above it. The netBIOS protocol allows applications to “talk” across the network. The session layer establishes the session, or logical connection, between communicating host processors. Name-to-address translation is another important function; most communicating processors are known by a common name, rather than a numerical address.

Multi-vendor problems often crop up in the session layer. Failures relating to gateway access usually fall into layer 5 for the OSI model, and are typically related to compatibility issues.

8.1.2f Presentation Layer

Layer 6 translates application layer commands into syntax understood throughout the network [2]. It also translates incoming transmissions for layer 7. The presentation layer masks other devices and software functions. Layer 6 software controls printers and other peripherals, and may handle encryption and special file formatting. Data compression, encryption, and translations are examples of presentation layer functions.

Failures in the presentation layer are often the result of products that are not compatible with the operating system, an interface card, a resident protocol, or another application.

8.1.2g Application Layer

At the top of the seven-layer stack is the application layer. It is responsible for providing protocols that facilitate user applications [2]. Print spooling, file sharing, and e-mail are components of the application layer, which translates local application requests into network application requests. Layer 7 provides the first layer of communications into other open systems on the network.

Failures at the application layer usually center around software quality and compatibility issues. The program for a complex network may include latent faults that will manifest only when a given set of conditions are present. The compatibility of the network software with other programs is another source of potential problems.

8.1.3 Network Classifications

Although there are no hard and fast rules for network categorization, some general parameters are usually accepted for most applications. The following are a few of the more common categories of networking [1]:

- **Remote connectivity:** A single remote user wishes to access local network resources. This type of networking is particularly important to mobile professionals such as sales representatives, service technicians, field auditors, and so on.
- **Local area networking:** Multiple users' computers are interconnecting for the purpose of sharing applications, data, or networked technology such as printers or mass storage. Local area networks (LANs) can have anywhere from two or three users to several hundred (or more). LANs are often limited to a single department or floor in a building, although technically any single-location corporation could be networked via a LAN.
- **Internetworking:** Also known as LAN-to-LAN networking or connectivity, internetworking involves the connection of multiple LANs and is common in corporations in which users on individual departmental LANs need to share data or otherwise communicate. The challenge of internetworking is in getting departmental LANs of different protocol stacks (as determined by use of the OSI model) to talk to each other, while only allowing authorized users

8-10 Audio Networking

access to the internetwork and other LANs. Variations of internetworking also deal with connecting LANs to mainframes or minicomputers rather than to other LANs.

- **Wide area networking:** Also known as *enterprise networking*, involves the connection of computers, network nodes, or LANs over a sufficient distance as to require the purchase of wide area network (WAN) service from the local phone company or an alternative carrier. In some cases, the wide area portion of the network may be owned and operated by the corporation itself. Nonetheless, the geographic distance between nodes is the determining factor in categorizing a wide area network. A subset of WANs, known as *metropolitan area networks* (MANs), are confined to a campus or metropolitan area of usually not more than a few miles in diameter.

The important thing to remember is that categorization of networking is somewhat arbitrary and that what really matters is that the proper networking technology (hardware and software) is specified in any given system in order to meet stated business objectives.

8.1.4 References

1. Goldman, James E.: "Network Communication," in *The Electronics Handbook*, Jerry C. Whitaker (ed.), CRC Press, Boca Raton, Fla., 1996.
2. Dahlgren, Michael W.: "Servicing Local Area Networks," *Broadcast Engineering*, Intertec Publishing, Overland Park, Kan., November 1989.

8.1.5 Bibliography

- Goldman, J: *Applied Data Communications: A Business Oriented Approach*, 2nd ed., Wiley, New York, N.Y., 1998.
- Goldman, J: *Local Area Networks: A Business Oriented Approach*, 2nd ed., Wiley, New York, N.Y., 2000.
- Held, G.: *Ethernet Networks: Design Implementation, Operation and Management*, Wiley, New York, N.Y., 1994.
- Held, G.: *Internetworking LANs and WANs*, Wiley, New York, N.Y., 1993.
- Held, G.: *Local Area Network Performance Issues and Answers*, Wiley, New York, N.Y., 1994.
- Held, G.: *The Complete Modem Reference*, Wiley, New York, N.Y., 1994.
- International Organization for Standardization: "Information Processing Systems—Open Systems Interconnection—Basic Reference Model," ISO 7498, 1984.
- Miller, Mark A.: *LAN Troubleshooting Handbook*, M&T Books, Redwood City, Calif., 1990.
- Miller, Mark A.: "Servicing Local Area Networks," *Microservice Management*, Intertec Publishing, Overland Park, Kan., February 1990.

Chapter
8.2

Serial Digital Video/Audio Systems

Jerry C. Whitaker, Editor-in-Chief

8.2.1 Introduction

Parallel connection of digital video equipment is practical only for relatively small installations. There is, then, a clear need to transmit data over a single coaxial or fiber line [1]. To reliably move large amounts of data from one location to another, it is necessary to modify the serial signal prior to transmission to ensure that there are sufficient edges (data transitions) for reliable clock recovery, to minimize the low frequency content of the transmitted signal, and to spread the transmitted energy spectrum so that radio frequency emission problems are minimized.

In the early 1980s, a serial interface for Rec. 601 signals was recommended by the EBU. This interface used 8/9 block coding and resulted in a bit rate of 243 Mbits/s. The interface did not support ten-bit precision signals, and there were some difficulties in producing reliable, cost effective integrated circuits for the protocol. The block coding-based interface was abandoned and replaced by an interface with a channel coding scheme that utilized scrambling and conversion to NRZI (*non return to zero inverted*). This serial interface was standardized as SMPTE 259M and EBU Tech. 3267, and is defined for both component and composite conventional video signals, including embedded digital audio.

8.2.2 Serial Digital Interface

Conceptually, the serial digital interface is much like a carrier system for studio applications. Baseband audio and video signals are digitized and combined on the serial digital “carrier.” (SDI is not strictly a carrier system in that it is a baseband digital signal, not a signal modulated on a carrier wave.) The bit rate (carrier frequency) is determined by the clock rate of the digital data: 143 Mbits/s for NTSC, 177 Mbits/s for PAL, and 270 Mbits/s for Rec. 601 component digital. The widescreen (16 × 9) component system defined in SMPTE 267 will produce a bit rate of 360 Mbits/s. This serial interface may be used with normal video coaxial cable or fiber optic cable, with the appropriate interface adapters.

Following serialization of the video information, the data stream is scrambled by a mathematical algorithm and then encoded. At the receiver, an inverse algorithm is used in the deserializer to recover the data. In the serial digital transmission system, the clock is contained in the data, as

8-12 Audio Networking

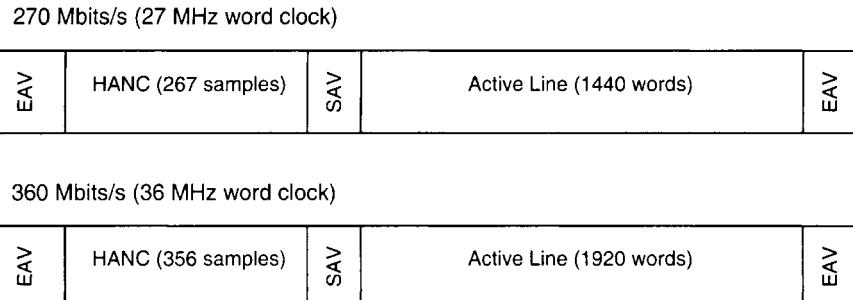


Figure 8.2.1 The basic SDI bitstream.

opposed to the parallel system where there is a separate clock line. By scrambling the data, an abundance of transitions is assured, which is required for reliable clock recovery.

Figure 8.2.1 shows the SDI bitstream for 270 Mbit/s and 360 Mbit/s operation. The EAV and SAV elements of the bitstream are reserved word sequences that indicate the start and end of a video line, respectively. For the 270 Mbit/s case, each line contains 1440 10 bit 4:2:2 video samples. The horizontal interval (HANC, *horizontal ancillary data*) contains ancillary data, error detection and control, embedded audio, and other information. Vertical ancillary data (VANC) also can be used.

8.2.2a Embedded Audio

One of the important features of the serial digital video interface is the facility to embed (multiplex) several channels of AES/EBU digital audio in the video bitstream. SDI with embedded audio is particularly helpful in large systems where a strict link between the video and its associated audio is an important feature. In smaller systems, such as a post-production suite, it is generally preferable to maintain a separate audio path.

SMPTE 272M defines the mapping of digital audio data, auxiliary data, and associated control information into the ancillary data space of the serial digital video format. Several modes of operation are defined and letter suffixes are used to help identify interoperability between equipment with differing capabilities. These descriptions are given in Table 8.2.1. (Note that modes *B* through *J* shown in the table require a special audio control packet.)

Some examples will help explain how Table 8.2.1 is used. A transmitter that can only accept 20 bit 48 kHz synchronous audio is said to conform to SMPTE 272M-A. A transmitter that supports 20 bit and 24 bit 48 kHz synchronous audio conforms to SMPTE 272M-ABC. A receiver that only uses the 20 bit data but can accept the level *B* sample distribution would conform to SMPTE 272M-AB because it can handle either sample distribution.

Table 8.2.1 SMPTE 272M Mode Definitions

A (default)	Synchronous 48 kHz, 20 bit audio, 48 sample buffer
B	Synchronous 48 kHz, composite video only, 64 sample buffer to receive 20 bits from 24 bit audio data
C	Synchronous 48 kHz, 24-bit audio and extended data packets
D	Asynchronous (48 kHz implied, other rates if so indicated)
E	44.1 kHz audio
F	32 kHz audio
G	32-48 kHz variable sampling rate audio
H	Audio frame sequence (inherent in 29.97 frame/s video systems, except 48 kHz synchronous audio—default A mode)
I	Time delay tracking
J	Non-coincident channel status Z bits in a pair.

8.2.2b Error Detection and Handling

SMPTE Recommended Practice RP 165-1994 describes the generation of error detection checkwords and related status flags to be used optionally in conjunction with the serial digital interface [2]. Although the RP on *error detection and handling* (EDH) recommends that the specified error-checking method be used in all serial transmitters and receivers, it is not required.

Two checkwords are defined: one based on a field of active picture samples and the other on a full field of samples. This two-word approach provides continuing error detection for the active picture when the digital signal has passed through processing equipment that has changed data outside the active picture area without re-calculating the full-field checkword.

Three sets of *flags* are provided to feed-forward information regarding detected errors to help facilitate identification of faulty equipment, and the type of fault. One set of flags is associated with each of the two field-related checkwords. A third set of flags is used to provide similar information based on evaluating all of the ancillary data checksums within a field. The checkwords and flags are combined in an *error detection data packet* that is included as ancillary data in the serial digital signal. At the receiver, a recalculation of check-words can be compared to the error detection data packet information to determine if a transmission error has occurred.

All error flags indicate only the status of the previous field; that is, each flag is set or cleared on a field-by-field basis. A logical *1* is the set state and a logical *0* is the unset state. The flags are defined as follows:

- EDH, *error detected here*: Signifies that a serial transmission data error was detected. In the case of ancillary data, this means that one or more ANC data blocks did not match its checksum.
- EDA, *error detected already*: Signifies that a serial transmission data error has been detected somewhere upstream. If device *B* receives a signal from device *A* and device *A* has set the

8-14 Audio Networking

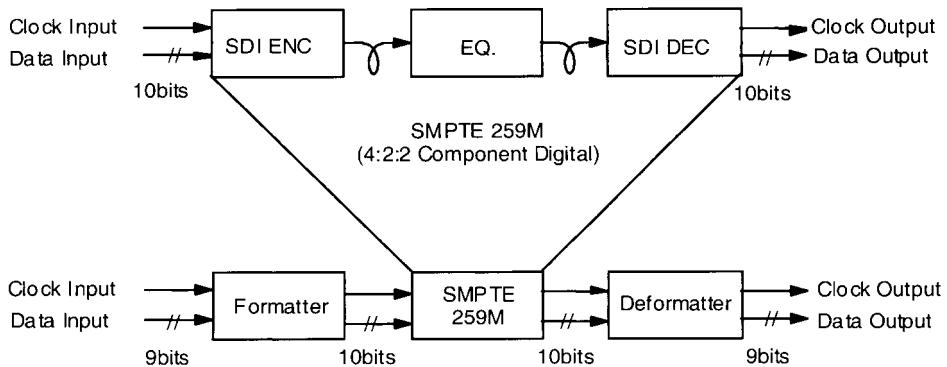


Figure 8.2.2 SMPTE 305M system block diagram. (After [3].)

EDH flag, when device *B* retransmits the data to device *C*, the EDA flag will be set and the EDH flag will be unset (if there is no further error in the data).

- IDH, *internal error detected here*: Signifies that a hardware error unrelated to serial transmission has been detected within a device. This feature is provided specifically for devices that have internal data error-checking facilities.
- IDA, *internal error detected already*: Signifies that an IDH flag was received and there was a hardware device failure somewhere upstream.
- UES, *unknown error status*: Signifies that a serial signal was received from equipment not supporting the RP 165 error-detection practice.

Individual error status flags (or all error status flags) may not be supported by all equipment.

8.2.2c Packetized Data Transport

The SMPTE 305M standard specifies a data stream used to transport packetized data within a studio/production center environment [3]. The data packets and synchronizing signals are compatible with SMPTE 259M, as illustrated in Figure 8.2.2. Other parameters of the protocol also are compatible with the 4:2:2 component SDI format.

The data stream is intended to transport any packetized data signal over the active lines that have a maximum data rate up to (approximately) 200 Mbits/s for a 270 Mbits/s system or (approximately) 270 Mbits/s for 360 Mbits/s system. The maximum data rate can be increased through use of a defined extended data space.

The SMPTE 305M standard describes the assembly of a stream of 10-bit words. The resulting word stream is serialized, scrambled, coded, and interfaced according to SMPTE 259M and ITU-R BT.656. The timing reference signals (EAV and SAV) occur on every line. The signal levels, specifications, and preferred connector type are as described in SMPTE 259M.

8.2.2d Optical Interconnect

SMPTE 297M-1997 defines an optical fiber system for transmitting bit-serial digital signals that conform to the SMPTE 259M serial digital format (143 through 360 Mbits/s). The standard's optical interface specifications and end-to-end system performance parameters are otherwise compatible with SMPTE 292M, which covers transmission rates of 1.3 through 1.5 Gbits/s.

During the Summer of 2000, a major revision was undertaken of SMPTE 297M. The document—proposed at this writing—updated applicable references and included new information on important elements covered by the standard. Among the new information were four appendices examining the following subjects:

- Maximum transmission distance range
- Minimum transmission distances
- Computing damage thresholds
- A comprehensive glossary of fiber optic terms

8.2.3 High-Definition Serial Digital Interface

In an effort to address the facility infrastructure requirements of HDTV, the SMPTE and BTA developed a standard for digital serial transmission of studio HDTV signals [4]. The overall transmission rate for transporting a digital studio HDTV signal (1125-line, 2:1 interlace, with 10-bit component sampling) is approximately 1.5 Gbits/s. The active payload is on the order of 1.2 Gbits/s (for 1035/1080 active lines). The transmission of video signals at these bit rates represents a far more difficult technical challenge than serial distribution at 270 Mbits/s used for conventional television signals.

The introduction of the serial digital interface for conventional video (SMPTE 259M) was well received by the television industry and has become the backbone of digital audio/video networking for broadcast and post production installations around the world. SDI is ideally suited to the task of transporting uncompressed component/composite digital video and multichannel audio signals over a single coaxial cable. To emulate the same level of operational usability and system integration of conventional television equipment in the HDTV world, the implementation of a *high-definition serial digital interface* (HD-SDI) system—based on an extension of SMPTE 259M—was essential.

Work on the HD-SDI system began in 1992 under the auspices of SMPTE and BTA. The end result of these efforts was the BTA document BTA S-004 (May 1995), followed closely by SMPTE 292M-1996, both with similar technical content.

The source formats of SMPTE 292M adhere to those signal characteristics specified in SMPTE 260M and 274M. In particular, the field frequencies of 59.94 Hz and 60.00 Hz, and active line numbers of 1035/1080 are used by HD-SDI. Table 8.2.2 lists the basic parameters of the input source formats.

Subsequently, a revision of SMPTE 292M was undertaken, resulting in SMPTE 292M-1998 [5]. The revised source format parameters are given in Table 8.2.3. Note that the total data rate is either 1.485 Gbits/s or 1.485/1.001 Gbits/s. In the table, the former is indicated by a rate of “1” and the latter by a rate of “M,” which is equal to 1.001.

8-16 Audio Networking

Table 8.2.2 SDI Reference Source Format Parameters (After [4].)

Reference Document	SMPTE-260M	SMPTE-274M	SMPTE-274M
Parallel word rate (each channel $Y, C_R/C_B$)	74.25 Mword/s	74.25 Mword/s	74.25/1.001 Mword/s
Lines per frame	1125	1125	1125
Words per active line (each channel $Y, C_R/C_B$)	1920	1920	1920
Total active line	1035	1080	1080
Words per total line (each channel $Y, C_R/C_B$)	2200	2200	2200
Frame rate	30 Hz	30 Hz	30/1.001 Hz
Total Fields per frame	2	2	2
Total data rate	1.485Gbits/s	1.485Gbits/s	1.485/1.001 Gbits/s
Field 1 EAV V = 1	Line 1121	Line 1124	Line 1124
Field 1 EAV V = 0	Line 41	Line 21	Line 21
Field 2 EAV V = 0	Line 558	Line 561	Line 561
Field 2 EAV V = 0	Line 603	Line 584	Line 584
EAV F = 0	Line 1	Line 1	Line 1
EAV F = 1	Line 564	Line 563	Line 563

Table 8.2.3 Source Format Parameters for SMPTE 292M (After [5].)

Reference SMPTE Standard	260M		295M	274M								296M	
Format	A	B	C	D	E	F	G	H	I	J	K	L	M
Lines per frame	1125	1125	1250	1125	1125	1125	1125	1125	1125	1125	1125	750	750
Words per active line (each channel $Y, C_B/C_R$)	1920	1920	1920	1920	1920	1920	1920	1920	1920	1920	1920	1280	1280
Total active lines	1035	1035	1080	1080	1080	1080	1080	1080	1080	1080	1080	720	720
Words per total line (each channel $Y, C_B/C_R$)	2200	2200	2376	2200	2200	2640	2200	2200	2640	2750	2750	1650	1650
Frame rate (Hz)	30	30/M	25	30	30/M	25	30	30/M	25	24	24/M	60	60/M
Fields per frame	2	2	2	2	2	2	1	1	1	1	1	1	1
Data rate divisor	1	M	1	1	M	1	1	M	1	1	M	1	M

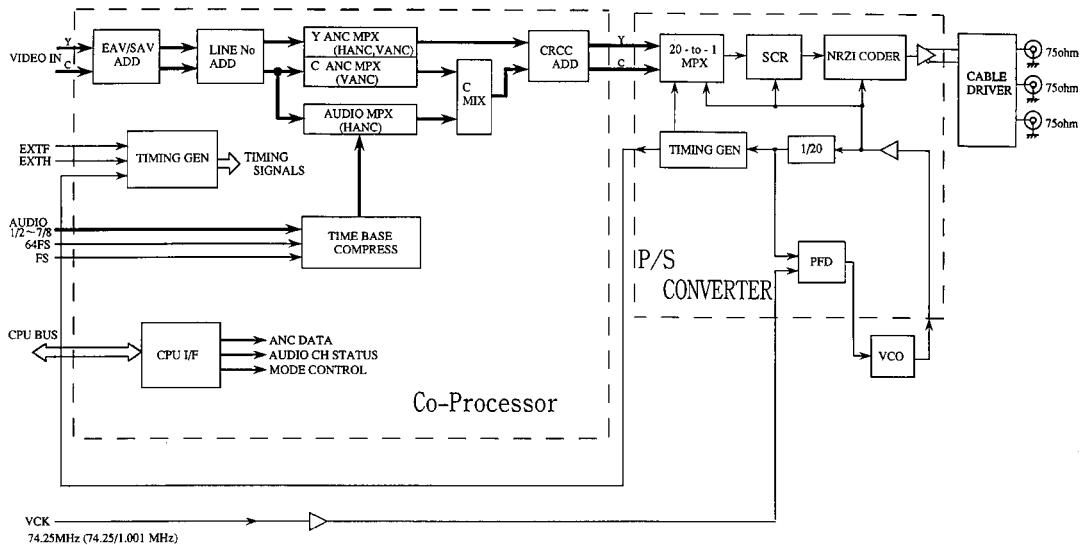


Figure 8.2.3 Block diagram of an HD-SDI transmission system. (From [4]. Courtesy of Sony.)

8.2.3a A Practical Implementation

To better understand the operation of the HD-SDI system, it is instructive to consider a practical example. A set of HD-SDI transmitter/receiver modules was developed (Sony) to provide the desired interconnection capabilities for HDTV [4].

Each module of the system makes use of a *coprocessor* IC that implements the data structure (protocol) specified in SMPTE 292M. This device has both transmitting and receiving functions, which makes it possible to transmit and receive video, audio, and ancillary data, as well as EAV/SAV, line number, and other parameters.

The transmission module consists of two main ICs. (A block diagram of the system is shown in Figure 8.2.3.) The first is the coprocessor, which is used to embed EAV/SAV, line number, and CRC information in the input digital video signal. This device also serves to multiplex the audio data and audio channel status information in the HANC area of the chrominance channel. Conventional ancillary data is multiplexed in the ancillary space (HANC and/or VANC of Y , and P_B / P_R signals).

The second element of the transport system is the *P/S converter* IC, which converts the two channels of parallel data (luminance and chrominance) into a single serial bit stream. This device also performs the encoding operation of scrambled NRZI, which is the channel coding technique stipulated in SMPTE 292M/BTA S-004. At the output of the P/S converter IC, the serial data rate is 1.485 Gbits/s. The input video signal is represented by 10 bits in parallel for each of the Y and P_B/P_R samples. External timing signals (EXTF/EXTH) are provided as additional inputs to the IC for cases when EAV/SAV information is not present in the input parallel video data.

8-18 Audio Networking

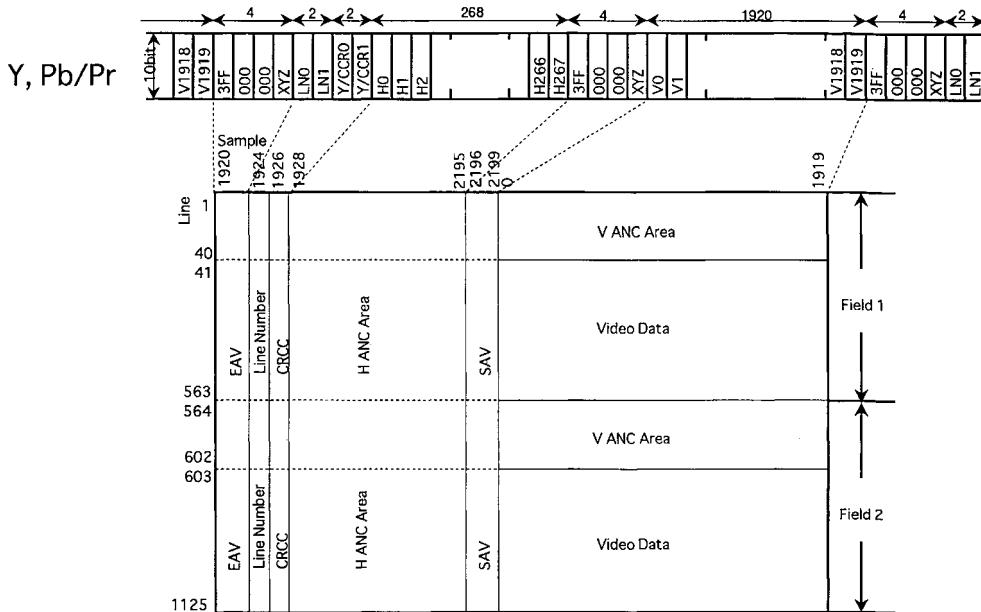


Figure 8.2.4 The HD-SDI signal format from the coprocessor IC to the P/S device. (From [4]. Courtesy of Sony.)

The video clock frequency is 74.25 MHz (or 74.25/1.001 MHz), which is synchronous with the video data. The audio data packet is multiplexed in the HANC area of the chrominance channel, with the exception of lines 8 and 570. All of the video and ancillary data present in the input digital signal, and the embedded data (line numbers, EAV/SAV, CRC, and other data) are transmitted to the P/S device without alteration, where the data is converted into a serial form. Figure 8.2.4 shows the data footprint for the system.

As mentioned previously, the interface between the coprocessor and P/S converter consists of two channels of I/O bit parallel data at 74.25 Mwords/s. The P/S IC converts these two parallel channels into a serial data stream at 1.485 Gbits/s by means of a 20-to-1 multiplexer. This serial data is then encoded using scrambled NRZI. The generator polynomial for the NRZI scrambler is [4]:

$$G_{(x)} = (x^9 + x^4 + 1) \cdot (x + 1) \quad (8.2.1)$$

From the P/S IC, the encoded serial data is output in ECL form and distributed by 3 channels of coaxial outputs. The signal amplitude of the coaxial output is 800 mV p-p into $75\ \Omega$.

Figure 8.2.5 shows a block diagram of the receiver module. Cable equalizer circuitry compensates for high-frequency losses of the coaxial cable, and the clock recovery circuit extracts the

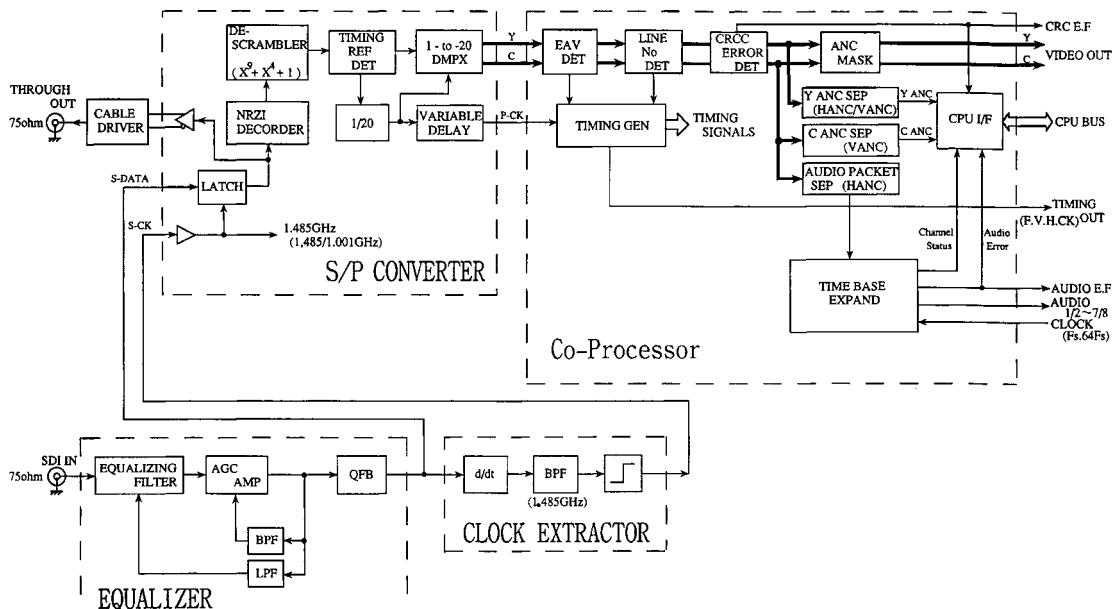


Figure 8.2.5 Block diagram of an HD-SDI receiver system. (From [4]. Courtesy of Sony.)

1.485 GHz serial clock. The serial-to-parallel converter section reconstructs the parallel data from the serial bitstream, while the coprocessor IC separates the video, audio, and ancillary data. The cable equalizer is an automatic system that compensates for the frequency-dependent attenuation of coaxial cable. The attenuation characteristic of a 5C-2V cable (75Ω comparable to Belden 8281) is shown in Figure 8.2.6.

By using a PIN diode, the compensation curve of the receiving system can be controlled while satisfying the condition of linear phase. The cable length is detected by a band-pass filter, and the equalizing filter is controlled to keep the amplitude constant. A dc restorer also is used. After a decision stage, the threshold-controlled signal is fed back through a low-pass filter, permitting recovery of the dc component. The data given in Figure 8.2.7 show the eye-patterns for two cable lengths. The bit error rate for 1.485 Gbits/s transmission over these lengths is less than 1 error in 10^{10} .

The clock recovery circuit is a clock frequency filter that works on the serial data recovered from the cable equalizer. The transition edges of the serial data contain the 1.485 GHz clock frequency that is selected by the filter. The HDTV source formats make use of two field frequencies: 60.00 Hz and 59.94 Hz. Hence, the HD-SDI system provides two values for the serial clock frequency, that is, 1.485 GHz and 1.4835 GHz (1.485 GHz/1.001). Figures 8.2.8a and 8.2.8b show the recovered clock frequencies of 1.485 GHz and 1.4835 GHz, respectively.

The serial-to-parallel converter reconstructs the parallel data from the serial bit stream. The serial data recovered from the cable equalizer is reclocked by the serial clock frequency that is

8-20 Audio Networking

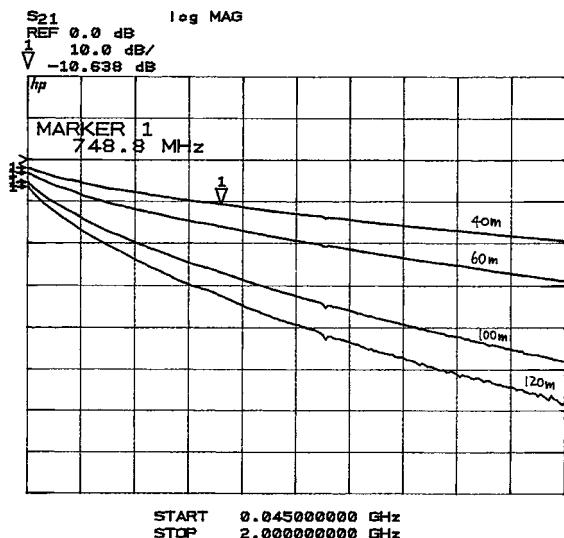


Figure 8.2.6 Attenuation characteristics of 5C-2V coaxial cable.
(From [4]. Courtesy of Sony.)

produced at the output of the clock recovery circuit. At this point, the channel coding scrambled NRZI is decoded and the unique pattern of the timing reference signal EAV/SAV is detected. In order to generate the 74.25 MHz clock frequency for the parallel data words, the 1.485 GHz serial clock is divided by 20 and synchronized using detection of the timing reference signals. The serial data is next latched by the parallel clock waveform to generate the 20 bits of parallel data. The serial data, latched by the serial high-frequency clock, also is provided as an active loop-through output.

The coprocessor IC separates video, audio, and ancillary data from the 20-bit parallel digital signal. EAV/SAV information and line number data are detected from the input signal, permitting the regeneration of F/V/H video timing waveforms. Transmission errors are detected by means of CRC coding.

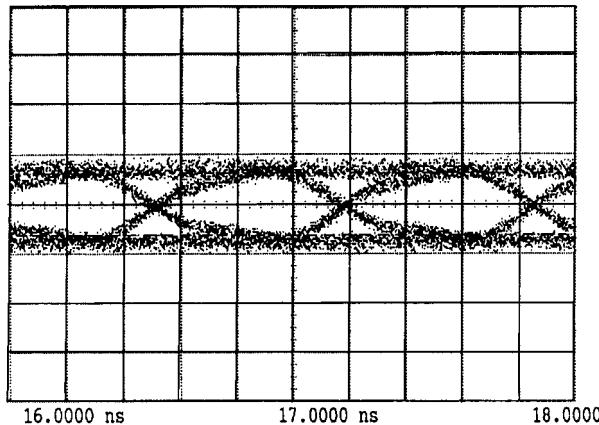
The embedded audio packet is extracted from the HANC space of the P_B/P_R channel and the audio data are written into memory. The audio data are then read out by an external audio clock frequency F_s , enabling the reconstruction of the received audio information. The coprocessor IC can receive up to 8 channels of embedded audio data.

8.2.3b Audio Interface Provisions

SMPTE 299M-1997 defines the mapping of 24-bit AES digital audio data and associated control information into the ancillary data space of a serial digital video stream conforming to SMPTE 292M [6]. The audio data are derived from ANSI S4.40, more commonly referred to as AES audio.

An audio signal, sampled at a clock frequency of 48 kHz locked (synchronous) to video, is the preferred implementation for intrastudio applications. As an option, this standard supports AES audio at synchronous or asynchronous sampling rates from 32 kHz to 48 kHz. The number

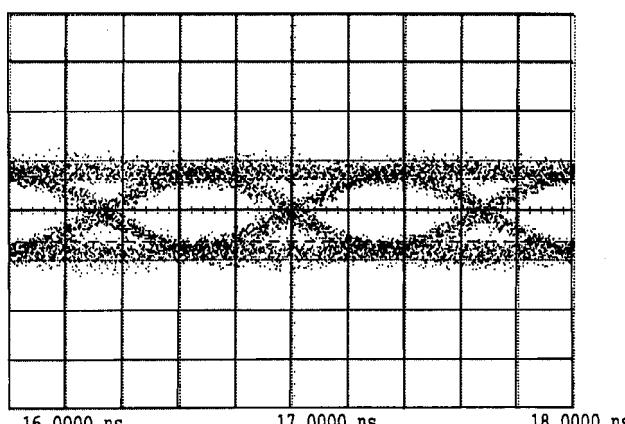
(a)



Ch. 1 = 80.00 mVolts/div Offset = -28.00 mVolts
 Timebase = 200 ps/div Delay = 16.0000 ns
 Delta V = 100.00 mVolts
 Vmarker1 = -77.500 mVolts Vmarker2 = 22.500 mVolts

(b)

Trigger on External at Pos. Edge at -1.902 Volts



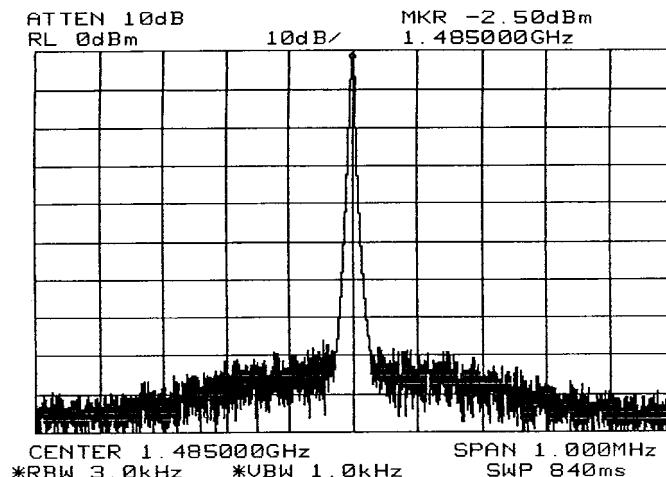
Ch. 1 = 80.00 mVolts/div Offset = -28.00 mVolts
 Timebase = 200 ps/div Delay = 16.0000 ns
 Delta V = 100.00 mVolts
 Vmarker1 = -77.500 mVolts Vmarker2 = 22.500 mVolts

Trigger on External at Pos. Edge at -1.902 Volts

Figure 8.2.7 Eye diagram displays of transmission over 5C-2V coaxial cable: (a) 3 m length, (b) 100 m length. (After [4]. Courtesy of Sony.)

8-22 Audio Networking

(a)



(b)

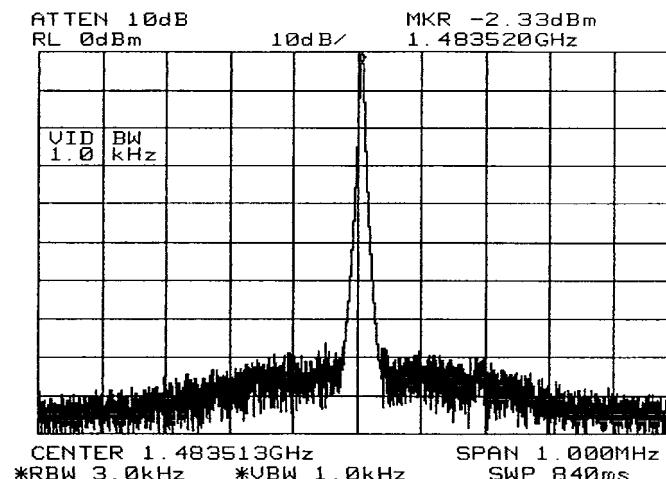


Figure 8.2.8 Recovered signal clocks: (a) 1.485 GHz 60 Hz system, (b) 1.4835 GHz 59.94 Hz system. (After [4]. Courtesy of Sony.)

of transmitted audio channels ranges from a minimum of two to a maximum of 16. Audio channels are transmitted in pairs, and where appropriate, in groups of four. Each group is identified by a unique ancillary data ID.

Audio data packets are multiplexed into the horizontal ancillary data space of the C_B/C_R parallel data stream, and audio control packets are multiplexed into the horizontal ancillary data space of the Y parallel data stream.

MPEG-2 Audio Transport

SMPTE 302M specifies the transport of AES3 data in an MPEG-2 transport stream for television applications at a sampling rate of 48 ksamples/s [7]. The MPEG audio standard itself defines compressed audio carriage, but does not define uncompressed audio for carriage in an MPEG-2 transport system. SMPTE 302M augments the MPEG standard to address the requirement to carry AES3 streams, which may consist of linear PCM audio or other data. MPEG-2 transport streams convey one or more programs of coded data, and may be constructed from one or more elementary coded data streams, program streams, or other transport streams.

The specifications are described in terms of a model that starts with AES3 data, constructs *elementary streams* (ES) from the AES3 data, then constructs *packetized elementary streams* (PES) from the elementary streams, and finally constructs *MPEG-2 transport streams* (MTS) from the packetized elementary streams. Although this model is used to describe the transport of AES3 streams in MPEG-2 transport streams, the model is not mandatory. MPEG-2 transport streams may be constructed by any method that results in a valid stream.

The SMPTE audio data elementary streams consists of audio sample words, which may be derived from AES3 digital audio subframes, together with validity, user, and channel status (V, U, C) bits and a framing (F) bit. There may be 2, 4, 6, or 8 channels of audio data conveyed in a single audio elementary stream and corresponding packetized elementary stream. Multiple packetized elementary streams may be used in applications requiring more channels.

8.2.3c Data Services

SMPTE 334M (proposed at this writing) defines a method of coding that allows data services to be carried in the vertical ancillary data space of a bit-serial component television signal conforming with SMPTE 292M or ANSI/SMPTE 259M [8]. This includes data broadcast services intended for the public as well as broadcaster internal control and communications. Despite the reference to the bit-serial interface, nothing in the specification precludes its use in a parallel digital interface for component digital video signals. The data described in the standard can also be transported in K-L-V format according to SMPTE 336M, or via other means.

The data packets are located in the active line portion of one or more lines in the vertical ancillary space. Data can be located in any lines in the area from the second line after the line specified for switching to the last line before active video, inclusively. Individual data services are not assigned to any specific data lines; receiving equipment should identify and select services on the basis of their ANC DID and SDID fields.

Because ANC data may be located in the lines immediately preceding active video, manufacturers of video compression equipment must ensure that these data bits are not included in video compression calculations.

The chrominance (C_b/C_r) and luminance (Y) data are carried in two separate streams within the 292M signal, complete with their own ANC data flags and CRCs. Defined data services are carried in the Y stream. Other data services can be inserted into either one of these streams without restrictions.

8-24 Audio Networking

In the 259M/125M signal, the chrominance and luminance data are carried in a single stream. In this case, all data services are carried in this stream with a single ANC data flag and CRC.

There is no specific provision in SMPTE 334M for ensuring that the relative timing between the video and its embedded VANC data is correct. The only timing relationship that exists is created when the data are embedded in the video. Once that relationship is established, the deterministic nature of 292M or 259M transport ensures that the relationship is preserved.

8.2.3d Time Division Multiplexing on SMPTE 292M

Given the wide variety of commercially accepted standards for both standard and high-definition television systems used in studio production, post-production, and distribution facilities, a large number of incompatible interfaces exists for the interconnection and routing of the various 4:2:2I, 4:4:4I, 4:2:2:4I, 4:4:4:4I, and 4:2:2P component and composite video signals throughout a given plant [9]. In addition, high-definition video signals used in a number of television facilities operate at a bandwidth of 1.485 Gbits/s.

A higher bandwidth interface such as that used to carry high-definition component video could carry several lower-bandwidth component or composite video signals over a single physical link, reducing the number of cable or fiber routing signals throughout the plant and simplifying the overall routing requirements. In fact, the lower-bandwidth signals could be formatted into digital active line areas of existing high-definition interfaces in such a way that the resulting signal would appear to most existing pieces of equipment as a regular high-definition video signal.

In addition, the low-bandwidth signals could be any generic data stream, including compressed standard- and high-definition video, not just component video signals. The obvious benefit of this multiplexing scheme is that existing equipment for serializing and deserializing high-definition bit-parallel data signals for distribution throughout the plant could also be used to serialize and deserialize the high-bandwidth multiplexed data stream without any additional hardware.

All that is needed to implement this system is hardware for multiplexing and demultiplexing between bit-parallel high- and low-bandwidth formats.

With these real-world needs in mind, the SMPTE developed an innovative solution aimed at bridging the video hardware of the present with that of the future. SMPTE 346M (proposed at this writing) defines the *time division multiplexing* (TDM) of various standard-definition digital video and generic 8-bit data signals over the high-definition serial digital interface specified in SMPTE 292M. The objective of this high-definition multiplexing interface is to use a single physical link to transmit, distribute, route, and switch a complete family of existing 10-bit video formats and various data formats.

Active video and vertical blanking areas in the HD SDI stream are time divided into 19 interleaved channels. The word in the first channel is used to indicate the data validation of the remaining 18 channels. A single video or data stream is multiplexed into one or multiple channels of the total 18 data channels. A control packet is multiplexed into the horizontal ancillary data space of the luminance parallel data area after the switching point of the HD stream for each video or data stream. The control packet indicates how and which channels are used for this video or data stream. It also contains *stream clock reference* information for clock recovery of the original clock signal.

Multiple standard-definition video or data streams can be multiplexed in and demultiplexed from a single HD SDI stream with a total delay of a fraction of a horizontal line. By dividing the

Table 8.2.4 Summary of SD Video Formats Referenced in SMPTE 346M (After [9])

SD System/Sampling Structure	525×60 or 625×50 , 4×3 13.5 MHz	525×59.94 or 625×50 , 16×9 18 MHz
4:0:0I	135 Mbits/s	180 Mbits/s
4:2:2I	270 Mbits/s ANSI/SMPTE 267M	360 Mbits/s ANSI/SMPTE 267M
4:2:2:4I	360 Mbits/s	540 Mbits/s
4:4:4I	360 Mbits/s ITU-R BT.601-5	540 Mbit/s ITU-R BT.601-5
4:2:0P	360 Mbits/s	540 Mbits/s
4:4:4:4I	540 Mbits/s SMPTE RP 174	720 Mbits/s
8:4:4I	540 Mbits/s	720 Mbits/s
4:2:2P	540 Mbits/s ANSI/SMPTE 293M	720 Mbits/s

payload into segments, which are subdivided into channels, time-division multiplexing can be applied to several data sources in such a way that very low latency is generated during the demultiplexing process. A fixed number of channels in each segment provides an efficient implementation of different data by reducing the complexity of the multiplexing and demultiplexing processes.

SMPTE 346M specifies the format for multiplexing multiple asynchronous standard-definition video streams and generic data into the high-definition system interfaces as defined in SMPTE 274M and ANSI/SMPTE 296M in the bit-parallel source format for the bit-serial interface defined in SMPTE 292M. The major standard-definition system interfaces are defined by ANSI/SMPTE 259M and ITU-R BT.601-5. They are the 4:2:2 digital component video signal interfaces defined in ANSI/SMPTE 125M, ANSI/SMPTE 267M, and ANSI/SMPTE 293M. These standards specify interfaces for the 270-, 360-, and 540-Mbits/s bit-parallel formats. The standard also covers other serial video standards such as 143 Mbits/s 525-line and 177 Mbits/s 625-line composite digital signals.

SMPTE 346M can be used for a 270-Mbits/s SDI interface to carry multiple streams of data and a single 8-bit or 10-bit composite digital video signal. In a similar manner, a 540-Mbits/s SDI system can carry multiple streams of video and data. SMPTE 346M can also be extended to future higher bit rate serial digital standards.

The standard definition video formats referred to in SMPTE 346M are listed in Table 8.2.4.

8.2.3e Packet Transport

SMPTE 348M (proposed at this writing) provides the mechanisms necessary to facilitate the transport of packetized data over a synchronous data carrier [10]. The HD-SDTI data packets and synchronizing signals provide a data transport interface that is compatible with SMPTE 292M such that it can be readily used by the infrastructure provided by the standard. SMPTE 348M uses a dual-channel technique where each line carries two data channels, each forming an independent HD-SDTI data transport mechanism. The two channels are word-multiplexed onto

8-26 Audio Networking

the HD-SDI stream such that one line-channel occupies the C data space and the other line-channel occupies the Y data space.

The standard provides for a baseline operation that supports a constant payload length per line-channel having a maximum payload data rate up to approximately 1.0 Gbits/s. It further provides for an extended operation mode that supports a variable payload length through the advancement of the SAV sequence to ensure a constant payload data rate regardless of the HD-SDI frame rate. The HD-SDTI protocol is compatible with SMPTE 305M.

SMPTE 348M describes the assembly of two channels of 10-bit words multiplexed into one HD-SDI line for the purpose of transporting the data streams in a structured framework. The HD-SDTI data blocks and synchronizing signals provide a data transport protocol that can readily be added to the infrastructure provided by SMPTE 292M.

SMPTE 292M requires a sequence of 10-bit words that define a television horizontal line comprising five areas in the following sequence (the first two areas are often described together):

- **EAV:** a four-word unique timing sequence defining the end of active video (of the previous line)
- **LN/CRC:** two words defining the line number followed by a two-word CRC error detection code
- Digital line blanking
- **SAV:** a four-word unique timing sequence defining the start of active video
- Digital active line

An associated television source format standard defines the rate of television horizontal lines by specifying the following parameters:

- The number of words per line
- The number of words in the digital active line (and hence the number of words in the digital line blanking period)
- The number of lines per frame
- The number of frames per second

SMPTE 292M currently defines four source format standards (1152, 1035, 1080, and 720 active lines per frame). SMPTE 125M describes the meaning of the EAV and SAV word sequences that can be applied to all relevant source formats.

A decoder operating under the SMPTE 348M standard is not required to decode all the source formats available to SMPTE 292M. The source formats that must be supported by the decoder are specified in the application document.

8.2.3f 540 Mbits/s Interface

SMPTE 344M (proposed at this writing) specifies a serial digital interface that operates at a nominal rate of 540 Mbits/s [11]. The standard is intended for applications in television studios over specified lengths of coaxial cable. Separate SMPTE documents specify the mapping of source image formats onto the 540 Mbits/s serial interface.

The connector has mechanical characteristics conforming to the $50\text{-}\Omega$ BNC type. Mechanical dimensions of the connector may produce either a nominal $50\text{-}\Omega$ or $75\text{-}\Omega$ impedance and are usable

at frequencies up to 540 MHz, assuming a return loss that is greater than 15 dB. However, the electrical characteristics of the connector and its associated interface circuitry must provide a resistive impedance of 75Ω . Where a $75\text{-}\Omega$ connector is used, its mechanical characteristics must reliably interface with the nominal $50\text{-}\Omega$ BNC type defined by IEC 60169-8.

The application of SMPTE 344M does not require a particular type of coax. It is necessary, however, for the coax to be a $75\text{-}\Omega$ type and for the frequency response of the coax, in dB, to be approximately proportional to $1/\sqrt{f}$ from 1 MHz to 540 MHz to ensure correct operation of automatic cable equalizers over moderate to maximum lengths.

The channel coding specified in SMPTE 344M is scrambled NRZI. The LSB of any data word is transmitted first. To maintain synchronization and word alignment at the serial receiver, EAV and SAV timing references are inserted into the data stream.

8.2.4 Serial Data Transport Interface

The serial data transport interface (SDTI) is a standard for transporting packetized audio, video, and data between cameras, VTRs, editing/compositing systems, video servers, and transmitters in professional and broadcast video environments [12]. SDTI builds on the familiar SDI standard that is now widely used in studios and production centers to transfer uncompressed digital video between video devices. SDTI provides for faster-than-real-time video transfers and a reduction in the number of decompression/compression generations required during the video production process, while utilizing the existing SDI infrastructure.

The SMPTE 305M SDTI specification evolved from a collaborative effort on the part of equipment vendors and interested parties, under the auspices of the SMPTE PT20.04 Workgroup on Packetized Television Interconnections, to define a common interchange interface for compressed audio and video.

Because SDTI is built upon the SMPTE 259M SDI specification, it shares the same mechanical, electrical, and transport mechanisms. BNC connectors and coaxial cables establish the mechanical link.

SDI transports uncompressed digital video using 10-bit words in the 4:2:2 Y, U, V component mode for 525- and 625-line applications. Words are serialized, scrambled, and coded into a 270-Mbits/s or 360-Mbits/s serial stream. In order to synchronize video timing between the transmitter and the receiver, SDI defines words in the bitstream called *end of active video* (EAV) and *start of active video* (SAV), as illustrated in Figure 8.2.9. At 270 Mbits/s, the active portion of each video line is 1440 words and at 360 Mbits/s, the active portion is 1920 words. The area between EAV and SAV can be used to transmit ancillary data such as digital audio and time code. The ancillary data space is defined by SMPTE 291M-1998.

SDI and SDTI can co-exist in a facility using the same cabling, distribution amplifiers, and routers. Cable lengths of more than 300 meters are supported. SDI repeaters can be used to reach longer distances. A versatile studio configuration that supports all the required point-to-point connections can be established using an SDI router.

8.2.4a SDTI Data Structure

SDTI uses the ancillary data space in SDI to identify that a specific video line carries SDTI information [12]. The packetized video is transported within the active video area, providing 200

8-28 Audio Networking

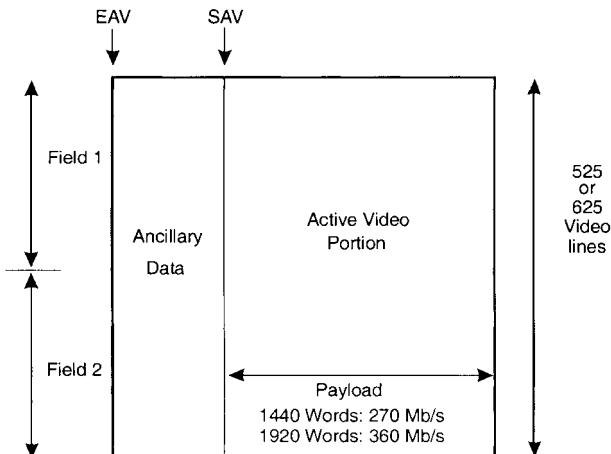


Figure 8.2.9 SMPTE 259M principal video timing parameters. (From [12]. Used with permission.)

Mbits/s of payload capacity on 270-Mbits/s links and 270 Mbits/s of payload capacity on 360-Mbits/s links.

A 53-word header data packet is inserted in the ancillary data space. The rest of the ancillary data space is left free to carry other ancillary data. The 53-word SDTI header data structure is in accordance with the SMPTE 291M ancillary data specification, shown in Figure 8.2.10. The specification contains an ancillary data flag (ADF), a data ID (DID) specified as code 140h for SDTI, secondary data ID (SDID) specified as code 101h for SDTI, and 46 words for header data. A checksum for data integrity checking also is included.

The 46 words of header data define source and destination addresses and the formatting of the payload (Figure 8.2.11). Line number and line number CRC bits are used to ensure data continuity. The code identifies the size of the payload to be either 1440 words or 1920 words long. An authorized address identifier (AAI) defines the addressing method utilized.

Currently, the Internet Protocol (IP) addressing scheme is defined for SDTI. The source and destination addresses are 128 bits long, allowing essentially a limitless number of addressable devices to be supported.

The block type identifies whether data is arranged in fixed-sized data blocks—with or without error correction code (ECC)—or variable-sized data blocks. In the case of fixed-sized blocks, the block type also defines the number of blocks that are transported in one video line. The block type depends on the data type of the payload.

Between SAV and EAV, the payload itself is inserted. The payload can be of any valid data type registered with SMPTE. The data structure of the payload includes a data type code preceding the data block, shown in Figure 8.2.12a. In addition, separator, word count, and end code are required for data types that feature variable-sized blocks (Figure 8.2.12b).

SMPTE 305M does not specify the data structure inside the data block, which is left to the registrant of the particular data type.

A significant revision of SMPTE 305M was undertaken in March 2000. The document, proposed at this writing, updated normative references and provided additional detail on the operation elements of the SDTI system. The revision was intended to keep the standard in conformance with related documents and to address certain implementation issues.

Ancillary Data Flag (ADF)	53 words
Data ID (DID)	
Secondary Data ID (SID)	
Data Count (DC)	
Header Data	
Checksum (CS)	46 words

Figure 8.2.10 Header data packet structure. (From [12]. Used with permission.)

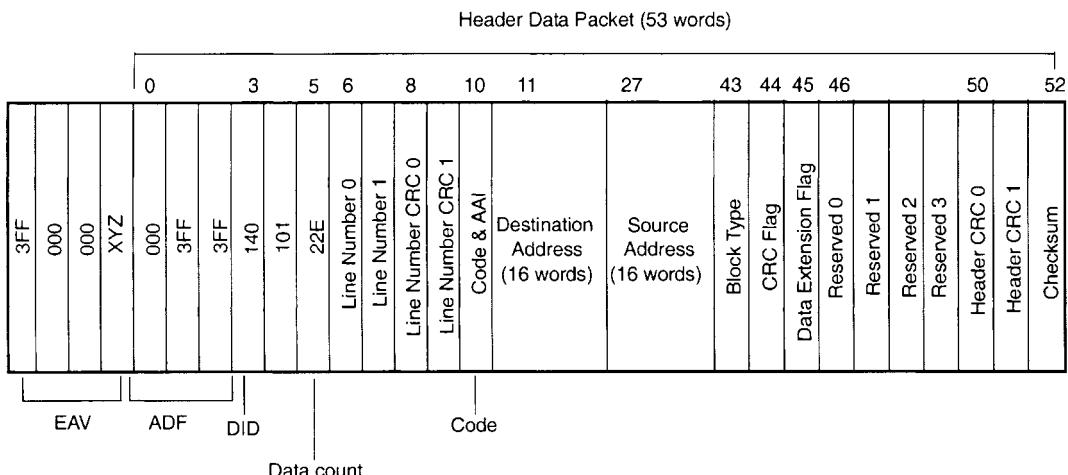


Figure 8.2.11 SDTI header data structure. (From [9]. Used with permission.)

8.2.4b SDTI in Computer-Based Systems

Transferring material to and from computer-based nonlinear editing (NLE)/compositing systems and video servers is one of the primary uses of SDTI [12]. Computers can interface with other SDTI devices through the use of an adapter. Typically, an SDTI-enabled computer-based NLE system will contain the following components:

- A CPU motherboard
- A/V storage and a controller to store the digital audio and compressed video material
- Editing board with audio/video I/O, frame buffer, digital video codecs, and processors
- Network and console display adapters
- SDI/SDTI adapter

8-30 Audio Networking

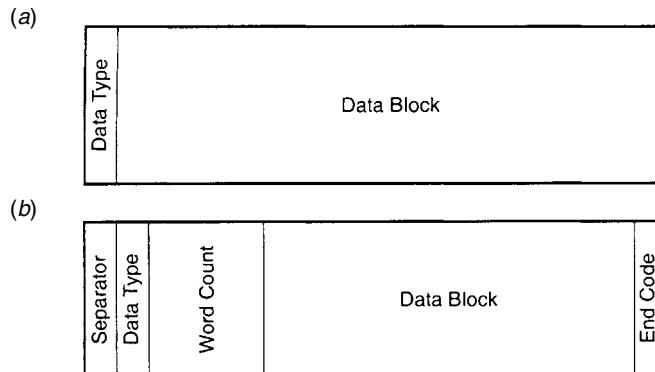


Figure 8.2.12 SDTI data structure: (a) fixed block size, (b) variable block size. (From [12]. Used with permission.)

As an example of a typical SDI/SDTI hardware adapter, consider the PCI-bus implementation illustrated in Figure 8.2.13. With the PCI adapter installed, the computer can share and move SDI and SDTI streams with and among other devices and terminals. Such a configuration also can be used to transcode from SDI to SDTI, and vice-versa.

8.2.4c SDTI Content Package Format

The SDTI *content package format* (SDTI-CP) is an extension of the basic SDTI system that allows for package-based delivery of data signals. The SMPTE developed a collection of standards that define the technology and protocol of SDTI-CP. The applicable documents include the following:

- SMPTE 326M: SDTI Content Package Format
- SMPTE 331M: Element and Metadata Definitions for the SDTI-CP
- SMPTE 332M: Encapsulation of Data Packet Streams over SDTI
- Recommended Practice RP 204: SDTI-CP MPEG Decoder Templates

The general parameters of these documents are described in the following sections.

SMPTE 326M

The SDTI-CP standard specifies the format for the transport of *content packages* (CP) over the serial digital transport interface [13]. Known as SDTI-CP, this format is a packaging structure for the assembly of system, picture, audio, and auxiliary data items in a specified manner. SMPTE 326M defines the structure of the content package mapped onto the SDTI transport; element and metadata formats are defined by SMPTE 331M [14].

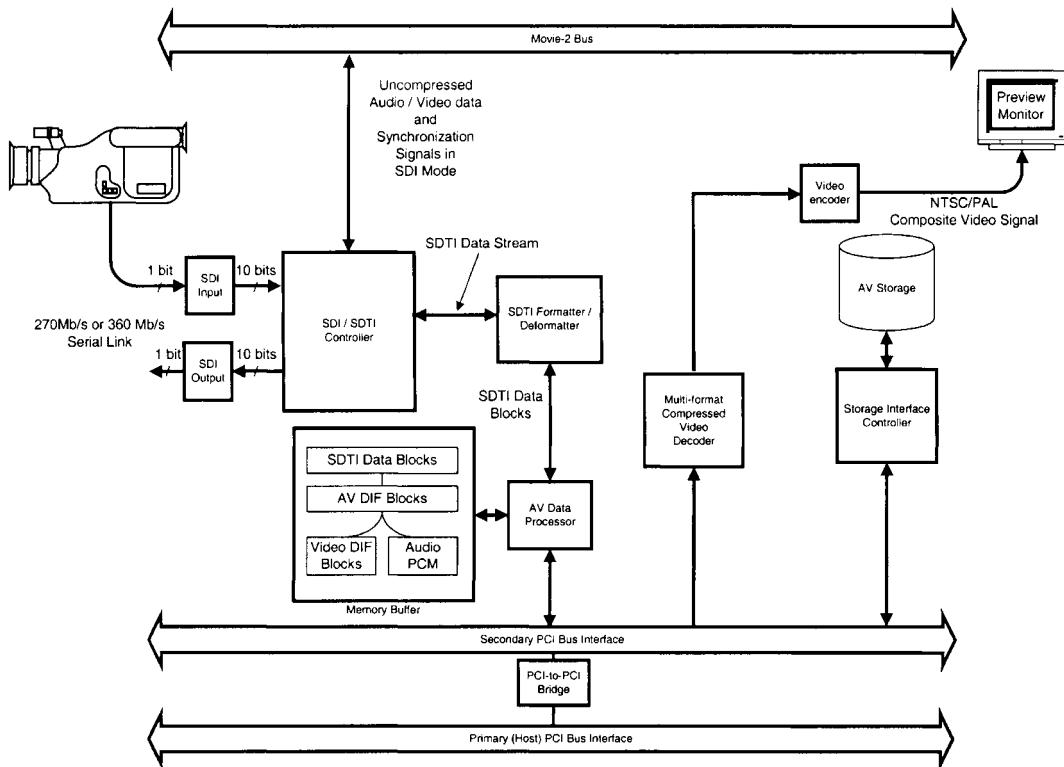


Figure 8.2.13 SDI/SDTI implementation for a computer-based editing/compositing system. (From [12]. Used with permission.)

The baseline operation of the SDTI-CP standard is defined by the transport of content packages locked to the SDTI transport frame rate. The standard additionally defines format extension capabilities that include:

- Content package transfers at higher and lower than the specified rate through isochronous and asynchronous transfer modes
- Provision of a timing mode to reduce delay and provision for two content packages in each SDTI transport frame
- Carriage of content packages in a low-latency mode
- Multiplexing of content packages from different sources onto one SDTI transport

8-32 Audio Networking

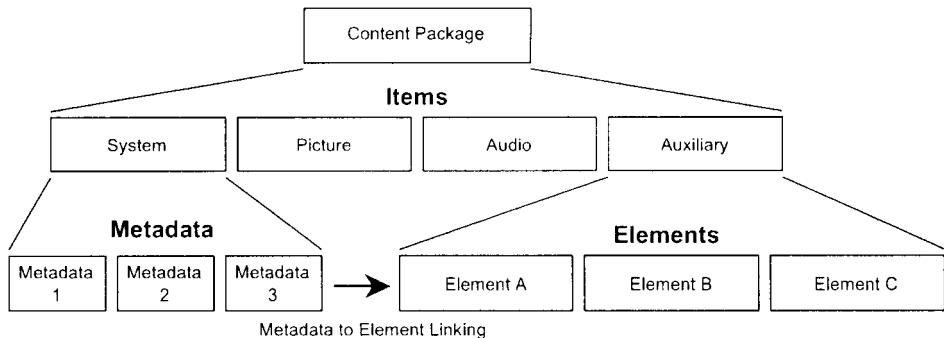


Figure 8.2.14 The basic content package structure of SMPTE 326M. (From [13]. Used with permission.)

The SMPTE 326M standard is limited to SDI systems operating at a bit rate of 270 Mbit/s and 360 Mbit/s as defined by SMPTE 305M.

An SDTI-CP compliant receiver must be capable of receiving and parsing the structure of the SDTI-CP format. An SDTI-CP compliant decoder is defined by the ability to both receive and decode a defined set of elements and metadata according to an associated decoder template document. The MPEG decoder template is detailed in RP 204 [15]. Other decoder template recommended practices can be defined as required for other applications of the SDTI-CP.

Figure 8.2.14 shows the basic layered structure of a *content package*. It is constructed of up to four *items*, where each item is constructed of one or more *elements*, which include:

- *System item*—carries content package metadata and may contain a control element. The system item also carries metadata that is related to elements in the other items.
 - *Picture item*—can consist of up to 255 picture stream elements.
 - *Audio item*—can consist of up to 255 audio stream elements.
 - *Auxiliary item*—can consist of up to 255 auxiliary data elements.

A content package contains the associated contents of one content package frame period starting with a system item and optionally containing picture, audio, and auxiliary items.

Data Packet Encapsulation

SMPTE 332M specifies an open framework for encapsulating data packet streams and associated control metadata over the SDTI transport (SMPTE 305M) [16]. Encapsulating data packet streams on SDTI allows them to be routed through conventional SDI (SMPTE 259M) equipment. The standard specifies a range of packet types that can be carried over SDTI, which may be expanded as requirements develop.

The standard does not attempt to specify the payload contained in any packet. It offers options to add specific information to each individual packet including localized user data space, forward error correction (FEC), and a mechanism for accurate packet retiming at the decoder.

The standard also offers a limited capability for metadata to be added, providing packet control information to aid the successful transfer of packets. The specification of the metadata follows the K-L-V approach of the SMPTE dynamic metadata dictionary and provides extensibility for future requirements.

Timing Issues

Most packet streams do not have critical timing requirements and a decoder can output packets in the order in which they were encoded, but with increased *packet jitter* resulting from the buffering of packets onto SDTI lines [16]. The result of the SDTI-PF packet encapsulation process is to introduce both delay and jitter to the packet stream. However, MPEG-2 transport stream (MPEG-2 TS) packets are one case where a relatively small packet jitter specification is required to ensure minimal impact on MPEG-2 transport stream clock recovery and buffering circuits. SMPTE 332M contains provisions to allow the packet jitter to be reduced to insignificant levels; the delay is an issue addressed by the method of packet buffering at the encoder. As a benchmark, the specification is defined so that a low packet jitter source can be carried through the SDTI-PF and be decoded to create an output with negligible packet jitter.

Although MPEG-2 TS packets are the most critical packet type for decoder timing accuracy, this standard also allows for other kinds of packets to be carried over the SDTI, with or without buffering, to reduce packet jitter. Such packets may be ATM cells and packets based on the *uni-directional Internet protocol* (Uni-IP).

MPEG Decoder Templates

SMPTE Recommended Practice RP 204 defines decoder templates for the encoding of SDTI content packages with MPEG coded picture streams [15]. The purpose of RP 204 is to provide appropriate limits to the requirements for a receiver/decoder in order to allow practical working devices to be supplied to meet the needs of defined operations. Additional MPEG templates are expected to be added to the practice as the SDTI-CP standard matures. The SMPTE document recommends that each new template be a superset of previous templates so that any decoder defined by a template in the document can operate with both the defined template and all subsets.

8.2.5 References

1. Fibush, David: *A Guide to Digital Television Systems and Measurement*, Tektronix, Beaverton, OR, 1994.
2. SMPTE RP 165-1994: *SMPTE Recommended Practice—Error Detection Checkwords and Status Flags for Use in Bit-Serial Digital Interfaces for Television*, SMPTE, White Plains, N.Y., 1994.
3. SMPTE Standard: SMPTE 305M-1998, *Serial Data Transport Interface*, SMPTE, White Plains, N.Y., 1998.
4. Gaggioni, H., M. Ueda, F. Saga, K. Tomita, and N. Kobayashi, “The Development of a High-Definition Serial Digital Interface,” Sony Technical Paper, Sony Broadcast Group, San Jose, Calif., 1998.

8-34 Audio Networking

5. "SMPTE Standard for Television—Bit-Serial Digital Interface for High-Definition Television Systems," SMPTE 292M-1998, SMPTE, White Plains, N.Y., 1998.
6. "SMPTE Standard for Television—24-Bit Digital Audio Format for HDTV Bit-Serial Interface," SMPTE 299M-1997, SMPTE, White Plains, N.Y., 1997.
7. "SMPTE Standard for Television—Mapping of AES3 Data into MPEG-2 Transport Stream," SMPTE 302M-1998, SMPTE, White Plains, N.Y., 1998.
8. "SMPTE Standard for Television (Proposed)—Vertical Ancillary Data Mapping for Bit-Serial Interface, SMPTE 334M, SMPTE, White Plains, N.Y., 2000.
9. "SMPTE Standard for Television (Proposed)—Signals and Generic Data over High-Definition Interfaces," SMPTE 346M, SMPTE, White Plains, N.Y., 2000.
10. "SMPTE Standard for Television (Proposed)—High Data-Rate Serial Data Transport Interface (HD-SDTI)," SMPTE 348M, SMPTE, White Plains, N.Y., 2000.
11. SMPTE 344M (Proposed), "540 Mb/s Serial Digital Interface," SMPTE, White Plains, N.Y., 2000.
12. Legault, Alain, and Janet Matey: "Interconnectivity in the DTV Era—The Emergence of SDTI," *Proceedings of Digital Television '98*, Intertec Publishing, Overland Park, Kan., 1998.
13. "SMPTE Standard for Television—SDTI Content Package Format (SDTI-CP)," SMPTE 326M-2000, SMPTE, White Plains, N.Y., 2000.
14. "SMPTE Standard for Television—Element and Metadata Definitions for the SDTI-CP," SMPTE 331M-2000, SMPTE, White Plains, N.Y., 2000.
15. "SMPTE Recommended Practice—SDTI-CP MPEG Decoder Templates," RP 204 (Proposed), SMPTE, White Plains, N.Y., 1999.
16. "SMPTE Standard for Television—Encapsulation of Data Packet Streams over SDTI (SDTI-PF)," SMPTE 332M-2000, SMPTE, White Plains, N.Y., 2000.

8.2.6 Bibliography

- "SMPTE Standard for Television—Ancillary Data Packet and Space Formatting," SMPTE 291M-1998, SMPTE, White Plains, N.Y., 1998.
- "SMPTE Standard for Television—Bit-Serial Digital Interface for High-Definition Television Systems," SMPTE 292M-1996, SMPTE, White Plains, N.Y., 1996.
- "SMPTE Standard for Television—Serial Data Transport Interface," SMPTE 305M-1998, SMPTE, White Plains, N.Y., 1998.
- Turow, Dan: "SDTI and the Evolution of Studio Interconnect," *International Broadcasting Convention Proceedings*, IBC, Amsterdam, September 1998.
- Wilkinson, J. H., H. Sakamoto, and P. Horne: "SDDI as a Video Data Network Solution," *International Broadcasting Convention Proceedings*, IBC, Amsterdam, September 1997.

Chapter
8.3

Video/Audio Networking Systems

Jerry C. Whitaker, Editor-in-Chief

8.3.1 Introduction

Video networking depends—to a large extent—on system interoperability, not only for basic functions but also for extended functionality. Interoperability has two aspects. The first is syntactic and refers only to the coded representation of the digital television information. The second relates to the delivery of the bit stream in real time.

Broadcast digital video systems, specifically DTV and DVB, support bit streams and services beyond basic compressed video and audio services, such as text-based services, emergency messages, and other future ancillary services. The MPEG-2 transport packet size is such that it can be easily partitioned for transfer in a link layer that supports ATM transmission. The MPEG-2 transport layer and the ATM layer serve different functions in a video delivery application: the MPEG-2 transport layer solves MPEG-2 presentation problems and performs the multimedia multiplexing function, and the ATM layer solves switching and network-adaptation problems.

In addition to ATM, a number of networking systems have been developed or refined to carry digital video signals, including IEEE 1394, Fibre Channel, and Gigabit Ethernet.

8.3.2 Architecture of ATM

Asynchronous transfer mode is a technology based on high-speed packet switching. It is an ideal protocol for supporting professional video/audio and other complex multimedia applications. ATM is capable of data rates of up to 622 Mbits/s.

ATM was developed in the early 1980s by Bell Labs as a backbone switching and transportation protocol. It is a high-speed integrated multiplexing and switching technology that transmits information using fixed-length cells in a connection-oriented manner. Physical interfaces for the *user-network interface* (UNI) of 155.52 Mbits/s and 622.08 Mbits/s provide integrated support for high-speed information transfers and various communications modes—such as *circuit* and *packet* modes—and constant, variable, or burst bit-rate communications. These capabilities lead to four basic types of service classes of interest to video users [1]:

- *Constant bit rate* (CBR), which emulates a leased line service, with fixed network delay
- *Variable bit rate* (VBR), which allows for bursts of data up to a predefined peak cell rate

8-36 Audio Networking

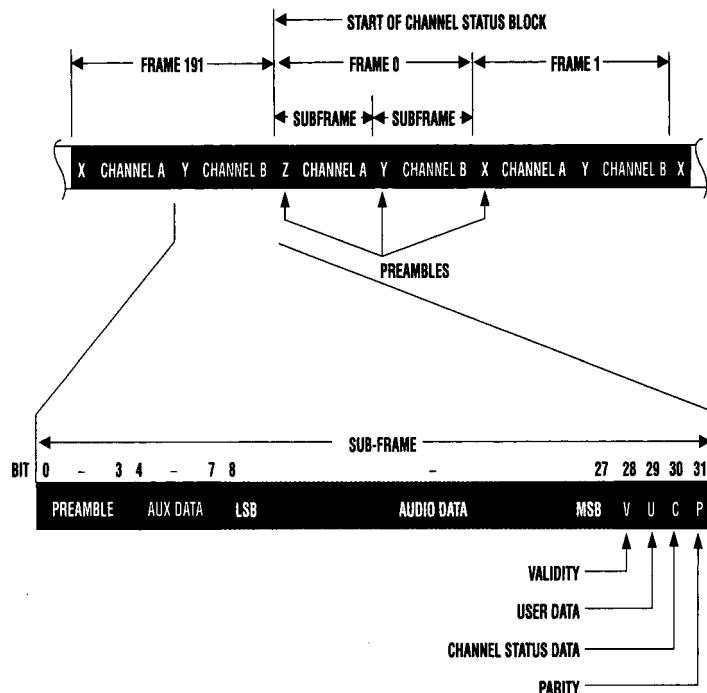


Figure 8.3.1 The typical packing of an internodal ATM trunk. (After [2].)

- *Available bit rate* (ABR), in which capacity is negotiated with the network to fill capacity gaps
- *Unspecified bit rate* (UBR), which provides unnegotiated use of available network capacity

These tiers of service are designed to maximize the traffic capabilities of the network. The CBR data streams are fixed and constant with time; the VBR and ABR systems vary. The bandwidth of the UBR class of service is a function of whatever network capacity is left over after all other users have claimed their stake to the bandwidth. Not surprisingly, CBR is usually the most expensive class of service, and UBR is the least expensive. Figure 8.3.1 illustrates typical packing of an ATM trunk.

One of the reasons ATM is attractive for video applications is that the transport of video and audio fits nicely into the established ATM service classes. For example, consider the following applications:

- Real-time video—which demands real-time transmission for scene capture, storage, processing, and relay—fits well into the CBR service class.
- Non-real-time video—such as recording and editing from servers, distributing edited masters, and other operations that can be considered essentially off-line—can use the ABR service.

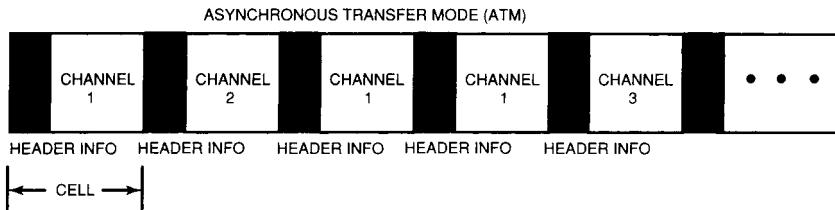


Figure 8.3.2 The ATM cell format. (After [2].)

- Machine control and file transfer—such as sending still clips from one facility to another—find the VBR service attractive.

ATM is growing and maturing rapidly. It already has been implemented in many industries, deployed by customers who anticipate such advantages as:

- Enabling high-bandwidth applications including desktop video, digital libraries, and real-time image transfer
- Coexistence of different types of traffic on a single network platform to reduce both the transport and operations costs
- Long-term network scalability and architectural stability

In addition, ATM has been used in both local- and wide-area networks. It can support a variety of high-layer protocols and will cope with future network speeds of gigabits per second.

8.3.2a ATM Cell Structure

It is worthwhile to explore the ATM channel format in some detail because its features are the key to the usefulness of ATM for video applications. ATM channels are represented by a set of fixed-size cells and are identified through the channel indicator in the *cell header* [2]. The ATM cell has two basic parts: the header (5 bytes) and the payload (48 bytes). This structure is shown in Figure 8.3.2. ATM switching is performed on a cell-by-cell basis, based on the routing information contained in the cell header.

Because the main function of the ATM layer is to provide fast multiplexing and routing for data transfer based on information included in the header, this element of the protocol includes not only information for routing, but also fields to indicate the type of information contained in the cell payload. Other data is included in the header to perform the following support functions:

- Assist in controlling the flow of traffic at the UNI
- Establish priority for the cell
- Facilitate header error-control and cell-delineation functions

One key feature of ATM is that the cells can be independently labeled and transmitted on demand. This allows facility bandwidth to be allocated as needed, without the fixed hierarchical

8-38 Audio Networking

Table 8.3.1 ATM Cell Header Fields (After [1].)

GFC	A 4-bit <i>generic flow control</i> field: used to manage the movement of traffic across the user network interface (UNI).
VPI	An 8-bit network <i>virtual path identifier</i> .
VCI	A 16-bit network <i>virtual circuit identifier</i> .
PT	A 3-bit <i>payload type</i> (i.e., user information type ID).
CLP	A 1-bit <i>cell loss priority flag</i> (eligibility of the cell for discard by the network under congested conditions).
HEC	An 8-bit <i>header error control</i> field for ATM header error correction.
AAL	ATM <i>adaptation-layer</i> bytes (user-specific header).

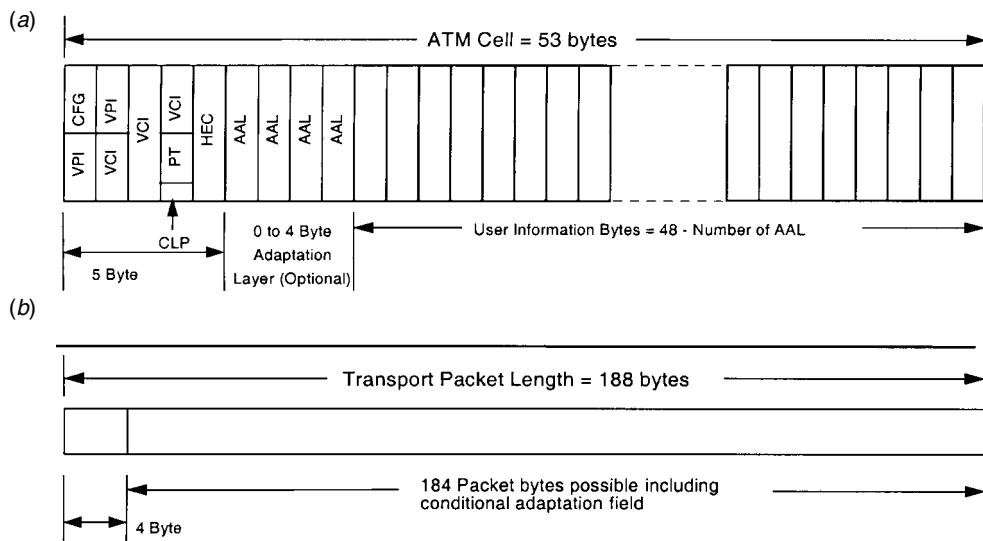


Figure 8.3.3 Comparison of the ATM cell structure and the MPEG-2 transport packet structure: (a) structure of the ATM cell, (b) structure of the transport packet. (After [3].)

channel rates required by other network protocols. Because the connections supported are either permanent or semipermanent and do not require call control, real-time bandwidth management, or processing capabilities, ATM has the flexibility for video/multimedia applications.

The ATM cell header, the key to the use of this technology for networking purposes, consists of the fields shown in Table 8.3.1. Figure 8.3.3 illustrates the differences between the format of an ATM cell and the format of the MPEG-2 transport packet.

8.3.3 IEEE 1394

IEEE 1394 is a serial bus interconnection and networking technology, and the set of protocols defining the communications methods used on the network [4]. IEEE 1394 development began in the 1980s at Apple Computer and was trademarked under the name *FireWire*. Sony has since trademarked its implementation as *iLINK*. The formal name for the standard is IEEE 1394-1995.

IEEE 1394 is widely supported by hardware, software, and semiconductor companies where it is implemented on computers, workstations, videotape recorders, cameras, professional audio, digital televisions and set top boxes, and consumer A/V equipment. 1394 enables transfers of video and data between devices without image degradation. Protocols for the transport of video, digital audio, and IP data are in place or under development at this writing. The EBU/SMPTE Task Force for Harmonized Standards for the Exchange of Program Material as Bit Streams has recognized IEEE 1394 as a recommended transport system for content.

Like other networking technologies, 1394 connects devices and transports information or data among the devices on a network. When IEEE 1394 was developed, the need to transport real-time media streams (video and audio, for example) was recognized. These signal types require consistent delivery of the data with a known latency, i.e., *isochronous* (constant or same time) delivery.

8.3.3a Operating Modes

When transmitting time sensitive material, such as real-time motion video, a slight impairment (a defective pixel, for example) is not as important as the delivery of the stream of pictures making up the program [4]. The IEEE 1394 designers recognized these requirements and defined 1394 from the outset with capabilities for isochronous delivery of data. This isochronous mode is a major differentiator of IEEE 1394 when compared with other networking technologies, such as Ethernet, Fibre Channel, or ATM.

IEEE 1394 divides the network bandwidth into 64 discreet channels per bus, including a special *broadcast channel* meant to transmit data to all users. This allows a single IEEE 1394 bus to carry up to 64 different independent isochronous streams. Each stream can carry a video stream, audio stream, or other types of data streams simultaneously. A network resource manager allocates bandwidth and transmission channels on the network to guarantee a fixed bandwidth channel for each stream.

IEEE 1394 also supports asynchronous transmissions. The asynchronous channels are used to transmit data that cannot suffer loss of information. This dual transmission scheme, supporting both isochronous and asynchronous modes, makes IEEE 1394 useful for a range of applications.

8.3.3b Data Rates and Cable Lengths

IEEE 1394 defines a set of parameters to assure reliable system operation. The initial standards called for data rates of 100 and 200 (nominal) Mbits/s. IEEE 1394a increased the data rate to 400 Mbits/s. IEEE 1394b increases data rates to 800 Mbits/s, 1.6 Gbits/s, and 3.2 Gbits/s. At 400 Mbits/s, the isochronous payload size is 4096 bytes—considerably larger than other technologies, such as ATM. As the clock speed increases, the size of the data packets increase linearly. Typical applications of IEEE 1394 are shown in Figures 8.3.4 and 8.3.5.

8-40 Audio Networking

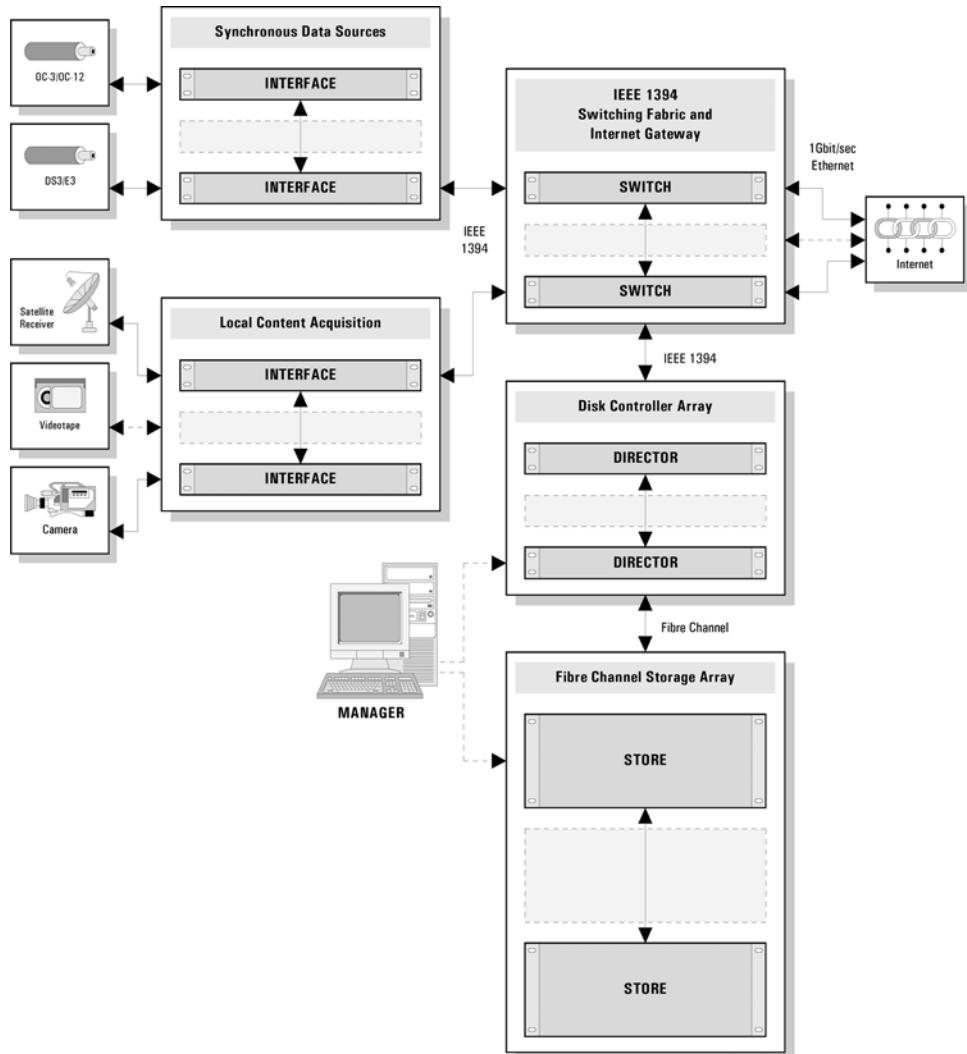


Figure 8.3.4 The application of IEEE 1394 in a large scale internet broadcast environment. (From [5]. Used with permission.)

8.3.3c Isochronous and Asynchronous Transmissions

At first glance, the terms *isochronous* and *synchronous*, as applied to network transmissions, describe essentially the same process. Upon further investigation, however, small but significant differences emerge [6, 7].

An *isochronous* channel has a constant time interval (or integer multiples thereof) between similar “significant instants.” This basic definition means that each isochronous packet in a

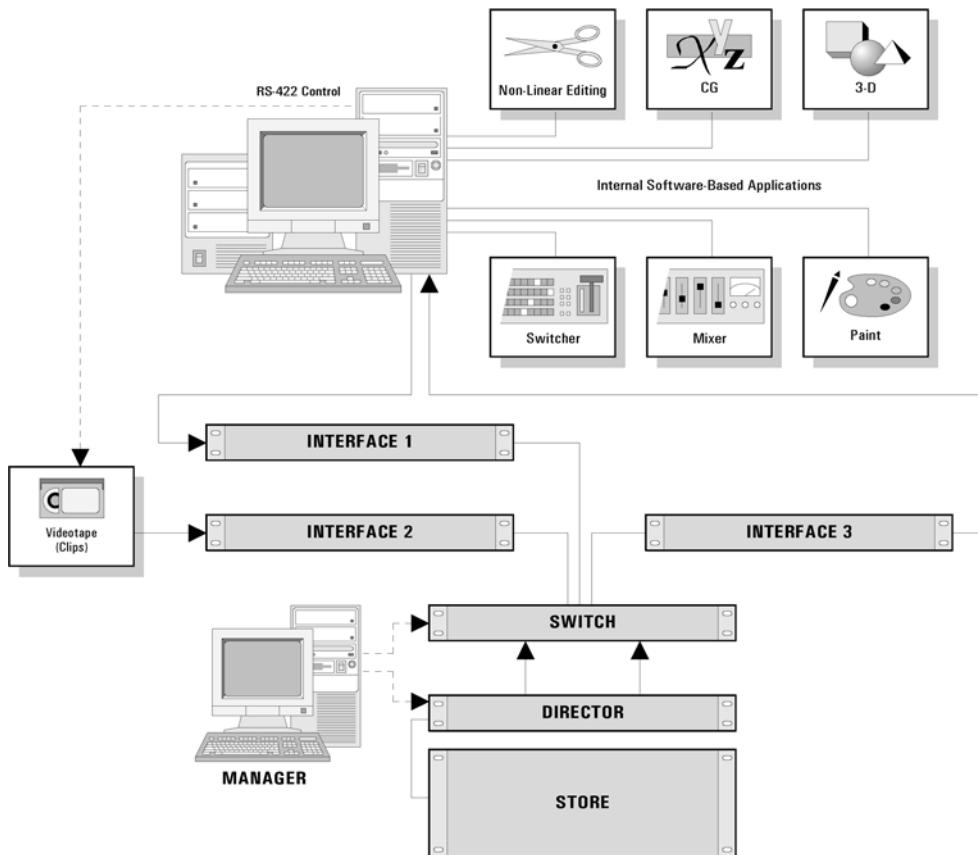


Figure 8.3.5 The application of IEEE 1394 in a nonlinear postproduction environment. (From [6]. Used with permission.)

stream begins a constant time interval after the previous packet in that stream. In IEEE 1394, isochronous packets begin 125 μ s apart (at least on average). This period is controlled by a clock packet, transmitted by the IEEE 1394 bus master, which signals devices on the network that it is time to send the next isochronous packet. Another example of an isochronous channel is the 270 Mbits/s serial digital video interface, where a line of video begins every 64 μ s (625/50) or 63.56 μ s (525/60).

The opposite of isochronous is *anisochronous*. An anisochronous channel does not maintain constant time intervals between significant instants. A good example of this is Ethernet, a technology that does not synchronize data transmission from nodes on the network, but instead relies on signaling when conflicts occur.

The term *synchronous* refers to a timing relationship between two different channels. If two “significant instants” (e.g., packet starts) on two different channels occur at the same time every time, the channels are synchronous. A good example of synchronous channels would be a 4:2:2

8-42 Audio Networking

4:4 SDI signal carried on two coaxial cables, where lines of video are completely time aligned on the two cables.

The opposite of synchronous is *asynchronous*. If “significant instants” on two channels do not consistently line up, then the channels are asynchronous.

8.3.4 Fibre Channel

Fibre Channel is a 1 Gbits/s data transfer interface technology that maps several common transport protocols, including IP and SCSI, allowing it to merge high-speed I/O and networking functionality in a single connectivity technology [4]. Fibre Channel is an open standard, as defined by ANSI and OSI standards, and operates over copper and fiber optic cabling at distances of up to 10 km. It is unique in its support of multiple inter-operable topologies—including point-to-point, arbitrated-loop, and switching—and it offers several qualities of service for network optimization.

Fibre Channel first appeared in enterprise networking applications and point-to-point RAID and mass storage subsystems. It has expanded to include video graphics networks, video editing, and visual imaging systems.

Fibre Channel Arbitrated Loop (FC-AL) was developed with peripheral connectivity in mind. It natively maps SCSI (as SCSI FCP), making it a powerful technology for high-speed I/O connectivity. Native FC-AL disk drives allow storage applications to take full advantage of Fibre Channel's gigabaud bandwidth, passing SCSI data directly onto the channel with access to multiple servers or nodes. FC-AL supports 127-node addressability and 10 km cabling ranges between nodes. (See Table 8.3.2.)

The current specification for FC-AL allows for 1, 2 and 4 Gbits/s speeds. At this writing, practical applications of the technology are at the 2 Gbits/s rate.

Fibre Channel is designed with many high-availability features, including dual ported disk drives, dual loop connectivity, and loop resiliency circuits. Full redundancy in Fibre Channel systems is achieved by cabling two fully independent, redundant loops. This cabling scheme provides two independent paths for data with fully redundant hardware. Most disk drives and disk arrays targeted for high-availability environments have dual ports specifically for this purpose.

8.3.5 Gigabit Ethernet

The Ethernet protocol is the dominant networking technology for data processing applications. Gigabit Ethernet adds a 1 Gbit/s variant to the existing 10 and 100 Mbits/s Ethernet family while retaining the Ethernet standard frame format, collision detection system, and flow control. For reasons of compatibility with traditional data processing systems and the wide bandwidth necessary for digital video, Gigabit Ethernet has emerged as a viable method of transporting video within a facility [8].

The basic Ethernet performance specifications are given in Table 8.3.3 for the minimum 46 data byte and the maximum 1500 data byte frames [9]. Gigabit Ethernet was standardized for fiber optic interconnection in July 1998, and a copper specification using 4 pairs of category 5 unshielded twisted pair (UTP) was released about a year later.

Table 8.3.2 Fibre Channel General Performance Specifications (After [4].)

Media	Speed		Distance
Electrical Characteristics			
Coax/twinax	ECL	1.0625 Gigabits/s	24 Meters
	ECL	266 Megabits/s	47 Meters
Optical Characteristics			
9 micrometer single mode fiber	Longwave laser	1.0625 Gigabits/s	10 Kilometers
50 micrometer multi-mode fiber	Shortwave laser	1.0625 Gigabits/s	300 Meters
	Shortwave Laser	266 Megabits/s	2 Kilometer
62.5 micrometer multi-mode fiber	Longwave LED	266 Megabits/s	1 Kilometer
	Longwave LED	132 Megabits/s	500 Meters
<p><i>Note:</i> In FC-AL configurations, the distance numbers represents the distance between nodes, not the total distance around the loop.</p> <p><i>Note:</i> In fabric configurations, the distance numbers represent the distance from the fabric to a node, not the distance between nodes.</p>			

Table 8.3.3 Basic Specifications of Ethernet Performance (After [9].)

Parameter	10 Mbits/s		100 Mbits/s		1000 Mbits/s	
Frame size	Minimum	Maximum	Minimum	Maximum	Minimum	Maximum
Frames/s	14.8 k	812	148 k	8.1 k	1.48 M	81 k
Data rate	5.5 Mbits/s	9.8 Mbits/s	55 Mbits/s	98 Mbits/s	550 Mbits/s	980 Mbits/s
Frame interval	67 µs	1.2 ms	6.7 µs	120 µs	0.7 µs	12 µs

Although Gigabit Ethernet networks can be built using shared media in a *broadcast architecture*, early implementations are typically *full duplex switched* (a point-to-point architecture). In a switched network, the full capacity of the network medium is available at each device. For such a configuration, the two key switch specifications are the *aggregate capacity*, which determines the total amount of throughput for the switch, and the number of frames/s that can be handled. Ethernet has a variable frame rate and so both parameters are important. Under best-case conditions with large frames, the aggregate capacity defines performance; under-worst case conditions with small frames, the frame throughput is the limiting factor. In practice, these switches are specified to achieve *wire speed* on all their ports simultaneously. For example, an eight port switch will have 8 Gbits/s aggregate capacity. Many Gigabit Ethernet switches offer backwards compatibility with existing Ethernet installations, allowing Fast Ethernet or even 10 BASE networks to be connected.

8-44 Audio Networking

8.3.5a Network Bottlenecks

In many cases the bandwidth available from a switched Gigabit network removes the network itself as a bottleneck [8]. In fact, the bottleneck moves to the devices themselves that must handle the Gigabit Ethernet data. It can be seen from Table 8.3.3 that even under best-case conditions, frames arrive only 12 µs apart. In this interval, the device must determine if the packet is addressed to it, verify the checksum, and move the data contents of the frame into memory. Modern network interfaces use dedicated hardware to take the load off the processor, maximizing the time it has for user applications. The network interface can verify addresses and checksums, only interrupting the processor with valid frames. It can write data to discontiguous areas of memory, removing the need for the host to move different parts of messages around in memory. The network interface can also dynamically manage interrupts so that when traffic is high, several frames will be handled with only a single interrupt, thus minimizing processor environment switching. When traffic is low, the processor will be interrupted for a single frame to minimize latency.

These measures allow high, real data rates over Gigabit Ethernet of up to 960 Mbits/s. For systems that are CPU bound, increasing the Ethernet frame size can raise throughput by reducing the processor overhead.

8.3.5b A Network Solution

Ethernet on its own does not provide complete network solutions; it simply provides the lower two layers of the Open Systems Interconnect (OSI) model, specifically [8]:

- Layer 7, Application
- Layer 6, Presentation
- Layer 5, Session
- Layer 4, Transport
- Layer 3, Network
- **Layer 2, Data Link:** *logical link*—framing and flow control; *media access*—controls access to the medium
- **Layer 1, Physical:** the cable and/or fiber

The most widely used protocols for the Transport and Network layers are the *Transmission Control Protocol* (TCP) and *Internet Protocol* (IP), more commonly referred to as TCP/IP. These layers provide for reliable transmission of messages between a given source and destination over the network. Using TCP/IP on Ethernet is sufficiently common that most network hardware, for example a *network interface card* (NIC) or switch, typically has built-in support for key aspects of layers three and four.

How messages are interfaced to user applications is the function of the higher OSI layers. Here again, there are many choices depending upon the application. The *Network File System* (NFS) is a collection of protocols (developed by Sun Microsystems) with multiplatform support that presents devices on the network as disk drives. The advantage of this approach is that applications do not need to be specially coded to take advantage of the network. After a network device is *mounted*, network access looks to the application exactly the same as accessing a local

drive. Familiar techniques such as “drag and drop” can continue to be used, only now working with media data over a network.

8.3.5c Quality of Service

A guaranteed *quality of service* (QoS) transfers allocate network bandwidth in advance and maintains it for the duration of a session [8]. After set up, such transfers are *deterministic* in that the time taken to transfer a given amount of data can be predicted. They can also be wasteful because the bandwidth is reserved, even if it is not being used, preventing it from being used by others. In some cases, a QoS transfer to a device can lock out other nodes communicating to the same device.

For established television practitioners, QoS brings familiarity to the world of data networking, however, its implications for complete system design may not be fully appreciated. One key issue is that of the capabilities of devices connected to the network, many of which have video and network interfaces. The allocation of device bandwidth among these different interfaces is an important consideration. Video transfers must be real-time and so bandwidth must be able to service them when needed. However, QoS transfers that are also deterministic need guaranteed device attention. Resolving this conflict is not a trivial matter, and different operational scenarios require different—sometimes creative—solutions.

8.3.6 References

1. Piercy, John: “ATM Networked Video: Moving From Leased-Lines to Packetized Transmission,” *Proceedings of the Transition to Digital Conference*, Intertec Publishing, Overland Park, Kan., 1996.
2. Wu, Tsong-Ho: “Network Switching Concepts,” *The Electronics Handbook*, Jerry C. Whitaker (ed.), CRC Press, Boca Raton, Fla., p. 1513, 1996.
3. ATSC, “Guide to the Use of the Digital Television Standard,” Advanced Television Systems Committee, Washington, D.C., Doc. A/54, Oct. 4, 1995.
4. “Technology Brief—Networking and Storage Strategies,” Omneon Video Networks, Campbell, Calif., 1999.
5. “Networking and Internet Broadcasting,” Omneon Video Networks, Campbell, Calif., 1999.
6. “Networking and Production,” Omneon Video Networks, Campbell, Calif., 1999.
7. Craig, Donald: “Network Architectures: What does Isochronous Mean?,” *IBC Daily News*, IBC, Amsterdam, September 1999.
8. Owen, Peter: “Gigabit Ethernet for Broadcast and Beyond,” *Proceedings of DTV99*, Intertec Publishing, Overland Park, Kan., November 1999.
9. Gallo and Hancock: *Networking Explained*, Digital Press, pp. 191–235, 1999.

Section 9

Audio Recording Systems

The audio tape recorders that first went on the air in the late 1940s were notable in their ability to store programs for later broadcast with a sound quality not available from disc recording. Nevertheless, the recorders then were extremely simple in comparison with today's machines. Even in this age of high technology, the progress that has taken place in audio tape recording is quite remarkable and may be attributed to two factors. One was the evolution in the field of electronics: transistor technology; integrated circuits and then large-scale integrated circuits—first in digital and then in linear devices; digital signal processing; and microprocessors. The second was the combination of foresight and creativity applied by engineers involved in improvements in the art (and in user operations as well) in anticipating the need for improved or new capabilities, and then bringing them to pass.

Most of the advances made in the field could be classified either as further development or as innovation of methods and techniques for meeting new requirements. Neither of the categories predominated over the other in milestones. The modern digital audio workstation and digital magnetic tape recorder represent the application of highly developed scientific technologies, the result of many innovations and refinements since the invention of recording by Valdemar Poulsen in 1898. Today, many technical and business disciplines depend on the audio recorder in one form or another as an information storage device. The advancements in audio compression, computer disk storage, recording media, heads, and signal-processing techniques have made it possible to achieve storage densities that rival or exceed most other information-storage systems.

In This Section:

Chapter 9.1: Audio/Video Server Systems	9-5
Introduction	9-5
Basic Architecture	9-5
Server Design	9-7
Archiving Considerations	9-9
Audio/Video Server Storage Systems	9-10
Basic Drive Technology	9-11
RAID Levels	9-12
RAID Level 0	9-12

9-2 Section Nine

RAID Level 1	9-13
RAID Level 2	9-14
RAID Level 3	9-15
RAID Level 4	9-15
RAID Level 5	9-16
Other RAID levels	9-16
Digital Media Applications	9-17
Optimizing Digital Media Disk Performance	9-19
Thermal Calibration	9-21
SCSI and Fibre Channel	9-21
FC-AL Topology	9-22
Connectivity Considerations	9-23
Bandwidth	9-23
Remote Online Storage	9-24
Array Implementations	9-24
Server-Based Audio/Video Editing	9-24
Perspective on Storage Options	9-25
References	9-25
Bibliography	9-25
Chapter 9.2: Properties of Magnetic Materials	9-27
Introduction	9-27
Basic Principles of Magnetism	9-28
Curie Point	9-30
Magnetic Induction	9-30
Initial Magnetization B-H Curve	9-32
Hysteresis Loop	9-33
Hysteresis Losses	9-33
Initial M-H Curve and M-H Hysteresis Loop	9-34
Demagnetization	9-35
References	9-37
Chapter 9.3: Recording Fundamentals	9-39
Introduction	9-39
Fundamental Principles	9-39
Recording Signal Parameters	9-40
The Recording Process	9-41
Bias Recording	9-44
Particle Orientation	9-45
Erasure	9-45
References	9-46
Bibliography	9-46
Chapter 9.4: Magnetic Tape	9-47
Introduction	9-47
Basic Construction	9-47
Magnetic Coating	9-49

Magnetic Materials	9-49
Binders	9-51
Miscellaneous Additives	9-52
Manufacturing Process	9-52
Dispersion	9-53
Coating	9-53
Metal Evaporation Process	9-53
Slot Die Coating	9-55
Orientation	9-56
Drying	9-56
Surface Finishing	9-56
Slitting	9-57
Testing	9-57
Assembly and Packaging	9-57
References	9-57
Bibliography	9-57

On the CD-ROM:



- “Analog Magnetic Tape Recording and Reproduction” by E. Stanley Busby, an archive chapter from the first edition of the *Audio Engineering Handbook*. This reference material provides considerable background information on analog audio recording technologies and techniques.
- “Digital Magnetic-Tape Recording and Reproduction” by W. J. van Gestel, et. al., an archive chapter from the first edition of the *Audio Engineering Handbook*. This reference material explains in detail the fundamental recording technologies applicable to digital audio systems.
- “Analog Disk Recording and Reproduction” by Gregory Bogantz and Joseph Ruda, an archive chapter from the first edition of the *Audio Engineering Handbook*. This reference material examines analog disk recording technologies, techniques, and applications.

Reference Documents for this Section

- Anderson, D: “Fibre Channel-Arbitrated Loop: The Preferred Path to Higher I/O Performance, Flexibility in Design,” Seagate Technology Paper #MN-24, Seagate, Scotts Valley, Calif., 1995.
- Bate, G.: “Recent Developments in Magnetic Recording Materials,” *J. Appl. Phys.* pg. 2447, 1981.
- Bertram, H. N.: “Long Wavelength ac Bias Recording Theory,” *IEEE Trans. Magnetics*, vol. MAG-10, pp. 1039–1048, 1974.
- Bozorth, Richard M.: *Ferromagnetism*, Van Nostrand, Princeton, N.J., 1961.
- Chikazumi, Soshin: *Physics of Magnetism*, Wiley, New York, N.Y., 1964.
- Goldberg, Thomas: “New Storage Technology,” *Proceedings of the Advanced Television Summit*, Intertec Publishing, Overland Park, Kan., 1996.

9-4 Section Nine

- Grega, Joe: "Magnetic and Optical Recording Media," in *NAB Engineering Handbook*, 9th ed., Jerry C. Whitaker (ed.), National Association of Broadcasters, Washington, D.C., pp. 893–906, 1999.
- Hawthorne, J. M., and C. J. Hefielinger: "Polyester Films," in *Encyclopedia of Polymer Science and Technology*, N. M. Bikales (ed.), vol. 11, Wiley, New York, N.Y., pg. 42, 1969.
- Heyn, T.: "The RAID Advantage," Seagate Technology Paper, Seagate, Scotts Valley, Calif., 1995.
- Jorgensen, F.: *The Complete Handbook of Magnetic Recording*, Tab Books, Blue Ridge Summit, Pa., 1980.
- Kalil, F. (ed): *Magnetic Tape Recording for the Eighties*, NASA References Publication 1975, April 1982.
- Kraus, John D.: *Electromagnetics*, McGraw-Hill, New York, N.Y., 1953.
- Lehtinen, Rick, "Editing Systems," *Broadcast Engineering*, Intertec Publishing, Overland Park, Kan., pp. 26–36, May 1996.
- Lueck, L. B. (ed): *Symposium Proceedings Textbook*, Symposium on Magnetic Media Manufacturing Methods, Honolulu, May 25–27, 1983.
- McConathy, Charles F.: "A Digital Video Disk Array Primer," *SMPTE Journal*, SMPTE, New York, N.Y., pp. 220–223, April 1998.
- McKnight, John G.: "Erasure of Magnetic Tape," *J. Audio Eng. Soc.*, Audio Engineering Society, New York, N.Y., vol. 11, no.3, pp. 223–232, 1963.
- Nylen, P., and E. Sunderland: *Modern Surface Coatings*, Interscience Publishers Division, Wiley, London, 1965.
- Pear, C. B.: *Magnetic Recording in Science and Industry*, Reinhold, New York, N.Y., 1967.
- Perry, R. H., and A. A. Nishimura: "Magnetic Tape," in *Encyclopedia of Chemical Technology*, 3d ed., Kirk Othmer (ed.), vol. 14, Wiley, New York, N.Y., pp. 732–753, 1981.
- Plank, Bob: "Video Disk and Server Operation," *International Broadcast Engineer*, September 1995.
- Robin, Michael, and Michel Poulin: "Multimedia and Television," in *Digital Television Fundamentals*, McGraw-Hill, New York, N.Y., pp. 455–488, 1997.
- Sharrock, Michael P., and D. P. Stubs: "Perpendicular Magnetic Recording Technology: A Review," *SMPTE J.*, SMPTE, White Plains, N.Y., vol. 93, pp. 1127–1133, December 1984.
- Smit, J., and H. P. J. Wijn: *Ferrite*, Wiley, New York, N.Y., 1959.
- Tochihara, S.: "Magnetic Coatings and Their Applications in Japan," *Prog. Organic Coatings*, vol. 10, pp. 195–204, 1982.
- Tyson, H: "Barracuda and Elite: Disk Drive Storage for Professional Audio/Video," Seagate Technology Paper #SV-25, Seagate, Scotts Valley, Calif., 1995.
- Whitaker, Jerry C.: "Data Storage Systems," in *The Electronics Handbook*, Jerry C. Whitaker (ed.), CRC Press, Boca Raton, Fla., pp. 1445–1459, 1996.

Chapter
9.1

Audio/Video Server Systems

Jerry C. Whitaker, Editor-in-Chief

9.1.1 Introduction

Audio/video servers, with their unique set of features and functions, have reshaped the way program segments are stored and played to air. These systems have emerged from the realm of limited-use, special-purpose devices to mainstream program production. Among the many attributes of servers are:

- They permit material from a single storage source to be used simultaneously by multiple users
- Provide a migration path to the all-digital facility that is not necessarily format-limited
- Result in a reduction in lost or misplaced materials
- Reduction in the size and space requirements relative to a tape environment
- Complete computer-control capabilities
- Near-instant access and playback of audio and video segments

The end-result of these attributes is an environment where multiple applications and/or services can be generated from a single system. This reduces the amount of playback equipment required, reduces tape consumption, and generally permits more efficient use of human resources.

9.1.2 Basic Architecture

The basic server architecture consists of three elements [1]:

- A multiple hard disk drive system capable of fast and simultaneous data access, with sufficient capacity and redundancy for the contemplated application. A disk array controller manages data distribution and communications among all drives.
- Fast data communication interfaces among disk drives and networks. Several approaches can be used to perform fast data transfer such as very fast CPUs, multiple CPUs and buses, and routing switchers. Interfaces may also include data compression encoding and decoding.

9-6 Audio Recording Systems

Table 9.1.1 Storage Space Requirements for Audio and Video Data Signals (After [1].)

Media Signals	Specifications	Data Rate
Voice-grade audio	1 ch; 8-bit @ 8 kHz	64 kbytes/s
MPEG audio Layer II	1 ch; 16-bit @ 48 kHz	128 kbytes/s
MPEG audio Layer III	1 ch; 16-bit @ 48 kHz	64 kbytes/s
AC-3	5.1 ch; 16-bit @ 48 kHz	384 kbytes/s
CD	2 ch; 16-bit @ 44.1 kHz	1.4 Mbytes/s
AES/EBU	2 ch; 24-bit @ 48 kHz	3.07 Mbytes/s
MPEG-1 (video)	352 × 288, 30 f/s, 8-bit	1.5 Mbytes/s
MPEG-2 (MP@ML)	720 × 576, 30 f/s, 8-bit	15 Mbytes/s, max.
MPEG-2 (4:2:2 P@ML)	720 × 608, 30 f/s, 8-bit	50 Mbytes/s, max.
ITU-R Rec. 601	720 × 480, 30 f/s, 8-bit	216 Mbytes/s
HDTV	1920 × 1080, 30 f/s, 8-bit	995 Mbytes/s

- An operating system capable of handling multiple digital audio and/or video data streams in any combination of record and playback modes, while ensuring correct file management and easy access.

The design and performance of the server involves trade-offs between the quality of compressed signals, storage capacity, data speed, play time, number of channels, access speed, and reliability. Table 9.1.1 shows the storage space requirements for different audio and video signals.

All servers have in common a large storage capacity and multiple channel capability; differences involve the basic architecture and performance targets, because different systems may be designed to meet specific requirements. The most common application groups for broadcasting include the following:

- **Transmission.** High compression ratios can be used for server-to-air applications, typically commercial and short-length program replay. The resulting bit rates and number of channels are usually low, therefore, with relatively low bandwidth requirements.
- **Data cache.** A caching system is basically a temporary random access disk buffer. It is commonly used in conjunction with tape library systems in on-air applications. Caches are well suited to commercial on-air insertion applications where elements are repeated several times a day and last-moment changes occur frequently. Bandwidth and channel requirements are usually not critical.
- **News.** For most news materials, moderate overall quality levels are acceptable. Audio/video compression is used to reduce the data file size and increase the transfer rate. Multiple access to multiple segments must be possible and a moderate bandwidth is necessary. Guaranteed availability of output ports might be required for direct on-air programming of news materials.

Table 9.1.2 Production and Broadcast Server Bandwidth Requirements for 4:2:2 Video Signals (After [1].)

Production Applications	Bandwidth, Mbits/s	Sample Resolution	Compression Ratio
High-end post-production	270	10	1
Typical post-production	90	10	2.3:1
Low-end post-production	25–50	8	6.6:1–3.3:1
News (compressed data)	18–25	8	9:1–6.6:1
HDTV broadcast	20	8	10:1–50:1
Good-quality SDTV broadcast	8	8	20:1
Medium-quality SDTV broadcast	3	8	55:1
Low-quality SDTV broadcast	1.5	8	110:1

- **Production.** Wide bandwidth must be provided to meet simultaneous transfer demands for large uncompressed files with real-time and random access capabilities. In a production facility, a server stores all compressed or uncompressed audio and video data files for use in post-production and distribution. The server is central to the production operation and, thus, determines the overall performance of the facility. Table 9.1.2 lists production server bandwidth requirements for 4:2:2 video signals.
- **Video on demand (VOD).** This server application must deliver a large number of channels, each of relatively low video quality, such as MPEG-1 (1.5 Mbits/s). A high overall bandwidth may be required to satisfy all demands. Short access times are necessary for VOD, whereas long access times permit *near-VOD* (NVOD) only.

Audio/video servers have redefined many of the common applications of video tape recorders. The VTR is being displaced in direct-to-air uses, complex post production, and desktop editing. As new VTR formats evolve, they will increasingly be aimed at acquisition and long-form storage applications. Table 9.1.3 lists the digital tape formats in common usage.

Among the technical frontiers for servers is the ability to transfer files at *faster than realtime*. Such features place considerable demands on the bandwidth and throughput of the system; however, they provide many operational benefits, such as accelerated non-linear editing and reduced program load-up times.

9.1.2a Server Design

A number of operational scenarios can be implemented with an audio/video server system. The simplest involves multiple streams of one or two video inputs, as illustrated in Figure 9.1.1. All the material is stored within the server and plays out to multiple channels as required.

In Figure 9.1.2, the basic system is expanded to allow for some material, particularly long form programs, to be held in an external storage area or device (most likely on tape) and loaded into the server on an as-needed basis to coexist with the short-form (typically commercial) library that is stored in the server. The server then manages the entire play-out process to one or more outputs.

9-8 Audio Recording Systems

Table 9.1.3 Digital Video Tape Formats

Format Designation	Format Description
D1	A format for digital video tape recording conforming to the ITU-R Rec. 601 (4:2:2) standard using 8-bit sampling. The tape is 19 mm wide and allows up to 94 minutes to be recorded on a cassette.
D2	The VTR standard for digital composite (encoded) NTSC or PAL signals. D2 uses 19 mm tape and records up to 208 minutes on a single cassette. Neither cassettes nor recording formats are compatible with D1. D2 has often been used as a direct replacement for 1-inch analog (Type C) VTRs.
D3	A VTR standard using 1/2-inch tape cassettes for recording digitized composite (encoded) NTSC or PAL signals sampled at 8 bits. Cassettes are available ranging from 50–245 minutes.
D4	There is no D4 designation; most DVTR formats were developed in Japan where "4" is regarded as an unlucky number.
D5	A VTR formal using the same cassette as D3 but recording component signals sampled to ITU-R Rec. 601 at 10-bit resolution. With internal decoding, D5 VTRs can play back D3 tapes and provide component video outputs.
D6	A helical-scan digital tape format that uses a 19 mm cassette to record uncompressed HDTV material at 1.88 Gbits/s. D6 accepts both the European 1250/50 interlaced format and the SMPTE 260M version of the 1125/60 interlaced format, which uses 1035 active lines.
D7	The designation assigned to the DVCRPO recording format.

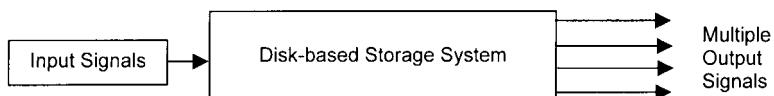


Figure 9.1.1 Simple application of a video server in a one-in, many-out configuration.

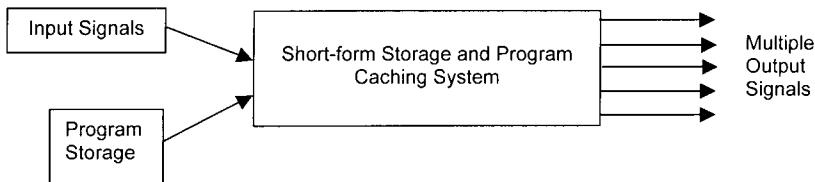


Figure 9.1.2 A server system expanded to utilize long-form program storage.

In Figure 9.1.3, a large-scale system is depicted typical of a newsroom environment. Feeds are brought into the system through a triage station permitting rough editing of material ("keep or discard" decisions). The selected material is then stored in a feed server, which is accessible

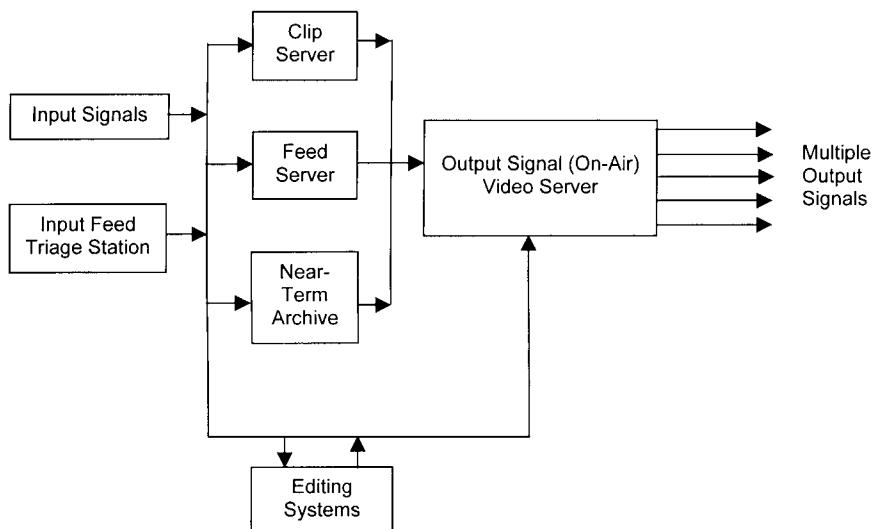


Figure 9.1.3 A server system designed for editing-intensive applications involving a variety of program inputs and distribution requirements.

for viewing or by editors for story composition. Completed program elements are forwarded to the on-air server, which stores them and plays-out as needed. Also shown in the Figure is a *clip server*, a low-resolution companion server that delivers video to multiple desktop viewing and editing systems.

Audio/video servers are often constructed from off-the-shelf computer components by integrating the basic hardware and software. This approach promotes a migration path for future growth in several ways, including:

- Utilization of standardized components
- Enhanced ability to expand or alter the system
- Ability to upgrade system core capabilities with new software releases

By utilizing standard components, the research and development cost are spread across a large market, providing faster and more cost-effective advancement. For example, all disk drives adhere to a similar footprint, respond to a similar set of commands, mount in a similar manner, and so forth. This makes it easy to take advantage of improvements in specific technologies.

9.1.2b Archiving Considerations

As a conventional tape library grows, a two-sided challenge unfolds. First, operators must identify and archive the contents of the tape. Next, they must be able to find the tape. Servers inherently must maintain a database of the materials within their contents and manage the recall of

9-10 Audio Recording Systems

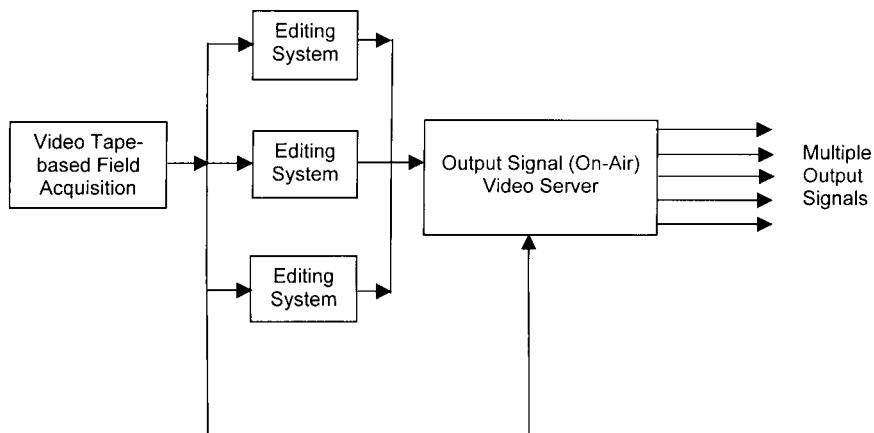


Figure 9.1.4 A video tape-to-server environment with editing capabilities.

that material on demand. In addition, because the server brings basic data storage concepts into play, more use can be made of asset management software that permits identification and location of all materials and their contents, regardless of the storage medium.

Tape, however, remains an exceptionally robust and inexpensive medium for field acquisition. It also is economical for many long-form materials and for archiving. Tape and servers can work together to provide a performance and economic model that offers maximum benefit to the user. As illustrated in Figure 9.1.4, raw footage is acquired on tape. The tape material is loaded into the editors for composition, and the finished programs are forwarded to a server, where it is held for play-out.

Despite the many benefits of server systems, there are a few drawbacks. Catastrophic failure is first among them. Tape, as a medium, is not susceptible to catastrophic failure (generally speaking). Disk drives can—and do—crash. However, advancements in drive technology have greatly increased the mean time between failure for disk drives. RAID (*redundant array of independent disks*)¹ schemes provides built-in back-up, particularly when linked with hot spare drives and hot-swappable components such as power supplies. With these additions, the risk of total failure of a server system is quite low.

9.1.2c Audio/Video Server Storage Systems

Because of the fundamental impact that storage technology has on any server, it is important to understand the various RAID configurations commonly in use. The typical solution to providing access to many gigabytes of data to users fast and reliably has been to assemble a number of drives together in a “gang” or array of disks. Simple RAID subsystems are basically a group of

1. Also known as redundant array of *inexpensive* disks

up to five or six disk drives assembled in a cabinet that are all connected to a single controller board. The RAID controller orchestrates read and write activities in the same way a controller for a single disk drive does, and treats the array as if it were—in fact—a single or “virtual” drive. RAID management software, which resides in the host system, provides the means to control the data being stored on the RAID subsystem.

Despite its multi-drive configuration, the individual disk drives of a RAID subsystem remain hidden from the user; the subsystem itself is the virtual drive, though it can be as large as 1,000 GB or more. The phantom virtual drive is created at a lower level within the host operating system through the RAID management software. Not only does the software set up the system to address the RAID unit as if it were a single drive, it allows the subsystem to be configured in ways that best suit the general needs of the host system.

RAID subsystems can be optimized for performance, highest capacity, fault tolerance, or a combination of these attributes. Different *RAID levels* have been defined and standardized in accordance with these general optimization parameters. There are six common standardized RAID levels, called RAID 0, 1, 2, 3, 4, and 5. The use of a particular level depends upon the performance, redundancy, and other attributes required by the host system.

The RAID controller board is the hardware element that serves as the backbone for the array of disks; it not only relays the input/output (I/O) commands to specific drives in the array, but provides the physical link to each of the independent drives so they may easily be removed or replaced. The controller also serves to monitor the integrity of each drive in the array to anticipate the need to move data should it be placed in jeopardy by a faulty or failing disk drive (a feature known as *fault tolerance*).

9.1.3 Basic Drive Technology

SCSI (*small computer system interface*) hard disk drives are the foundation of most RAID systems. The drives use various numbers of magnetic coated disks that rotate at 5,400, 7,200, or 10,000 RPM [2]. Each disk has two recording surfaces. The disks are logically divided into concentric circles (*tracks*). A set of tracks at a given position on the disks is known as a *cylinder*. The number of cylinders is the same as the number of tracks across the disks. Tracks are then divided into varying numbers of *sectors*. The number of sectors varies according to the position of the track located on the disk. This technology is known as *zone bit recording* (ZBR).

Data are written to one track of a given surface at a time. If more space is needed, a head switch takes place and data then are written to the next surface. Head switches continue down the cylinder until the last track in the cylinder is filled. If additional space is needed, the heads are stepped to a new cylinder and the process of head switching and track stepping continues until the file is completed. Using ZBR, the outside tracks contain more sectors than the inside tracks because there is more physical space on the outside tracks. The transfer rate is greater from the outside as well because a greater amount of data is available from a single rotation of the disks. The first partition created uses the outside tracks and moves to inside tracks as the drive or drives are fully partitioned.

SCSI is a general-purpose interface. In its basic configuration, a maximum of eight or 16 devices can be connected to a single bus. SCSI specifies a cabling standard, a protocol for sending and receiving commands, and the format for those commands. SCSI is intended as a device-

9-12 Audio Recording Systems

independent interface so the host computer requires no details about the peripherals that it controls.

As the SCSI standard has evolved, various levels have been produced, the most common being:

- Standard SCSI, with a 5 MB/s transfer rate
- Fast SCSI, with a 10 MB/s transfer rate
- Ultra SCSI, with a 20 MB/s transfer rate
- Ultra SCSI-2, with a 40 MB/s transfer rate

For each of these schemes, there is the 8-bit normal or so-called narrow bus (1 byte per transfer) or the 16-bit wide bus (2 bytes per transfer). Therefore, for example, Ultra SCSI-2 is designed to transfer data at a maximum rate of 80 MB/s. Actual continuous rates achieved from the disk drive will typically be considerably less, however.

9.1.3a RAID Levels

The RAID 0 through 5 standards offer users a host of configuration options [3]. These options permit the arrays to be tailored to their application environments, for our purposes an audio/video server. Each of the various configurations focus on maximizing the abilities of an array in one or more of the following areas:

- Capacity
- Data availability
- Performance
- Fault tolerance

RAID Level 0

An array configured to RAID Level 0 is an array optimized for performance, but at the expense of fault tolerance or data integrity [3]. RAID Level 0 is achieved through a method known as *striping*. The collection of drives (*virtual drive*) in a RAID Level 0 array has data laid down in such a way that it is organized in stripes across the multiple drives. A typical array can contain any number of stripes, usually in multiples of the number of drives present in the array. As an example, a four-drive array configured with 12 stripes (four stripes of designated “space” per drive). Stripes 0, 1, 2, and 3 would be located on corresponding hard drives 0, 1, 2, and 3. Stripe 4, however, appears on a segment of drive 0 in a different location than Stripe 0; Stripes 5 through 7 appear accordingly on drives 1, 2, and 3. The remaining four stripes are allocated in the same even fashion across the same drives such that data would be organized in the manner depicted in Figure 9.1.5. Practically any number of stripes can be created on a given RAID subsystem for any number of drives. Two hundred stripes on two disk drives is just as feasible as 50 stripes across 50 hard drives (data management “overhead” notwithstanding). Most RAID subsystems, however, tend to have between three and 10 stripes.

The reason RAID 0 is a performance-enhancing configuration is that striping enables the array to access data from multiple drives at the same time. In other words, because the data is spread out across a number of drives in the array, it can be accessed faster because it is not bot-

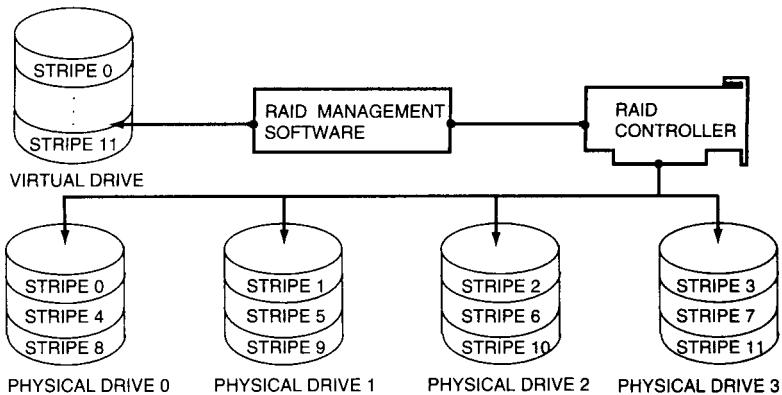


Figure 9.1.5 In a RAID Level 0 configuration, a virtual drive is comprised of several stripes of information. Each consecutive stripe is located on the next drive in the chain, evenly distributed over the number of drives in the array. (From [3]. Used with permission.)

tled up on a single drive. This is especially beneficial for retrieving a very large file, because it can be spread out effectively across multiple drives and accessed as if it were the size of any of the fragments it is organized into on the data stripes. By any measure of comparison, video files qualify as “very large files.”

The downside to the RAID Level 0 configuration is that it sacrifices fault tolerance, raising the risk of data loss because no room is made available to store redundant information. If one of the drives in the RAID 0 fails for any reason, there is no way of retrieving the lost data, as can be done in other RAID implementations.

RAID Level 1

The RAID Level 1 configuration employs what is known as *disk mirroring*, and is done to ensure data reliability (a high degree of fault tolerance) [3]. RAID 1 also enhances read performance, but the improved performance and fault tolerance come at the expense of available capacity in the drives used. In a RAID Level 1 scheme, the RAID management software instructs the subsystem’s controller to store data redundantly across a number of the drives (*mirrored set*) in the array. In other words, the same data is copied and stored on different disks to ensure that, should a drive fail, the data is available somewhere else within the array. In fact, all but one of the drives in a mirrored set could fail and the data stored to the RAID 1 subsystem would remain intact. A RAID Level 1 configuration can consist of multiple mirrored sets, whereby each mirrored set can be of a different capacity. Usually, the drives making up a mirrored set are of the same capacity. If drives within a mirrored set are of different capacities, the capacity of a mirrored set within the RAID 1 subsystem is limited to the capacity of the smallest-capacity drive in the set, hence the sacrifice of available capacity across multiple drives.

The read performance gain can be realized if the redundant data is distributed evenly on all of the drives of a mirrored set within the subsystem. The number of read requests and total wait

9-14 Audio Recording Systems

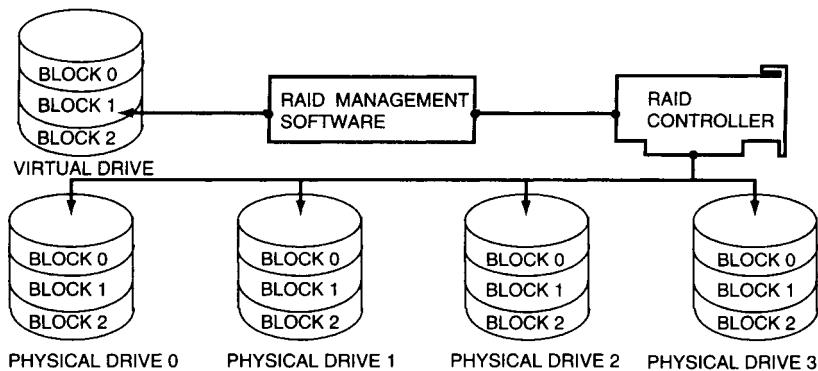


Figure 9.1.6 A RAID Level 1 subsystem provides high data reliability by replicating (*mirroring*) data between physical hard drives. In addition, I/O performance is boosted as the RAID management software allocates simultaneous read requests among several drives. (From [3]. Used with permission.)

state times both drop significantly; inversely proportional to the number of hard drives in the RAID, in fact. To illustrate, suppose three read requests are made to the RAID Level 1 subsystem (see Figure 9.1.6). The first request looks for data in the first block of the virtual drive; the second request goes to block 0, and the third seeks from block 2. The host-resident RAID management software can assign each read request to an individual drive. Each request is then sent to the various drives, and now—rather than having to handle the flow of each data stream one at a time—the controller can send three data streams almost simultaneously, which in turn reduces the data access time.

RAID Level 2

RAID Level 2, rarely used in audio/video applications, is another means of ensuring that data is protected in the event drives in the subsystem incur problems or otherwise fail [3]. This level builds fault tolerance around *Hamming error correction code* (ECC), which is often used in modems and solid-state memory devices as a means of maintaining data integrity. ECC tabulates the numerical values of data stored on specific blocks in the virtual drive using a formula that yields a checksum. The checksum is then appended to the end of the data block for verification of data integrity when needed.

As data is read back from the drive, ECC tabulations are again computed, and specific data block checksums are read and compared against the most recent tabulations. If the numbers match, the data is intact; if there is a discrepancy, the lost data can be recalculated using the first or earlier checksum as a reference.

This form of ECC is actually different from the ECC technologies employed within the drives themselves. The topological formats for storing data in a RAID Level 2 array is somewhat limited, however, compared to the capabilities of other RAID implementations, which is why it is not commonly used in commercial applications.

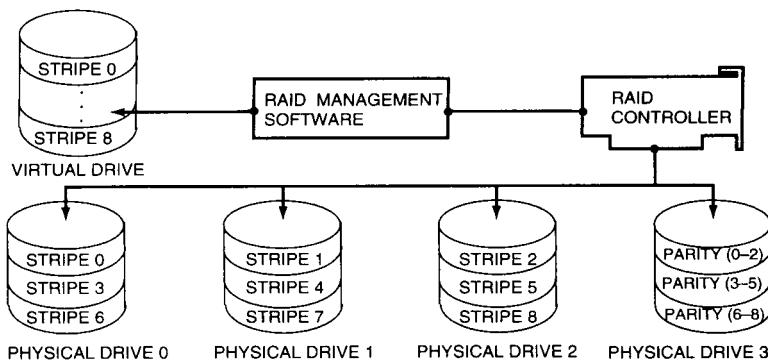


Figure 9.1.7 A RAID Level 3 configuration is similar to a RAID Level 0 in its utilization of data stripes dispersed over a series of hard drives to store data. In addition to these data stripes, a specific drive is configured to hold parity information for the purpose of maintaining data integrity throughout the RAID subsystem. (From [3]. Used with permission.)

RAID Level 3

This RAID level is essentially an adaptation of RAID Level 0 that sacrifices some capacity, for the same number of drives, but achieves a high level of data integrity or fault tolerance [3]. It takes advantage of RAID Level 0 data striping methods, except that data is striped across all but one of the drives in the array. This drive is used to store parity information for maintenance of data integrity across all drives in the subsystem. The parity drive itself is divided into stripes, and each parity drive stripe is used to store parity information for the corresponding data stripes dispersed throughout the array. This method achieves high data transfer performance by reading from or writing to all of the drives in parallel or simultaneously, but retains the means to reconstruct data if a given drive fails, maintaining data integrity for the system. This concept is illustrated in Figure 9.1.7. RAID Level 3 is an excellent configuration for moving very large sequential files, such as video, in a timely manner.

The stripes of parity information stored on the dedicated drive are calculated using the Exclusive OR function. By using Exclusive OR with a series of data stripes in the RAID, lost data can be recovered. Should a drive in the array fail, the missing information can be determined in a manner similar to solving for a single variable in an equation.

RAID Level 4

This level of RAID is similar in concept to RAID Level 3, but emphasizes performance for particular applications, e.g. database files versus large sequential files [3]. Another difference between the two is that RAID Level 4 has a larger stripe depth, usually of two blocks, which allows the RAID management software to operate the disks more independently than RAID Level 3 (which controls the disks in unison). This essentially replaces the high data throughput capability of RAID Level 3 with faster data access in read-intensive applications. (See Figure 9.1.8.)

9-16 Audio Recording Systems

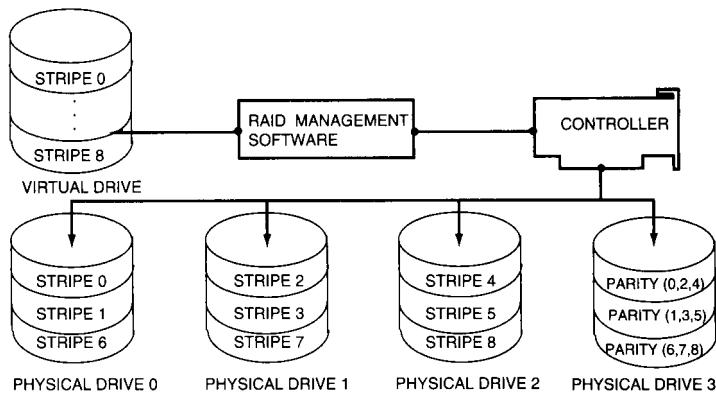


Figure 9.1.8 RAID Level 4 builds on RAID Level 3 technology by configuring parity stripes to store data stripes in a non-consecutive fashion. This enables independent disk management, ideal for multiple-read-intensive environments. (From [3]. Used with permission.)

A shortcoming of RAID level 4 is rooted in an inherent bottleneck on the parity drive. As data is written to the array, the parity encoding scheme tends to be more tedious in write activities than with other RAID topologies. This more or less relegates RAID Level 4 to read-intensive applications with little need for similar write performance. As a consequence, like Level 2, Level 4 does not see much common use in commercial applications.

RAID Level 5

Level 5 is the last of the common RAID levels in general use, and is probably the most frequently implemented [3]. RAID Level 5 minimizes the write bottlenecks of RAID Level 4 by distributing parity stripes over a series of hard drives. In so doing, it provides relief to the concentration of write activity on a single drive, which in turn enhances overall system performance. (See Figure 9.1.9.)

The way RAID Level 5 reduces parity write bottlenecks is relatively simple. Instead of allowing any one drive in the array to assume the risk of a bottleneck, all of the drives in the array assume write activity responsibilities. This distribution eliminates the concentration on a single drive, improving overall subsystem throughput. The RAID Level 5 parity encoding scheme is the same as Levels 3 and 4, and maintains the system's ability to recover lost data should a single drive fail. This recovery capability is possible as long as no parity stripe on an individual drive stores the information of a data stripe on the same drive. In other words, the parity information for any data stripe must always be located on a drive other than the one on which the data resides.

Other RAID levels

Other, less common, RAID levels have been developed as custom solutions by independent vendors, including:

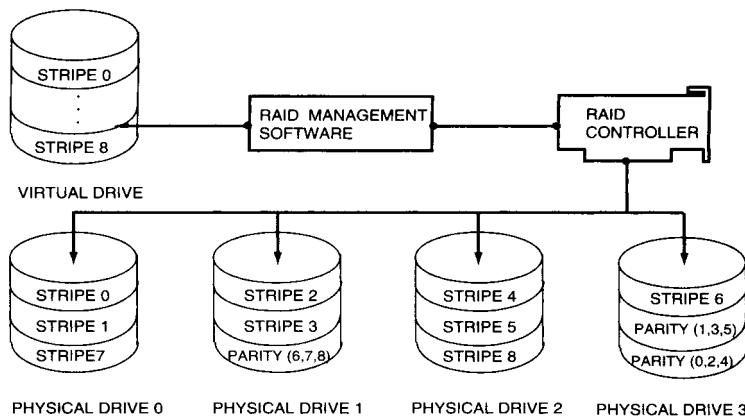


Figure 9.1.9 RAID Level 5 overcomes the RAID Level 4 write bottleneck by distributing parity stripes over two or more drives within the system. This better allocates write activity over the RAID drive members, thus enhancing system performance. (From [3]. Used with permission.)

- RAID Level 6, which emphasizes ultra-high data integrity
- RAID Level 10, which focuses on high I/O performance and very high data integrity
- RAID Level 53, which combines RAID Level 0 and 3 for uniform read and write performance

Perhaps the greatest advantage of RAID technology is the sheer number of possible adaptations available to users and systems designers. RAID offers the ability to customize an array subsystem to the requirements of its environment and the applications demanded of it. These attributes make RAID systems an integral element in applications requiring mass storage of audio and video signals, such as servers. Table 9.1.4 lists the relative attributes of the various RAID implementations.

9.1.4 Digital Media Applications

Speed and storage capacity are particularly important to digital media applications (those applications that record or play audio and/or video data) [3,4]. One minute of uncompressed CD-quality audio requires over 4.5 megabytes of storage. That is over 150 times the storage required to hold most word processing documents. The typical data rate requirement for a single channel of uncompressed digitally encoded audio is 76 KB/s.

For digital video, the capacity requirements jump exponentially. A single frame of 70 mm film, if stored in uncompressed format, requires 40 MB of disk storage. Video is typically compressed to increase the amount of information that can be stored on the disk and to reduce the high data rate requirements. The data rate varies depending upon the video compression techniques used (JPEG, MPEG, or other proprietary schemes) and the desired quality of the stored

9-18 Audio Recording Systems

Table 9.1.4 Summary of RAID Level Properties (After [3].)

RAID Level	Capacity	Data Availability	Data Throughput	Data Integrity
0	High	Read/Write High	High I/O Transfer Rate	
1		Read/Write High		Mirrored
2	High		High I/O Transfer Rate	ECC
3	High		High I/O Transfer Rate	Parity
4	High	Read High		Parity
5	High	Read/Write High		Parity
6		Read/Write High		Double Parity
10		Read/Write High	High I/O Transfer Rate	Mirrored
53			High I/O Transfer Rate	Parity

video image. One minute of broadcast-quality video using JPEG compression requires at least 120 MB of storage. The data rate can vary from 150 KB/s of video for MPEG compression, up to 27 MB/s for full frame, uncompressed, interlaced video.

Simply put, the higher the video resolution the more space and higher data rate it requires. Where a 4 GB capacity drive might have been big enough for a mid-quality video production, higher quality video may require 9 GB of storage for a given segment.

Storing audio and video information on disk drives is a logical step on the digital media road because it provides instant access to stored data. However, not all disk storage is created equal. Optimizing disk drives for digital media requires more than just high capacity and fast data rates. Data processing applications read and write small chunks of data stored more or less randomly around the disk. Consequently, conventional disk drives have been optimized for small block, random data transfers. Yet, while contemporary data processing disk drives deliver data to traditional applications quickly, they may have difficulties satisfying the requirements for real-time audio and video applications. Because of the nature of audio and video playback, digital media applications require that data be delivered on time, at the required rate, with no delay. Any delay in the delivery of data from the disk will cause noticeable interruption in the playback, resulting in audio breaks and/or missing video frames.

The linear nature of sound and video playback dictates that these files must be organized contiguously on the disk in much the same way as they are on tape. With tape, the media containing the information travels past the read head at the velocity required to present the sound and images at a real-time speed. In this case, the next word, musical note, or video frame always resides adjacent to the current sound or picture. It is the sequential linearity of tape that ideally suits it for these purely “playback” applications. When the application demands manipulation or editing of the audio or video information, the utility of linear tape storage methods quickly become dubious and cumbersome.

For video servers and non-linear audio and video production systems, a storage device is required that can instantly access any video frame, sound, or word, and record or play back with the same or better measured precision as tape. This is a type of storage for which standard data processing disk drives were not designed. Figure 9.1.10 charts the speed and smoothness of a disk’s throughput by graphing worst-case data access over time. The performance of the tested

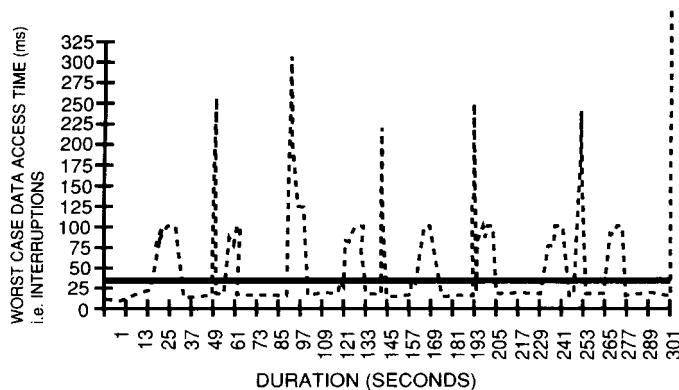


Figure 9.1.10 Disk drive data access as a function of play time for a conventional hard drive. Host demand rate = 4.0 MB/s. (From [3]. Used with permission.)

drive is indicated by the broken line. The straight, solid line above the X axis indicates the slowest acceptable access time to deliver video data. Every time the disk drive data response time exceeds this value, the video playback will experience dropouts or lost frames, which result in visual and/or audible discontinuities in the playback.

Previously, data interruptions were solved by adding large memory buffers in the host system to hold several frames of data while the disk performed the requisite housekeeping chores. Although host memory buffers cannot be completely eliminated, their size and subsequent cost can be significantly reduced by minimizing the time the disk spends attending to certain housekeeping functions. A disk drive optimized for continuous recording and playback will maintain a virtually constant data rate to ensure that every picture and every sound is recorded and played back without any evident errors. The broken line in Figure 9.1.11 illustrates the performance of a drive optimized for audio/video applications.

Different manufacturers approach the challenges of designing drives optimized for audio/video applications in different ways. One design achieves the desired relatively flat data rate performance by changing the way disk drives work. By designing disk storage with the requirements of digital media in mind—including the development of new on-disk caching schemes optimized for the file characteristics of audio and video applications and the implementation of “intelligent housekeeping” techniques to hide necessary tasks in the background—the overall performance is improved considerably.

9.1.4a Optimizing Digital Media Disk Performance

Although the error recovery methods used in conventional disk drives are adequate to keep disks operating within desired parameters, they are not optimized for speed [3,4]. To maintain the disk drive at its optimum operating range, or to correct errors, some additional issues are important. Issues that can interrupt the flow of data include the following:

9-20 Audio Recording Systems

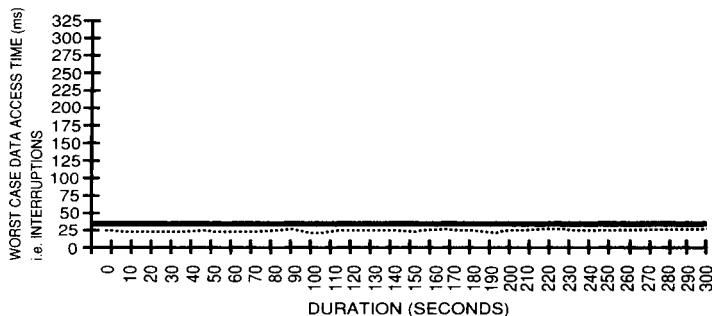


Figure 9.1.11 Disk drive data access as a function of play time for a hard drive optimized for A/V applications. Host demand rate = 4.0 MB/s. (From [3]. Used with permission.)

- **Error recovery procedure overhead.** The error recovery procedure within most conventional disk drives is very robust and thorough. However, it is a major contributor to data flow interruptions due to “soft” data errors. For example, the time required for a worst case recovery from a correctable data error can exceed 750 ms. Most digital media applications cannot tolerate a delay of this magnitude without a disruption in the audio or video playback. One solution to this situation is to implement sophisticated and exceptionally reliable means of error correction that maintains data integrity while rapidly completing housekeeping and recovery tasks. The time required to perform error recovery procedures can, in most cases, be held to 10 ms or less.
- **Hardware ECC on the fly.** By adding a dedicated hardware correction engine, hardware-implemented error correcting code can be implemented “on the fly.” In other words, a soft read error can—in most cases—be corrected by a dedicated hardware correction engine. This operation is done literally “on the fly” within a few hundred microseconds, without incurring the overhead of waiting a full rotation in order to read the same data again. The more convention approach uses a software error correcting code technique which is many times slower than the hardware ECC approach.
- **Rotational retries.** If hardware ECC is not present in the drive or is unsuccessful in properly recovering the data, the next step in the error recovery process is to simply retry the operation. This approach, however, is not optimized for speed. When a disk tries to re-read data it could not retrieve on the first pass, at least one revolution is lost. The ECC on-the-fly techniques described previously can usually recover soft read errors without extra disk revolutions.
- **Thermal calibration (T-Cal).** Thermal calibration is a periodic (e.g., every 10 min) house-keeping function that is necessary for disk drives incorporating a dedicated servo system. Dedicated servo designs are commonly used by many disk drive manufacturers. The T-Cal operation calibrates the servo system to ensure that the disk heads remain precisely over the data tracks by compensating for temperature changes during normal operations. The T-Cal operation, depending upon the exact design, can require hundreds of milliseconds to complete. During all or at least part of this process, the drive will not be responsive to data

requests. Drives optimized for audio/video applications can employ a hybrid servo system, which combines the best features of both dedicated and embedded servo system designs and eliminates the requirement for periodic T-Cal operations. Through clever design, it is possible to eliminate T-Cal operations, not simply defer them.

- **Data head degaussing.** During the normal course of disk operation, the data heads accumulate a slight magnetic orientation. This residual magnetism can adversely affect operation. To prevent this condition, the drive must degauss each head. Because head degaussing involves at least one head seek plus some rotational latency, this activity can cost more than 40 ms and will interrupt the flow of data from the disk. Refined degaussing schemes minimize head seeking and rotational latency.

By combining the foregoing techniques, it is possible to significantly improve the overall sustainable throughput of disk drives so that they can reliably be used for demanding audio/video applications.

Thermal Calibration

It is worthwhile to explore the issue of T-Cal in greater detail. To understand the need for thermal calibration, the physics of disk drives must be examined [5]. All metals expand and contract to a certain degree under varying temperatures. The same is true with disk drive parts, including platters and the actuator arms that read the data. As a drive spins, the components heat up and expand. Although the size variance may at first seem minimal, it is potentially rather significant as tracks are spaced closer together on disk drives.

Initially, drives circumvented this problem by delaying read/write operations long enough to re-calibrate the location of the actuator arm over the platter. This would take a few milliseconds, and would be unnoticeable in standard applications. With A/V projects, however, a pause lasting several milliseconds can result in skipped frames and lost audio output.

Several approaches have been developed to solve this problem. One involves the repositioning of the heads. Normally, all heads of a disk drive would re-calibrate at the same time, bringing a complete stop to the data flow. To minimize the duration of data flow interruptions, newer drives reposition one head at a time. Further, the drive can complete a thermal calibration by taking advantage of its onboard cache. A segment of data is read ahead of time and stored in the drive buffer. When that data is requested from the computer, the drive dumps it from the cache back to the system. While the buffer is being read, the actuator arms take the opportunity to re-calibrate.

9.1.4b SCSI and Fibre Channel

Although not specifically a function of disk drive technology, new variations of SCSI and *Fibre Channel Arbitrated-Loop* (FC-AL) are important in the broad view of storage technology because of their impact on data transfer rates [3]. For this reason, they will be examined briefly here.

Since its inception in the early 1980s, SCSI has evolved to become a widely-accepted and successful disk drive interface [6]. Despite its vulnerability, however, parallel SCSI still has its drawbacks, including cable distance limitations, a confusing mix of variants to the bus (*Fast*, *Fast&Wide*, *Differential*, *Ultra*, and so on) and a ceiling on the number of peripherals it can support. SCSI throughput performance has also reached a point where disk drive capacities and data

9-22 Audio Recording Systems

Table 9.1.5 Fibre Channel Features and Capabilities

Parameter	Range of Capabilities
Line rate	266, 531, or 1062.5 Mbits/s
Data transfer rate (maximum)	640–720 Mbits/s @ 1062.5 line rate
Frame size	2112 byte payload
Protocol	SCSI, IP, ATM, SDI, HIPPI, 802.3, 802.5
Topology	Loop, switch
Data integrity	10E–12 BER
Distance	Local and campus; up to 10 km

rates are likely to render current SCSI variants as bottlenecks in future systems built around fast new microprocessors. These and other limitations of SCSI are creating demand for a better solution to the high-performance I/O needs of the computer systems market, and at least one solution is a serial data interface called Fibre Channel Arbitrated Loop. FC-AL is a subset of the box-to-box standard created by original members of the Fibre Channel Association.¹ Like its Fibre Channel superset, FC-AL is an industry-standard interface endorsed by the American National Standards Institute (ANSI). Fibre Channel is usually thought of as a system-to-system or system-to-subsystem interconnection standard that uses optical cable in a point-to-point or switch configuration. This is what it was originally designed to do, in fact, since HIPPI and the Internet Protocol were among the protocols defined for it. One of the goals in the development of the Fibre Channel interface was to improve or eliminate SCSI shortcomings, particularly in the areas of connectivity, performance, and physical robustness.

In 1991 the Fibre Channel box-to-box interconnect standard was enhanced to include support for copper (nonoptical) media and multidrop configurations, both of which enable the low-cost connection of many devices to a host port. This subset of the Fibre Channel standard is called Fibre Channel-Arbitrated Loop, and is what made it possible for Fibre Channel to be used as a direct-disk-attachment interface (SCSI-3 has been defined as the disk interface protocol, specifically SCSI-FCP). The implications of that capability are enormous in terms of the cost savings and ease with which users can migrate to standard systems with performance capabilities heretofore only found in expensive proprietary systems at the workstation or mainframe level.

The basic features of Fibre Channel are listed in Table 9.1.5. The interface features of FC-AL are given in Table 9.1.6.

FC-AL Topology

The Fibre Channel-Arbitrated Loop interface is a loop topology, not a bus in the conventional sense like SCSI [3]. It can have any combination of hosts and peripherals, up to a loop maximum of 126 devices.

Using a connector based on the 80-pin parallel SCSI *single connector attachment* (SCA), Fibre Channel disk drives attach directly to a backplane. This not only eliminates cable congest-

1. Hewlett-Packard, IBM, and Sun Microsystems

Table 9.1.6 Basic Parameters of FC-AL (After [3].)

Parameter	Range of Capabilities
Number of devices	126
Data rate	100 MB/s (1.062 GHz using an 8B/10B code)
Cable distance	30 m between each device using copper (longer, with other cabling options)
Cable types	Backplane, twinaxial, coaxial, optical
Fault tolerance	Dual porting, hot plugging

tion, it makes hot drive insertions practical and simplifies mechanical designs. Fewer cables and components translate to lower-cost systems and higher reliability.

Connectivity Considerations

As discussed previously, applications such as video and image processing have pushed the demand for huge increases in disk capacity per system [3]. In some cases, the capacity requirements for these types of applications are such that it is difficult to configure a sufficient number of SCSI buses on a system so that enough drive addresses are available to attach the necessary disk storage. Moreover, simply increasing the addressability of SCSI—making it possible to have more than 15 devices per Wide SCSI bus, for example—would not be a solution, because more bus bandwidth is needed to support the additional drives. Besides, protocol overhead is already rather high. FC-AL can address up to 126 devices, but practical usage is another matter. A loop can practically support about 60 drives in a UNIX (with 8 KB I/Os) environment, as an example. With an 18 GB FC-AL interface disk drive, a loop of 50 drives would make more than 900 GB available on a single FC-AL host adapter. This makes it possible for any workstation or system with a Fibre Channel port to become an incredibly large file server.

Bandwidth

A Fibre Channel loop, as stated previously, supports data rates up to 100 MB/s in single-port applications and up to 200 MB/s in dual-port configurations [3]. Applications such as digital video data storage and retrieval, computer modeling, and image processing are growing in popularity and demanding ever-increasing improvements to disk data transfer performance. Moreover, file servers are increasingly looked upon as replacements for mainframe computers. In order to fulfill that promise, they will need to deliver higher transaction rates to provide a mainframe level of service.

Magnetic disk drive areal densities are known to be increasing at about 60 percent per year in production products. Because the number of bits per inch—one of the two components of disk areal density—must increase at about 30 percent per year, the data rate performance must also increase proportionately. In addition, drive spindle speeds (or spin rates) continue to increase, which directly contribute to the need for improved data rate performance.

9-24 Audio Recording Systems

Remote Online Storage

Because the FC-AL interface is part of, and fully compatible with the Fibre Channel standard, optical cabling can be used in any portion of a subsystem (with the exception that optical signals, of course, cannot be used on a backplane) [3]. This makes it possible to have a disk subsystem a significant distance from the computer system to which it is attached. For example, using single-mode fibre optics, on-line disk storage could be as far as 10 km away from its host. A computer system could have disks within the system connected via a Fibre Channel loop. That loop can be extended by using an electronic-to-optic signal adapter and lengths of fibre optic cable. On the same loop running the internal disks, remote disks would appear to the system to be directly attached exactly like the local disks, even though they might be five miles away. This can be an advantage in many ways, including the shadowing the local disks in the event a disaster destroys the on-site data banks. That capability, in particular, offers an attractive means for having a remote and secure on-line copy of critical data that could be used to continue operations should anything happen to the facility housing the primary computer system.

Array Implementations

Array controllers have traditionally been designed with multiple SCSI interfaces for drive attachment [3]. This enables the controller to supply data and I/O rates equal to several times those achievable from a single interface. This is sometimes referred to as an *orthogonal array*, because the disks comprising the arrays elements, are across, or orthogonal to, the SCSI controllers. Unfortunately, the decision to design a specific number of drive interfaces into a given controller forces on the customer the parity amortization, granularity, and controller cost associated with that decision. This limits the customer's choices for configuring the optimal combination of economy—that is, maximizing the *granularity*, or number of data drives per parity drive, the total capacity per array, and overall performance. With Fibre Channel, it is possible to configure an array along a single interface instead of across many. Because the drives constituting the array unit are organized along an interface, it is sometimes described as a *longitudinal array*.

9.1.4c Server-Based Audio/Video Editing

Audio/video editing is mostly a cutting and pasting process, and as such, lends itself to synergistic applications involving the server [6]. The digital recorder enabled users to re-record the same video several times with minimal generation loss. Digital recording also offered the ability to read-before-write. With tape, this required judicious care, because the underlying video track was erased by the subsequent one. Disk-based systems, either optical or magnetic hard-disk drive designs, also offer this capability, but allow nondestructive read-out.

Server-based editing resembles digital disk-based editing, except that files from one user can be instantly available to another. Facilities need only to endure the time penalty of transfer and digitizing (as necessary) the input source material once. Thereafter; all potential users can access the material simultaneously. For example, assume an important piece for the five o'clock news is being produced edit room "A." Via the server, the producer in edit room "B" can start putting together the same story for the six o'clock news, accessing the same digitized elements.

Not every facility puts out back-to-back newscasts. Many that do, reuse stories with minimal updating. There are, however, a number of facilities that not only produce multiple newscasts, but provide separate news programming to cable channels or that sell news to other stations. As

Internet broadcasting increases in importance, it will begin to consume editing resources as well. With this many hands fighting for a field tape, conflict is inevitable. A unique economic advantage of the server-based facility is that a multitude of users can cherry pick off the main storage system, without disrupting the work flow of other operation.

9.1.4d Perspective on Storage Options

The professional audio/video industry has become accustomed to an ever increasing number of tape formats, each targeting a particular market segment. Disk storage of audio and video has been an effective refuge from such “format wars.” This is not to say that disk stores do not vary greatly. They do. The only standardization from vendor to vendor is their interface: SCSI, SCSI-1, SCSI Wide, Fiber Channel, IDE, IEEE 1394, and so on. Because audio/video disk interchange is not a major requirement today, there is no specific need to define the internal format. With no standards of measurement except the ultimate performance of the system to fall back upon, it is important to carefully examine the system architecture to be certain that it addresses the requirements for performance and reliability for a given application.

9.1.5 References

1. Robin, Michael, and Michel Poulin: “Multimedia and Television,” in *Digital Television Fundamentals*, McGraw-Hill, New York, N.Y., pp. 455–488, 1997.
2. McConathy, Charles F.: “A Digital Video Disk Array Primer,” *SMPTE Journal*, SMPTE, New York, N.Y., pp. 220–223, April 1998.
3. Whitaker, Jerry C.: “Data Storage Systems,” in *The Electronics Handbook*, Jerry C. Whitaker (ed.), CRC Press, Boca Raton, Fla., pp. 1445–1459, 1996.
4. Portions of this section based on background information provided by Micropolis, Chatsworth, Calif., 1996.
5. Tyson, H: “Barracuda and Elite: Disk Drive Storage for Professional Audio/Video,” Seagate Technology Paper #SV-25, Seagate, Scotts Valley, Calif., 1995.
6. Lehtinen, Rick, “Editing Systems,” *Broadcast Engineering*, Intertec Publishing, Overland Park, Kan., pp. 26–36, May 1996.

9.1.6 Bibliography

- Anderson, D: “Fibre Channel-Arbitrated Loop: The Preferred Path to Higher I/O Performance, Flexibility in Design,” Seagate Technology Paper #MN-24, Seagate, Scotts Valley, Calif., 1995.
- Goldberg, Thomas: “New Storage Technology,” *Proceedings of the Advanced Television Summit*, Intertec Publishing, Overland Park, Kan., 1996.
- Heyn, T.: “The RAID Advantage,” Seagate Technology Paper, Seagate, Scotts Valley, Calif., 1995.

9-26 Audio Recording Systems

Plank, Bob: "Video Disk and Server Operation," *International Broadcast Engineer*, September 1995.

Chapter
9.2

Properties of Magnetic Materials

Beverley R. Gooch

9.2.1 Introduction

The performance of a magnetic tape recorder depends heavily on the properties of the magnetic materials used to make the recording heads and tapes. Today's magnetic materials are the product of sophisticated metallurgy and advanced manufacturing techniques, which in large measure are responsible for the advancement of the magnetic recording technology.

Magnetic materials are classified as either magnetically *hard* or magnetically *soft*. Both types are used in magnetic tape recorders.

The hard magnetic materials are so-called because of their ability to retain magnetism after being exposed to a magnetic field. The measure of this property is called *remanence*. These materials may be further characterized by high coercivity and low permeability. *Coercivity* is the resistance of the material to being magnetized or demagnetized. *Permeability* is a measure of the magnetic conductivity relative to air.

In magnetic recording, hard magnetic materials are used chiefly in the manufacturing of recording tape and other related media. Some examples are gamma ferric oxide (γ -ferric oxide), iron oxide, and chromium dioxide. Hard materials are also used to make permanent magnets for use in loudspeakers, electric motors, and other applications.

On the other hand, soft magnetic materials such as Alresil, hot-pressed ferrite, and Permalloy exhibit low coercivity, low remanence, and relatively high permeability. These materials are used to make cores for magnetic heads.

Ferromagnetic materials have permeabilities much greater than unity and show a strong magnetic effect. Ferromagnetism is exhibited mostly by metallic elements such as iron, cobalt, nickel, and magnetic metals that are alloys of these elements. With the exception of ferrites [1, 2], most magnetic materials used in tape recorders are ferromagnetic.

Paramagnetic substances have permeabilities that lie between 1.000 and 1.001. These materials do not show *hysteresis*, and their permeabilities are independent of field strength. Some examples of paramagnetic materials are sodium, potassium, oxygen, platinum, and ferromagnetic metals above the *Curie temperature* [1].

Diamagnetic materials have a relative permeability slightly less than 1. Many of the metals and most nonmetals are diamagnetic [1].

Magnetic anisotropy is the term applied to magnetic materials that exhibit preferred directions of magnetization. These preferred and nonpreferred directions are referred to as the *easy*

9-28 Audio Recording Systems

Table 9.2.1 Properties of Soft and Hard Magnetic Materials

Material	M_s , G	$B_s = 4\pi M_s$, G	H_c , O	B_r , G	μ (dc) initial	Resistivity, $\Omega \cdot \text{cm}$	Thermal expn.	Curie temp.	Vicker hardne
Soft magnetic materials									
Iron Fe	1700	21,362	1	20,000				
Hi-Mu 80 80% Ni, 20% Fe	661	8,300	0.02	50,000	65	12.9×10^{-6} cm/(cm°C)	733 °C	127
Alfesil (Sendust) 85% Fe, 6% Al, 9% Si	796	10,000	0.06	10,000	90	11.3×10^{-6} cm/(cm°C)	773 °C	496
Mn Zn, Hot-pressed ferrite	358	4,500	0.02–0.2	≈900	2000–5000	10^4	$10\text{--}15 \times 10^{-6}$ cm/(cm°C)	100–300 °C	650–75
Ni Zn, Hot-pressed ferrite	238	3,000	0.15–3	≈1800	100–2000	10^{10}	$7\text{--}9 \times 10^{-6}$ cm/(cm°C)	150–200 °C	700–75
Hard magnetic materials									
Squareness ratio									
γ-Ferric oxide	400	5,026	300–350	1300†	0.75†				
Chromium dioxide	470	6,000	300–700	1600†	0.9†				
Metal particles	800	10,000	1000	3500†	0.8†				

†Value typical for finished tape.

and *hard* axes of magnetization, respectively. The higher the magnetic anisotropy, the harder it is to change the magnetization away from the preferred direction. In most polycrystalline materials, the crystals are randomly oriented and are magnetically isotropic. Single crystal ferrites and magnetic particles used in tape coating are examples of magnetic materials that are anisotropic [1, 3].

Table 9.2.1 shows properties of materials commonly used in magnetic heads and tapes.

9.2.2 Basic Principles of Magnetism

Magnetism results from two sources: orbital motion of electrons around the nucleus and the spinning of the electrons on their own axes (see Figure 9.2.1). Both the orbital and spin motions contribute to the *magnetic moment* of the atom, although in most magnetic substances almost all the magnetic moment is due to the spin motion. As the electron spins on its axis, the charge on its surface moves in a circular pattern. This moving charge, in turn, produces a current that creates a magnetic field. This phenomenon occurs in all substances. However, the electrons of the atoms in nonmagnetic materials occur in pairs with the spins in opposite directions, balancing each other and rendering the atom magnetically neutral. The atoms can produce the external effect of a magnet only when the electron spins are unbalanced.

The iron atom, for example, has 26 electrons in rotation around its nucleus (Figure 9.2.1). These orbiting electrons occur in regions called *shells*. According to quantum theory, the maximum number of electrons that can exist in each shell is $2N^2$, where N is the number of the shell. Starting from the nucleus, the first, second, third, and fourth shells could have a maximum num-

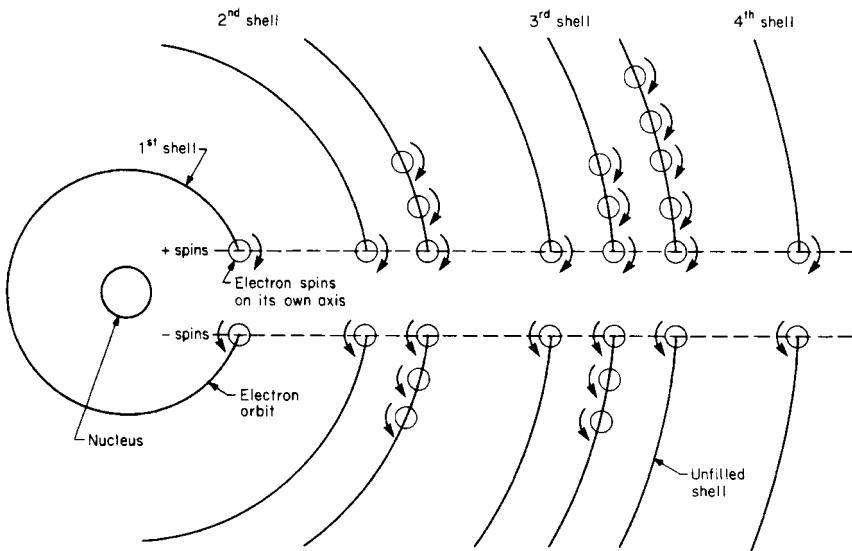


Figure 9.2.1 Schematic diagram of an iron atom.

ber of 2, 8, 18, and 32, respectively. The maximum number of electrons in each shell may not be reached before the next shell begins to form. The iron atom actually has two electrons in the first shell, the second has eight, the third and fourth shells have fourteen and two, respectively. The plus and minus signs show the direction of the electron spins. The electron spins in the first, second, and fourth shells balance each other, and produce no magnetic effect. It is the third shell that is of particular interest in the iron atom. In this shell there are five electrons with positive spins and one with a negative spin, which gives the atom a net magnetic effect.

Thermal agitation energy, even at low temperatures, would prevent the atomic magnets from being aligned sufficiently to produce a magnetic effect. However, powerful forces hold the electron spins in tight parallel alignment against the disordering effect of thermal energy. These forces are called *exchange forces*.

The parallel alignment of the electron spins, due to the exchange forces, occurs over large regions containing a great number of atoms. These regions are called *domains*. Each domain is magnetized to saturation by the aligned electron spins. Because this magnetization occurs with no external field applied, it is referred to as *spontaneous magnetization*. When the magnetic material is in the demagnetized state, the direction of the magnetization of the saturated domains is distributed in a random order, bringing the net magnetization of the material to zero. The domains are separated from each other by partitions called *Bloch walls* [1, 3]. The domain wall pattern is determined by the strains within the material and its composition.

In soft magnetic materials the magnetization takes place by the displacement of the domain walls [1, 3]. The wall movement is not continuous but occurs in discrete steps called *Barkhausen*

9-30 Audio Recording Systems

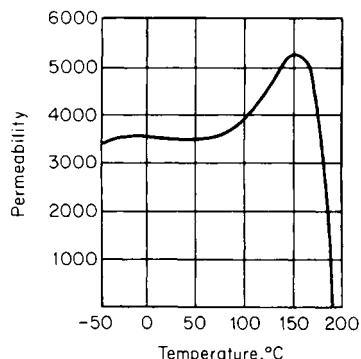


Figure 9.2.2 Effect of temperature on the permeability of typical ferrite.

steps or jumps that are related to imperfections or inclusions in the crystalline structure of the material.

The particles used in magnetic tape coating are so small that Bloch walls do not form. They behave as single-domain particles that are spontaneously magnetized to saturation. Irreversible magnetization is achieved only through irreversible rotation of the individual particle magnetizations [4, 5].

9.2.2a Curie Point

The Curie point is the temperature at which the thermal agitation energy overcomes the exchange forces. The spontaneous magnetization disappears and the material is rendered nonmagnetic. This process is reversible; when the temperature is lowered below the Curie point, the spontaneous magnetization returns and the material is again magnetic. Figure 9.2.2 shows the effect of temperature on the permeability of a typical ferrite.

9.2.2b Magnetic Induction

When a current I is connected to a solenoid coil of N turns, a magnetic field H is created that has direction as well as strength, and is defined by:

$$H = \frac{0.4\pi NI}{l} \quad (9.2.1)$$

Where:

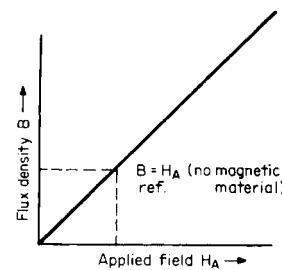
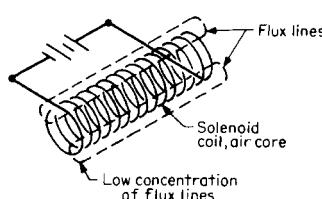
H = magnetic field in oersteds

l = length of the solenoid in centimeters

I = current in amperes

As a result of the field H , flux lines are produced in the surrounding space (see Figure 9.2.3). The flux lines form closed loops that flow from one end of the solenoid coil, into the air, and

(a)



(b)

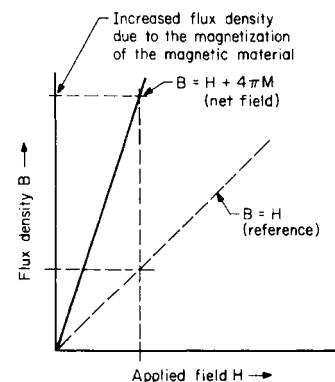
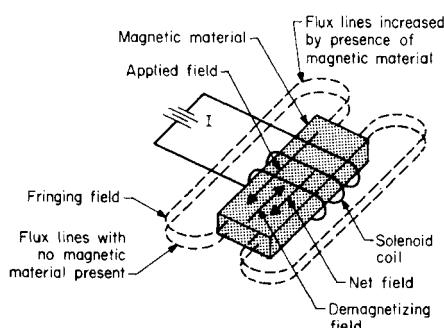


Figure 9.2.3 Properties of flux increase: (a) air core, (b) magnetic core material.

reenter the coil at the opposite end. The measure of the intensity or the concentration of the flux lines per unit area is called the *flux density*, or the *induction B*.

Figure 9.2.3a shows that with no magnetic material present in the solenoid coil the flux density B is relatively low and is equal to the applied field H . When a piece of magnetic material is placed in the solenoid coil, the flux density is increased (Figure 9.2.3b). This results from the magnetic moments of the electron spins aligning themselves with the applied field H , causing the magnetic material to become a magnet [1, 6]. The sum of the magnetic moments per unit volume is the magnetization M . The magnetization of a material creates magnetic fields. Inside the material these fields are called *demagnetization fields* because they oppose the magnetization. Outside the material, they are called *stray* or *fringing fields*. The net field acting on the material is the vectorial sum of the demagnetization field and the applied field. The flux density is the net field plus the magnetization M , that is:

$$B = H + 4\pi M \quad (9.2.2)$$

where H = net field and M and B are in gauss

9-32 Audio Recording Systems

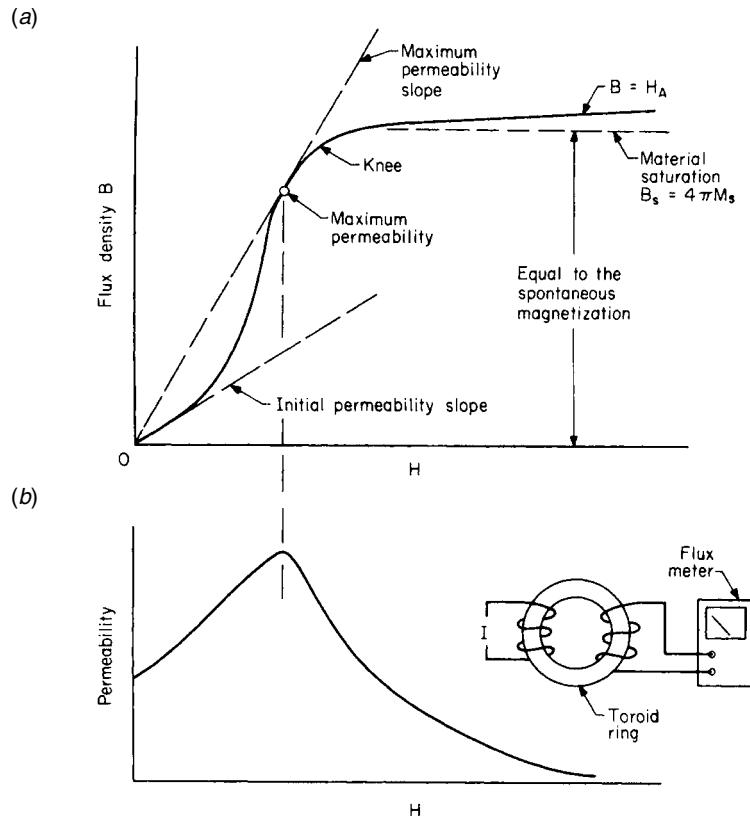


Figure 9.2.4 Permeability effects: (a) initial B - H curve, (b) permeability versus H field.

Initial Magnetization B - H Curve

The relationship of the induced flux density B and the net field H of soft magnetic material is typically described by the initial B - H magnetization curves and the B - H hysteresis loop.

Figure 9.2.4a shows the initial magnetization curve of a typical soft magnetic material. This curve is obtained by starting with a toroid ring in the demagnetized state and plotting the flux density B against the field H . The demagnetization field in a toroid ring is zero; the net field is therefore equal to the applied field. The slope of the initial magnetization curve is the permeability μ , defined by:

$$\mu = \frac{B}{H} \quad (9.2.3)$$

In CGS units, the permeability is a dimensionless ratio and represents the increase in flux density relative to air caused by the presence of the magnetic material. The permeability can also be defined in terms of the magnetization M as:

$$\mu = 1 + \frac{4\pi M}{H} \quad (9.2.4)$$

Starting at the origin, the curve has a finite slope which is the initial permeability. As the field H is increased, the slope becomes steeper. This is the maximum *permeability region*. The value of the maximum permeability is determined with a straight line of the steepest slope that passes through the origin and also contacts the magnetization curve. Finally, as H is further increased, a point is reached on the initial B - H curve where the magnetization approaches a finite limit indicated by the dotted line. At this point, the magnetization of the material does not increase with further increases in the field. This is the saturation flux density B , which is equal to the spontaneous magnetization of the magnetic material. After the material has reached saturation, the slope of the B - H curve changes and the flux density B continues to rise indefinitely at the rate of $B = H_A$ as if the magnetic material were not present. Figure 9.2.4b shows a plot of the permeability as a function of the field.

9.2.2c Hysteresis Loop

If the H field is decreased after the initial magnetization curve reaches the saturated state, it is found that the induction does not follow the same initial curve back to the origin but traces a curve called the *hysteresis loop*, shown by Figure 9.2.5. As the magnetization is gradually decreased from the saturation point C , it follows along the lines CD and reaches a finite value B_r (the *remanence*), which is the flux density remaining after removal of the applied field. In order to reduce the remanence to zero, a negative field—the coercive force H_c —must be applied. The curve from D to E is the demagnetization curve. As H is further increased in the negative direction, the magnetization will proceed from E to F , and the material will eventually become saturated in the opposite direction. If at this point the field is again reversed to the positive direction, the magnetization will trace the line F, G, C and the hysteresis loop is completed.

Hysteresis Losses

The area of the hysteresis loop is the energy necessary to magnetize a magnetic substance. This energy is expended as heat. The loop area is a measure of the heat energy expended per cycle, per unit volume, and is called the *hysteresis loss*:

$$W_h = \frac{A}{4\pi} \text{ ergs}/(\text{cm}^3 \cdot \text{cycle}) \quad (9.2.5)$$

A practical expression for power loss P in watts is given by:

$$P = \frac{f a l}{4\pi} \times A \times 10^{-7} \quad (9.2.6)$$

9-34 Audio Recording Systems

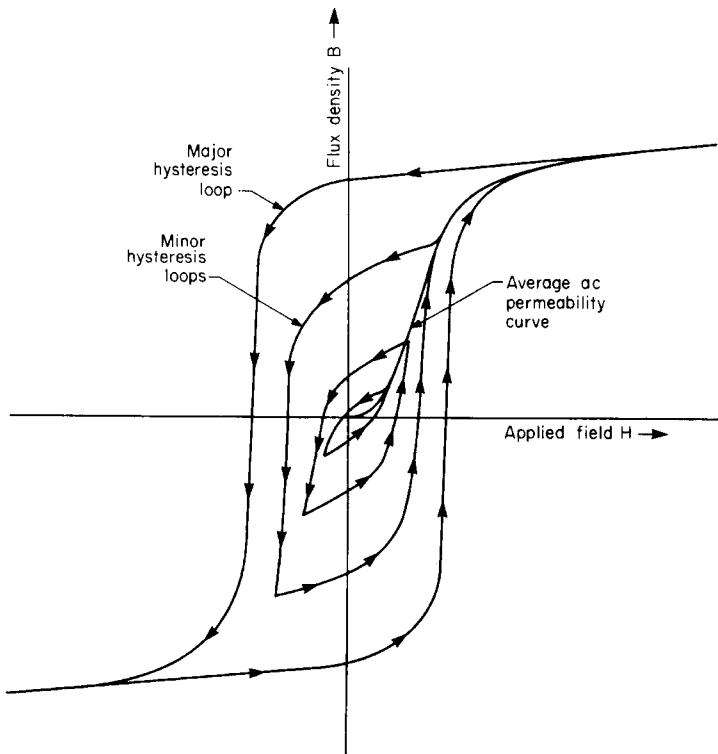


Figure 9.2.5 *B-H* loops for hard and soft materials.

Where:

A = area of the loop, gauss-oersteds

f = frequency, Hz

a = cross-sectional area of core, cm^2

l = magnetic path length, cm

Figure 9.2.5 shows a comparison between the hysteresis loops for hard and soft magnetic materials. As indicated by the difference in the areas of the loops, more energy is required to magnetize the hard magnetic materials.

Initial *M-H* Curve and *M-H* Hysteresis Loop

The initial *M-H* curve and *M-H* hysteresis loops are plots of the magnetization M versus the net field H and are typically used to describe the intrinsic properties of hard magnetic materials such as those used in recording media. An initial *M-H* curve is shown in Figure 9.2.6. The slope of the *M-R* curve is the *susceptibility* x and is defined by:

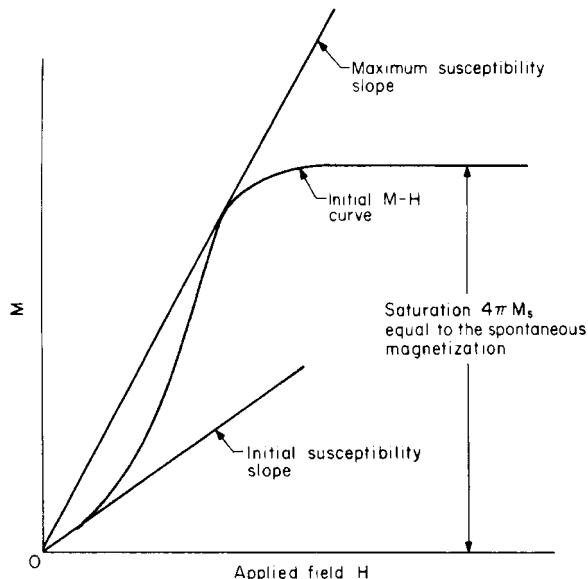


Figure 9.2.6 Initial M - H curve.

$$\chi = \frac{M}{H} \quad (9.2.7)$$

The permeability may be related to the susceptibility by:

$$\mu = 1 + 4\pi\chi \quad (9.2.8)$$

When the saturation magnetization M_s is reached, the M - H curve approaches a finite limit and does not increase indefinitely as in the case of the B - H curve.

If, at the saturation point of the initial M - H curve, the applied field is made to follow the same sequence as previously outlined for the B - H loop, an M - H hysteresis loop will be traced (see Figure 9.2.7).

The ratio of the remanent magnetization M_r to the saturation magnetization M_s is called the *squareness ratio* and is an important parameter in evaluating the magnetic orientation of the particles in magnetic tape. The squareness ratio is 1.0 for perfectly oriented particles. More practical values for oriented particles range from 0.7 to 0.9. Randomly oriented particles are approximately 0.5.

9.2.2d Demagnetization

If a short bar of magnetic material is magnetized by an applied field H , poles are created at each end. These poles in turn create a magnetic field in the opposite direction to the applied field.

9-36 Audio Recording Systems

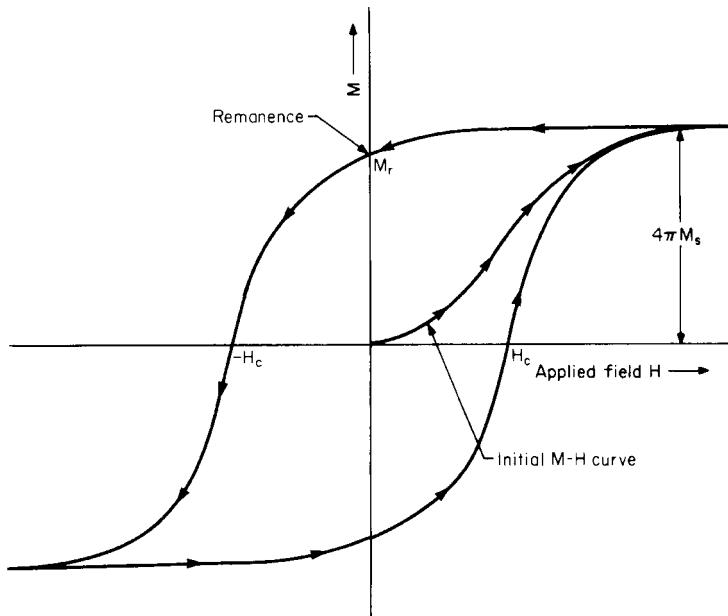


Figure 9.2.7 *M-H* hysteresis loop.

This opposition field is called the *demagnetizations field* H_d (see Figure 9.2.8). The net field H acting on the bar is:

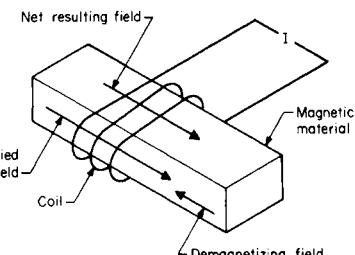
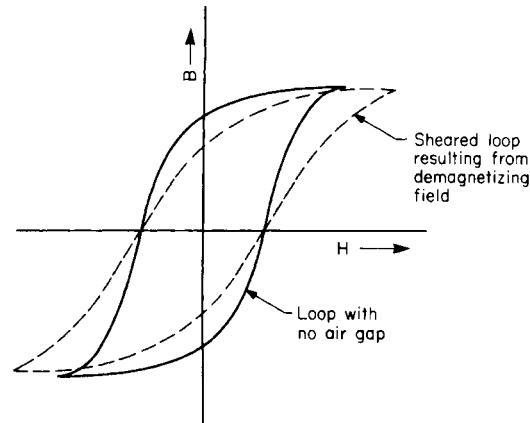
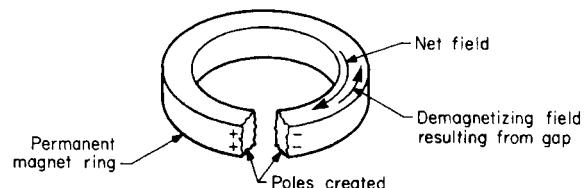
$$H = H_A - H_d \quad (9.2.9)$$

The demagnetizing field H_d is dependent on the shape of the magnetic object and the magnetization M [1, 3].

The demagnetization field is zero in a ring core with no air gap. However, when an air gap is cut, creating poles at the gap-confronting surfaces, the resulting demagnetization field shears the hysteresis loop from the original position. This effect is shown by Figure 9.2.9.

To bring a magnetic substance to a demagnetized state, a field that is equal to the coercive force H_c must be applied. However, upon removal of this field, the residual flux density will rise to a value B_1 , as illustrated by Figure 9.2.10. It is possible to reduce this residual flux density to zero by increasing demagnetization field to a value greater than H_c and then decreasing it to zero as shown by the dashed lines. This technique requires knowledge of the magnetic history of the material.

A more effective method to completely demagnetize a magnetic material is demagnetization by reversals. In this method, the material is first saturated by an ac field, then cycled through a series of diminishing field reversals as shown by Figure 9.2.11. The magnetic material will be left in a demagnetized state when zero field is reached regardless its magnetic history. This tech-

**Figure 9.2.8** Demagnetization field.**Figure 9.2.9** Sheared hysteresis loop.

nique is used to bulk erase magnetic tape and other recording media, by exposing it to a strong ac field and then slowly removing the magnetic media from the field.

9.2.3 References

1. Chikazumi, Soshin: *Physics of Magnetism*, Wiley, New York, N.Y., 1964.
2. Smit, J., and H. P. J. Wijn: *Ferrite*, Wiley, New York, N.Y., 1959.
3. Bozorth, Richard M.: *Ferromagnetism*, Van Nostrand, Princeton, N.J., 1961.
4. Pear, C. B.: *Magnetic Recording in Science and Industry*, Reinhold, New York, N.Y., 1967.

9-38 Audio Recording Systems

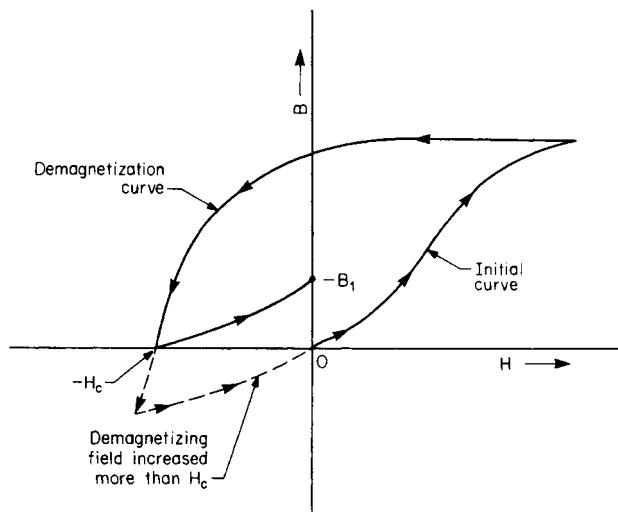


Figure 9.2.10 Demagnetization curve.

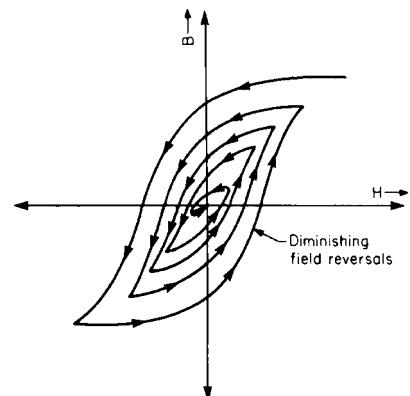


Figure 9.2.11 Demagnetization by reversals.

5. Jorgensen, Finn: *The Complete Handbook of Magnetic Recording*, Tab Books, Blue Ridge Summit, Pa., 1980.
6. Kraus, John D.: *Electromagnetics*, McGraw-Hill, New York, N.Y., 1953.

Chapter
9.3

Recording Fundamentals

Charles P. Ginsburg, Beverley R. Gooch, H. Neal Bertram

9.3.1 Introduction

The basic elements of a magnetic tape recorder are illustrated in Figure 9.3.1. A magnetic tape is moved in the direction indicated by a tape-drive device or *transport*. The magnetic coating of the tape contacts the magnetic heads in a prescribed sequence, starting with the erase head and ending with the reproduce head.

The erase head demagnetizes the tape coating by exposing the magnetic particles to a high-frequency field that is several times greater in strength than the coercivity of the particles. As the tape is drawn past the erase head, the erasing field gradually decays, leaving the magnetic coating in a demagnetized state.

The tape then moves into contact with the record head, which consists of a ring-shaped core made of a relatively high-permeability material, and having a nonmagnetic gap. A magnetic field fringes from the gap, varying in accordance with the magnitude of the current signal flowing in the head coil. With low-level signals the field is small, and some magnetic particles in the tape coating will be forced into alignment with the field. As the signal field is increased, a larger number of particles will become oriented in the direction of the recording field. As the tape is moved past the record gap, the magnetic coating acquires a net surface magnetization having both magnitude and direction. This magnetization is a function of the recording field at the instant the tape leaves the *recording zone*, a small region in the vicinity of the trailing edge of the gap.

The magnetization of the fundamental recording system just described is not necessarily linear with respect to the head current. Linear magnetization can be achieved by adding a high-frequency ac bias current to the signal current. Audio recorders use such a scheme to linearize the tape and reduce the distortion. In video recorders, the signal information in the form of a frequency-modulated carrier is recorded directly, without ac bias.

9.3.2 Fundamental Principles

When the tape approaches the nonmagnetic gap of the reproduce head, the flux $d\Phi/dt$ from the magnetized particles is forced to travel through the high-permeability core to link the signal windings and produce an output voltage. The output voltage is proportional to $d\Phi/dt$, the rate of

9-40 Audio Recording Systems

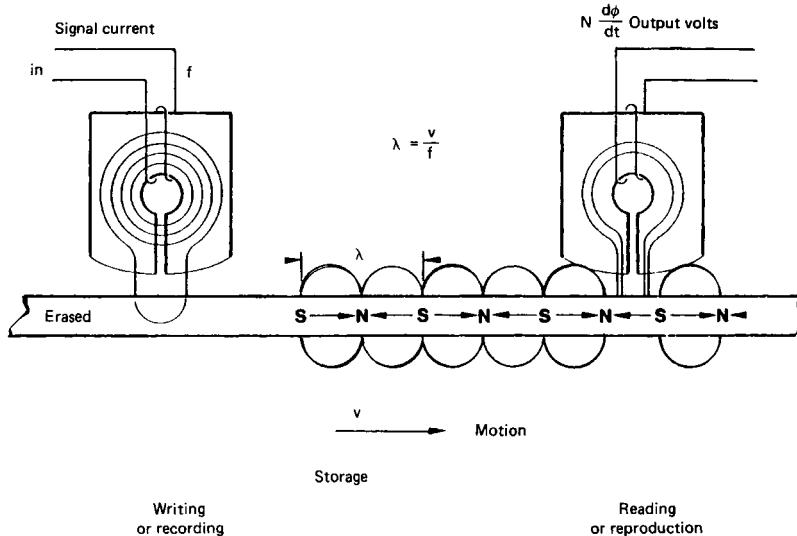


Figure 9.3.1 Fundamental recording and reproduction process.

change of the induced flux, and therefore will rise at the rate of 6 dB per octave until a wavelength is reached where the gap and spacing losses begin to reduce the head output.

9.3.2a Recording Signal Parameters

The physical distance that one cycle of the recorded signal occupies along the tape is called the *wavelength*, which is directly proportional to the relative velocity between the head and the tape, and inversely proportional to the frequency of the recorded signal. It may be expressed as:

$$\lambda = \frac{v}{f} \quad (9.3.1)$$

Where:

λ = wavelength, in

v = velocity, in/s

f = frequency, Hz

The *linear packing density* is the number of flux reversals per unit length along the recording medium. Because there are two flux reversals, or bits, per cycle, the linear packing density may be expressed as:

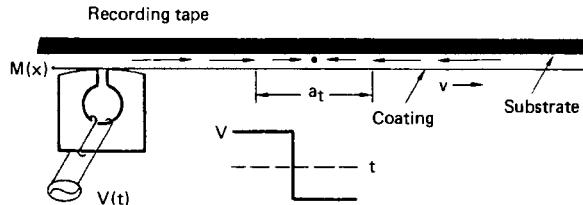


Figure 9.3.2 Resultant magnetization from a step change in head voltage.

$$\frac{\text{bits}}{\text{in}} = 2 \frac{2}{\lambda} = \frac{2}{\lambda} \quad (9.3.2)$$

The *area packing density* is the number of bits per unit area and is, therefore, equal to the number of recorded tracks per inch times the linear packing density, or:

$$\frac{\text{bits}}{\text{in}^2} = \frac{2}{\lambda} \frac{\text{tracks}}{\text{in}} \quad (9.3.3)$$

When the magnetization is oriented in the direction of relative motion between the head and tape, the process is referred to as *longitudinal recording*. If the magnetization is aligned perpendicular to the surface of the tape, it is called *vertical* or *perpendicular* recording.

Transverse recording exists when the magnetization is oriented at right angles to the direction of relative head-to-tape motion. From these definitions, longitudinal magnetization patterns are produced by both rotary- and stationary-head recorders.

9.3.2b The Recording Process

The recording process consists of applying a temporally changing signal voltage to a record head as the tape is drawn by the head. The magnetic field that results from the energized head records a magnetization pattern that spatially approximates the voltage waveform. In *saturation* or *direct recording*, the signal consists of polarity changes with modulated *transition times* or zero crossings. Strict linear replication of this signal is not required because the information to be recovered depends only on a knowledge of when the polarity transitions occur. Examples are digital recording, where the transitions are synchronized with a bit time interval and occur at bit positions depending upon the coded pattern, and FM video recording, where a modulated sine wave is applied so that the transitions occur not regularly but according to the signal information contained in the modulation.

The essential process in direct recording, therefore, is the writing of a transition or polarity change of magnetization. In Figure 9.3.2, the resulting magnetization from a step change in head voltage is shown. In saturation recording, the spatial variation of magnetization will not be a perfect replica of the time variation of signal voltage. Even if the head field change is perfectly abrupt, the magnetization will gradually change from one polarity to another. In Figure 9.3.2, this is indicated by a gradual change in vector lengths; the notation a_t denotes an estimate of the distance along tape over which the magnetization reverses. The nonzero distance between polar-

9-42 Audio Recording Systems

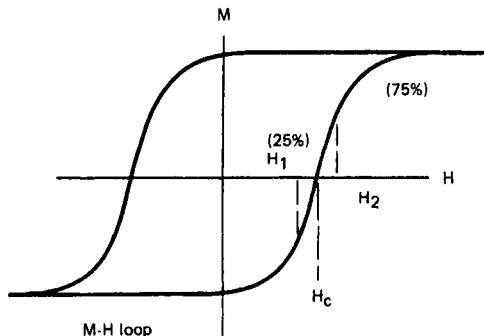


Figure 9.3.3 M - H hysteresis loop for a typical tape.

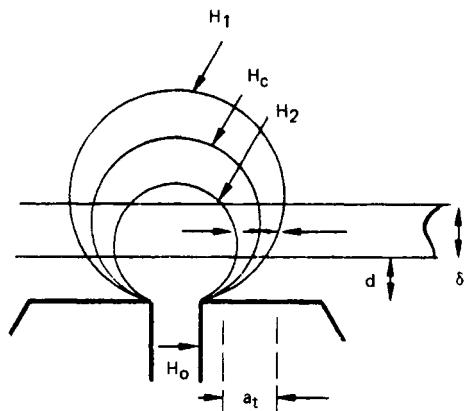


Figure 9.3.4 Head-field contours showing recording zone in tape.

ity changes of magnetization is due to the finite-loop slope at the coercivity, combined with the gradual decrease of the head field away from gap center. This process is illustrated in Figures 9.3.3 and 9.3.4. In Figure 9.3.3 the M - H remanence loop is shown for a well-oriented tape sample. The magnetization M is the remanence magnetization that results from the application of a field H , which is subsequently removed. If the tape is saturated in one direction, for example, $-M$, and a positive field is applied, then the magnetization will start to switch toward the positive direction when the field is close to the remanent coercivity H_c' . Because the slope of this M - H loop is not infinitely steep for fields near H_c' , the switching will take place gradually. H_1 denotes the field that switches 25 percent of the particles to leave the magnetization at $-M/2$; H_c' is, in fact, the 50 percent reversing field that leaves $M = 0$. H_2 denotes the 75 percent switching field that leaves the magnetization halfway to positive saturation ($+M/2$). During recording, a finite transition width will occur, depending on how H_1 and H_2 are spatially separated. In Figure 9.3.4, three contours of recording field are plotted for the three fields H_1 , H_c' , and H_2 . In plots of head fields, larger fields are closer to the surface of the head and toward the gap center. Thus, along the midplane of the tape the field magnitudes H_2 , H_c' , H_1 are in decreasing order away from the gap centerline. Therefore, if the tape is initially magnetized negatively, a positively energized head (H_0) will switch the magnetization according to Figure 9.3.3 following the spatial change of

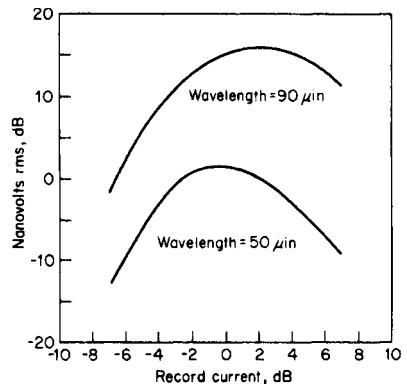


Figure 9.3.5 Reproduce voltage versus record current for typical video tape.

the fields. This yields a finite transition width a_r . The transition width can be narrowed by using tape with a steeper loop gradient, making H_1 closer to H_2 in magnitude (a narrower spread in switching fields) and decreasing the head-to-tape spacing, which moves H_1 and H_2 closer together spatially (a larger head field gradient) as indicated in Figure 9.3.4. In addition, spatial changes in magnetization cause demagnetization fields in the tape that further broaden the transition. This demagnetization broadening can be reduced by increasing the tape coercivity.

In saturation recording, the signal current is held fixed for all wavelengths. The current level is set to optimize the short-wavelength output, and complete saturation of the tape does not occur. In Figure 9.3.5 reproduce voltage versus input current is shown for two different wavelengths in square-wave recording for video tape (Ampex 196). If true saturation were to occur, the curves would increase initially with current as the tape is recorded, and then level, representing a magnetization saturated to full remanence and recorded fully through the tape thickness of 200 μin (5 μm). However, at short wavelengths these curves are peaked, and the current that yields the maximum output represents recording only a very small distance into the tape. For video recording on a type C format machine optimized at 10 MHz ($\lambda \approx 100 \mu\text{in}$), this is a record depth of approximately 50 μin . A mechanism for optimization of this parameter can be seen by considering the change in transition with record current. As the current is raised, the point of recording shifts continuously downstream from the gap center. The transition width depends upon the head-field gradient at the recording point. This field derivative, $H_2 - H_1$, divided by the separation between them, increases with distance along the head surface, as shown in Figure 9.3.4, reaching a maximum near the gap edge and thereafter decreasing. Because the reproduce voltage increases with decreasing transition width, a maximum voltage will occur as the current is increased. This peaking becomes more pronounced as the wavelength is reduced.

A form of linearity known as *linear superposition* is found in saturation recording. For constant-current recording (strictly, *constant-field amplitude*) the reproduce voltage from a complicated pattern can be shown to closely resemble the linear superposition of isolated transition voltage pulses, according to the timing and polarity change of the series of transitions. The lack of complete linear superposition is believed due to demagnetization fields. This accompanies large head-to-medium separations, as in rigid-disc applications, where the increase in the demagnetization fields can cause significant nonlinearities.

9-44 Audio Recording Systems

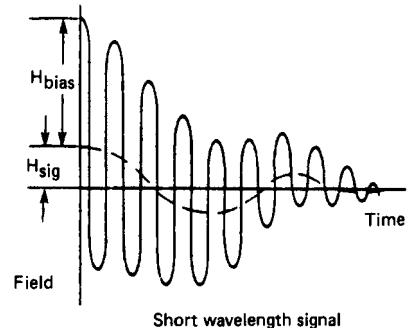


Figure 9.3.6 Signal-field history for ac-bias recording.

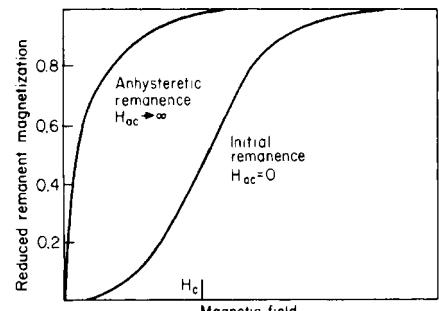


Figure 9.3.7 Comparison of sensitivities of ac-bias and direct recording.

9.3.2c Bias Recording

In some applications, predominantly audio recording, strict linearity is required between the reproduced voltage and the input signal. This may be achieved by superimposing a high-frequency, large-amplitude bias current on the signal (Figure 9.3.6). The physical process is called *anhysteresis* [1]. The bias field supplies the energy to switch the particles while the resulting remanent magnetization is a balance between the signal field and the interparticle magnetization interactions. In Figure 9.3.7 a comparison is shown between the magnetic sensitivity of ac bias recording and direct recording. Alternating-current bias or anhysteresis results in an extremely linear characteristic with a sensitivity an order of magnitude greater than that for unbiased recording. In typical audio applications, the bias current is somewhat greater than that of the signal in direct recording. The signal current is approximately an order of magnitude less than the bias current and is set to maintain the harmonic distortion below 1 to 2 percent. For complete anhysteresis there should be many field reversals as the tape passes the recording point where the bias field equals the tape coercivity. This is achieved if the bias wavelength is substantially less than the record-gap length. In fact, to avoid reproducing the bias signal itself, the bias wavelength is usually less than the reproduce-gap length.

As in direct recording, a current optimization occurs, but in bias recording it is with respect to the bias current. In Figure 9.3.8, reproduce voltage is shown versus bias at short and long wavelengths. At long wavelengths, the optimum bias occurs approximately when the bias field has recorded through to the back of the medium. This is often taken to be the usable bias current

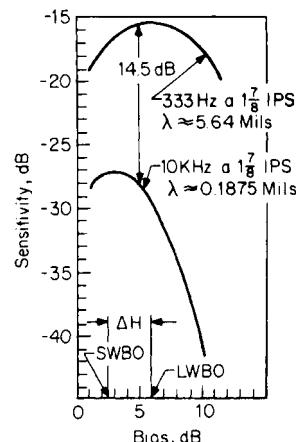


Figure 9.3.8 Low-level output sensitivity versus bias for long and short audio wavelengths.

since close to this optimization a minimum in the distortion occurs. For shorter-wavelength machines, the bias is chosen as a compromise between *short-wavelength bias optimization* (SWBO) and *long-wavelength bias optimization* (LWBO).

9.3.2d Particle Orientation

The previous discussion applies generally to all types of recording tape; the most common is that composed of uniaxial elliptical particles oriented *longitudinally* in the direction of head-tape motion. However, *isotropic* tape does exist and is composed of particles of cubic (threefold) symmetry that exhibit high remanence in all directions. In addition, it is conceptually possible to *vertically* orient the grains to result in a tape isotropic in the plane, but capable of recording signals perpendicular to the surface. These last two would be advantageous for transverse recording since the difficult process of orienting elliptical particles along the tape cross direction could be avoided.

During the tape-coating process, elliptical particles will naturally orient along the tape-coating direction. A field applied during coating improves the orientation even further. It is quite difficult to orient these particles vertically because the hydroscopic coating forces overwhelm the magnetic force from a vertical-orienting field. Success has, however, been achieved with systems that *inherently* yield vertical orientation. As an example, barium-ferrite platelets have been successfully coated to yield perpendicular media since the magnetization anisotropy axis is perpendicular to the plane of the particles. Other vertical medium has been made by sputtering CoCr on either a tape or substrate.

9.3.2e Erasure

The writing of new information on previously recorded media requires that the previous information be completely removed. Erasure requirements, in terms of the previous-signal to new-sig-

9-46 Audio Recording Systems

nal ratio, vary from -30 dB for digital systems to as much as -90 dB for professional analog audio. Analog video recorders require about -60 dB.

Erasure is the ac-bias or anhysteretic process with zero-signal field. If a reel of tape is placed in a large ac field that is slowly reduced so that many field cycles occur when the field is near the coercivity, then complete erasure is easily obtained. In addition, the largest amplitude of the ac-erasing field must be sufficient to reverse at least 99.9 percent of the particles (for -60 dB), and in practice that field is about three times the coercivity.

Most tape recorders utilize erase heads to remove old information before recording new data. Similar to bias recording, the requirement of the erase frequency is that the wavelength be much less than the erase-head gap to provide sufficient reversals of the particles. However, one important problem occurs with an erase head. As the erase current increases, the erasure level does not continuously increase as more of the $M-H$ -loop tail is switched. There is an erasure plateau of about -40 dB for erase-gap lengths of 1 to 2 mils (0.0254 to 0.0508 mm) and tape thicknesses of 200 to 400 μ in (5.08 to 10.16 μ m). This leveling is believed to be due to the phenomenon of rerecording [2]. As the tape passes the erase head, the field from the portion yet to be erased (entering the gap region) acts as a signal for the bias-erase field to record a residual signal at the recording zone on the far side of the gap. This effect is seen only at long wavelengths where the field is high. The problem is eliminated with double erasure by using a double-gap erase head. The erasure level can be decreased by decreasing the ratio of the tape thickness (or recording depth) to erase-gap length; this reduces the rerecording field.

9.3.3 References

1. Bertram, H. N.: "Long Wavelength ac Bias Recording Theory," *IEEE Trans. Magnetics*, vol. MAG-10, pp. 1039–1048, 1974.
2. McKnight, John G.: "Erasure of Magnetic Tape," *J. Audio Eng. Soc.*, Audio Engineering Society, New York, N.Y., vol. 11, no.3, pp. 223–232, 1963.

9.3.4 Bibliography

Roters, Herbert C.: *Electromagnetic Devices*, Wiley, New York, N.Y., 1961.

Sharrock, Michael P., and D. P. Stubs: "Perpendicular Magnetic Recording Technology: A Review," *SMPTE J.*, SMPTE, White Plains, N.Y., vol. 93, pp. 1127–1133, December 1984.

Chapter
9.4
Magnetic Tape

Robert H. Perry

9.4.1 Introduction

Magnetic tape includes a multiplicity of products used for magnetic recording, all consisting of a magnetizable medium on a flexible substrate. Because of the great variety of machine types and recording formats in use and being developed, magnetic tape is designed and produced with widely different magnetic media, widths, thicknesses, lengths, and other properties optimized for each application. Media are used in either strip form in reels, cartridges, cassettes, and cards or in discs of different diameters. Similar technologies are used to produce all these products.

9.4.2 Basic Construction

Magnetic tape consists of a magnetic film or coating supported by a flexible substrate, or base film, which in many applications is coated on the back with a nonmagnetic coating (Figure 9.4.1).

Backcoatings are used primarily in the most demanding tapes, such as professional and some consumer video, professional audio, instrumentation, and data products, where special winding and handling characteristics are required. This coating contains a conductive pigment, usually carbon black, which reduces buildup of static charge and therefore minimizes the accumulation of dirt and debris on the tape, factors which can cause *drop-outs*, or loss of signal with attendant loss of stored information. The backcoating also provides better frictional characteristics than raw base film does, and air is more easily eliminated from adjacent layers during winding. This reduces the tendency of the tape to cinch or form *pop strands*, and there is less likelihood of uneven stacking, edge damage, and creasing of the tape.

The *base film* is an integral and significant part of the whole tape system and is largely responsible for its mechanical strength and stability. Other factors such as stiffness and surface smoothness have a profound influence on tape performance in many applications, and base films having the proper characteristics for a given application must be carefully selected.

The principal substance used in the great preponderance of magnetic tapes today is poly(ethylene terephthalate), or simply *polyester*; abbreviated PET. PET has an excellent combination of properties including chemical stability and mechanical properties, such as tensile strength, elon-

9-48 Audio Recording Systems

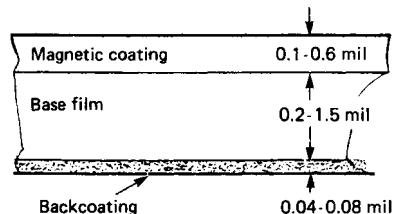


Figure 9.4.1 Magnetic-tape base-film and coating thickness.

Table 9.4.1 Physical Properties of Poly(ethyleneterephthalate) Base

Property	Balanced	Tensilized
Tensile strength, lb/in ² N/m ²	25,000 172.38×10^6	40,000 275.8×10^6
Force to elongate 5%, lb/in ² N/m ²	14,000 96.53×10^6	22,000 151.69×10^6
Elastic modulus, lb/in ² N/m ²	550,000 3.79×10^9	1,100,000 7.58×10^9
Elongation, %	130	40
Thermal coefficient of linear expansion, per deg C	1.7×10^{-5}	1.7×10^{-5}
Shrinkage at 100°C, % (per 30 min interval)	0.4	2.5

Note: Measurements in machine direction

gation and modulus, tear resistance, availability, and cost. Some typical properties are shown in Table 9.4.1.

Many different types and grades of PET are on the market for both magnetic tape-related and unrelated applications. In all cases, mechanical strength in the plastic film is achieved during its manufacture by a process of biaxial, and sometimes uniaxial, orientation of polymer chains in the hot film after extrusion of the melt. Biaxial orientation is achieved by stretching in both the machine and the transverse directions, and the resulting film has a balance of properties in the two directions. Balanced film is adequate for many magnetic tape applications, especially those employing gauge thicknesses greater than 0.5 mil (0.0127 mm). In thinner gauges greater resistance to stretching is needed, and PET is used that is *tensitized*, i.e., oriented by drawing additionally in the machine direction.

Base films for magnetic tape range in thickness from about 0.2 to 1.5 mils (3 mils for flexible disks). They are employed by tape manufacturers in widths ranging from 12 to 60 in (0.3 to 1.5 m) and in lengths up to 15,000 ft (4572 m). The base-film manufacturer must ensure that the base film has the right balance of surface smoothness for recording performance and roughness for runability in the coating and processing steps. Small-particle-size, inorganic additives are incorporated in the PET to provide slip properties in film that would be otherwise unmanageable. These surface asperities must be critically controlled, especially for short-wavelength

recording applications, since the base-film surface-roughness profile can be carried—to a degree—through the magnetic coating and reflected in the tape-surface roughness. An asperity of 10 μin , for example, in a typical 100- μin -wavelength video recording can result in a loss of signal of 5.5 dB due to head-to-tape separation, as seen from the Wallace formula:

$$\text{Spacing loss (dB)} = 54.6 \frac{d}{\lambda} \quad (9.4.1)$$

where d = the head-tape spacing and λ = the wavelength.

9.4.2a Magnetic Coating

There are two types of magnetic coatings used in magnetic tape. Most of them use magnetic particles bound in a matrix of organic, polymeric binder that is applied to the substrate from a dispersion in solvents. Other types are made by vapor deposition of thin films of metal alloys.

Most magnetic coatings contain a single layer, although some tapes are made with dual-layer magnetic coatings having different coercivities and are designed to have flat response over a range of frequencies. Magnetic-tape performance is a function of both the formulation of ingredients in the coating and the process by which the coating is applied and processed. The most important component in the formulation is the magnetic material itself.

Magnetic Materials

A wide variety of single-domain magnetic particles is used having different properties, depending on the electrical requirements of each tape application. Retentivities range from about 1000 to 3000 G, and coercivities range from about 300 to 1500 Oe. Size and shape are important because they relate to how well the particles pack in the coating; the signal-to-noise ratio achievable is proportional to the number of particles per unit volume in the coating. The length of the particles is about 8 to 40 μin (0.2 to 1.0 μm), and they are acicular with aspect ratios of 5/1 to 10/1. Acicularity makes the particles magnetically anisotropic, and thus it governs magnetic properties not inherent in the material. In general, magnetic pigments are loaded to as high a level as possible commensurate with retention of desirable physical properties and avoidance of shedding. The limiting factor is the amount of pigment the binder can retain without loss of cohesion and, hence, durability.

There are four basic types of magnetic particles used in tape:

- γ -ferric oxide
- Doped iron oxides
- Chromium dioxide
- Metallic particles, which usually consist of elemental iron, cobalt, and/or nickel

γ -ferric oxide has been by far the most widely used material (H_c 300 to 360 Oe) and is useful for many of the lower-energy applications in which the ultimate in recording density or short-wavelength recording capability is not required. The sequence of steps used in the commercial production of γ -ferric oxide is as follows:

9-50 Audio Recording Systems

- Precipitation of seeds of $\alpha\text{-FeOOH}$ (goethite) from solutions of scrap iron dissolved in sulfuric acid, or from copperas (ferrous sulfate obtained as a by-product from titanium dioxide manufacture)
- Growth of more goethite on the seeds
- Dehydration to $\alpha\text{-Fe}_2\text{O}_3$ (hematite)
- Reduction to Fe_3O_4 (magnetite)
- Oxidation to γ -ferric oxide (maghemite)

An improved γ -ferric oxide is produced starting with ferrous chloride rather than ferrous sulfate and precipitating $\gamma\text{-FeOOH}$ (lepidocrocite) rather than $\alpha\text{-FeOOH}$ in the initial step.

Cobalt doping of iron oxide affords particles with higher coercivities (500 to 1200 Oe). The older process involves precipitation of cobaltous salts with alkali in the presence of yellow iron oxide ($\alpha\text{-FeOOH}$), dehydration, reduction to cobalt-doped magnetite, and oxidation to cobalt-doped magnetite containing varying amounts of FeO. The resulting particles have cobalt ions within the lattice of the oxide, and they exhibit a marked magnetocrystalline anisotropy. This gives rise both to a strong temperature dependence of the coercivity and to magnetostrictive effects, which can cause problems of greatly increased print-through, increased noise, and loss of output resulting from stress on the tape through head contact. Improved stability can be achieved by using other additives, such as zinc, manganese, or nickel, with cobalt.

Epitaxial cobalt-doped particles can be used that largely overcome these problems because cobalt ion adsorption is limited to the surface of the oxide. Epitaxial particles have superseded lattice-doped particles in most applications.

Chromium dioxide provides a range of coercivities similar to that of cobalt-doped iron oxide (450 to 650 Oe) and possesses a slightly higher saturation magnetization, that is, 80 to 85 emu/g compared with 70 to 75 emu/g for γ -ferric oxide. It has uniformly good shape and high acicularity and lacks voids and dendrites, factors that account for the excellent rheological properties of coating mixes made with it. Its low Curie temperature (128°C) has been exploited in thermal contact duplication, a process which was largely developed in the late 1960s but because of problems in obtaining high-quality duplicates was not commercialized.

A disadvantage of chromium dioxide is its abrasiveness, which can cause excessive head wear. Also, it is chemically less stable than iron oxide, and under conditions of high temperature and humidity it can degrade to nonmagnetic chromium compounds, resulting in loss of output of the tape. Chromium dioxide and cobalt-doped iron oxide yield tapes having 5 to 7 dB higher S/N ratio than those made from γ -ferric oxide.

The presence of metallic particles results in tapes that have a 10 to 12-dB higher S/N ratio than those made from γ -ferric oxides because of their much higher saturation magnetization (150 to 200 emu/g), retentivity (2000 to 3000 G), coercivity (1000 to 1500 Oe), and smaller particle size. These factors, together with a square shape of the hysteresis loop, permit recording at shorter wavelengths with less self-erasure. Thus, recordings can be made at slower speeds without sacrifice in dynamic range, and higher bit-packing densities can be achieved.

Metallic particles are made by several different kinds of processes, the more important commercial ones being reduction of iron oxide with hydrogen, and chemical reduction of aqueous ferrous salt solutions with borohydrides. Metallic particles are more difficult to disperse than iron oxides because of their smaller size and higher remanence, and they are highly reactive. Processes such as partial oxidation of the surface or treatment with chromium compounds are used

Table 9.4.2 Physical Properties of a Poly(esterurethane)

Parameter	Value
Tensile strength, lb/in ² N/m ²	8000 55.16×10^6
Stress at 100% elongation, lb/in ² N/m ²	300 2.07×10^6
Ultimate elongation, %	450
Glass transition temperature, deg C	12
Hardness, shore A	76
Density, g/cm ³	1.17
Viscosity at 15% solids/tetrahydrofuran, cP	800
Pa· s	0.8

in their preparation to stabilize them for handling during tape manufacture. The corresponding tapes are more stable, but their susceptibility to corrosion at an elevated temperature and humidity is a disadvantage.

Magnetic-tape manufacturers have also developed products consisting of thin films (100 to 150 nm) of metal alloys deposited on the substrate under vacuum or by sputtering. The retentivity of these tapes (1.2×10^4 G) is almost an order of magnitude higher than that of γ -ferric oxides, with a corresponding increase in recording density.

In other areas, research is being devoted to very small, isotropic particles, which have aspect ratios of 1/1 to 2/1, because of advantages that can be taken of magnetization vectors in more than one direction, i.e., vertical as well as longitudinal recording, and because of the increased number of particles that can be packed in a coating per unit volume. Particles having the shape of rice grains have been used with success.

9.4.2b Binders

Binders must be capable of holding the magnetic pigment together in a flexible film that adheres to the base film with a high degree of toughness and chemical stability, and with thermoplastic properties enabling the pigmented film to be compacted to give smooth surfaces. It should also be soluble in suitable solvents. These requirements are not met by many substances available today for producing magnetic tapes.

Polyurethanes, either used as such or prepared in situ, represent the most important class of polymers for this purpose because of their affinity for pigments, their toughness and abrasion resistance, and their availability in soluble forms. Of the two types in use, poly(esterurethanes) are preferred over poly(etherurethanes) because of their superior mechanical properties in tape. Some physical properties of a typical poly(esterurethane) are shown in Table 9.4.2.

Other polymers can be used alone or in combination with one or two other polymers to obtain the desired properties. Although a great many types are claimed in the patent literature, the other

9-52 Audio Recording Systems

most important polymers include poly(vinyl chloride-co-vinyl acetate/vinyl alcohol), poly(vinylidene dichloride-co-acrylonitrile), polyesters, cellulose nitrate, and phenoxy resin.

Most magnetic-tape coatings are cross-linked with *isocyanates* to provide durability. Isocyanate-curing chemistry is rather complex and difficult to control, and for this reason the industry has explored curing with electron-beam radiation. A whole new field of binders has been developed for this purpose that polymerize extremely rapidly to high polymers in a much more controllable fashion.

Dispersants are surface-active agents that aid in the separation of magnetic particles, a process necessary for achieving the desired electrical performance of the tape. They facilitate separation of charges on the particles and stabilize particle separation. Common dispersants are lecithin, organic esters of phosphoric acid, quaternary ammonium compounds, fatty acids, and sulfosuccinates.

Conductive materials are often added to tape formulations to reduce electrostatic charge buildup on tape as it is run on machines. Conductive carbon blacks are commonly used to reduce the resistivity of tape by about four to six orders of magnitude.

Lubricants are necessary to prevent *stiction* of the tape as it comes in contact with the record or playback head. A great many different materials are effective as lubricants, including:

- Silicones; fatty acids, esters, and amides
- Hydrocarbon oils
- Triglycerides
- Perfluoroalkyl polyethers
- Related materials, often from natural products

Lubricants can be either incorporated in the tape coating formulation or added topically at the end of the tape process.

9.4.2c Miscellaneous Additives

Small amounts of other materials are included in many tape products to achieve special properties. For example, fine-particle alumina, chromia, or silica is often added to prevent debris obtained during use of the tape from accumulating on the heads and clogging them. This is not normally a requirement in tapes containing chromium dioxide as a magnetic pigment. Other additives include fungicides, which are used in certain limited applications.

Solvent choice is determined by chemical inertness, binder solubility and mix rheology, evaporation rate, availability, toxicity, ease of recovery, and cost. The most commonly used solvents for magnetic tape processes are tetrahydrofuran, methyl ethyl ketone, cyclohexanone, methyl isobutyl ketone, and toluene. Many common types of coating defects can be avoided by the combinations of solvents to provide differential evaporation rates from the coating during the drying process. Finished tape normally has very low levels of residual solvent.

9.4.3 Manufacturing Process

The following sequence of steps is employed in manufacturing magnetic tape:

- Mix preparation
- Dispersion, or milling
- Coating
- Drying
- Surface finishing
- Slitting
- Rewind and/or assembly
- Testing
- Packaging

9.4.3a Dispersion

The magnetic particles must be deagglomerated without reducing the size of individual particles. This step is accomplished by agitating the combined ingredients as a wet mix in one of several types of mills, such as pebble, steel ball, sand, or Sweco, which produce high shear between agglomerates. Milling efficiency in a given system is controlled by mix solids content, viscosity, mix-to-media ratio, and temperature. The end point is reached when visual examination of a drawdown sample under magnification shows the absence of agglomerates or that it meets a pre-determined standard of dispersion quality. Another method is to mill until a maximum in the derivative of the B - H loop is attained. Some commercial dispersion testers are available based on dc noise measurements.

9.4.3b Coating

The coater is perhaps the most critical processing step in the entire operation. There are trade-offs between advantages and disadvantages among the different types of coating methods used, principal among which are *reverse roll* (Figure 9.4.2b) and *gravure* (Figure 9.4.2c). *Reverse roll* is the most widely used, general-purpose method. *Gravure* is especially suited for very thin coatings (0.2 mil or less). *Knife coating* (Figure 9.4.2a), one of the oldest methods, is disappearing with the advent of thin coatings on thin films and high-speed, precision coatings. *Extrusion* and *curtain* coating are increasingly important because they afford high-quality coatings at high speeds. Coaters vary in width from 12 to 60 in as do the base films, and operate at speeds of approximately 250 to 1000 ft/min.

Metal Evaporation Process

The process of depositing a metal layer onto a flexible substrate through evaporation of a metal in a vacuum is referred to as *metal evaporation* (ME) coating [1]. This technique was developed to increase packing density.

One means of increasing the volumetric packing density of magnetic tape is to make the magnetic layer thinner while preserving or increasing its magnetic capabilities. A means of achieving these seemingly contradictory objectives is to deposit a continuous pure metal film onto the base film, instead of using the typical dispersed particulate coating. The most direct way to maximize

9-54 Audio Recording Systems

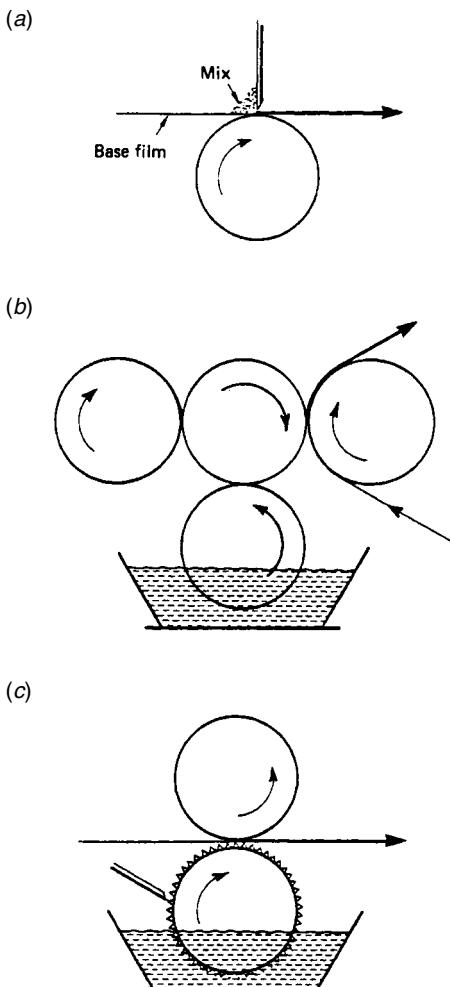


Figure 9.4.2 Tape manufacturing process: (a) knife coating, (b) reverse-roll coating, (c) gravure coating.

the magnetic volume of the recording layer is to remove all non-magnetic components (oxygen) of the particle itself and by removing the binding and resin matrix from the formulation. A nearly ideal magnetic recording medium is achievable through physical deposition of magnetic material on the base film through metal evaporation.

As pressure is decreased (vacuum increased) liquids evaporate more rapidly and at a lower temperature. A vacuum chamber used for metal evaporation onto a base film consists of the following components (shown in diagram form in Figure 9.4.3):

- Transport system for unwinding and rewinding the base film
- A high temperature crucible to hold molten metal after it has been heated, either by resistance or with an electron beam gun

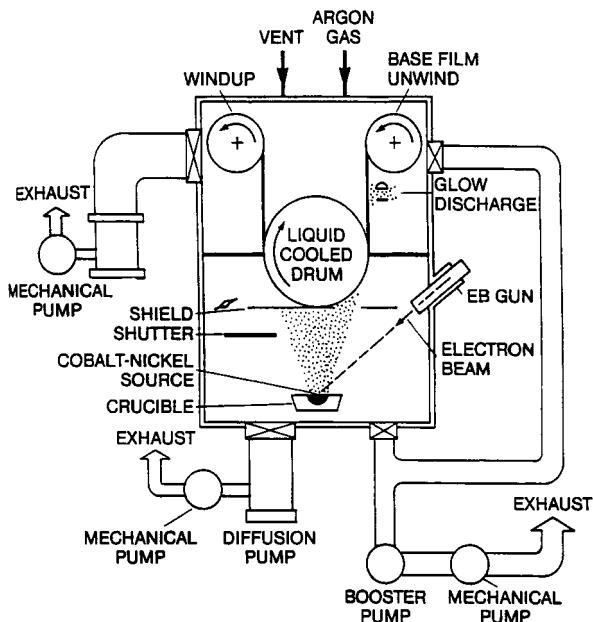


Figure 9.4.3 The metal evaporation process. (After [1]. Courtesy of Quantegy.)

- An internally chilled cooling drum around which the base film is wrapped to remove heat from the condensing metal
- A vacuum pump to quickly remove the air and water vapor from the vacuum chamber

The magnetic coating thickness of a metal evaporated tape is typically $4\mu\text{in}$. This is more than 20 times thinner than the metal particle layer of D-2 tape. Most of the thickness of the tape, therefore, is basically that of the base film being used. The advanced magnetic properties of metal evaporated tape have allowed equipment manufacturers to significantly increase packing densities for recorders of all types.

Slot Die Coating

As shown in Figure 9.4.4, slot die coating utilizes a coating head that has a slot cut into it [1]. The geometry of this slot and its position relative to the base film determines the parameters of the coating surface. Slot die coating is a relatively new coating technology and is typically used for the thinnest and most critical of coatings. The slot die head usually has one slot but is capable of two or three slots, which results in multiple and different layers being applied to the base film. These can be applied wet on wet, or wet on dry.

9-56 Audio Recording Systems

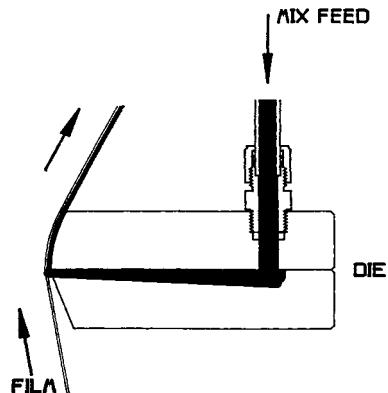


Figure 9.4.4 Simplified slot die coating process. (After [1]. Courtesy of Quantegy.)

9.4.3c Orientation

Maximum S/N performance is obtained when the magnetic particles are aligned, before drying, to the maximum extent possible in the direction of the intended recording. Accordingly, immediately after the wet-coating mix is applied, the web is passed through the field of an orienting magnet having a field strength (500 to 2000 G) optimized for the particular magnetic particle being used. Most tapes are longitudinally oriented, although some are oriented transversely to some degree. The coater itself exerts shearing forces on the mix and thus often imparts some longitudinal orientation in the particles even in this stage.

9.4.3d Drying

The web is next passed through an oven containing circulating forced hot air. Many oven designs use air bearings at web-turnaround points to avoid rubbing between plastic and metal surfaces, and to minimize the formation of abrasion products, which can cause drop-outs. After the coating is dried, the magnetic particles are no longer free to move. During the eventual recording process, only magnetization vectors, or aligned spins of electrons within the molecular species of the particle domains, rotate.

9.4.3e Surface Finishing

Surface finishing is generally required to produce an extremely smooth surface to maximize head-to-tape contact, an absolute necessity for short-wavelength recording. This is accomplished by calendering the tape, or passing the web one or more times through a *nip*, or line of contact, between a highly polished metal roll and a plastic or cellulosic compliant roll. This compaction process also reduces voids in the coating and increases the magnetic pigment volume concentration, and in turn the retentivity of the tape.

9.4.3f Slitting

The web is slit into strands of the desired width, from 150 mils to 3 in (3.8 mm to 7.6 cm). Tolerances in width variation are about ± 1.0 mil to less than ± 0.4 mil, depending on the application. Edge weave, or width waviness (*country laning*) over extended length, should not vary more than about 1 to 2 mils in 1-in video tapes and 0.4 mil in video cassette tape. Tape must be free from jagged edges and debris. Additional tape cleaning processes are sometimes used to ensure that loosely held dropout contributors are effectively removed.

9.4.3g Testing

Professional magnetic tape manufacturers test every component of tape in every step in the process, from individual raw materials through packaging. The most exacting specifications are set forth and followed. Electrical tests, including those for dropouts, are especially stringent, and in professional audio, instrumentation, video, and computer tapes, each reel of tape is tested, in some cases end to end, before shipment. In addition, warehouse audits are performed to ensure maintenance of quality.

9.4.3h Assembly and Packaging

Tape is assembled in various formats but mainly in reels, pancakes, cassettes, and cartridges of different sizes. The same standards of precision, cleanliness, and quality exist in these areas as in tape making per se, and final assemblies of tape components and packages are all performed in ultra-clean-room environments.

9.4.4 References

1. Grega, Joe: "Magnetic and Optical Recording Media," in *NAB Engineering Handbook*, 9th ed., Jerry C. Whitaker (ed.), National Association of Broadcasters, Washington, D.C., pp. 893–906, 1999.

9.4.5 Bibliography

- Bate, G.: "Recent Developments in Magnetic Recording Materials," *J. Appl. Phys.* pg. 2447, 1981.
- Hawthorne, J. M., and C. J. Hefielinger: "Polyester Films," in *Encyclopedia of Polymer Science and Technology*, N. M. Bikales (ed.), vol. 11, Wiley, New York, N.Y., pg. 42, 1969.
- Jorgensen, F.: *The Complete Handbook of Magnetic Recording*, Tab Books, Blue Ridge Summit, Pa., 1980.
- Kalil, F. (ed): *Magnetic Tape Recording for the Eighties*, NASA References Publication 1975, April 1982.

9-58 Audio Recording Systems

- Lueck, L. B. (ed): *Symposium Proceedings Textbook*, Symposium on Magnetic Media Manufacturing Methods, Honolulu, May 25–27, 1983.
- Nylen, P., and E. Sunderland: *Modern Surface Coatings*, Interscience Publishers Division, Wiley, London, 1965.
- Perry, R. H., and A. A. Nishimura: "Magnetic Tape," in *Encyclopedia of Chemical Technology*, 3d ed., Kirk Othmer (ed.), vol. 14, Wiley, New York, N.Y., pp. 732–753, 1981.
- Tochihara, S.: "Magnetic Coatings and Their Applications in Japan," *Prog. Organic Coatings*, vol. 10, pp. 195–204, 1982.

Section**10**

Production Standards, Equipment, and Facility Design

The distinction between an audio system intended for production applications and a system intended for the transmission sounds and other information is an important one. Because of their closed-loop characteristics, production systems can be of any practical design. The process of developing a production system can focus simply on those who will directly use the system. It is not necessary to consider the larger issues of compatibility and public policy, which drive the design and implementation of over-the-air broadcast systems.

Although the foregoing is certainly correct, in the abstract, it is obvious that the economies of scale argue in favor of the development of a production system—even if only closed-loop—that meets multiple applications. The benefits of expanded markets and interoperability between systems are well documented. It was into this environment that production systems intended for digital audio applications were born.

A system intended for broadcast applications must—by necessity—strictly adhere to established standards and practices, usually determined by governmental licensing authorities. Production-oriented systems, however, are not bound by such restrictions. This flexibility is a two-edged sword. While it permits systems of any practical design to be implemented—the system need only communicate with itself and the equipments that directly interface with it—this situation permits—and even encourages—a diversity of product development. In some cases, such proprietary systems have benefited the end-user and the industry; in other cases, it has led to wasted time and money through investment in a technology that held promise at the outset but wound up going nowhere.

This issue is an important one as the professional audio industry moves ahead in the realm of digital broadcasting. Audio industry organizations—most notably the AES and its members—have devoted enormous energies to the development of standards that help end-users chart their paths into the digital domain. With the AES's historic focus on production issues, the Society's activities have naturally focused there.

Industry veterans will remember that in the not too distant past, standards were usually difficult to develop and often required many years to formalize. Standardization efforts in the digital era are—in many cases—leading product development, ensuring that products will communicate with each other and work as promised when integrated into larger systems.

10-2 Section Ten**In This Section:** _____

Chapter 10.1: Audio Production System Standards	10-5
Introduction	10-5
The History of Modern Standards	10-5
American National Standards Institute (ANSI)	10-6
Professional Society Engineering Committees	10-7
Advanced Television Systems Committee	10-7
Electronics Industries Alliance	10-8
Institute of Electrical and Electronic Engineers	10-8
Society of Motion Picture and Televisions Engineers	10-8
Audio Engineering Society	10-9
AES Documents	10-9
Informational Documents	10-9
Project Reports	10-10
Standards and Recommended Practices	10-11
Bibliography	10-15
Chapter 10.2: Production Facility Design	10-17
Introduction	10-17
Project Management	10-17
Feasibility Study and Technology Assessment	10-18
Project Tracking and Control	10-18
Change Order	10-19
Electronic System Design	10-20
Developing a Flow Diagram	10-21
Estimating Cable Lengths	10-22
Signal Timing Considerations	10-22
Cable Loss and Equalization	10-22
Facility Design Process	10-23
Preliminary Space Planning	10-24
Design Models and Mockups	10-26
Construction Considerations	10-26
Component Selection and Installation	10-27
Technical Documentation	10-28
Documentation Tracking	10-29
Symbols	10-29
Cross-Referencing Documentation	10-30
Specifications	10-30
Working with the Contractors	10-30
Computer-Based Tools	10-31
Bibliography	10-31

**On the CD-ROM:**

- “Studio Production Systems” by Ernst-Joacham Voelker, an archive chapter from the first edition of the *Audio Engineering Handbook*. This chapter examines design principles and techniques for a variety of audio studio facilities.
- “Postproduction Systems and Editing” by Tomlinson Holman, an archive chapter from the first edition of the *Audio Engineering Handbook*. This chapter provides a wealth of information on audio postproduction equipment, systems, and techniques.
- “Noise Reduction Systems” by Ray Dolby, et. al., an archive chapter from the first edition of the *Audio Engineering Handbook*. This chapter explains the noise reduction techniques used with analog audio systems.
- “Film Recording and Reproduction” by Ronald Uhlig, an archive chapter from the first edition of the *Audio Engineering Handbook*. This chapter discusses original production recording, postproduction editing, and duplication of film-based audio.

Reference Documents for this Section

DeSantis, Gene, Jerry C. Whitaker, and C. Robert Paulson: *Interconnecting Electronic Systems*, CRC Press, Boca Raton, Fla., 1992.

Whitaker, Jerry C.: *Facility Design Handbook*, CRC Press, Boca Raton, Fla., 2000.

Chapter

10.1

Audio Production System Standards

Donald C. McCroskey

Jerry C. Whitaker, editor

10.1.1 Introduction

Standardization usually starts within a company as a way to reduce costs associated with parts stocking, design drawings, training, and retraining of personnel. The next level might be a cooperative agreement between firms making similar equipment to use standardized dimensions, parts, and components. Competition, trade secrets, and the *NIH factor* (not invented here) often generate an atmosphere that prevents such an understanding. Enter the professional engineering society, which promises a forum for discussion between users and engineers while down playing the commercial and business aspects.

10.1.2 The History of Modern Standards

In 1836, the U. S. Congress authorized the Office of Weights and Measures (OWM) for the primary purpose of ensuring uniformity in custom house dealings. The Treasury Department was charged with its operation. As advancements in science and technology fueled the industrial revolution, it was apparent that standardization of hardware and test methods was necessary to promote commercial development and to compete successfully with the rest of the world. The industrial revolution in the 1830s introduced the need for interchangeable parts and hardware. Economical manufacture of transportation equipment, tools, weapons, and other machinery were possible only with mechanical standardization.

By the late 1800's professional organizations of mechanical, electrical, chemical, and other engineers were founded with this aim in mind. The Institute of Electrical Engineers developed standards between 1890 and 1910 based on the practices of the major electrical manufacturers of the time. Such activities were not within the purview of the OWM, so there was no government involvement during this period. It took the pressures of war production in 1918 to cause the formation of the American Engineering Standards Committee (AESC) to coordinate the activities

10-6 Audio Production Standards, Equipment, and System Design

of various industry and engineering societies. This group became the American Standards Association (ASA) in 1928.

Parallel developments would occur worldwide. The International Bureau of Weights and Measures was founded in 1875, the International Electrotechnical Commission (IEC) in 1904, and the International Federation of Standardizing Bodies (ISA) in 1926. Following World War II (1946) this group was reorganized as the International Standards Organization (ISO) comprised of the ASA and the standardizing bodies of 25 other countries. Present participation is approximately 55 countries and 145 technical committees. The stated mission of the ISO is to *facilitate the internationalization and unification of industrial standards*.

The International Telecommunications Union (ITU) was founded in 1865 for the purpose of coordinating and interfacing telegraphic communications worldwide. Today, its member countries develop regulations and voluntary recommendations, and provide coordination of telecommunications development. A sub-group, the International Radio Consultative Committee (CCIR) (which no longer exists under this name), is concerned with certain transmission standards and the compatible use of the frequency spectrum, including geostationary satellite orbit assignments. Standardized transmission formats to allow interchange of communications over national boundaries are the purview of this committee. Because these standards involve international treaties, negotiations are channeled through the U. S. State Department.

10.1.2a American National Standards Institute (ANSI)

ANSI coordinates policies to promote procedures, guidelines, and the consistency of standards development. Due process procedures ensure that participation is open to all persons who are materially affected by the activities without domination by a particular group. Written procedures are available to ensure that consistent methods are used for standards developments and appeals. Today, there are more than 1000 members who support the U.S. voluntary standardization system as members of the ANSI federation. This support keeps the Institute financially sound and the system free of government control.

The functions of ANSI include: (1) serving as a clearinghouse on standards development and supplying standards-related publications and information, and (2) the following business development issues:

- Provides national and international standards information necessary to market products worldwide.
- Offers American National Standards that assist companies in reducing operating and purchasing costs, thereby assuring product quality and safety.
- Offers an opportunity to voice opinion through representation on numerous technical advisory groups, councils, and boards.
- Furnishes national and international recognition of standards for credibility and force in domestic commerce and world trade.
- Provides a path to influence and comment on the development of standards in the international arena.

Prospective standards must be submitted by an ANSI accredited standards developer. There are three methods which may be used:

- **Accredited organization method.** This approach is most often used by associations and societies having an interest in developing standards. Participation is open to all interested parties as well as members of the association or society. The standards developer must fashion its own operating procedures, which must meet the general requirements of the ANSI procedures.
- **Accredited standards committee method.** Standing committees of directly and materially affected interests develop documents and establish consensus in support of the document. This method is most often used when a standard affects a broad range of diverse interests or where multiple associations or societies with similar interests exist. These committees are administered by a *secretariat*, an organization that assumes the responsibility for providing compliance with the pertinent operating procedures. The committee can develop its own operating procedures consistent with ANSI requirements, or it can adopt standard ANSI procedures.
- **Accredited canvass method.** This approach is used by smaller trade association or societies that have documented current industry practices and desire that these standards be recognized nationally. Generally, these developers are responsible for less than five standards. The developer identifies those who are directly and materially affected by the activity in question and conducts a letter ballot *canvass* of those interests to determine consensus. Developers must use standard ANSI procedures.

Note that all methods must fulfill the basic requirements of public review, voting, consideration, and disposition of all views and objections, and an appeals mechanism.

The introduction of new technologies or changes in the direction of industry groups or engineering societies may require a mediating body to assign responsibility for a developing standard to the proper group. The Joint Committee for Intersociety Coordination (JCIC) operates under ANSI to fulfill this need.

10.1.2b Professional Society Engineering Committees

The engineering groups that collate and coordinate activities that are eventually presented to standardization bodies encourage participation from all concerned parties. Meetings are often scheduled in connection with technical conferences to promote greater participation. Other necessary meetings are usually scheduled in geographical locations of the greatest activity in the field. There are no charges or dues to be a member or to attend the meetings. An interest in these activities can still be served by reading the reports from these groups in the appropriate professional journals. These wheels may seem to grind exceedingly slowly at times, but the adoption of standards that may have to endure for 50 years or more should not be taken lightly.

Advanced Television Systems Committee

The Advanced Television Systems Committee (ATSC) is an international, non-profit organization developing voluntary standards for digital television. The ATSC has over 200 member organizations representing the broadcast, broadcast equipment, motion picture, consumer electronics, computer, cable, satellite, and semiconductor industries. The following is a partial list of ATSC Standards and technical activities. Additional information is available at www.atsc.org.

10-8 Audio Production Standards, Equipment, and System Design

Electronics Industries Alliance

Many of the early standards relating to radio and television broadcasting were developed by equipment manufacturers under the banner of the Radio Manufacturers Association (RMA), later the RETMA (add Electronics and Television), then the Electronics Industries Association (EIA), and now the Electronic Industries Alliance (www.eia.org).

The EIA is a national trade organization made up of a number of product divisions. Some of the best known EIA standards activities are in the areas of data communications, instrumentation, broadcast transmitters, video transmission, video cameras, test charts, video monitors, and RF interference.

With the proliferation and expansion of electronics, the EIA is now divided into many sectors and groups. Of particular interest to the broadcaster are the Consumer Electronics Association (CEA), Electronic Components, Assemblies, Equipment, & Supplies Association (ECA), and Telecommunications Industry Association (TIA). For a comprehensive, searchable listing of all electronic international standards visit <http://global.ihs.com/>.

Institute of Electrical and Electronic Engineers

The IEEE has many branches (professional groups) that serve the standardization needs of the electrical, electronic, and computer industries. Presently available standards relate to definitions, measurement techniques, and test methods. The Institute of Radio Engineers (long since joined the IEEE) was responsible for measurement standards and techniques in televisions early years. The Standards Coordinating Committees publish books and documents covering definitions of electric and electronics terms, graphic symbols and reference designations for engineering drawings, and letter symbols for measurement units. Anyone concerned with power wiring and distribution should be interested in the National Electrical Safety Code books. Nearly all of these documents are available from IEEE (www.ieee.org) or ANSI.

Society of Motion Picture and Televisions Engineers

Organizations such as the SMPTE (www.smpte.org), composed primarily of users of equipment and processes, are able to accomplish what is nearly impossible in the manufacturing community. Namely, to provide a forum where users and manufacturers can distill the best of current technology to promote basic interchangeability in hardware and software. A chronology of the development of this engineering society provides insights as to how such organizations adapt to the needs of advancing technologies.

Around 1915 it became obvious that the rapidly expanding motion picture industry must standardize basic dimensions and tolerances of film stocks and transport mechanisms. After two unsuccessful attempts to form industry based standardizing committees, the Society of Motion Picture Engineers was formed. The founding goals were to standardize the nomenclature, equipment, and processes of the industry; to promote research and development in the industry's science and technology; and to remain independent of, while cooperating with, its business partners. It is this independent quality of a professional society that makes it possible to mediate strongly held opinions of business competitors.

By the late 1940's it was apparent that the future of motion pictures and television would involve sharing technology, techniques, and the market for visual education and entertainment. SMPE became SMPTE. In comparatively recent times the Society has been assigned more responsibility for television standards. The recording and reproduction of television signals has

become the province of SMPTE standardization efforts. An index of the work of the engineering committees is published yearly. The basic SMPTE documents issued as a part of the organization's standardization efforts are:

- **Engineering Guidelines**—guidelines for the implementation of test materials and equipment operation.
- **Recommended Practices**—these include specifications for test materials, generic equipment setup and operating techniques, and mechanical dimensions involving operational procedures.
- **Standards**—mechanical specifications for film, tape, cassettes, and transport mechanisms; electrical recording and reproduction characteristics; and protocol and software issues for digital video systems.

Audio Engineering Society

The AES (www.aes.org) was organized in 1948 primarily to serve the needs of the high quality audio recording and reproduction community. The Society maintains a standards committee (AESSC) that supervises the work of subcommittees and working groups. Drafts of proposed standards are published in the *Journal of the Audio Engineering Society* (JAES) for review and comment by all interested parties. Any substantive comments (as opposed to editorial) are then considered by the committee before submitting the final document to a vote. Current AES standards address measurement methods, commercial loudspeaker specifications, and digital audio recording/transmission systems.

10.1.3 AES Documents

The following documents relating to audio technologies in general, and digital audio in particular, are available from the Audio Engineering Society. For more information, contact the AES at www.aes.org.

10.1.3a Informational Documents

AES-2id-1996 (r2001): Guidelines for the Use of the AES3 Interface

This document provides guidelines for the use of AES3, AES Recommended Practice for Digital Audio Engineering—Serial transmission format for two-channel linearly represented digital audio data, together with AES5, AES Recommended Practice For professional digital audio applications employing pulse-code modulation—Preferred sampling frequencies, AES11, AES Recommended Practice for Digital Audio Engineering—Synchronization of digital audio equipment in studio operations, and AES18, AES Recommended Practice for Digital Audio Engineering—Format for the user data channel of the AES digital audio interface.

AES-3id-2001 (Revision of AES-3id-1995): Transmission of AES3 Formatted Data by Unbalanced Coaxial Cable

This document contains information regarding cables, cable equalizers, and receiver circuits including adaptors to or from standard AES3 equipment and cabling where it is required to trans-

10-10 Audio Production Standards, Equipment, and System Design

mit AES3 formatted signals over long distances (up to 1000 m), or in a video installation using analog video distribution equipment. It is not intended to be an alternative electrical specification to AES3, which is based on balanced, shielded, twisted-pair cable transmission over distances of up to 100 m. The information is based on studies and laboratory experiments discussed in a series of technical reports that have been partly summarized and included.

AES-4id-2001: Characterization and Measurement of Surface Scattering Uniformity

This document provides guidelines for characterizing the uniformity of scattering produced by surfaces from measurements or predictions of scattered polar responses. In this context, the surface scattering is quantified in terms of a single diffusion coefficient. The diffusion coefficient is a measure of quality designed to be used by producers and users of surfaces that, either deliberately or accidentally, diffuse sound. It is also intended for use when needed by developers and users of geometric room acoustic models. The diffusion coefficient is not intended, however, to be blindly used as an input to current diffusion algorithms in geometric room acoustic models. The diffusion coefficient characterizes the sound reflected from a surface in terms of the uniformity of the scattered polar distribution. The information document details a free-field characterization method.

AES-5id-1997: Loudspeaker Modeling and Measurement—Frequency and Angular Resolution for Measuring, Presenting, and Predicting Loudspeaker Polar Data

This document provides guidelines for measuring, presenting and predicting polar data from a single acoustic source or from an array of acoustic sources. It describes and quantifies measurement resolution, presentation resolution, prediction techniques and measurement environments. The information presented here is based on objective measurements and does not take subjective or psychoacoustic criteria into account.

AES-6id-2000: Personal Computer Audio Quality Measurements

This document focuses on the measurement of audio quality specifications in a PC environment. Each specification listed has a definition and an example measurement technique. Also included is a detailed description of example test setups to measure each specification.

AES-10id-1995: Engineering Guidelines for the Multichannel-Audio Digital Interface (MADI) AES10

This document provides guidance for areas of application of the MADI standard (AES10) that might be unclear. It is not intended to replace AES10, but to supplement it in those areas that are not suitable for definition in a standards document.

10.1.3b Project Reports

AES-R1-1997: Specifications for Audio on High-Capacity Media

This document is the report of Task group SC-02-M, a task force within the Audio Engineering Society Standards Committee (AESSC), which studied the future of high-capacity audio media (such as DVD) for over one year with input from more than 80 persons and organizations.

AES-R3-2001: Compatibility for Patch Panels of Tip-Ring-Sleeve Connectors

This report covers the concentric connectors known as phone plugs and jacks that are widely used in the audio industry for the interconnection of sound system components as general-use consumer tip-ring-sleeve (TRS) types as well as for the interconnection of broadcast and other professional systems, such as professional TRS types. Because differing dimensions specified by various standards often result in mechanical and, as a consequence, electrical incompatibilities among the various TRS jacks and plugs, this report surveys the standards, connectors, and possible incompatibility problems.

10.1.3c Standards and Recommended Practices**AES2-1984 r1997:** Specification of Loudspeaker Components used in Professional Audio and Sound Reinforcement

This document is a recommended practice for describing and specifying loudspeaker components used in professional audio and sound-reinforcement systems. These components include high-frequency drivers, high- and mid- frequency horns, low-frequency drivers, and low-frequency enclosures. For drivers, specifications are given for describing frequency response, impedance, distortion, and power handling. For horns and enclosures, specifications are given for describing directional characteristics and additional pertinent performance data. For all components, specifications as are given for describing necessary physical and mechanical characteristics, such as hardware, mounting data, size, and weight. Appendices supporting the text give guidelines for making proper free-field measurements, sizing of baffles for low-frequency measurements, a method for producing the specified noise signal used in power testing, and a summary of required information.

AES3-1992 (r1997): Serial Transmission Format for Two-Channel Linearly Represented Digital Audio Data

The format provides for the serial digital transmission of two channels of periodically sampled and uniformly quantized audio signals on a single shielded twisted wire pair. The transmission rate is such that samples of audio data, one from each channel, are transmitted in time division multiplex in one sample period. Provision is made for the transmission of both user and interface related data as well as of timing related data, which may be used for editing and other purposes. It is expected that the format will be used to convey audio data that have been sampled at any of the sampling frequencies recognized by the AES5, Recommended Practice for Professional Digital Audio Applications Employing Pulse-Code Modulation—Preferred Sampling Frequencies.

AES5-1998: Preferred Sampling Frequencies for Applications Employing Pulse-Code Modulation

A sampling frequency of 48 kHz is recommended for the origination, processing, and interchange of audio programs employing pulse-code modulation. Recognition is also given to the use of a 44.1-kHz sampling frequency related to certain consumer digital applications, the use of a 32-kHz sampling frequency for transmission-related applications, and the use of a 96-kHz sampling frequency for applications requiring a higher bandwidth or more relaxed anti-alias filtering.

10-12 Audio Production Standards, Equipment, and System Design

AES6-1982 r1997: Method for Measurement of Weighted Peak Flutter of Sound Recording and Reproducing Equipment

Weighted peak flutter is measured using a 3150-Hz tone transmitted through the equipment. The tone is frequency demodulated, frequency-response weighted, peak-to-peak detected, time-response weighted, and read on a meter as the zero-to-peak (one-half of peak-to-peak) values. Results are reported as “weighted peak of the recorder (or reproducer, or recording/reproducing system): +/- ____ percent.” A toleranced graph and table give the frequency-response weighting (approximately a 6-dB-per-octave drop above and below 4 Hz, with an additional drop below 0.5 Hz.) A toleranced table gives the time-response weighting (the reference is the peak-to-peak amplitude of a 4-Hz sine wave; the test signal is a series of unidirectional pulses of the same peak-to-peak amplitude, spaced 1 second apart; a pulse length of 60 ms gives 90-percent response; the fall between 100-ms pulses is a 40-percent reading.)

Good engineering practices are given for the meter design. The rationale for this standard is given in an appendix. This standard, originally published as IEEE Std-193, has technical requirements identical to standards IEC Pub. 386, CCIR 409-2, and DIN 45 507. An ANSI version is available as S4.3-1982.

AES7-2000: Method of Measuring Recorded Fluxivity of Magnetic Sound Records at Medium Wavelengths

This standard specifies a method of measuring the recorded flux per unit track width, called *fluxivity*, of a magnetically recorded sinusoidal test signal at medium wavelengths for all magnetic sound record formats, by using a high-efficiency magnetic reproducing head. It also specifies the equipment needed to implement this method. An ANSI version is available as S4.6-1982.

AES10-1991 (r1997): Serial Multichannel Audio Digital Interface (MADI)

This standard describes the data organization for a multi-channel-audio digital interface. It includes a bit level description, features in common with the AES3-1985 two channel format, and the data rates required for its utilization. The specification provides for the serial digital transmission of 56 channels of linearly represented digital audio data at a common sampling frequency within the range 32 kHz to 48 kHz (+/- 12.5%) having a resolution of up to 24 bits per channel.

The format makes possible the transmission and reception of the complete 28-bit channel word (excluding preamble) as specified in the document AES3-1985, providing for the validity, user, channel status, and parity information allowable under that Standard. The transmission format is of the asynchronous simplex type and is specified for a single 75 ohm coaxial cable point-to-point interconnection, although the use of fibre-optic medium is also possible. An ANSI version is available as S4.43-1991.

AES11-1997: Synchronization of Digital Audio Equipment in Studio Operations

This standard provides a systematic approach to the synchronization of digital audio signals. Recommendations are made concerning the accuracy of sample clocks as embodied in the interface signal and the use of this format as a convenient synchronization reference where signals must be rendered co-timed for digital processing. Synchronism is defined, and limits are given which take account of relevant timing uncertainties encountered in an audio studio.

AES14-1992 (r1998): Application of Connectors, Part 1—XLR-Type Polarity and Gender

This standard specifies a common scheme for wiring the connectors used in audio systems, particularly to avoid the inversion of absolute polarity among the items in the analog signal chain. An ANSI version is available as S4.48-1992.

AES15-1991 (r1997): Sound-Reinforcement Systems Communications Interface (PA-422)

This standard specifies the electrical characteristics of a balanced-voltage circuit for the interchange of serial binary signals for the control of sound-reinforcement systems. It provides for interchange among data terminal equipment (DTE), that is, computers and microprocessors, and data circuit terminating equipment (DCE). PA-422 is a mnemonic, signifying professional audio implementation of Electronics Industries Association EIA-422-A. Device control language is provided in an annex. An ANSI version is available as S4.49-1991.

AES17-1998 (revision of AES17-1991): Measurement of Digital Audio Equipment

This standard provides methods for specifying and verifying the performance of digital audio equipment. Many tests are substantially identical to those used when testing analog equipment. However, because of the unique requirements of digital audio equipment and the effects of its imperfections, additional tests are necessary.

AES18-1996 (revision of AES18-1992): Format for the User Data Channel of the AES Digital Audio Interface

This standard describes a method of formatting the user data channels provided within the digital audio serial interface format (AES3). The transmission format is an adaptation of the packet-based high-level data link control (HDLC) communications protocol and provides for the transmission of ancillary data that may or may not be time related to the audio signal. The data rate is constant within a range of +/- 12.5 percent of a sampling frequency of 48 kHz. The standard also provides a data priority and management strategy to ensure that adequate capacity is available for downstream data insertion.

AES19-1992 r1998: Measurement of the Lowest Resonance Frequency of Loudspeaker Cones

This standard test method is intended to determine the frequency of lowest resonance of a loudspeaker cone. Such information is used for engineering design and for quality control. The method has been developed to improve correlation of measurement between cone manufacturers and loudspeaker manufacturers. An ANSI version is available as S4.30-1992.

AES20-1996: Subjective Evaluation of Loudspeakers

This standard is a set of recommendations for subjective evaluation of high-performance loudspeaker systems. It is believed that, for certain audio components including loudspeakers, subjective evaluation is a necessary adjunct to objective measurements. The strong influence of listening conditions, program material and of individual evaluators is recognized. This document seeks, therefore, to assist in avoiding testing errors rather than to attempt to establish a correct procedure.

10-14 Audio Production Standards, Equipment, and System Design

AES22-1997: Storage of Polyester-Base Magnetic Tape

This standard provides recommendations concerning the storage conditions, storage facilities, enclosures, and inspection for recorded polyester-base magnetic tapes in roll form. It covers analog and digital tape and includes tape made for audio, video, instrumentation, and computer use.

AES24-1-1999, (Revision of AES24-1-1995): Application Protocol for Controlling and Monitoring Audio Devices via Digital Data Networks—Part 1: Principles, Formats, and Basic Procedures

This document describes the architecture of AES24, the name assigned to an extensible application protocol for controlling and monitoring audio devices via local area networks, and, when possible in the future, devices designed for other media.

AES26-1995: Conservation of the Polarity of Audio Signals

This document standardizes the polarity of the signals at the various interface points between different items of equipment, in particular from the acoustical, electrical, and the magnetic aspects. Each item of equipment complies separately with the polarity requirements for the input and output signals.

AES27-1996: Managing Recorded Audio Materials Intended for Examination

This document specifies recommended practices for safekeeping, conveyance, inspection, description, and labeling of audio recordings offered as evidence in criminal investigations, in criminal or civil proceedings, or in other forensic applications. It does not cover analysis of magnetic tapes or other recording media for the purposes of authenticity determination, talker identification, copyright violation, enhancement of oral conversations or other signals, or otherwise characterizing signals recorded on such tapes.

AES28-1997: Method for Estimating Life Expectancy of Compact Discs (CD-ROM), Based on Effects of Temperature and Relative Humidity

This standard specifies test methods for estimating the storage life expectancy (LE) of information stored on compact discs (CD-ROM). Only the effects of temperature and relative humidity are considered. Block error rate (BLER) is the measured response and the end-of-life criterion. An Eyring model is developed from accelerated test results. Data are normalized to 25 degree C and 50 percent relative humidity, and the LE, percent compliance, and confidence intervals at these conditions are calculated.

AES31-3-1999: Audio-File Transfer and Exchange—Part 3: Simple Project Interchange

This standard provides a convention for expressing edit data in text form in a manner that enables simple and accurate computer parsing while retaining human readability. It also describes a method for expressing time-code information in character notation.

AES33-1999: Database of Multiple-Program Connection Configuration

This document specifies the purpose and procedures for the maintenance of data on an AEESC database of connectors for multiple-program applications. Purchasers of this standard obtain full access to the database using their names and order numbers for identification.

AES35-2000: Method for Estimating Life Expectancy of Magneto-Optical (MO) Disks, Based on Effects of Temperature and Relative Humidity

This standard specifies test methods for estimating the life expectancy of information stored on magneto-optical (M-O) disks. Only the effects of temperature and relative humidity on the media are considered.

AES38-2000: Life Expectancy of Information Stored in Recordable Compact Disc Systems—Method for Estimating, Based on Effects of Temperature and Relative Humidity

This standard specifies test methods for estimating the life expectancy of information stored in recordable compact disc systems. Only the effects of temperature and relative humidity on the media are considered.

AES41-2000: Recoding Data Set for Audio Bit-Rate Reduction

This document describes a format for the data to be transmitted to identify a number of popular audio bit-rate reduction techniques. Provision is also made for the transmission of additional ancillary information. This document also describes a mechanism whereby the data derived from the coded signal can be transported with the decoded signal so that the data may be recovered and used to aid a subsequent re-encoding process.

AES42-2001: Digital Interface for Microphones

This standard describes an extension of the existing digital audio interface AES3 to provide a digital interface for microphones.

AES43-2000: Criteria for the Authentication of Analog Audio Tape Recordings

The purpose of this standard is to formulate a standard scientific procedure for the authentication of audio tape recordings intended to be offered as evidence or otherwise utilized in civil, criminal, or other fact finding proceedings.

AES45-2001: Connectors for Loudspeaker-Level Patch Panels

This standard complements IEC 60268-12 including amendments 1 and 2, extending the standardization of application of AES45-series connectors to their use for loudspeaker-level patch panels.

10.1.4 Bibliography

McCroskey, Donald C.: "Standardization: History and Purpose," in *The Electronics Handbook*, Jerry C. Whitaker (ed.), CRC Press, Boca Raton, Fla., 1996.

Chapter

10.2

Production Facility Design

Gene DeSantis

Jerry C. Whitaker, Editor-in-Chief

10.2.1 Introduction

The implementation of digital broadcasting has launched an industry-wide upgrade program—both on the RF side and the studio side. Planning how the audio facility will function is a difficult exercise, and one that will have far-reaching effects. The success of any upgrade project depends in large part on the planning that goes into the effort before any concrete is poured or any cables are pulled.

10.2.2 Project Management

The need for a project may originate with management, operations staff, technicians, or engineers. In any case, some sort of logical reasoning or a specific production requirement will justify the need. On small projects, such as the addition of one piece of equipment, money must only be available for the purchase and installation costs. When the need justifies a large project, the final cost is not always immediately apparent. The project must be analyzed by dividing it into its constituent parts or elements:

- Equipment and parts
- Materials
- Resources, including money and time needed for project completion

An executive summary or capital project budget request, which contains a detailed breakdown of these elements, can provide the information management needs to determine the return on investment, and to make an informed decision on whether or not to authorize the project.

A capital project budget request, which contains the minimum information, may consist of the following items:

10-18 Audio Production Standards, Equipment, and System Design

- Project name—a name that describes the result of the project, such as “control room upgrade.”
- Project number (if required). A large organization that does many projects will use some kind of project numbering system, or it may use a budget code assigned by the accounting department.
- Project description—a brief description of what the project will accomplish, such as “design the technical system upgrade for the renovation of production control room 2.”
- Initiation date—the date the request will be submitted.
- Completion date—the date the project will be completed.
- Justification—the reason the project is needed.
- Material cost breakdown—a list of equipment, parts, and materials required for construction, fabrication, and equipment installation.
- Total material cost.
- Labor cost breakdown—a list of personnel required to complete the project, their hourly pay rates, the number of hours they will spend on the project, and the total cost for each person.
- Total project cost—the sum of material and labor costs.
- Payment schedule—an estimation of individual amounts that will be paid out during the course of the project, and the approximate dates that each will be payable.
- Preparer’s name and the date prepared.
- Approval signature(s) and date(s) approved.

More detailed analysis, such as return on investment, can be carried out by an engineer. Financial analysis, however, should be left to the accountants, who have access to company financial data.

10.2.2a Feasibility Study and Technology Assessment

In cases where an attempt must be made to implement new technology, and where a determination must be made as to whether certain equipment can perform a desired function, a feasibility study should be conducted. The system engineer may be called upon to assess the state of the art in order to develop a new application. In addition to a capital project budget request, an executive summary or a more detailed report of evaluation test results may be required to help management make its decision.

10.2.2b Project Tracking and Control

A project team member may be selected by the project manager to report the status of work during the course of the project. A standardized project status report form can provide consistent and complete information to the project manager. The purpose is to supply information to the project manager regarding work completed and money spent on resources and materials.

A project status report containing minimum information should contain the following items:

- Project number (if required)

- Date prepared
- Project name
- Project description
- Start date
- Completion date (the date this part of the project was completed)
- Total material cost
- Labor cost breakdown
- Preparer's name

10.2.2c Change Order

After all or part of a project design has been approved and money has been allocated, any changes may increase or decrease the cost. Several factors can affect the cost:

- Material
- Resources, such as labor and special tools or construction equipment
- Costs incurred because of manufacturing or construction delays

Management should know about such changes, and will want to control them. For this reason, a method of reporting changes to management and soliciting its approval should be instituted. The best way to do this is with a change order request or change order. A change order includes a brief description and reason for the change and a summary of the effect it will have on costs and the project schedule.

Management will exercise its authority to approve or disapprove each change, based upon its understanding of the cost and benefits and the perceived need for the modification of the original plan. Therefore, the system engineer should provide as much information and explanation as may be necessary to make the change clear and understandable to management.

A change order form, containing the minimum information, should include the following items:

- Project number
- Date prepared
- Project name
- Labor cost breakdown
- Preparer's name
- Description of the change
- Reason for the change
- Equipment and materials to be added or deleted
- Material costs or savings
- Labor costs or savings

10-20 Audio Production Standards, Equipment, and System Design

- Total cost of this change (increase or decrease)

10.2.3 Electronic System Design

Performance standards and specifications must be established in advance for a technical facility project. This will set the performance level of equipment that is acceptable for the system and affect the size of the budget. Signal quality, stability, reliability, and accuracy are examples of the kinds of parameters that must be specified. Access and processor speeds are important parameters when dealing with computer-driven products. The system engineer must confirm whether selected equipment conforms to the standards.

At this point, it must be determined what functions each component in the system will be required to fulfill, and how each will perform with other components in the system. The management and operation staff usually know what they would like the system to do, and how they can best accomplish the task. They should select equipment that they think will do the job. With a familiarity of the capabilities of different equipment, the system engineer should be able to contribute to this function/definition stage. The following is a list of questions that must be answered:

- What functions must be available to the operators?
- What functions are secondary and, therefore, not necessary?
- What level of automation should be required to perform a function?
- How accessible should the controls be?

Over-engineering or over-design must be avoided. Such serious and costly mistakes are often made by engineers and company staff when planning technical system requirements. A staff member may, for example, ask for a feature or capability without fully understanding its complexity or the additional cost it may impose. Other portions of the system may have to be compromised to implement the additional feature. An experienced system engineer will be able to spot this and determine whether the trade-offs and added engineering time and cost are really justified.

When existing equipment is used, an inventory list should be made. This is the preliminary part of a final equipment list. Normally, when confronted with a mixture of acceptable and unacceptable equipment, the system engineer must determine what meets current standards and what should be replaced. Then, after soliciting input from facility technical personnel, the system engineer develops a summary of equipment needs, including future acquisitions. One of the system engineer's most important contributions is the ability to identify and meet these needs within the facility budget.

A list of major equipment is then prepared. The system engineer selects equipment based on experience with the products and on owner preferences. Existing equipment is often reused. A number of considerations are discussed with the facility owner to determine the best product selection. Some major points include:

- Budget restrictions
- Space limitations
- Performance requirements

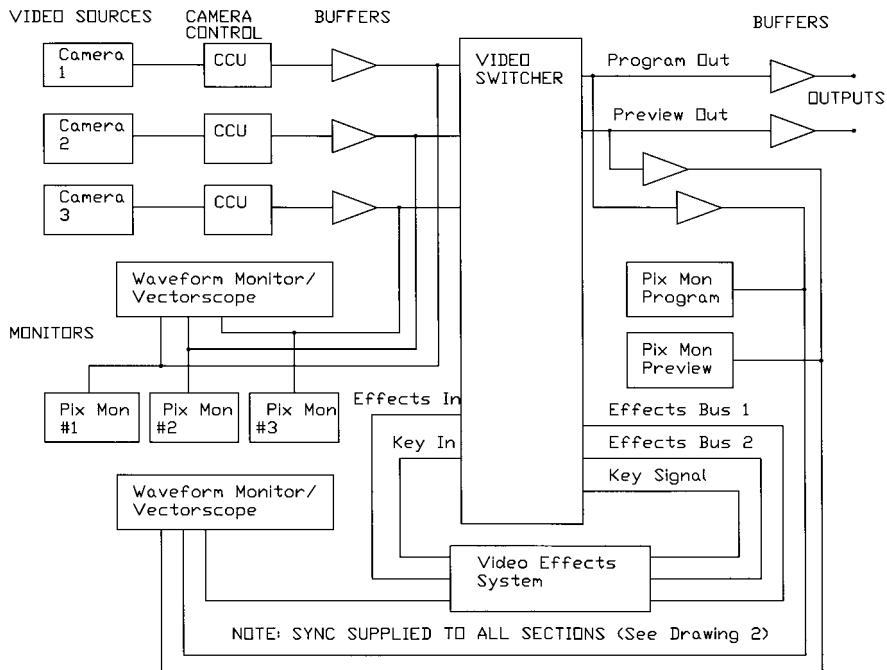


Figure 10.2.1 Example flow diagram of a small postproduction facility.

- Ease of operation
- Flexibility
- Functions and features
- Past performance history
- Manufacturer support

The system engineer's goal is to choose and install equipment that will meet the project's functional requirements efficiently and economically. Simplified block diagrams of the video, audio, control, data, RF, and communication systems are drawn and then discussed with the owner and presented for approval.

10.2.3a Developing a Flow Diagram

The flow diagram is a schematic drawing used to show the interconnections among all equipment that will be installed. It differs from a block diagram because it contains much more detail. Every wire and cable must be included on these drawings. Figure 10.2.1 illustrates a typical flow diagram of a postproduction facility.

10-22 Audio Production Standards, Equipment, and System Design

The starting point for preparing a flow diagram can vary depending on the information available from the design development phase of the project, and on the similarity of the project to previous projects. If a similar system has been designed previously, the diagrams from that project can be modified to include the equipment and functionality required for the new system. New equipment models can be shown on the diagram in place of their counterparts, and minor wiring changes can be made to reflect the new equipment connections and changes in functional requirements. This method is efficient and easy to complete.

If the facility requirements do not fit any previously completed design, the block diagram and equipment list are used as a starting point. Essentially, the block diagram is expanded and details are added to show all of the equipment and interconnections, and to show any details necessary to describe the installation and wiring completely.

An additional design feature that is desirable for specific applications is the capability to disconnect a rack assembly easily from the system and relocate it. This would be used if a system was pre-built at a system integration facility and later moved and installed at the client's site. With this type of situation, the interconnecting cable harnessing scheme must be well planned and identified on the drawings and cable schedules.

10.2.3b Estimating Cable Lengths

Cable lengths are calculated using dimensions taken from the floor plans and rack elevations and should be included on the cable schedule. The quantity of each cable type can then be estimated for pricing and purchasing. A typical cable schedule database printout is shown in Figure 10.2.2.

10.2.3c Signal Timing Considerations

For certain signal paths, the length of the cable connecting two pieces of equipment may critically affect the timing relationship of that signal as it relates to others in the system. Calculate these critical cable lengths and include them on the cable schedule.

Electrical signals travel through cable at a velocity determined by the physical properties of the cable. Using the published value for the velocity of propagation, calculate the amount of delay in a given length of cable. The velocity for video cables may vary from 66 percent to 78 percent, depending on the manufacturer. The delay may also be determined experimentally by simply measuring the amount of delay produced in a given length of cable being used.

When building a facility that has critical timing requirements, keep cables as short as possible to minimize signal attenuation and crosstalk. This requires keeping interconnected equipment as close together as possible. It is best to locate all of the distribution equipment in the same or adjacent racks. Because most video cabling among distribution elements must be timed or of matching lengths, short cables make the job manageable, and, at the same time, cable costs are kept low.

Cable Loss and Equalization

The high frequency response of a cable decreases with increasing frequency. The loss can be compensated for by using an equalizing amplifier with a response curve that complements the cable loss. For video applications, a typical distribution amplifier (DA) has six outputs isolated from one another by fan-out resistors. Because the equalization is adjusted to produce a flat

WIRE NO.	DESCRIPTION	CONN	SOURCE	DESTINATION	CONN	CABLE TYPE	TIME LENGTH	CABLE NO. OF WIRES	COLOR CODE
1031 BB1		BNC	DA01 OUT 1	CCU1 GENLOCK IN	BNC	RK7560	26'	1	YEL
1032 BB1		BNC	DA01 OUT 2	CCU2 GENLOCK IN	BNC	RK7560	26'	1	YEL
1033 BB1		BNC	DA01 OUT 3	CCU3 GENLOCK IN	BNC	RK7560	26'	1	YEL
1034 BB1		BNC	DA01 OUT 4	CCU4 GENLOCK IN	BNC	RK7560	26'	1	YEL
1035 BB		BNC	DA02 IN LOOP	DA03 IN	BNC	RK7560	1'	1	YEL
1036 BB1		BNC	DA02 OUT 1	P1 RRF VID IN	BNC	RK7560	21'	1	YEL
1037 BB1		BNC	DA02 OUT 2	P2 RRF VID IN	BNC	RK7560	25'	1	YEL
1038 BB1		BNC	DA02 OUT 3	P3 RRF VID IN	BNC	RK7560	27'	1	YEL
1039 BB1		BNC	DA02 OUT 4	P4 RRF VID IN	BNC	RK7560	29'	1	YEL
1040 BB		BNC	DA03 IN LOOP	DA04 IN	BNC	RK7560	1'	1	YEL
1041 BB1		BNC	DA03 OUT 1	P5 EXT RRF IN	BNC	RK7560	29'	1	YEL
1042 BB1		BNC	DA03 OUT 2	P6 REF VID IN (FUTURE)	BNC	RK7560			
1043 BB1		BNC	DA03 OUT 3	R1 RRF VID IN	BNC	RK7560	29'	1	YEL
1044 BB1		BNC	DA03 OUT 4	VTR H TBC COMP VID IN	BNC	RK7560	34'	1	YEL
1045 BB		BNC	DA04 IN LOOP	DA05 IN	BNC	RK7560	1'	1	YEL
1046 BB1		BNC	DA04 OUT 1	VTR I TBC GENLOCK IN	BNC	RK7560	34'	1	YEL
1047 BB1		BNC	DA04 OUT 2	CG1 GEN LOCK VID IN	BNC	MINI	47'	10	1
1048 BB1		BNC	DA04 OUT 3	CG2 ENC PGM IN	BNC	MINI	39'	10	1
1049 BB1		BNC	DA04 OUT 4	CG3 GEN LOCK VID IN	BNC	RK7560	10' 6"	1	YEL
1050 BB		BNC	DA05 IN LOOP	DA51 IN	BNC	RK7560	6'	1	YEL
1051 BB1		BNC	DA05 OUT 1	DVE VBS/S IN	BNC	RK7560	8' 9"	1	YEL
1052 BB1		BNC	DA05 OUT 2	TBC GENLOCK IN	BNC	RK7560	5' 6"	1	YEL
1053 BB1		BNC	DA05 OUT 3	TCG 2 VID IN	BNC	RK7560	5' 5"	1	YEL
1054 BB1		BNC	DA05 OUT 4	CG3 CC SYNC IN	BNC	RK7560			WHT
1055 TBC ADV SYNC OUT		BNC	VTR H TBC ADV SYNC OUT	VTR H SYNC IN	BNC	RK7560	4' 6"	1	WHT
1056 TBC ADV SYNC OUT		BNC	VTR I TBC ADV SYNC OUT	VTR I VID IN 1	BNC	RK7560	4' 6"	1	WHT
1057 SC TO VCR		BNC	TBC VTR SC OUT	VTR I SC IN	BNC	VP618PE	4' 6"	1	BLK
1058 BB1		BNC	P5 REF VID IN LOOP OUT	P5 TBC COMP VID IN	BNC	RK7560	6'	1	YEL
1059									
1060 BB2		BNC	DA51 OUT 1	TEST SRCE SWR IN 1	BNC	RK7560	E 7' 6"	1	YEL
1061 BB2		BNC	DA51 OUT 2	PS EXT RRF IN	BNC	RK7560	G 25' 7"	1	YEL

Figure 10.2.2 Wiring database printout for a portion of the facility illustrated in Figure 10.2.1.

response at the end of a length of a specific type of cable, all of the cables being driven by the amplifier must be the same type and length.

10.2.3d Facility Design Process

The best way to design a facility is to begin with the architectural drawings of the existing building or planned construction. If architectural drawings are not available, it is necessary to have the architect prepare them. For small renovation projects, the system engineer may prepare the needed drawings to plan equipment layout.

Before any details are confirmed, a site visit should be made to record and confirm building space dimensions, clearances, and access to building services. Also, existing rack and console dimensions and locations should be measured. If the site is some distance away, photograph important elements, such as existing construction details or current equipment configurations, to reduce the need to travel back to the site.

10-24 Audio Production Standards, Equipment, and System Design

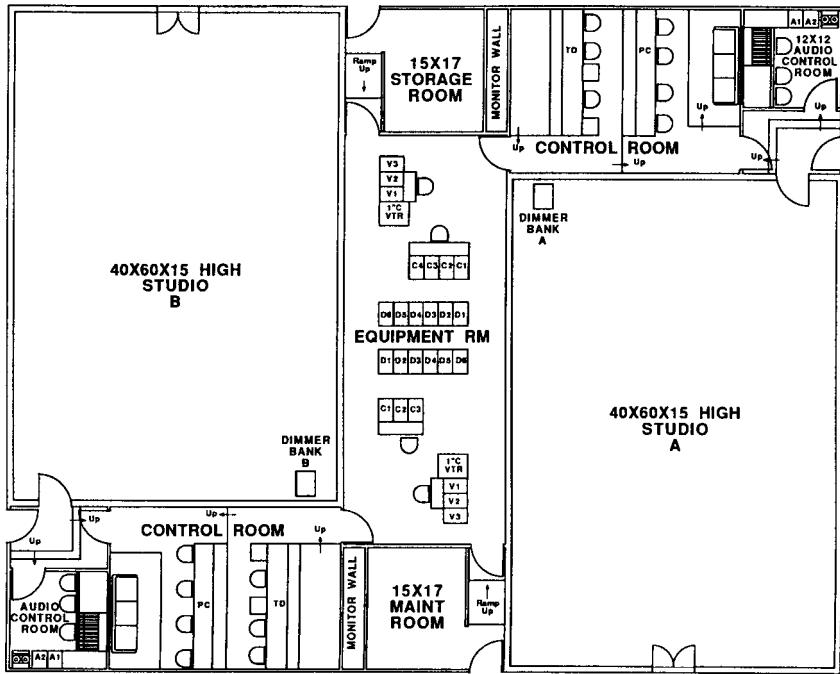


Figure 10.2.3 Architectural floor plan of a new facility.

Preliminary Space Planning

Whether the project involves new construction or renovation of an existing building, current facilities and equipment are reviewed to determine a starting point for the planning process. Building and room layouts are determined by studying each function and its relationship to all others. Functional requirements of each operational department are assessed to determine the gross space requirements of areas to be expanded or renovated. Key facility personnel are interviewed to determine past experiences, future trends, operational requirements for immediate use, and future needs of the facility. This should include the number of present employees and those anticipated in the future.

Environmental factors, such as noise, vibration, RF interference, power line interference, temperature, and humidity also must be considered. Accessibility to utilities, such as communications, power, air supply, fuel, and water, must be calculated. Air conditioning is a major concern in all large facilities that employ a lot of equipment or lighting.

After management approves the equipment list, a rough schematic layout is prepared in conjunction with the architect's preliminary drawings. The system engineer examines this layout to be certain that it provides adequate space for present and future equipment and for maintenance and operation. Equipment identification is then added to the architect's schematic, and the proce-

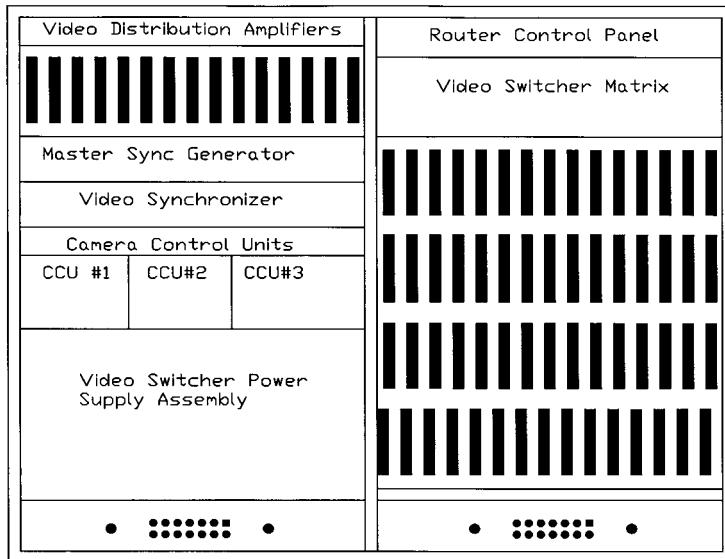


Figure 10.2.4 Equipment placement in a rack assembly.

dure continues to the design-development phase. An example of an architectural floor plan is shown in Figure 10.2.3. Equipment placement in rack assemblies is illustrated in Figure 10.2.4.

Design renderings (drawings or paintings created by an artist or drafter to show a realistic flat or perspective view of a design) are then produced. Full-color 3-D models can be generated by a computer for viewing from different perspectives. The printout of any view can be used as the rendering.

A color and materials presentation board is usually prepared for review by decision makers. The presentation may include the following:

- Artist or computer renderings
- Color chips
- Wood types
- Work surface laminates
- Metal samples
- Samples of carpeting, furniture fabrics, and wall coverings

Several different combinations may be prepared. The samples and renderings are attached to a board or heavy paper stock for easy presentation.

10-26 Audio Production Standards, Equipment, and System Design

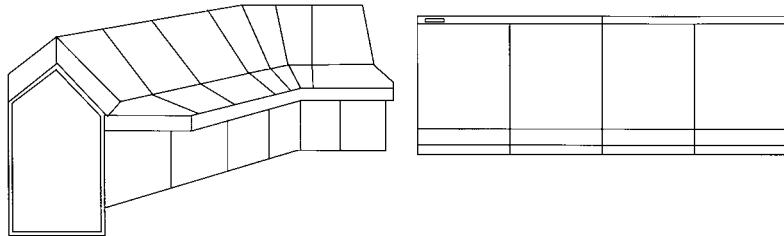


Figure 10.2.5 Equipment rack and console mockup for planning purposes.

Design Models and Mockups

When a drawing cannot be interpreted easily by the owner and/or staff, a scale model or full-size mockup of the facility (or portions of it) can be constructed. This will help familiarize them with the design, allowing them to make decisions and changes. Models can also be used to present a design concept to company executives. Models can provide a cost-effective way to evaluate new ideas. Inexpensive materials can be formed to represent racks, consoles, or equipment. For example, the top and four side views of an enclosure can be drawn or plotted at a reduced scale on stiff paper so that the drawings touch at adjoining surfaces. When cut out, they are glued together to form a 3-D model. Flaps added to the drawing make it easier to join the surfaces. The more detail provided in the mockup, the better.

Blocks of wood can also be cut to the shape of the equipment being modeled. Cut-out drawings of the equipment features are pasted on the block's surfaces to add realism. Plastic scale models of structural components, piping, furniture, and other elements are available from model manufacturers and can be used to enhance the presentation.

Full-scale mockups, like models, can be built using any combination of construction materials. Stiff foam board is a relatively easy and inexpensive material to use to prepare full-size models. Pieces can be cut to any shape and joined to form 3-D models of racks, consoles, and equipment. (See Figure 10.2.5.) Actual-size drawings of equipment outlines, or more detailed representations, can then be pasted in place on the surfaces of the mockup.

10.2.3e Construction Considerations

Demolition and construction of existing structures may have to be specified by the system engineer. Electrical power, lighting, and air conditioning requirements must be identified and layout drawings prepared for use by the electrical and mechanical engineering consultants and the architect.

During preparation of final construction documents, the architect and the system engineer can confirm the layout of technical equipment wire ways, including access to flooring, conduits, trenches, and overhead raceways. At this point, the system engineer also provides layouts of cable runs and connections. This makes equipment installation and future changes much easier. An overhead cable routing plan is shown in Figure 10.2.6.

When it is necessary to install ac power cables in conduit, follow National Electrical Code (NEC) requirements for conduit fill and the number of pull boxes. More pull boxes or larger con-

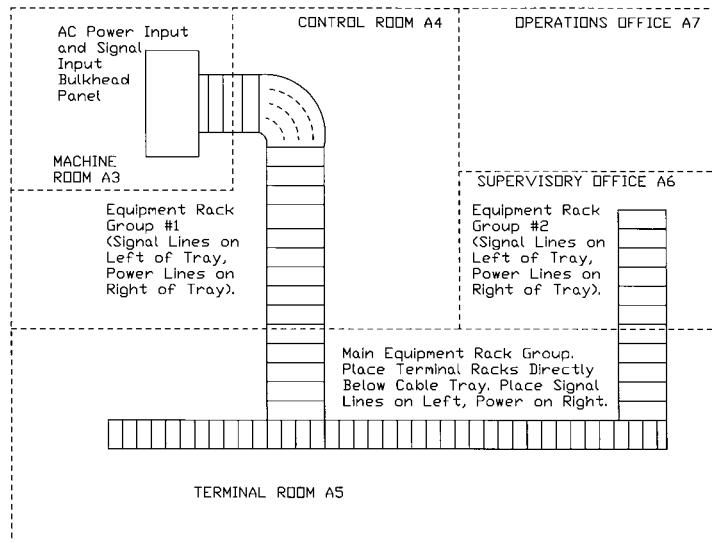


Figure 10.2.6 Cable wire routing plan for a new facility.

duit is required in conduit runs that have many bends. Specify direct-burial-type cable when the conduit or cable trays are underground, and where there is a possibility of standing water. Conduit and cable trays should be designed to accommodate the minimum bend radius requirements for the cables being used.

Debur and remove all sharp edges and splinters from installed conduit. Remove construction debris from inside the lines to prevent damage to the cable jacket during pulling. Cover openings to the conduit to prevent contamination or damage from other construction activities. If the cable is damaged during pulling, moisture could enter the cable.

10.2.3f Component Selection and Installation

Equipment selection is normally based on the function it will perform. User input about operational ease and flexibility of certain models is also important. However, to ensure that the most cost-effective choice is made, consider certain technical issues before making a decision. The system engineer should research, test (when required), and provide the technical input needed for selecting hardware and software.

Equipment features and functional capabilities are probably the main concerns of the users and management. Technical performance data and specifications are important considerations that should be contributed by the system engineer during the selection process. For a piece of equipment to qualify, technical specifications must be checked to ensure that they meet set standards for the overall system. Newly introduced products should be tested and compared before a

10-28 Audio Production Standards, Equipment, and System Design

decision is made. The experienced system engineer can test and measure the equipment to evaluate its performance.

A simple visual inspection inside a piece of equipment by an experienced technician or engineer can uncover possible weaknesses, design flaws, and problem areas that may affect reliability or make maintenance difficult. It is therefore advisable to request a sample of the equipment for evaluation before committing to its use.

The availability of replacement parts is another important consideration when specifying products. A business that depends on its equipment functioning to specifications requires that the service technician be able to make repairs when needed in a timely manner. Learn about manufacturer replacement parts policies and their reliability. When possible, select equipment that uses standard off-the-shelf components that are available from multiple sources. Avoid equipment that incorporates custom components that are available only from the equipment manufacturer. This will make it easier to acquire replacements, and the cost of the parts will—most likely—be less.

When possible, specify products manufactured by the same company. Avoid mixing brands. Maintenance technicians will more easily become familiar with equipment maintenance, and experience gained while repairing one piece of hardware can be directly applied to another of the same model. Service manuals published by the same manufacturer will be similar and therefore easier to understand and use to locate the information or diagram needed for a repair.

Commonality of replacement parts will keep the parts inventory requirements and the inventory cost low. Because the technical staff will be dealing with the manufacturer on a regular basis, familiarity with the company's representatives makes it easier to get technical support quickly.

Sometimes components are selected that are not really compatible, such as differing signal levels or impedances. The responsibility then falls on the system engineer to devise a fix to make that component compatible with the rest of the system. The component may have been originally selected because of its low price, but additional components, engineering, and labor costs often offset the expected savings. Extra wiring and components can also clutter the equipment enclosure, hampering access to the equipment inside. Nonstandard mounting facilities on equipment can add unnecessary cost and can result in a less than elegant solution.

10.2.3g Technical Documentation

Engineering documentation describes the practices and procedures used within the industry to specify a design and communicate the design requirements to technicians and contractors. Documentation preparation should include, but not be limited to, the generation of technical system flow diagrams, material and parts lists, custom item fabrication drawings, and rack and console elevations. The required documents include the following:

- Documentation schedule
- Signal flow diagram
- Equipment schedule
- Cable schedule
- Patch panel assignment schedule
- Rack elevation drawing

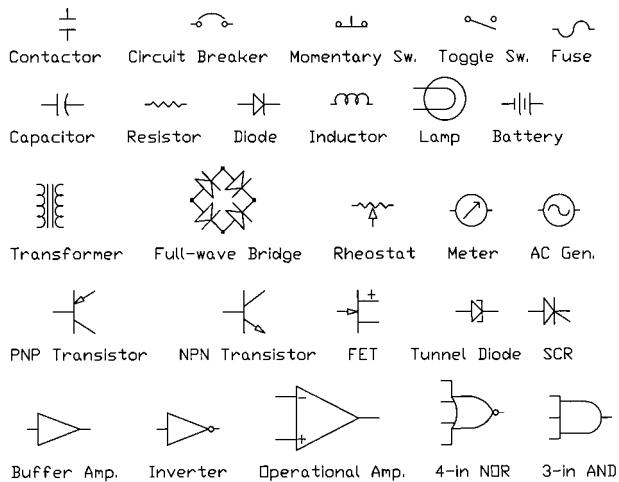


Figure 10.2.7 Schematic representations of common electrical components and devices.

- Construction detail drawing
- Console fabrication mechanical drawing
- Duct and conduit layout drawing
- Single-line electrical flow diagram

Documentation Tracking

The documentation schedule provides a means of keeping track of the project's paperwork. During engineering design, drawings are reviewed, and changes are made. A system for efficiently handling changes is essential, especially on big projects that require a large amount of documentation.

Completed drawings are submitted for management approval. A set of originals is signed by the engineers and managers who are authorized to check the drawings for correctness and to approve the plans.

Symbols

Because there are only limited informal industry standards for the design of electronic component symbols to represent equipment and other elements in a system, custom symbols are usually created by the designer. Each organization develops its own symbols. The symbols that exist apply to component-level devices, such as integrated circuits, resistors, and diodes. Some common symbols apply to system-level components, such as amplifiers and speakers. Figure 10.2.7 shows some of the more common component-level symbols currently used in electronics.

10-30 Audio Production Standards, Equipment, and System Design

The proliferation of manufacturers and equipment types makes it impractical to develop a complete library, but, by following basic rules for symbol design, new component symbols can be produced easily as they are added to the system.

For small systems built with a few simple components, all of the input and output signals can be included on one symbol. However, when the system uses complex equipment with many inputs and outputs with different types of signals, it is usually necessary to draw different diagrams for each type of signal. For this reason, each component requires a set of symbols, with a separate symbol assigned for each signal type, showing its inputs and outputs only. For example, a videotape recorder will require a set of symbols for audio, video, sync, time code, and control signal.

If abbreviations are used, be consistent from one drawing to the next, and develop a dictionary of abbreviations for the drawing set. Include the dictionary with the documentation.

Cross-Referencing Documentation

In order to tie all of the documentation together and to enable fabricators and installers to understand the relationships between the drawings, the documents should include reference designations common to the project. That way, items on one type of document can be located on another type. For example, the location of a piece of equipment can be indicated at its symbol on the flow diagram so that the technician can identify it on the rack elevation drawing and in the actual rack.

A flow diagram is used by the installation technicians to assemble and wire the system components together. All necessary information must be included to avoid confusion and delays. When designing a symbol to represent a component in a flow diagram, include all of the necessary information to identify, locate, and wire that component into the system. The information should include the following:

- Generic description of the component or its abbreviation. When no abbreviation exists, create one. Include it in the project manual reference section and in the notes on the drawing.
- Model number of the component.
- Manufacturer of the component.
- All input and output connections with their respective name and/or number.

Specifications

Specifications are a compilation of knowledge about how something should be done. An engineer condenses years of personal experience, and that of others, into the specification. The more detailed the specification, the higher the probability that the job will be done right.

The project manual is the document where specifications and other printed project documentation is compiled.

10.2.3h Working with the Contractors

The system engineer must provide support and guidance to contractors during the procurement, construction, installation, testing, and acceptance phases of a project. The system engineer can assist in ordering equipment and can help coordinate the move to a new or renovated facility. This can be critical if a great deal of existing equipment is being relocated. In the case of new

equipment, the system engineer's knowledge of prices, features, and delivery times is invaluable to the facility owner.

The steps to assure quality workmanship from contractors on a job include the following:

- Clarify details.
- Clarify misunderstandings about the requirements.
- Resolve problems that may arise.
- Educate contractors about the requirements of the project.
- Assure that the work conforms to the specifications.
- Evaluate and approve change requests.
- Provide technical support to the contractors when needed.

10.2.3i Computer-Based Tools

Technology is evolving so rapidly that it takes a great deal of time just to keep up with the changes. Competition forces change and improvements that would otherwise not take place at such a rapid pace. In this environment, engineering skills must be augmented with tools to speed the design process. Many of the tasks required of the system engineer can be accelerated and the resulting documentation enhanced with the aid of computer programs. Computer aided design (CAD) tools include application software from simple word processing and spreadsheet programs to complex simulation, 3-D graphic modeling, and artificial intelligence. Software is commonly used in new construction and renovations to perform the following tasks:

- Document tracking
- Documentation preparation
- Correspondence
- Report generation
- Technical manual publication
- List management
- Mechanical design
- Electrical design
- Schematic capture

10.2.4 Bibliography

DeSantis, Gene, Jerry C. Whitaker, and C. Robert Paulson: *Interconnecting Electronic Systems*, CRC Press, Boca Raton, Fla., 1992.

Whitaker, Jerry C.: *Facility Design Handbook*, CRC Press, Boca Raton, Fla., 2000.

Section**11**

Broadcast Transmission Systems

Radio broadcasting has been around for a long time. Amplitude modulation (AM) was the first system that permitted voice communications to take place. In this scheme, the magnitude of the carrier wave is varied in accordance with the amplitude and frequency of the applied audio input. The magnitude of the incoming signal determines the magnitude of the carrier wave, while the frequency of the modulating signal determines the rate at which the carrier wave is varied. This simple modulation system was predominant throughout the 1920s and 1930s. Frequency modulation (FM) came into regular broadcast service during the 1940s.

These two basic approaches to communicating over long distances have served the broadcast industry well. While the basic modulating schemes still exist today, numerous enhancements have been made, including stereo operation and subcarrier programming. New technologies have given radio broadcasters new ways to serve their listeners.

Technology has also changed the rules by which the AM and FM broadcasting systems developed. AM radio, as a technical system, offered limited audio fidelity but provided design engineers with a system that allowed uncomplicated transmitters and simple, inexpensive receivers. FM radio, on the other hand, offered excellent audio fidelity but required a more complex and unstable transmitter (in the early days) and complex, expensive receivers. It is, therefore, no wonder that AM radio flourished and FM remained relatively unchanged for at least two decades following its introduction.

Advancements in transmitter and receiver technology during the 1960s changed the picture dramatically. Designers found ways of producing FM transmission equipment that was reliable and capable of high power output. Receiver manufacturers then began producing compact, high-quality receivers at prices consumers could afford. These developments began a significant change in listening habits that is still unfolding today. As late as the mid-1970s, AM radio stations as a whole commanded the majority of all listeners in the U.S. By 1980, however, the tables had turned. FM is now king, and AM is still struggling to keep up with its former stepchild.

As new digital radio systems and services emerge, the radio playing field will again be changed—just how much no one is certain. Stay tuned.

11-2 Section Eleven**In This Section:**

Chapter 11.1: The Electromagnetic Spectrum	11-9
Introduction	11-9
Spectral Sub-Regions	11-10
Optical Spectrum	11-10
Visible Light Band	11-10
IR Band	11-11
UV Band	11-12
DC to Light	11-12
Microwave Band	11-12
Radio Frequency (RF) Band	11-13
Power Frequency (PF)/Telephone Band	11-13
Frequency Band Designations	11-13
Light to Gamma Rays	11-16
X Ray Band	11-16
Gamma Ray Band	11-18
Bibliography	11-18
Chapter 11.2: Propagation	11-21
Introduction	11-21
Propagation in Free Space	11-21
Transmission Loss Between Antennas in Free Space	11-23
Propagation Over Plane Earth	11-25
Field Strengths Over Plane Earth	11-25
Transmission Loss Between Antennas Over Plane Earth	11-29
Propagation Over Smooth Spherical Earth	11-30
Propagation Beyond the Line of Sight	11-30
Effects of Hills, Buildings, Vegetation, and the Atmosphere	11-33
Effects of Hills	11-35
Effects of Buildings	11-38
Effects of Trees and Other Vegetation	11-41
Effects of the Lower Atmosphere (Troposphere)	11-41
Stratification and Ducts	11-42
Tropospheric Scatter	11-42
Atmospheric Fading	11-43
Effects of the Upper Atmosphere (Ionosphere)	11-43
References	11-44
Chapter 11.3: Radio Broadcasting	11-47
Introduction	11-47
Standard Broadcasting	11-47
Modulation Techniques	11-47
Modulation of the AM Carrier Wave	11-48
Channel and Station Classifications	11-49
Field Strength	11-49
Propagation	11-50

Transmitters	11-50
High-Level AM Modulation	11-50
Pulse-Width Modulation	11-51
Solid-State Transmitters	11-53
FM Broadcasting	11-55
Frequency Modulation	11-56
FM Broadcast Modulation Techniques	11-56
Modulation Circuits	11-56
Subsidiary Communications Authorization (SCA)	11-57
Frequency Allocations	11-58
Transmitters	11-59
FM Power Amplifiers	11-59
Tuning the Cavity	11-60
Transmitter Performance and Maintenance	11-60
Key System Measurements	11-61
Synchronous AM in FM Systems	11-61
Incidental Phase Modulation	11-62
Bibliography	11-63

Chapter 11.4: Radio STL Systems **11-65**

Introduction	11-65
Radio vs. Landline Links	11-65
STL System Configuration	11-67
Digital STL vs. Analog STL Link	11-69
Digital STL Performance	11-71
Coding System	11-72
Analog Composite STL Transmitter and Receiver Characteristics	11-73
Components of an STL System	11-73
Transmitter and Receiver Requirements	11-74
Transmission Lines	11-77
Antenna System	11-80
Mounting Structures	11-81
Hardware Considerations	11-81
System Planning	11-82
Frequency Selection	11-83
Spectrum Considerations	11-83
Path Layout	11-83
Line of Sight	11-84
STL Site Selection	11-84
Terrain Considerations	11-84
Path Reliability	11-86
Dealing with Problem Paths	11-91
Hot-Standby Redundant Systems	11-91
Pre-Installation Checkout	11-93
Planning Cable Runs	11-97
Installation	11-97
Checking System Operation	11-99
Troubleshooting	11-99

11-4 Section Eleven

Operating the System	11-99
References	11-100
Bibliography	11-100
Chapter 11.5: Transmitting Antennas	11-103
Introduction	11-103
Operating Characteristics	11-103
Antenna Bandwidth	11-104
Polarization	11-105
Antenna Beamwidth	11-105
Antenna Gain	11-105
Space Regions	11-106
Impedance Matching	11-107
L Network	11-107
Tee Network	11-107
Pi Network	11-108
Line Stretcher	11-109
Antenna Types	11-111
Quarter-Wave Monopole	11-112
Log-Periodic Antenna	11-112
Yagi-Uda Antenna	11-113
Waveguide Antenna	11-113
Horn Antenna	11-114
Reflector Antenna	11-114
Spiral Antenna	11-115
Array Antenna	11-115
AM Broadcast Antenna Systems	11-116
Directional AM Antenna Design	11-116
Antenna Pattern Design	11-117
Bearing	11-120
Antenna Monitoring System	11-120
Phase/Current Sample Loop	11-120
Sample lines	11-121
The Common Point	11-121
Monitor Points	11-122
Folded Unipole Antenna	11-123
FM Broadcast Antenna Systems	11-123
Antenna Types	11-124
Bibliography	11-126
Chapter 11.6: RF Combiner and Diplexer Systems	11-127
Introduction	11-127
Passive Filters	11-127
Filter Type	11-128
Filter Alignment	11-128
Filter Order	11-129
Four-Port Hybrid Combiner	11-130
Microwave Combiners	11-134

Hot Switching Combiners	11-136
High-Power Isolators	11-138
Theory of Operation	11-139
Applications	11-140
Hot Switch	11-141
Diplexer	11-141
Multiplexer	11-142
References	11-142
Bibliography	11-143

On the CD-ROM:



- “Audio Tests and Measurements” by Richard Cabot, an archive chapter from the first edition of the *Audio Engineering Handbook*. This chapter provides an extensive overview of audio test and measurement techniques.

Reference Documents for this Section

Bean, B. R., and E. J. Dutton: “Radio Meteorology,” National Bureau of Standards Monograph 92, March 1, 1966.

Benson, B., and Whitaker, J.: *Television and Audio Handbook for Technicians and Engineers*, McGraw-Hill, New York, N.Y., 1989.

Bingeman, Grant: “AM Tower Impedance Matching,” *Broadcast Engineering*, Intertec Publishing, Overland Park, Kan., July 1985.

Bixby, Jeffrey: “AM DAs—Doing it Right,” *Broadcast Engineering*, Intertec Publishing, Overland Park, Kan., February 1984.

Bordonaro, Dominic: “Reducing IPM in AM Transmitters,” *Broadcast Engineering*, Intertec Publishing, Overland Park, Kan. May 1988.

Bullington, K.: “Radio Propagation at Frequencies above 30 Mc,” *Proc. IRE*, pg. 1122, October 1947.

Bullington, K.: “Radio Propagation Variations at VHF and UHF,” *Proc. IRE*, pg. 27, January 1950.

Burrows, C. R., and M. C. Gray: “The Effect of the Earth’s Curvature on Groundwave Propagation,” *Proc. IRE*, pg. 16, January 1941.

Chick, Elton B.: “Monitoring Directional Antennas,” *Broadcast Engineering*, Intertec Publishing, Overland Park, Kan., July 1985.

Collocott, T. C., A. B. Dobson, and W. R. Chambers (eds.): *Dictionary of Science & Technology*.

Crutchfield, E. B. (ed): *NAB Engineering Handbook*, 8th ed., National Association of Broadcasters, Washington, D.C., 1992.

DeComier, Bill: “Inside FM Multiplexer Systems,” *Broadcast Engineering*, Intertec Publishing, Overland Park, Kan., May 1988.

de Lisle, E. W.: “Computations of VHF and UHF Propagation for Radio Relay Applications,” RCA, Report by International Division, New York, N.Y.

11-6 Section Eleven

- Dickson, F. H., J. J. Egli, J. W. Herbstreit, and G. S. Wickizer: "Large Reductions of VHF Transmission Loss and Fading by the Presence of a Mountain Obstacle in Beyond-Line-of-Sight Paths," *Proc. IRE*, vol. 41, no. 8, pg. 96, August 1953.
- "Documents of the XVth Plenary Assembly," CCIR Report 238, vol. 5, Geneva, 1982.
- "Documents of the XVth Plenary Assembly," CCIR Report 563, vol. 5, Geneva, 1982.
- "Documents of the XVth Plenary Assembly," CCIR Report 881, vol. 5, Geneva, 1982.
- Dougherty, H. T., and E. J. Dutton: "The Role of Elevated Ducting for Radio Service and Interference Fields," NTIA Report 81-69, March 1981.
- Eckersley, T. L.: "Ultra-Short-Wave Refraction and Diffraction," *J. Inst. Elec. Engrs.*, pg. 286, March 1937.
- Epstein, J., and D. Peterson: "An Experimental Study of Wave Propagation at 850 Mc," *Proc. IRE*, pg. 595, May 1953.
- Fink, D. G., (ed.): *Television Engineering Handbook*, McGraw-Hill, New York, N.Y., 1957.
- Fink, D., and D. Christiansen (eds.): *Electronics Engineer's Handbook*, 3rd ed., McGraw-Hill, New York, N.Y., 1989.
- Handbook of Physics*, McGraw-Hill, New York, N.Y., 1958.
- Harrison, Cecil: "Passive Filters," in *The Electronics Handbook*, Jerry C. Whitaker (ed.), CRC Press, Boca Raton, Fla., pp. 279-290, 1996.
- Hauptstuek, Jim: "Interconnecting the Digital Chain," *NAB 1996 Broadcast Engineering Conference Proceedings*, National Association of Broadcasters, Washington, D.C., pp. 360-358, 1996.
- Heymans, Dennis: "Hot Switches and Combiners," *Broadcast Engineering*, Overland Park, Kan., December 1987.
- Jordan, Edward C. (ed.): *Reference Data for Engineers—Radio, Electronics, Computer and Communications*, 7th ed., Howard W. Sams, Indianapolis, Ind., 1985.
- Judd, D. B., and G. Wyszecki: *Color in Business, Science and Industry*, 3rd ed., John Wiley and Sons, New York, N.Y.
- Kaufman, Ed: *IES Illumination Handbook*, Illumination Engineering Society.
- Lapedes, D. N. (ed.): *The McGraw-Hill Encyclopedia of Science & Technology*, 2nd ed., McGraw-Hill, New York, N.Y.
- Longley, A. G., and P. L. Rice: "Prediction of Tropospheric Radio Transmission over Irregular Terrain—A Computer Method," ESSA (Environmental Science Services Administration), U.S. Dept. of Commerce, Report ERL (Environment Research Laboratories) 79-ITS 67, July 1968.
- McClanahan, M. E.: "Aural Broadcast Auxiliary Links," in *NAB Engineering Handbook*, 8th ed., E. B. Crutchfield (ed.), National Association of Broadcasters, Washington, D.C., pp. 671-678, 1992.

- McPetrie, J. S., and L. H. Ford: "An Experimental Investigation on the Propagation of Radio Waves over Bare Ridges in the Wavelength Range 10 cm to 10 m," *J. Inst. Elec. Engrs.*, pt. 3, vol. 93, pg. 527, 1946.
- Megaw, E. C. S.: "Some Effects of Obstacles on the Propagation of Very Short Radio Waves," *J. Inst. Elec. Engrs.*, pt. 3, vol. 95, no. 34, pg. 97, March 1948.
- Mullaney, John H.: "The Folded Unipole Antenna for AM Broadcast," *Broadcast Engineering*, Intertec Publishing, Overland Park, Kan., January 1960.
- Mullaney, John H.: "The Folded Unipole Antenna," *Broadcast Engineering*, Intertec Publishing, Overland Park, Kan., July 1986.
- National Bureau of Standards Circular 462, "Ionospheric Radio Propagation," June 1948.
- NIST: *Manual of Regulations and Procedures for Federal Radio Frequency Management*, September 1995 edition, revisions for September 1996, January and May 1997, NTIA, Washington, D.C., 1997.
- Norgard, John: "Electromagnetic Spectrum," *NAB Engineering Handbook*, 9th ed., Jerry C. Whitaker (ed.), National Association of Broadcasters, Washington, D.C., 1999.
- Norgard, John: "Electromagnetic Spectrum," *The Electronics Handbook*, Jerry C. Whitaker (ed.), CRC Press, Boca Raton, Fla., 1996.
- Norton, K. A.: "Ground Wave Intensity over a Finitely Conducting Spherical Earth," *Proc. IRE*, pg. 622, December 1941.
- Norton, K. A.: "The Propagation of Radio Waves over a Finitely Conducting Spherical Earth," *Phil. Mag.*, June 1938.
- Parker, Darryl: "TFT DMM92 Meets STL Requirements," *Radio World*, Falls Church, VA, October 21, 1992.
- "Radio Wave Propagation," Summary Technical Report of the Committee on Propagation of the National Defense Research Committee, Academic Press, New York, N.Y., 1949.
- "Report of the Ad Hoc Committee, Federal Communications Commission," vol. 1, May 1949; vol. 2, July 1950.
- Rollins, William W., and Robert L. Band: "T1 Digital STL: Discrete vs. Composite Transmission," *NAB 1996 Broadcast Engineering Conference Proceedings*, National Association of Broadcasters, Washington, D.C., pp. 356–359, 1996.
- Salek, Stanley: "Analysis of FM Booster System Configurations," *Proceedings of the 1992 NAB Broadcast Engineering Conference*, National Association of Broadcasters, Washington, DC, April 1992.
- Selvidge, H.: "Diffraction Measurements at Ultra High Frequencies," *Proc. IRE*, pg. 10, January 1941.
- Smith, E. E., and E. W. Davis: "Wind-induced Ions Thwart TV Reception," *IEEE Spectrum*, pp. 52–55, February 1981.
- Stemson, A: *Photometry and Radiometry for Engineers*, John Wiley and Sons, New York, N.Y.

11-8 Section Eleven

- Stenberg, James T.: "Using Super Power Isolators in the Broadcast Plant," *Proceedings of the Broadcast Engineering Conference*, Society of Broadcast Engineers, Indianapolis, IN, 1988.
- Surette, Robert A.: "Combiners and Combining Networks," in *The Electronics Handbook*, Jerry C. Whitaker (ed.), CRC Press, Boca Raton, Fla., pp. 1368–1381, 1996.
- The Cambridge Encyclopedia*, Cambridge University Press, 1990.
- The Columbia Encyclopedia*, Columbia University Press, 1993.
- "The Propagation of Radio Waves through the Standard Atmosphere," *Summary Technical Report of the Committee on Propagation*, vol. 3, National Defense Research Council, Washington, D.C., 1946, published by Academic Press, New York, N.Y.
- van der Pol, Balth., and H. Bremmer: "The Diffraction of Electromagnetic Waves from an Electrical Point Source Round a Finitely Conducting Sphere, with Applications to Radiotelegraphy and to Theory of the Rainbow," pt. 1, *Phil. Mag.*, July, 1937; pt. 2, *Phil. Mag.*, November 1937.
- Vaughan, T., and E. Pivit: "High Power Isolator for UHF Television," *Proceedings of the NAB Engineering Conference*, National Association of Broadcasters, Washington, D.C., 1989.
- Webster's New World Encyclopedia*, Prentice Hall, 1992.
- Westberg, J. M.: "Effect of 90° Stub on Medium Wave Antennas," *NAB Engineering Handbook*, 7th ed., National Association of Broadcasters, Washington, D.C., 1985.
- Whitaker, Jerry C., (ed.): *A Primer: Digital Aural Studio to Transmitter Links*, TFT, Santa Clara, CA, 1994.
- Whitaker, Jerry C., and Skip. Pizzi: "Radio Electronic News Gathering and Field Production," in *NAB Engineering Handbook*, 8th ed., E. B. Crutchfield (ed.), National Association of Broadcasters, Washington, D.C., pp. 1051–1072, 1992.
- Wyszecki, G., and W. S. Stiles: *Color Science, Concepts and Methods, Quantitative Data and Formulae*, 2nd ed., John Wiley and Sons, New York, N.Y.

Chapter**11.1****The Electromagnetic Spectrum****John Norgard****11.1.1 Introduction**

The electromagnetic (EM) spectrum consists of all forms of EM radiation—EM waves (radian energy) propagating through space, from DC to light to gamma rays. The EM spectrum can be arranged in order of frequency and/or wavelength into a number of regions, usually wide in extent, within which the EM waves have some specified common characteristics, such as characteristics relating to the production or detection of the radiation. A common example is the spectrum of the radiant energy in white light, as dispersed by a prism, to produce a “rainbow” of its constituent colors. Specific frequency ranges are often called *bands*; several contiguous frequency bands are usually called *spectrums*; and sub-frequency ranges within a band are sometimes called *segments*.

The EM spectrum can be displayed as a function of frequency (or wavelength). In air, frequency and wavelength are inversely proportional, $f = c/\lambda$ (where $c \approx 3 \times 10^8$ m/s, the speed of light in a vacuum). The MKS unit of frequency is the Hertz and the MKS unit of wavelength is the meter. Frequency is also measured in the following sub-units:

- Kilohertz, 1 kHz = 10^3 Hz
- Megahertz, 1 MHz = 10^6 Hz
- Gigahertz, 1 GHz = 10^9 Hz
- Terahertz, 1 THz = 10^{12} Hz
- Petahertz, 1 PHz = 10^{15} Hz
- Exahertz, 1 EHz = 10^{18} Hz

Or for very high frequencies, *electron volts*, 1 ev $\sim 2.41 \times 10^{14}$ Hz

Wavelength is also measured in the following sub-units:

- Centimeters, 1 cm = 10^{-2} m
- Millimeters, 1 mm = 10^{-3} m
- Micrometers, 1 μm = 10^{-6} m (microns)
- Nanometers, 1 nm = 10^{-9} m

11-10 Broadcast Transmission Systems

- Ångstroms, $1 \text{ \AA} = 10^{-10} \text{ m}$
- Picometers, $1 \text{ pm} = 10^{-12} \text{ m}$
- Femtometers, $1 \text{ fm} = 10^{-15} \text{ m}$
- Attometers, $1 \text{ am} = 10^{-18} \text{ m}$

11.1.2 Spectral Sub-Regions

For convenience, the overall EM spectrum can be divided into three main sub-regions:

- *Optical spectrum*
- *DC to light spectrum*
- *Light to gamma ray spectrum*

These main sub-regions of the EM spectrum are next discussed. Note that the boundaries between some of the spectral regions are somewhat arbitrary. Certain spectral bands have no sharp edges and merge into each other, while other spectral segments overlap each other slightly.

11.1.2a Optical Spectrum

The optical spectrum is the “middle” frequency/wavelength region of the EM spectrum. It is defined here as the visible and near-visible regions of the EM spectrum and includes:

- The *infrared (IR)* band, circa $300 \mu\text{m}$ – $0.7 \mu\text{m}$ (circa 1 THz–429 THz)
- The *visible light* band, $0.7 \mu\text{m}$ – $0.4 \mu\text{m}$ (429 THz–750 THz)
- The *ultraviolet (UV)* band, $0.4 \mu\text{m}$ –circa 10 nm (750 THz–circa 30 PHz), approximately 100 eV

These regions of the EM spectrum are usually described in terms of their wavelengths.

Atomic and molecular radiation produce radiant light energy. Molecular radiation and radiation from hot bodies produce EM waves in the IR band. Atomic radiation (outer shell electrons) and radiation from arcs and sparks produce EM waves in the UV band.

Visible Light Band

In the “middle” of the optical spectrum is the visible light band, extending approximately from $0.4 \mu\text{m}$ (violet) up to $0.7 \mu\text{m}$ (red), i.e. from 750 THz (violet) down to 429 THz (red). EM radiation in this region of the EM spectrum, when entering the eye, gives rise to visual sensations (colors), according to the spectral response of the eye, which responds only to radiant energy in the visible light band extending from the extreme long wavelength edge of red to the extreme short wavelength edge of violet. (The spectral response of the eye is sometimes quoted as extending from $0.38 \mu\text{m}$ (violet) up to 0.75 or $0.78 \mu\text{m}$ (red); i.e., from 789 THz down to 400 or 385 THz.) This visible light band is further subdivided into the various colors of the rainbow, in decreasing wavelength/increasing frequency:

- Red, a primary color, peak intensity at 700.0 nm (429 THz)

- Orange
- Yellow
- Green, a primary color, peak intensity at 546.1 nm (549 THz)
- Cyan
- Blue, a primary color, peak intensity at 435.8 nm (688 THz)
- Indigo
- Violet

IR Band

The IR band is the region of the EM spectrum lying immediately below the visible light band. The IR band consists of EM radiation with wavelengths extending between the longest visible red (circa 0.7 μm) and the shortest microwaves (300 μm –1 mm), i.e., from circa 429 THz down to 1 THz–300 GHz.

The IR band is further subdivided into the “near” (shortwave), “intermediate” (midwave), and “far” (longwave) IR segments as follows¹:

- *Near* IR segment, 0.7 μm up to 3 μm (429 THz down to 100 THz)
- *Intermediate* IR segment, 3 μm up to 7 μm (100 THz down to 42.9 THz)
- *Far* IR segment, 7 μm up to 300 μm (42.9 THz down to 1 THz)
- Sub-millimeter band, 100 μm up to 1 mm (3 THz down to 300 GHz). Note that the sub-millimeter region of wavelengths is sometimes included in the very far region of the IR band.

EM radiation is produced by oscillating and rotating molecules and atoms. Therefore, all objects at temperatures above absolute zero emit EM radiation by virtue of their thermal motion (warmth) alone. Objects near room temperature emit most of their radiation in the IR band. However, even relatively cool objects emit some IR radiation; hot objects, such as incandescent filaments, emit strong IR radiation.

IR radiation is sometimes incorrectly called “radian heat” because warm bodies emit IR radiation and bodies that absorb IR radiation are warmed. However, IR radiation is not itself “heat”. This radiant energy is called “black body” radiation. Such waves are emitted by all material objects. For example, the background cosmic radiation (2.7K) emits microwaves; room temperature objects (293K) emit IR rays; the Sun (6000K) emits yellow light; the Solar Corona (1 million K) emits X rays.

IR astronomy uses the 1 μm to 1 mm part of the IR band to study celestial objects by their IR emissions. IR detectors are used in night vision systems, intruder alarm systems, weather forecasting, and missile guidance systems. IR photography uses multilayered color film, with an IR sensitive emulsion in the wavelengths between 700–900 nm, for medical and forensic applications, and for aerial surveying.

1. Some reference texts use 2.5 mm (120 THz) as the breakpoint between the near and the intermediate IR bands and 10 mm (30 THz) as the breakpoint between the intermediate and the far IR bands. Also, 15 mm (20 Thz) is sometimes considered as the long wavelength end of the far IR band.

11-12 Broadcast Transmission Systems

UV Band

The UV band is the region of the EM spectrum lying immediately above the visible light band. The UV band consists of EM radiation with wavelengths extending between the shortest visible violet (circa 0.4 μm) and the longest X rays (circa 10 nm), i.e., from 750 THz—approximately 3 ev—up to circa 30 PHz—approximately 100 ev.²

The UV band is further subdivided into the “near” and the “far” UV segments as follows:

- “Near” UV segment, circa 0.4 μm down to 100 nm (circa 750 THz up to 3 PHz, approximately 3 ev up to 10 ev)
- “Far” UV segment, 100 nm down to circa 10 nm, (3 PHz up to circa 30 PHz, approximately 10 ev up to 100 ev)

The far UV band is also referred to as the *vacuum UV band*, since air is opaque to all UV radiation in this region.

UV radiation is produced by electron transitions in atoms and molecules, as in a mercury discharge lamp. Radiation in the UV range is easily detected and can cause fluorescence in some substances, and can produce photographic and ionizing effects.

In UV astronomy, the emissions of celestial bodies in the wavelength band between 50–320 nm are detected and analyzed to study the heavens. The hottest stars emit most of their radiation in the UV band.

11.1.2b DC to Light

Below the IR band are the lower frequency (longer wavelength) regions of the EM spectrum, subdivided generally into the following spectral bands (by frequency/wavelength):

- *Microwave* band, 300 GHz down to 300 MHz (1 mm up to 1 m). Some reference works define the lower edge of the microwave spectrum at 1 GHz.
- *Radio frequency (RF)* band, 300 MHz down to 10 kHz (1 m up to 30 Km)
- *Power (PF)/telephony* band, 10 kHz down to dc (30 Km up to ∞)

These regions of the EM spectrum are usually described in terms of their frequencies.

Radiations whose wavelengths are of the order of millimeters & centimeters are called *microwaves*, and those still longer are called radio frequency (RF) waves (or *Hertzian waves*).

Radiation from electronic devices produces EM waves in both the microwave and RF bands. Power frequency energy is generated by rotating machinery. Direct current (dc) is produced by batteries or rectified alternating current (ac).

Microwave Band

The microwave band is the region of wavelengths lying between the far IR/sub-millimeter region and the conventional RF region. The boundaries of the microwave band have not been definitely fixed, but it is commonly regarded as the region of the EM spectrum extending from about 1 mm up to 1 m in wavelengths, i.e. from 300 GHz down to 300 MHz. The microwave band is further sub-divided into the following segments:

2. Some references use 4, 5, or 6 nm as the upper edge of the UV band.

The Electromagnetic Spectrum 11-13

- *Millimeter* waves, 300 GHz down to 30 GHz (1 mm up to 1 cm); the EHF band. (Some references consider the top edge of the millimeter region to stop at 100 GHz.)
- *Centimeter* waves, 30 GHz down to 3 GHz (1 cm up to 10 cm); the SHF band.

The microwave band usually includes the UHF band from 3 GHz down to 300 MHz (from 10 cm up to 1 m).

Microwaves are used in radar, in communication links spanning moderate distances, as radio carrier waves in television broadcasting, for mechanical heating, and cooking in microwave ovens.

Radio Frequency (RF) Band

The RF range of the EM spectrum is the wavelength band suitable for utilization in radio communications extending from 10 kHz up to 300 MHz (from 30 Km down to 1 m). (Some references consider the RF band as extending from 10 kHz to 300 GHz, with the microwave band as a subset of the RF band from 300 MHz to 300 GHz.)

Some of the radio waves in this band serve as the carriers of low-frequency audio signals; other radio waves are modulated by video and digital information. The *amplitude modulated* (AM) broadcasting band uses waves with frequencies between 550–1640 kHz; the *frequency modulated* (FM) broadcasting band uses waves with frequencies between 88–108 MHz.

In the U.S., the Federal Communications Commission (FCC) is responsible for assigning a range of frequencies to specific services. The International Telecommunications Union (ITU) coordinates frequency band allocation and cooperation on a worldwide basis.

Radio astronomy uses radio telescopes to receive and study radio waves naturally emitted by objects in space. Radio waves are emitted from hot gases (*thermal radiation*), from charged particles spiraling in magnetic fields (*synchrotron radiation*), and from excited atoms and molecules in space (*spectral lines*), such as the 21 cm line emitted by hydrogen gas.

Power Frequency (PF)/Telephone Band

The PF range of the EM spectrum is the wavelength band suitable for generating, transmitting, and consuming low frequency power, extending from 10 kHz down to dc (zero frequency), i.e., from 30Km up in wavelength. In the US, most power is generated at 60 Hz (some military and computer applications use 400 Hz); in other countries, including Europe, power is generated at 50 Hz.

Frequency Band Designations

The combined microwave, RF (Hertzian Waves), and power/telephone spectra are subdivided into the specific bands given in Table 11.1.1, which lists the international radio frequency band designations and the numerical designations. Note that the band designated (12) has no commonly used name or abbreviation.

The radar band often is considered to extend from the middle of the HF (7) band to the end of the EHF (11) band. The current US Tri-Service radar band designations are listed in Table 11.1.2.

An alternate and more detailed sub-division of the UHF (9), SHF (10), and EHF (11) bands is given in Table 11.1.3.

Several other frequency bands of interest (not exclusive) are listed in Tables 11.1.4–11.1.6.

11-14 Broadcast Transmission Systems

Table 11.1.1 Frequency Band Designations

Description	Band Designation	Frequency	Wavelength
Extremely Low Frequency	ELF (1) Band	3 Hz up to 30 Hz	100 Mm down to 10 Mm
Super Low Frequency	SLF (2) Band	30 Hz up to 300 Hz	10 Mm down to 1 Mm
Ultra Low Frequency	ULF (3) Band	300 Hz up to 3 kHz	1 Mm down to 100 Km
Very Low Frequency	VLF (4) Band	3 kHz up to 30 kHz	100 Km down to 10 Km
Low Frequency	LF (5) Band	30 kHz up to 300 kHz	10 Km down to 1 Km
Medium Frequency	MF (6) Band	300 kHz up to 3 MHz	1 Km down to 100 m
High Frequency	HF (7) Band	3 MHz up to 30 MHz	100 m down to 10 m
Very High Frequency	VHF (8) Band	30 MHz up to 300 MHz	10 m down to 1 m
Ultra High Frequency	UHF (9) Band	300 MHz up to 3 GHz	1 m down to 10 cm
Super High Frequency	SHF (10) Band	3 GHz up to 30 GHz	10 cm down to 1 cm
Extremely High Frequency	EHF (11) Band	30 GHz up to 300 GHz	1 cm down to 1 mm
—	(12) Band	300 GHz up to 3 THz	1 mm down to 100 μ

Table 11.1.2 Radar Band Designations

Band	Frequency	Wavelength
A Band	0 Hz up to 250 MHz	∞ down to 1.2 m
B Band	250 MHz up to 500 MHz	1.2 m down to 60 cm
C Band	500 MHz up to 1 GHz	60 cm down to 30 cm
D Band	1 GHz up to 2 GHz	30 cm down to 15 cm
E Band	2 GHz up to 3 GHz	15 cm down to 10 cm
F Band	3 GHz up to 4 GHz	10 cm down to 7.5 cm
G Band	4 GHz up to 6 GHz	7.5 cm down to 5 cm
H Band	6 GHz up to 8 GHz	5 cm down to 3.75 cm
I Band	8 GHz up to 10 GHz	3.75 cm down to 3 cm
J Band	10 GHz up to 20 GHz	3 cm down to 1.5 cm
K Band	20 GHz up to 40 GHz	1.5 cm down to 7.5 mm
L Band	40 GHz up to 60 GHz	7.5 mm down to 5 mm
M Band	60 GHz up to 100 GHz	5 mm down to 3 mm
N Band	100 GHz up to 200 GHz	3 mm down to 1.5 mm
O Band	200 GHz up to 300 GHz	1.5 mm down to 1 mm

Table 11.1.3 Detail of UHF, SHF, and EHF Band Designations

Band	Frequency	Wavelength
L Band	1.12 GHz up to 1.7 GHz	26.8 cm down to 17.6 cm
LS Band	1.7 GHz up to 2.6 GHz	17.6 cm down to 11.5 cm
S Band	2.6 GHz up to 3.95 GHz	11.5 cm down to 7.59 cm
C(G) Band	3.95 GHz up to 5.85 GHz	7.59 cm down to 5.13 cm
XN(J, XC) Band	5.85 GHz up to 8.2 GHz	5.13 cm down to 3.66 cm
XB(H, BL) Band	7.05 GHz up to 10 GHz	4.26 cm down to 3 cm
X Band	8.2 GHz up to 12.4 GHz	3.66 cm down to 2.42 cm
Ku(P) Band	12.4 GHz up to 18 GHz	2.42 cm down to 1.67 cm
K Band	18 GHz up to 26.5 GHz	1.67 cm down to 1.13 cm
V(R, Ka) Band	26.5 GHz up to 40 GHz	1.13 cm down to 7.5 mm
Q(V) Band	33 GHz up to 50 GHz	9.09 mm down to 6 mm
M(W) Band	50 GHz up to 75 GHz	6 mm down to 4 mm
E(Y) Band	60 GHz up to 90 GHz	5 mm down to 3.33 mm
F(N) Band	90 GHz up to 140 GHz	3.33 mm down to 2.14 mm
G(A)	140 GHz p to 220 GHz	2.14 mm down to 1.36 mm
R Band	220 GHz up to 325 GHz	1.36 mm down to 0.923 mm

Table 11.1.4 Low Frequency Bands of Interest

Band	Frequency
Sub-sonic band	0 Hz–10 Hz
Audio band	10 Hz–10 kHz
Ultra-sonic band	10 kHz and up

Table 11.1.5 Applications of Interest in the RF Band

Band	Frequency
Longwave broadcasting band	150–290 kHz
AM broadcasting band	550–1640 kHz (1.640 MHz), 107 channels, 10 kHz separation
International broadcasting band	3–30 MHz
Shortwave broadcasting band	5.95–26.1 MHz (8 bands)
VHF TV (Channels 2 - 4)	54–72 MHz
VHF TV (Channels 5 - 6)	76–88 MHz
FM broadcasting band	88–108 MHz
VHF TV (Channels 7 - 13)	174–216 MHz
UHF TV (Channels 14 - 69)	512–806 MHz

11-16 Broadcast Transmission Systems

11.1.2c Light to Gamma Rays

Above the UV spectrum are the higher frequency (shorter wavelength) regions of the EM spectrum, subdivided generally into the following spectral bands (by frequency/wavelength):

- *X ray* band, approximately 10 ev up to 1 Mev (circa 10 nm down to circa 1 pm), circa 3 PHz up to circa 300 EHz
- *Gamma ray* band, approximately 1 Kev up to ∞ (circa 300 pm down to 0 m), circa 1 EHz up to ∞

These regions of the EM spectrum are usually described in terms of their photon energies in electron volts. Note that the bottom of the gamma ray band overlaps the top of the X ray band.

It should be pointed out that *cosmic “rays”* (from astronomical sources) are not EM waves (rays) and, therefore, are not part of the EM spectrum. Cosmic “rays” are high energy charged particles (electrons, protons, and ions) of extraterrestrial origin moving through space, which may have energies as high as 10^{20} ev. Cosmic “rays” have been traced to cataclysmic astrophysical/cosmological events, such as exploding stars and black holes. Cosmic “rays” are emitted by supernova remnants, pulsars, quasars, and radio galaxies. Cosmic “rays” that collide with molecules in the Earth’s upper atmosphere produce secondary cosmic “rays” and gamma rays of high energy that also contribute to natural background radiation. These gamma rays are sometimes called “cosmic” or *secondary* gamma rays. Cosmic rays are a useful source of high-energy particles for certain scientific experiments.

Radiation from atomic inner shell excitations produces EM waves in the X ray band. Radiation from naturally radioactive nuclei produces EM waves in the gamma ray band.

X Ray Band

The X ray band is further sub-divided into the following segments:

- *Soft X rays*, approximately 10 ev up to 10 Kev (circa 10 nm down to 100 pm), circa 3 PHz up to 3 EHz
- *Hard X rays*, approximately 10 Kev up to 1Mev (100 pm down to circa 1 pm), 3 EHz up to circa 300 EHz

Because the physical nature of these rays was first unknown, this radiation was called X rays. The more powerful X rays are called hard X rays and are of high frequencies and, therefore, are more energetic; less powerful X rays are called soft X rays and have lower energies.

X rays are produced by transitions of electrons in the inner levels of excited atoms or by rapid deceleration of charged particles (*Bremsstrahlung* or breaking radiation). An important source of X rays is *synchrotron radiation*. X rays can also be produced when high energy electrons from a heated filament cathode strike the surface of a target anode (usually tungsten) between which a high alternating voltage (approximately 100 kV) is applied.

X rays are a highly penetrating form of EM radiation and applications of X rays are based on their short wavelengths and their ability to easily pass through matter. X rays are very useful in crystallography for determining crystalline structure and in medicine for photographing the body. Because different parts of the body absorb X rays to a different extent, X rays passing through the body provide a visual image of its interior structure when striking a photographic plate. X rays are dangerous and can destroy living tissue. They can also cause severe skin burns. X rays are useful in the diagnosis and non-destructive testing of products for defects.

Table 11.1.6 Applications of Interest in the Microwave Band (up to 40 GHz):

Application	Frequency
Aero Navigation	0.96–1.215 GHz
GPS Down Link	1.2276 GHz
Military COM/Radar	1.35–1.40 GHz
Miscellaneous COM/Radar	1.40–1.71 GHz
L-Band Telemetry	1.435–1.535 GHz
GPS Down Link	1.57 GHz
Military COM (Troposcatter/Telemetry)	1.71–1.85 GHz
Commercial COM & Private LOS	1.85–2.20 GHz
Microwave Ovens	2.45 GHz
Commercial COM/Radar	2.45–2.69 GHz
Instructional TV	2.50–2.69 GHz
Military Radar (Airport Surveillance)	2.70–2.90 GHz
Maritime Navigation Radar	2.90–3.10 GHz
Miscellaneous Radars	2.90–3.70 GHz
Commercial C-Band SAT COM Down Link	3.70–4.20 GHz
Radar Altimeter	4.20–4.40 GHz
Military COM (Troposcatter)	4.40–4.99 GHz
Commercial Microwave Landing System	5.00–5.25 GHz
Miscellaneous Radars	5.25–5.925 GHz
C-Band Weather Radar	5.35–5.47 GHz
Commercial C-Band SAT COM Up Link	5.925–6.425 GHz
Commercial COM	6.425–7.125 GHz
Mobile TV Links	6.875–7.125 GHz
Military LOS COM	7.125–7.25 GHz
Military SAT COM Down Link	7.25–7.75 GHz
Military LOS COM	7.75–7.9 GHz
Military SAT COM Up Link	7.90–8.40 GHz
Miscellaneous Radars	8.50–10.55 GHz
Precision Approach Radar	9.00–9.20 GHz
X-Band Weather Radar (& Maritime Navigation Radar)	9.30–9.50 GHz
Police Radar	10.525 GHz
Commercial Mobile COM (LOS & ENG)	10.55–10.68 GHz
Common Carrier LOS COM	10.70–11.70 GHz
Commercial COM	10.70–13.25 GHz
Commercial Ku-Band SAT COM Down Link	11.70–12.20 GHz
DBS Down Link & Private LOS COM	12.20–12.70 GHz
ENG & LOS COM	12.75–13.25 GHz
Miscellaneous Radars & SAT COM	13.25–14.00 GHz
Commercial Ku-Band SAT COM Up Link	14.00–14.50 GHz
Military COM (LOS, Mobile, & Tactical)	14.50–15.35 GHz
Aero Navigation	15.40–15.70 GHz
Miscellaneous Radars	15.70–17.70 GHz
DBS Up Link	17.30–17.80 GHz
Common Carrier LOS COM	17.70–19.70 GHz
Commercial COM (SAT COM & LOS)	17.70–20.20 GHz
Private LOS COM	18.36–19.04 GHz
Military SAT COM	20.20–21.20 GHz
Miscellaneous COM	21.20–24.00 GHz
Police Radar	24.15 GHz
Navigation Radar	24.25–25.25 GHz
Military COM	25.25–27.50 GHz
Commercial COM	27.50–30.00 GHz
Military SAT COM	30.00–31.00 GHz
Commercial COM	31.00–31.20 GHz

11-18 Broadcast Transmission Systems

Gamma Ray Band

The gamma ray band is sub-divided into the following segments:

- *Primary* gamma rays, approximately 1 Kev up to 1 Mev (circa 300 pm down to 300 fm), circa 1 EHHz up to 1000 EHHz
- *Secondary* gamma rays, approximately 1 Mev up to ∞ (300 fm down to 0 m), 1000 EHHz up to ∞

Secondary gamma rays are created from collisions of high energy cosmic rays with particles in the Earth's upper atmosphere.

The primary gamma rays are further sub-divided into the following segments:

- *Soft* gamma rays, approximately 1 Kev up to circa 300 Kev (circa 300 pm down to circa 3 pm), circa 1 EHHz up to circa 100 EHHz
- *Hard* gamma rays, approximately 300 Kev up to 1 Mev (circa 3 pm down to 300 fm), circa 100 EHHz up to 1000 EHHz

Gamma rays are essentially very energetic X rays. The distinction between the two is based on their origin. X rays are emitted during atomic processes involving energetic electrons; gamma rays are emitted by excited nuclei or other processes involving sub-atomic particles.

Gamma rays are emitted by the nucleus of radioactive material during the process of natural radioactive decay as a result of transitions from high energy excited states to low energy states in atomic nuclei. Cobalt 90 is a common gamma ray source (with a half-life of 5.26 years). Gamma rays are also produced by the interaction of high energy electrons with matter. "Cosmic" gamma rays cannot penetrate the Earth's atmosphere.

Applications of gamma rays are found both in medicine and in industry. In medicine, gamma rays are used for cancer treatment, diagnoses, and prevention. Gamma ray emitting radioisotopes are used as tracers. In industry, gamma rays are used in the inspection of castings, seams, and welds.

11.1.3 Bibliography

- Collocott, T. C., A. B. Dobson, and W. R. Chambers (eds.): *Dictionary of Science & Technology. Handbook of Physics*, McGraw-Hill, New York, N.Y., 1958.
- Judd, D. B., and G. Wyszecki: *Color in Business, Science and Industry*, 3rd ed., John Wiley and Sons, New York, N.Y.
- Kaufman, Ed: *IES Illumination Handbook*, Illumination Engineering Society.
- Lapedes, D. N. (ed.): *The McGraw-Hill Encyclopedia of Science & Technology*, 2nd ed., McGraw-Hill, New York, N.Y.
- Norgard, John: "Electromagnetic Spectrum," *NAB Engineering Handbook*, 9th ed., Jerry C. Whitaker (ed.), National Association of Broadcasters, Washington, D.C., 1999.
- Norgard, John: "Electromagnetic Spectrum," *The Electronics Handbook*, Jerry C. Whitaker (ed.), CRC Press, Boca Raton, Fla., 1996.

The Electromagnetic Spectrum 11-19

- Stemson, A: *Photometry and Radiometry for Engineers*, John Wiley and Sons, New York, N.Y.
The Cambridge Encyclopedia, Cambridge University Press, 1990.
The Columbia Encyclopedia, Columbia University Press, 1993.
Webster's New World Encyclopedia, Prentice Hall, 1992.
Wyszecki, G., and W. S. Stiles: *Color Science, Concepts and Methods, Quantitative Data and Formulae*, 2nd ed., John Wiley and Sons, New York, N.Y.

Chapter

11.2

Propagation

William Daniel, Edward W. Allen, Donald G. Fink

11.2.1 Introduction

The portion of the electromagnetic spectrum currently used for radio transmissions lies between approximately 10 kHz and 40 GHz. The influence on radio waves of the medium through which they propagate is frequency-dependent. The lower frequencies are greatly influenced by the characteristics of the earth's surface and the ionosphere, while the highest frequencies are greatly affected by the atmosphere, especially rain. There are no clear-cut boundaries between frequency ranges but instead considerable overlap in propagation modes and effects of the path medium.

In the U.S., those frequencies allocated for television-related use include the following:

- 54–72 MHz: TV channels 2–4
- 76–88 MHz: TV channels 5–6
- 174–216 MHz: TV channels 7–13
- 470–806 MHz: TV channels 14–69
- 0.9–12.2 GHz: nonexclusive TV terrestrial and satellite ancillary services
- 12.2–12.7 GHz: direct satellite broadcasting
- 12.7–40 GHz: direct satellite broadcasting

11.2.2 Propagation in Free Space

For simplicity and ease of explanation, propagation in space and under certain conditions involving simple geometry, in which the wave fronts remain coherent, may be treated as *ray propagation*. It should be kept in mind that this assumption may not hold in the presence of obstructions, surface roughness, and other conditions which are often encountered in practice.

For the simplest case of propagation in space, namely that of uniform radiation in all directions from a point source, or *isotropic radiator*, it is useful to consider the analogy to a point source of light. The radiant energy passes with uniform intensity through all portions of an imaginary spherical surface located at a radius r from the source. The area of such a surface is $4\pi r^2$

11-22 Broadcast Transmission Systems

and the power flow per unit area $W = P_t / 4\pi r^2$, where P_t is the total power radiated by the source and W is represented in W/m^2 . In the engineering of broadcasting and of some other radio services, it is conventional to measure the intensity of radiation in terms of the strength of the electric field E_o rather than in terms of power density W . The power density is equal to the square of the field strength divided by the impedance of the medium, so for free space:

$$W = \frac{E_o^2}{120\pi} \quad (11.2.1)$$

and:

$$P_t = \frac{4\pi r^2 E_o^2}{120\pi} \quad (11.2.2)$$

or:

$$P_t = \frac{r^2 E_o^2}{30} \quad (11.2.3)$$

Where:

P_t = watts radiated

E_o = the free space field in volts per meter

r = the radius in meters

A more conventional and useful form of this equation, which applies also to antennas other than isotropic radiators, is:

$$E_o = \frac{\sqrt{30 g_t P_t}}{r} \quad (11.2.4)$$

where g_t is the power gain of the antenna in the pertinent direction compared to an isotropic radiator.

An isotropic antenna is useful as a reference for specifying the radiation patterns for more complex antennas but does not in fact exist. The simplest forms of practical antennas are the *electric doublet* and the *magnetic doublet*, the former a straight conductor that is short compared with the wavelength and the latter a conducting loop of short radius compared with the wavelength. For the doublet radiator, the gain is 1.5 and the field strength in the equatorial plane is:

$$E_o = \frac{\sqrt{45 P_t}}{r} \quad (11.2.5)$$

For a half-wave dipole, namely, a straight conductor one-half wave in length, the power gain is 1.64 and:

$$E_o = \frac{7\sqrt{P_t}}{r} \quad (11.2.6)$$

From the foregoing equations it can be seen that for free space:

- The radiation intensity in watts per square meter is proportional to the radiated power and inversely proportional to the square of the radius or distance from the radiator.
- The electric field strength is proportional to the square root of the radiated power and inversely proportional to the distance from the radiator.

11.2.2a Transmission Loss Between Antennas in Free Space

The maximum useful power P_r that can be delivered to a matched receiver is given by [1]:

$$P_r = \left(\frac{E\lambda}{2\pi}\right)^2 \frac{g_r}{120} \text{ W} \quad (11.2.7)$$

Where:

E = received field strength in volts per meter

λ = wavelength in meters, $300/F$

F = frequency in MHz

g_r = receiving antenna power gain over an isotropic radiator

This relationship between received power and the received field strength is shown by scales 2, 3, and 4 in Figure 11.2.1 for a half-wave dipole. For example, the maximum useful power at 100 MHz that can be delivered by a half-wave dipole in a field of 50 dB above 1 $\mu\text{V}/\text{m}$ is 95 dB below 1 W.

A general relation for the ratio of the received power to the radiated power obtained from Equations (11.2.4) and (11.2.7) is:

$$\frac{P_r}{P_t} = \left(\frac{\lambda}{4\pi r}\right)^2 g_t g_r \left(\frac{E}{E_o}\right)^2 \quad (11.2.8)$$

When both antennas are half-wave dipoles, the power-transfer ratio is:

$$\frac{P_r}{P_t} = \left(\frac{1.64\lambda}{4\pi r}\right)^2 \left(\frac{E}{E_o}\right)^2 = \left(\frac{0.13\lambda}{r}\right)^2 \left(\frac{E}{E_o}\right)^2 \quad (11.2.9)$$

and is shown on scales 1 to 4 of Figure 11.2.2. For free-space transmission, $E/E_o = 1$.

11-24 Broadcast Transmission Systems

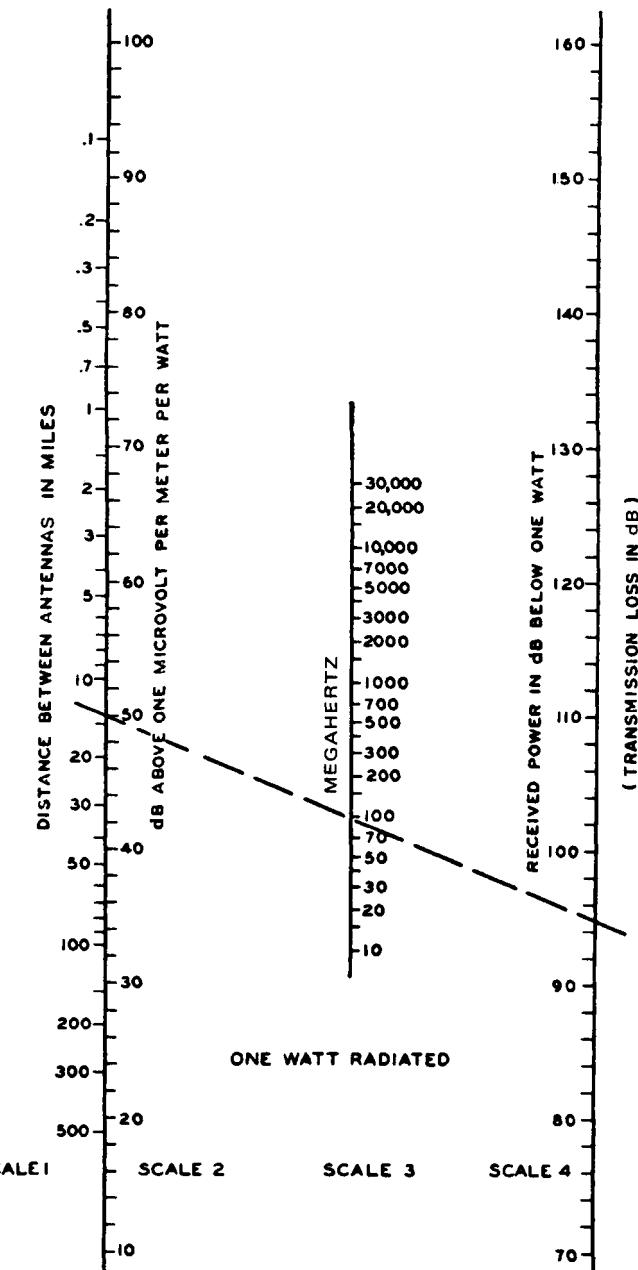


Figure 11.2.1 Free-space field intensity and received power between half-wave dipoles. (From [2]. Used with permission.)

When the antennas are horns, paraboloids, or multielement arrays, a more convenient expression for the ratio of the received power to the radiated power is given by the following:

$$\frac{P_r}{P_t} = \frac{B_t B_r}{(\lambda r)^2} \left(\frac{E}{E_o} \right)^2 \quad (11.2.10)$$

where B_t and B_r are the effective areas of the transmitting and receiving antennas, respectively. This relation is obtained from Equation (11.2.8) by substituting as follows:

$$g = \frac{4\pi B}{\lambda^2} \quad (11.2.11)$$

This is shown in Figure 11.2.2 for free-space transmission when $B_t = B_r$. For example, the free-space loss at 4000 MHz between two antennas of 10 ft^2 (0.93 m^2) effective area is about 72 dB for a distance of 30 mi (48 km).

11.2.3 Propagation Over Plane Earth

The presence of the ground modifies the generation and propagation of radio waves so that the received field strength is ordinarily different than would be expected in free space [3, 4]. The ground acts as a partial reflector and as a partial absorber, and both of these properties affect the distribution of energy in the region above the earth.

11.2.3a Field Strengths Over Plane Earth

The geometry of the simple case of propagation between two antennas each placed several wavelengths above a plane earth is shown in Figure 11.2.3. For isotropic antennas, for simple magnetic-doublet antennas with vertical polarization, or for simple electric-doublet antennas with horizontal polarization the resultant received field is [4, 5]:

$$E = \frac{E_o d}{r_1} + \frac{E_o d R e^{j\Delta}}{r_2} = E_o \left(\cos \theta_1 + R \cos \theta_2 e^{j\Delta} \right) \quad (11.2.12)$$

For simple magnetic-doublet antennas with horizontal polarization or electric-doublet antennas with vertical polarization at both the transmitter and receiver, it is necessary to correct for the cosine radiation and absorption patterns in the plane of propagation. The received field is:

$$E = E_o \left(\cos^3 \theta_1 + R \cos^3 \theta_2 e^{j\Delta} \right) \quad (11.2.13)$$

Where:

E_o = the free-space field at distance d in the equatorial plane of the doublet

11-26 Broadcast Transmission Systems

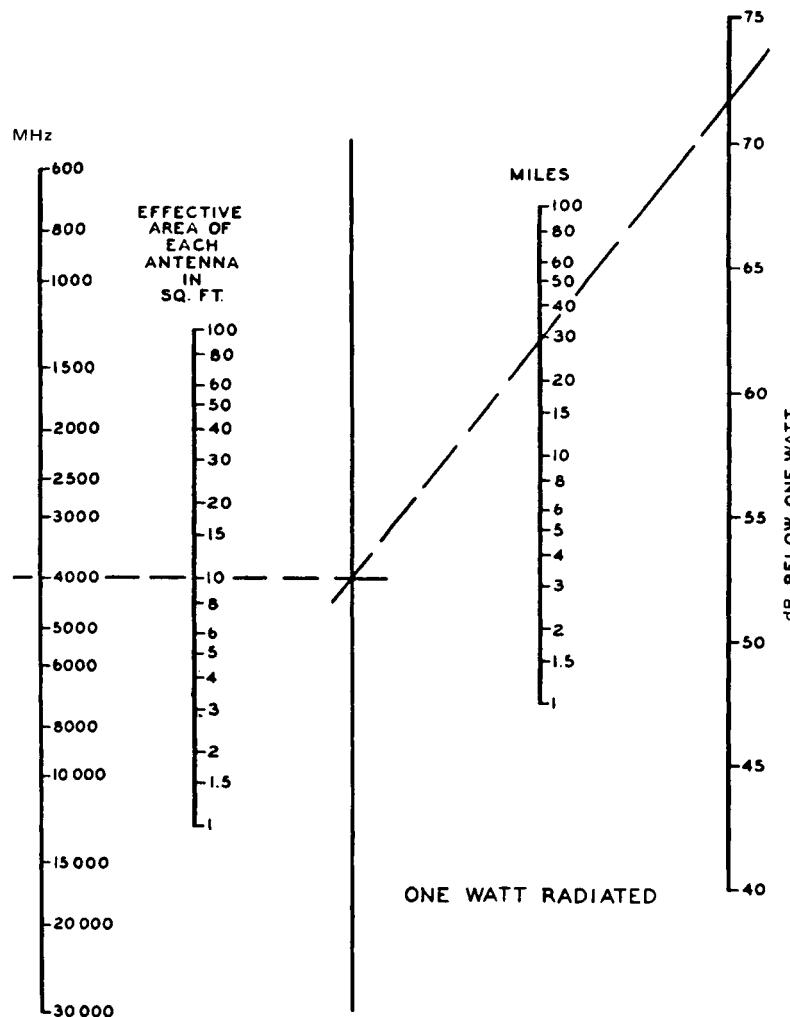


Figure 11.2.2 Received power in free space between two antennas of equal effective areas. (From [2]. Used with permission.)

R = the complex reflection coefficient of the earth

j = the square root of -1

$$e^{j\Delta} = \cos \Delta + j \sin \Delta$$

Δ = the phase difference between the direct wave received over path r_1 and the ground-reflected wave received over path r_2 , which is due to the difference in path lengths

For distances such that θ is small and the differences between d and r_1 and r_2 can be neglected, Equations (11.2.12) and (11.2.13) become:

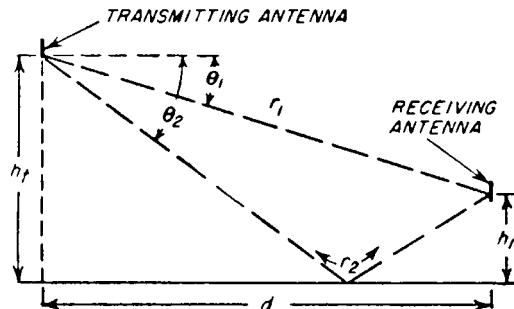


Figure 11.2.3 Ray paths for antennas above plane earth. (From [2]. Used with permission.)

$$E = E_o \left(1 + R e^{j\Delta} \right) \quad (11.2.14)$$

When the angle θ is very small, R is approximately equal to -1 . For the case of two antennas, one or both of which may be relatively close to the earth, a surface-wave term must be added and Equation (11.2.14) becomes [3, 6]:

$$E = E_o [1 + R e^{j\Delta} + (1 - R)A e^{j\Delta}] \quad (11.2.15)$$

The quantity A is the *surface-wave attenuation factor*, which depends upon the frequency, ground constants, and type of polarization. It is never greater than unity and decreases with increasing distance and frequency, as indicated by the following approximate equation [1]:

$$A \approx \frac{-1}{1 + j \left(\frac{2\pi d}{\lambda} \right) (\sin \theta + z)^2} \quad (11.2.16)$$

This approximate expression is sufficiently accurate as long as $A < 0.1$, and it gives the magnitude of A within about 2 dB for all values of A . However, as A approaches unity, the error in phase approaches 180° . More accurate values are given by Norton [3] where, in his nomenclature, $A = f(P, B) e^{j\phi}$.

The equation (14.4.15) for the absolute value of field strength has been developed from the successive consideration of the various components that make up the ground wave, but the following equivalent expressions may be found more convenient for rapid calculation:

$$E = E_o \left\{ 2 \sin \frac{\Delta}{2} + j [(1 + R) + (1 - R)A] e^{j\Delta/2} \right\} \quad (11.2.17)$$

When the distance d between antennas is greater than about five times the sum of the two antenna heights h_t and h_r , the phase difference angle Δ (rad) is:

11-28 Broadcast Transmission Systems

$$\Delta = \frac{4\pi h_t h_r}{\lambda d} \quad (11.2.18)$$

Also, when the angle Δ is greater than about 0.5 rad, the terms inside the brackets of Equation (11.2.17)—which include the surface wave—are usually negligible, and a sufficiently accurate expression is given by the following:

$$E = E_o \left(2 \sin \frac{2\pi h_t h_r}{\lambda d} \right) \quad (11.2.19)$$

In this case, the principal effect of the ground is to produce interference fringes or *lobes*, so that the field strength oscillates about the free-space field as the distance between antennas or the height of either antenna is varied.

When the angle Δ is less than about 0.5 rad, there is a region in which the surface wave may be important but not controlling. In this region, $\sin \Delta/2$ is approximately equal to $\Delta/2$ and:

$$E = E_o \frac{4\pi h'_t h'_r}{\lambda d} \quad (11.2.20)$$

In this equation $h' = h + jh_o$, where h is the actual antenna height and $h_o = \lambda/2\pi z$ has been designated as the minimum effective antenna height. The magnitude of the minimum effective height h_o is shown in Figure 11.2.4 for seawater and for “good” and “poor” soil. “Good” soil corresponds roughly to clay, loam, marsh, or swamp, while “poor” soil means rocky or sandy ground [1].

The surface wave is controlling for antenna heights less than the minimum effective height, and in this region the received field or power is not affected appreciably by changes in the antenna height. For antenna heights that are greater than the minimum effective height, the received field or power is increased approximately 6 dB every time the antenna height is doubled, until free-space transmission is reached. It is ordinarily sufficiently accurate to assume that h' is equal to the actual antenna height or the minimum effective antenna height, whichever is the larger.

When translated into terms of antenna heights in feet, distance in miles, effective power in kilowatts radiated from a half-wave dipole, and frequency F in megahertz, Equation (11.2.20) becomes the following very useful formula for the rapid calculation of approximate values of field strength for purposes of prediction or for comparison with measured values:

$$E \approx F \frac{h'_t h'_r \sqrt{P_t}}{3d^2} \quad (11.2.21)$$

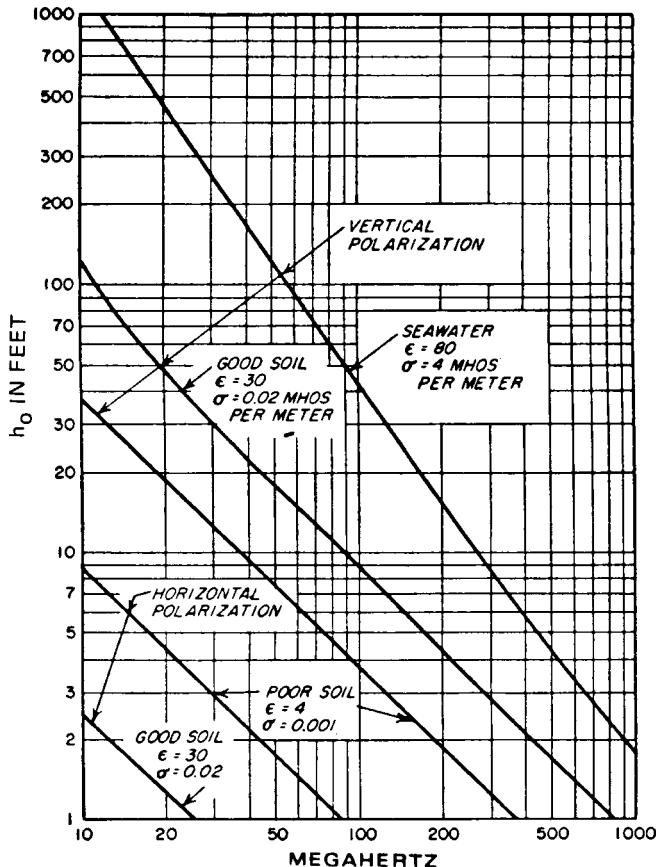


Figure 11.2.4 Minimum effective antenna height. (From [2]. Used with permission.)

11.2.3b Transmission Loss Between Antennas Over Plane Earth

The ratio of the received power to the radiated power for transmission over plane earth is obtained by substituting Equation (11.2.20) into (11.2.8), resulting in the following:

$$\frac{P_r}{P_t} = \left(\frac{\lambda}{4\pi d} \right)^2 g_t g_r \left(\frac{4\pi h'_t h'_r}{\lambda d} \right) = \left(\frac{h'_t h'_r}{d^2} \right)^2 g_t g_r \quad (11.2.22)$$

This relationship is independent of frequency, and is shown on Figure 11.2.5 for half-wave dipoles ($g_t = g_r = 1.64$). A line through the two scales of antenna height determines a point on

11-30 Broadcast Transmission Systems

the unlabeled scale between them, and a second line through this point and the distance scale determines the received power for 1 W radiated. When the received field strength is desired, the power indicated on Figure 11.2.5 can be transferred to scale 4 of Figure 11.2.1, and a line through the frequency on scale 3 indicates the received field strength on scale 2. The results shown on Figure 11.2.5 are valid as long as the value of received power indicated is lower than that shown on Figure 11.2.3 for free-space transmission. When this condition is not met, it means that the angle Δ is too large for Equation (11.2.20) to be accurate and that the received field strength or power oscillates around the free-space value as indicated by Equation (11.2.19) [1].

11.2.3c Propagation Over Smooth Spherical Earth

The curvature of the earth has three effects on the propagation of radio waves at points within the line of sight:

- The *reflection coefficient* of the ground-reflected wave differs for the curved surface of the earth from that for a plane surface. This effect is of little importance, however, under the circumstances normally encountered in practice.
- Because the ground-reflected wave is reflected against the curved surface of the earth, its energy diverges more than would be indicated by the inverse distance-squared law, and the ground-reflected wave must be multiplied by a divergence factor D .
- The heights of the transmitting and receiving antennas h'_t and h'_r , above the plane that is tangent to the surface of the earth at the point of reflection of the ground-reflected wave, are less than the antenna heights h_t and h_r above the surface of the earth, as shown in Figure 11.2.6.

Under these conditions, Equation (11.2.14), which applies to larger distances within the line of sight and to antennas of sufficient height that the surface component may be neglected, becomes:

$$E = E_o \left(1 + D R' e^{j\Delta} \right) \quad (11.2.23)$$

Similar substitutions of the values that correspond in Figures 11.2.3 and 11.2.6 can be made in Equations (11.2.15) through (11.2.22). However, under practical conditions, it is generally satisfactory to use the plane-earth formulas for the purpose of calculating smooth-earth values. An exception to this is usually made in the preparation of standard reference curves, which are generally calculated by the use of the more exact formulas [1, 4–9].

11.2.3d Propagation Beyond the Line of Sight

Radio waves are bent around the earth by the phenomenon of *diffraction*, with the ease of bending decreasing as the frequency increases. Diffraction is a fundamental property of wave motion, and in optics it is the correction to apply to geometrical optics (*ray theory*) to obtain the more accurate *wave optics*. In wave optics, each point on the wave front is considered to act as a radiating source. When the wave front is coherent or undisturbed, the resultant is a progression of the front in a direction perpendicular thereto, along a path that constitutes the ray. When the front is disturbed, the resultant front can be changed in both magnitude and direction with resulting

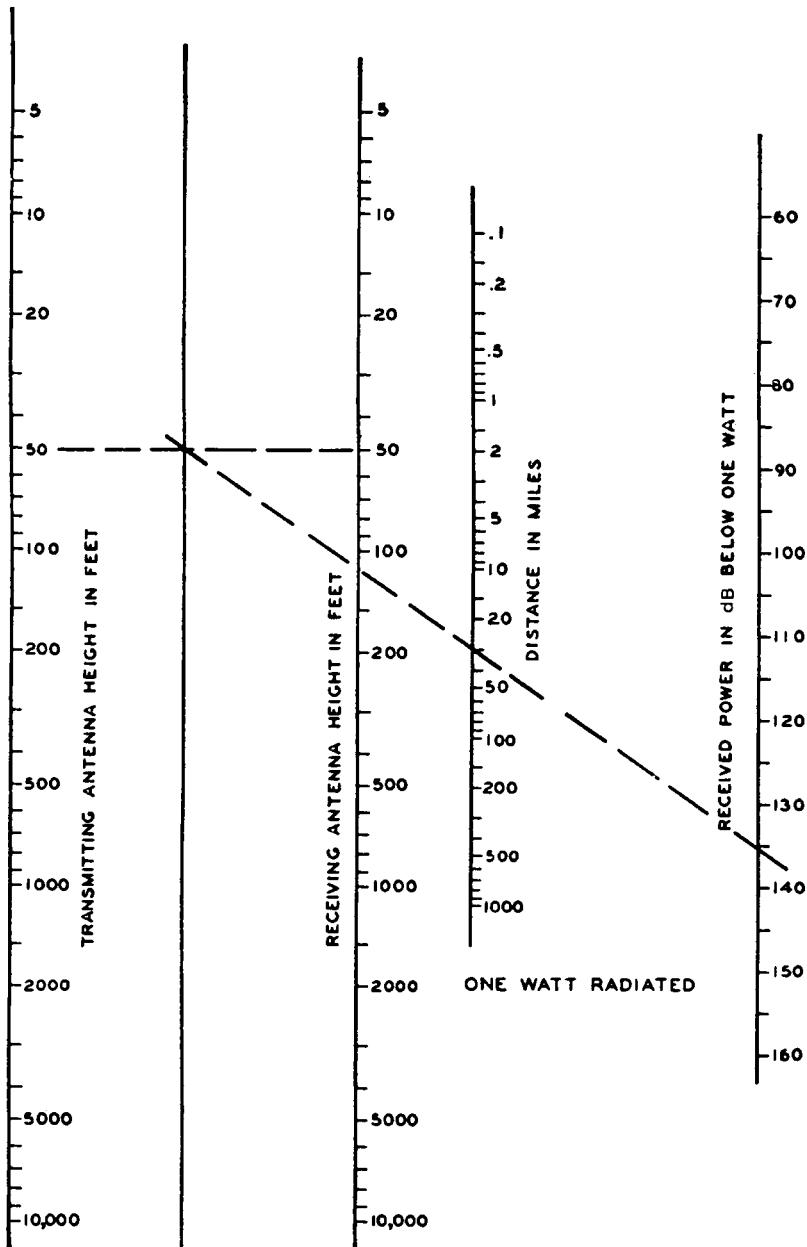


Figure 11.2.5 Received power over plane earth between half-wave dipoles. Notes: (1) This chart is not valid when the indicated received power is greater than the free space power shown in Figure 11.2.1. (2) Use the actual antenna height or the minimum effective height shown in Figure 11.2.4, whichever is the larger. (*From [2]. Used with permission.*)

11-32 Broadcast Transmission Systems

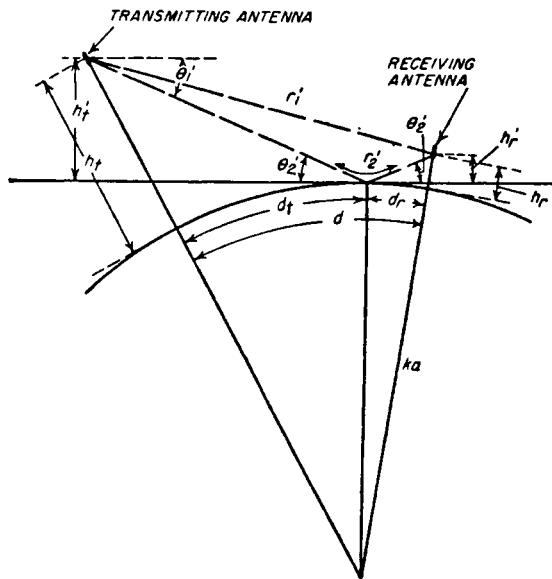


Figure 11.2.6 Ray paths for antennas above spherical earth.
(From [2]. Used with permission.)

attenuation and bending of the ray. Thus, all shadows are somewhat “fuzzy” on the edges and the transition from “light” to “dark” areas is gradual, rather than infinitely sharp.

The effect of diffraction around the earth’s curvature is to make possible transmission beyond the line of sight, with somewhat greater loss than is incurred in free space or over plane earth. The magnitude of this loss increases as either the distance or the frequency is increased and it depends to some extent on the antenna height.

The calculation of the field strength to be expected at any particular point in space beyond the line of sight around a spherical earth is rather complex, so that individual calculations are seldom made except with specially designed software. Rather, nomograms or families of curves are usually prepared for general application to large numbers of cases. The original wave equations of Van der Pol and Bremmer [6] have been modified by Burrows [7] and by Norton [3, 5] so as to make them more readily usable and particularly adaptable to the production of families of curves. Such curves have been prepared by a variety of organizations. These curves have not been included herein, in view of the large number of curves that are required to satisfy the possible variations in frequency, electrical characteristics of the earth, polarization, and antenna height. Also, the values of field strength indicated by smooth-earth curves are subject to considerable modification under actual conditions found in practice. For VHF and UHF broadcast purposes, the smooth-earth curves have been to a great extent superseded by curves modified to reflect average conditions of terrain.

Figure 11.2.7 is a nomogram to determine the additional loss caused by the curvature of the earth [1]. This loss must be added to the free-space loss found from Figure 11.2.1. A scale is included to provide for the effect of changes in the effective radius of the earth, caused by atmospheric refraction. Figure 11.2.7 gives the loss relative to free space as a function of three distances; d_1 is the distance to the horizon from the lower antenna, d_2 is the distance to the horizon

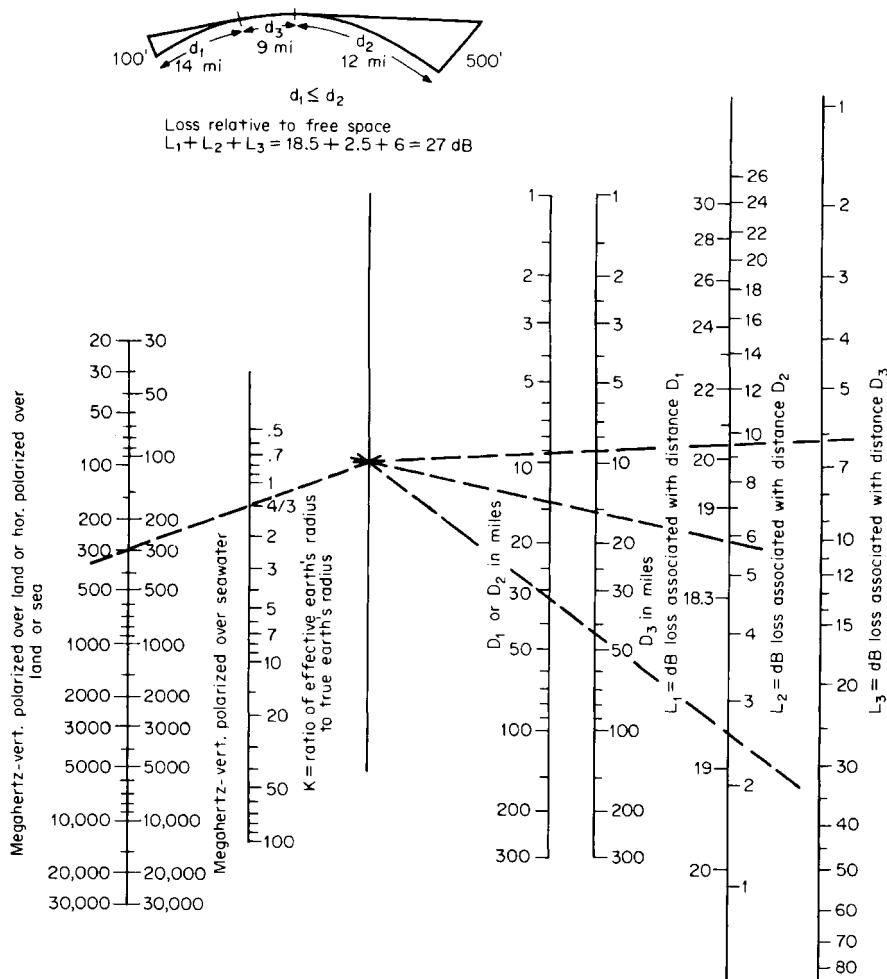


Figure 11.2.7 Loss beyond line of sight in decibels. (From [2]. Used with permission.)

from the higher antenna, and d_3 is the distance between the horizons. The total distance between antennas is $d = d_1 + d_2 + d_3$.

The horizon distances d_1 and d_2 for the respective antenna heights h_1 and h_2 and for any assumed value of the earth's radius factor k can be determined from Figure 11.2.8 [1].

11-34 Broadcast Transmission Systems

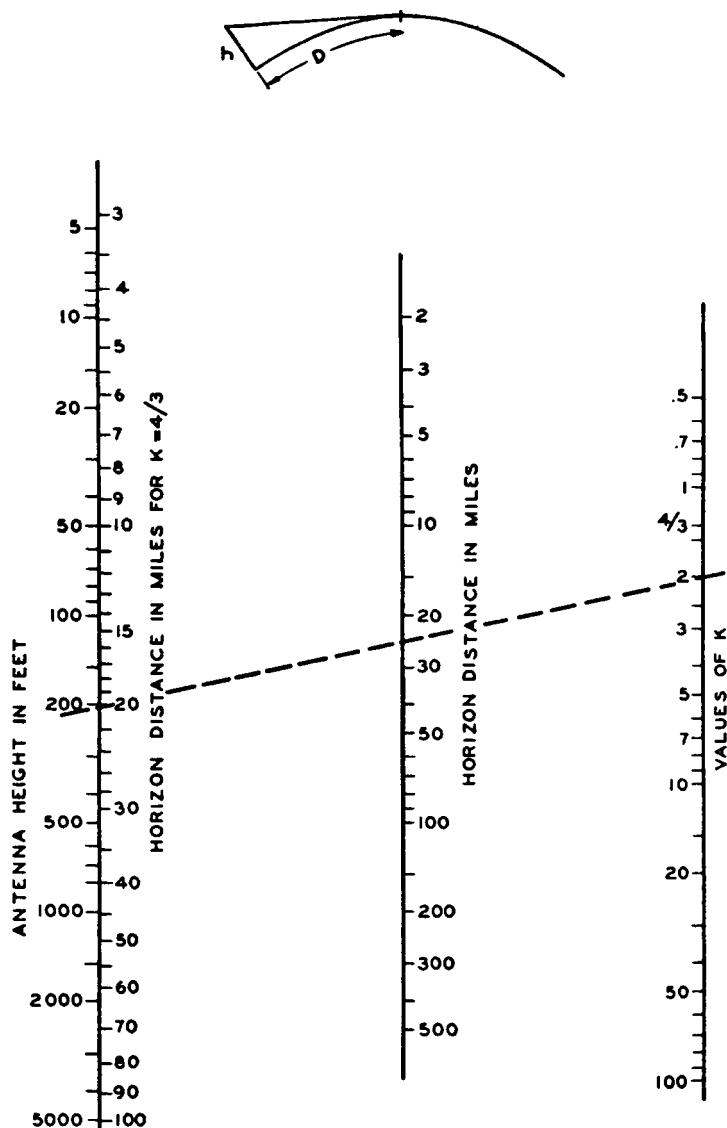


Figure 11.2.8 Distance to the horizon. (From [2]. Used with permission.)

11.2.3e Effects of Hills, Buildings, Vegetation, and the Atmosphere

The preceding discussion assumes that the earth is a perfectly smooth sphere with a uniform or a simple atmosphere, for which condition calculations of expected field strengths or transmission

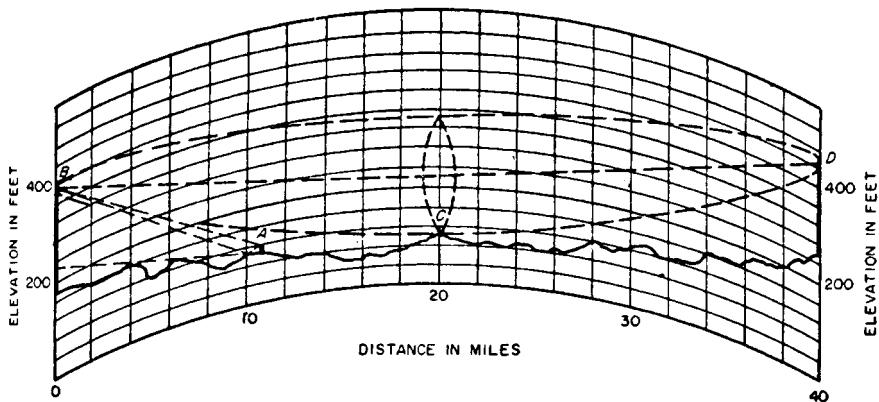


Figure 11.2.9 Ray paths for antennas over rough terrain. (From [2]. Used with permission.)

losses can be computed for the regions within the line of sight and regions well beyond the line of sight, and interpolations can be made for intermediate distances. The presence of hills, buildings, and trees has such complex effects on propagation that it is impossible to compute in detail the field strengths to be expected at discrete points in the immediate vicinity of such obstructions or even the median values over very small areas. However, by the examination of the earth profile over the path of propagation and by the use of certain simplifying assumptions, predictions that are more accurate than smooth-earth calculations can be made of the median values to be expected over areas representative of the gross features of terrain.

Effects of Hills

The profile of the earth between the transmitting and receiving points is taken from available topographic maps and is plotted on a chart that provides for average air refraction by the use of a 4/3 earth radius, as shown in Figure 11.2.9. The vertical scale is greatly exaggerated for convenience in displaying significant angles and path differences. Under these conditions, vertical dimensions are measured along vertical parallel lines rather than along radii normal to the curved surface, and the propagation paths appear as straight lines. The field to be expected at a low receiving antenna at *A* from a high transmitting antenna at *B* can be predicted by plane-earth methods, by drawing a tangent to the profile at the point at which reflection appears to occur with equal incident and reflection angles. The heights of the transmitting and receiving antennas above the tangent are used in conjunction with Figure 14.4.5 to compute the transmission loss, or with Equation (11.2.21) to compute the field strength. A similar procedure can be used for more distantly spaced high antennas when the line of sight does not clear the profile by at least the first *Fresnel zone* [10].

Propagation over a sharp ridge, or over a hill when both the transmitting and receiving antenna locations are distant from the hill, may be treated as diffraction over a knife edge, shown schematically in Figure 11.2.10a [1, 9–14]. The height of the obstruction *H* is measured from the line joining the centers of the two antennas to the top of the ridge. As shown in Figure 11.2.11,

11-36 Broadcast Transmission Systems

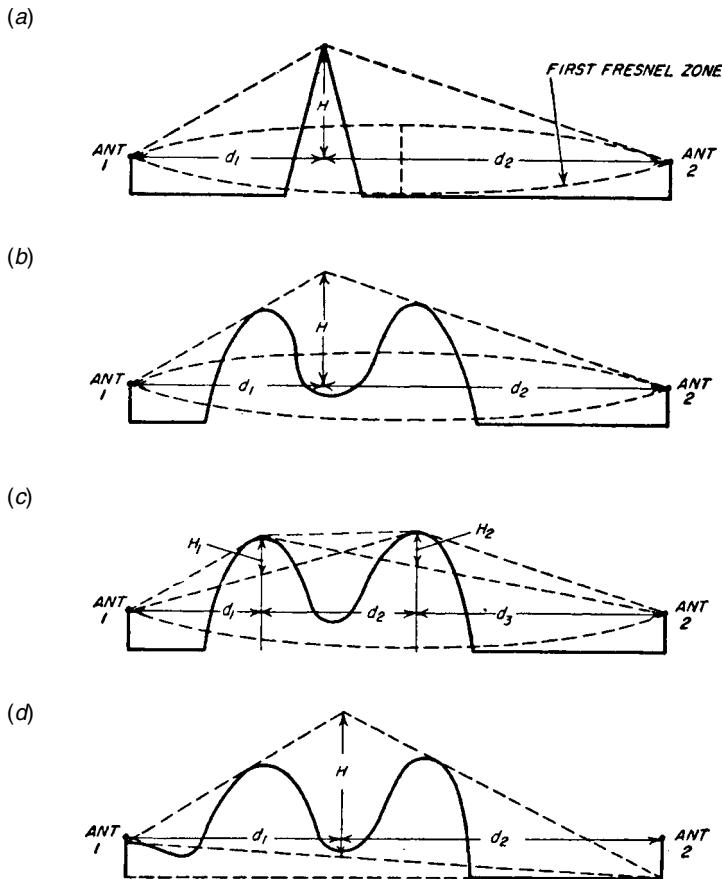


Figure 11.2.10 Ray paths for antennas behind hills: (a-d), see text. (From [2]. Used with permission.)

the shadow loss approaches 6 dB as H approaches 0—*grazing incidence*—and it increases with increasing positive values of H . When the direct ray clears the obstruction, H is negative, and the shadow loss approaches 0 dB in an oscillatory manner as the clearance is increased. Thus, a substantial clearance is required over line-of-sight paths in order to obtain free-space transmission. There is an optimum clearance, called the first Fresnel-zone clearance, for which the transmission is theoretically 1.2 dB better than in free space. Physically, this clearance is of such magnitude that the phase shift along a line from the antenna to the top of the obstruction and from there to the second antenna is about one-half wavelength greater than the phase shift of the direct path between antennas.

The locations of the first three Fresnel zones are indicated on the right-hand scale on Figure 11.2.11, and by means of this chart the required clearances can be obtained. At 3000 MHz, for example, the direct ray should clear all obstructions in the center of a 40 mi (64 km) path by

Note: When accuracy greater than ± 1.5 dB is required, values on the d_1 scale should be:

$$d_1 - \frac{\sqrt{2}}{1 + d_1/d_2}$$

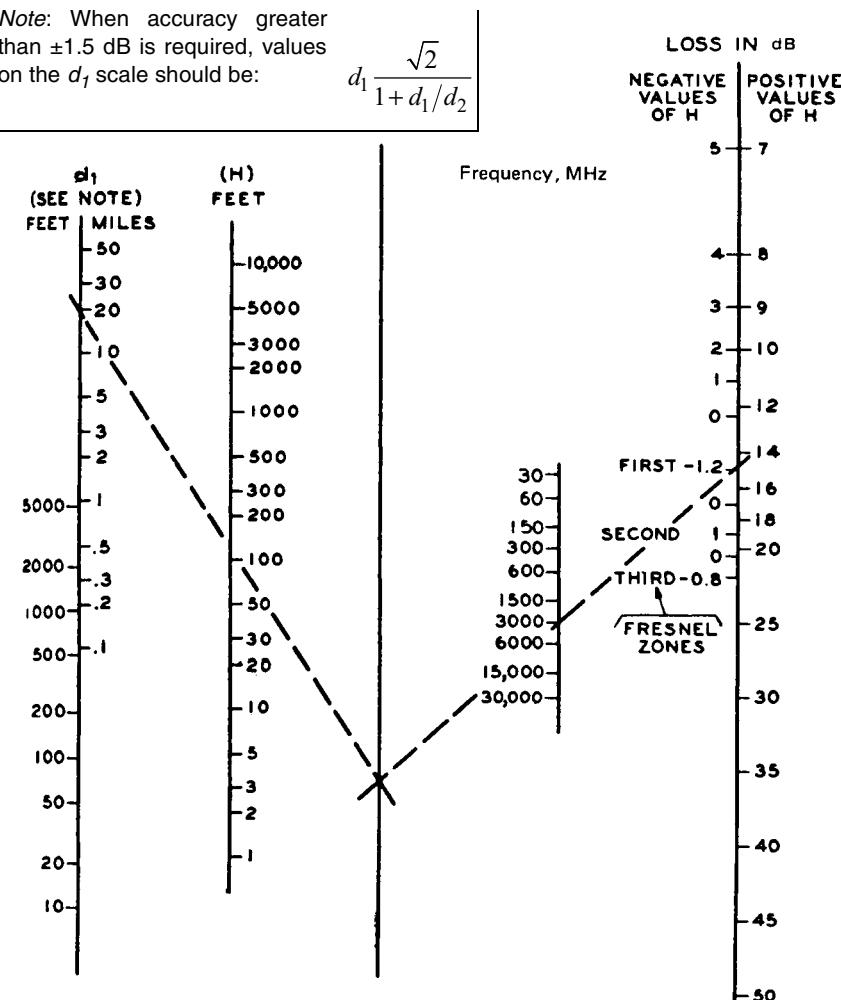


Figure 11.2.11 Shadow loss relative to free space. (From [2]. Used with permission.)

about 120 ft (36 m) to obtain full first-zone clearance, as shown at "C" in Figure 11.2.9. The corresponding clearance for a ridge 100 ft (30 m) in front of either antenna is 4 ft (1.2 m). The locus of all points that satisfy this condition for all distances is an ellipsoid of revolution with foci at the two antennas.

When there are two or more knife-edge obstructions or hills between the transmitting and receiving antennas, an equivalent knife edge can be represented by drawing a line from each antenna through the top of the peak that blocks the line of sight, as in Figure 11.2.10b.

11-38 Broadcast Transmission Systems

Alternatively, the transmission loss can be computed by adding the losses incurred when passing over each of the successive hills, as in Figure 11.2.10c. The height H_1 is measured from the top of hill 1 to the line connecting antenna 1 and the top of hill 2. Similarly, H_2 is measured from the top of hill 2 to the line connecting antenna 2 and the top of hill 1. The nomogram in Figure 11.2.11 is used for calculating the losses for terrain conditions represented by Figure 11.2.10a–c.

This procedure applies to conditions for which the earth-reflected wave can be neglected, such as the presence of rough earth, trees, or structures at locations along the profile at points where earth reflection would otherwise take place at the frequency under consideration; or where first Fresnel-zone clearance is obtained in the foreground of each antenna and the geometry is such that reflected components do not contribute to the field within the first Fresnel zone above the obstruction. If conditions are favorable to earth reflection, the base line of the diffraction triangle should not be drawn through the antennas, but through the points of earth reflection, as in Figure 11.2.10d. H is measured vertically from this base line to the top of the hill, while d_1 and d_2 are measured to the antennas as before. In this case, Figure 11.2.12 is used to estimate the shadow loss to be added to the plane-earth attenuation [1].

Under conditions where the earth-reflected components reinforce the direct components at the transmitting and receiving antenna locations, paths may be found for which the transmission loss over an obstacle is less than the loss over spherical earth. This effect may be useful in establishing VHF relay circuits where line-of-sight operation is not practical. Little utility, however, can be expected for mobile or broadcast services [14].

An alternative method for predicting the median value for all measurements in a completely shadowed area is as follows [15]:

1. The roughness of the terrain is assumed to be represented by height H , shown on the profile at the top of Figure 11.2.13.
2. This height is the difference in elevation between the bottom of the valley and the elevation necessary to obtain line of sight with the transmitting antenna.
3. The difference between the measured value of field intensity and the value to be expected over plane earth is computed for each point of measurement within the shadowed area.
4. The median value for each of several such locations is plotted as a function of sq. rt. (H/λ).

These empirical relationships are summarized in the nomogram shown in Figure 11.2.13. The scales on the right-hand line indicate the median value of shadow loss, compared with plane-earth values, and the difference in shadow loss to be expected between the median and the 90 percent values. For example, with variations in terrain of 500 ft (150 m), the estimated median shadow loss at 4500 MHz is about 20 dB and the shadow loss exceeded in 90 percent of the possible locations is about $20 + 15 = 35$ dB. This analysis is based on large-scale variations in field intensity, and does not include the standing-wave effects that sometimes cause the field intensity to vary considerably within a matter of a few feet.

Effects of Buildings

Built-up areas have little effect on radio transmission at frequencies below a few megahertz, since the size of any obstruction is usually small compared with the wavelength, and the shadows caused by steel buildings and bridges are not noticeable except immediately behind these obstructions. However, at 30 MHz and above, the absorption of a radio wave in going through an

Note: When accuracy greater than ± 1.5 dB is required, values on the d_1 scale should be:

$$d_1 \frac{\sqrt{2}}{1 + d_1/d_2}$$

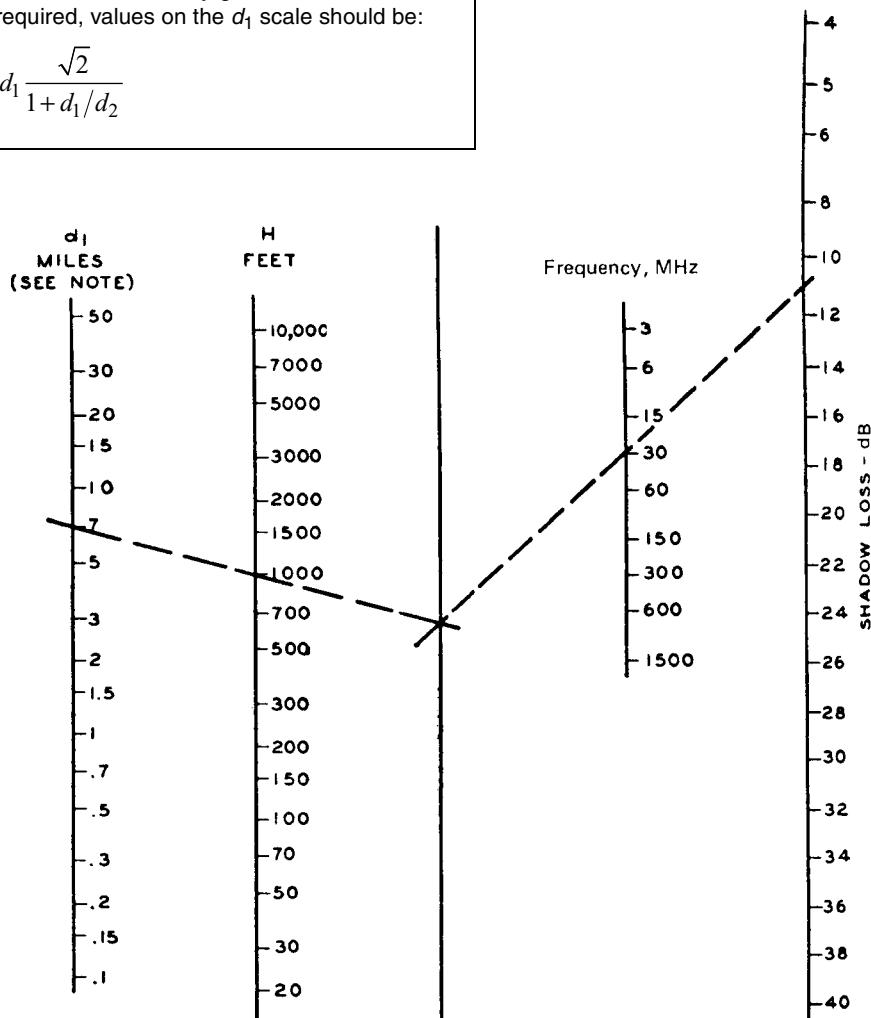


Figure 11.2.12 Shadow loss relative to plane earth. (From [2]. Used with permission.)

obstruction and the shadow loss in going over it are not negligible, and both types of losses tend to increase as the frequency increases. The attenuation through a brick wall, for example, can vary from 2 to 5 dB at 30 MHz and from 10 to 40 dB at 3000 MHz, depending on whether the wall is dry or wet. Consequently, most buildings are rather opaque at frequencies of the order of thousands of megahertz.

For radio-relay purposes, it is the usual practice to select clear sites; but where this is not feasible the expected fields behind large buildings can be predicted by the preceding diffraction

11-40 Broadcast Transmission Systems

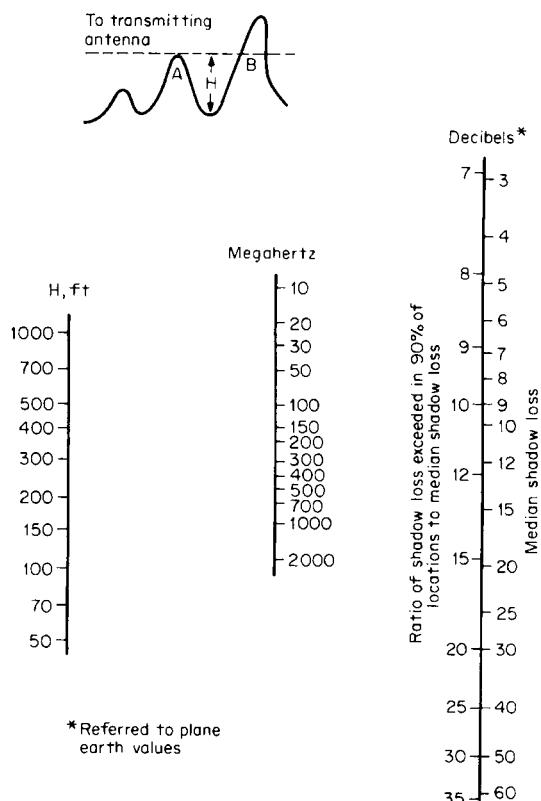


Figure 11.2.13 Estimated distribution of shadow loss for random locations (referred to plane-earth values). (From [2]. Used with permission.)

methods. In the engineering of mobile- and broadcast-radio systems it has not been found practical in general to relate measurements made in built-up areas to the particular geometry of buildings, so that it is conventional to treat them statistically. However, measurements have been divided according to general categories into which buildings can readily be classified, namely, the tall buildings typical of the centers of cities on the one hand, and typical two-story residential areas on the other.

Buildings are more transparent to radio waves than the solid earth, and there is ordinarily much more backscatter in the city than in the open country. Both of these factors tend to reduce the shadow losses caused by the buildings. On the other hand, the angles of diffraction over or around the buildings are usually greater than for natural terrain, and this factor tends to increase the loss resulting from the presence of buildings. Quantitative data on the effects of buildings indicate that in the range of 40 to 450 MHz there is no significant change with frequency, or at least the variation with frequency is somewhat less than the square-root relationship noted in the case of hills. The median field strength at street level for random locations in New York City is about 25 dB below the corresponding plane-earth value. The corresponding values for the 10 percent and 90 percent points are about -15 and -35 dB, respectively [1, 15]. Measurements in congested residential areas indicate somewhat less attenuation than among large buildings.

Effects of Trees and Other Vegetation

When an antenna is surrounded by moderately thick trees and below treetop level, the average loss at 30 MHz resulting from the trees is usually 2 or 3 dB for vertical polarization and negligible with horizontal polarization. However, large and rapid variations in the received field strength can exist within a small area, resulting from the standing-wave pattern set up by reflections from trees located at a distance of as much as 100 ft (30 m) or more from the antenna. Consequently, several nearby locations should be investigated for best results. At 100 MHz, the average loss from surrounding trees may be 5 to 10 dB for vertical polarization and 2 or 3 dB for horizontal polarization. The tree losses continue to increase as the frequency increases, and above 300 to 500 MHz they tend to be independent of the type of polarization. Above 1000 MHz, trees that are thick enough to block vision present an almost solid obstruction, and the diffraction loss over or around these obstructions can be obtained from Figures 11.2.9 or 11.2.11.

There is a pronounced seasonal effect in the case of deciduous trees, with less shadowing and absorption in the winter months when the leaves have fallen. However, when the path of travel through the trees is sufficiently long that it is obscured, losses of the above magnitudes can be incurred, and the principal mode of propagation may be by diffraction over the trees.

When the antenna is raised above trees and other forms of vegetation, the prediction of field strengths again depends upon the proper estimation of the height of the antenna above the areas of reflection and of the applicable reflection coefficients. For growth of fairly uniform height and for angles near grazing incidence, reflection coefficients will approach -1 at frequencies near 30 MHz. As indicated by Rayleigh's criterion of roughness, the apparent roughness for given conditions of geometry increases with frequency so that near 1000 MHz even such low and relatively uniform growth as farm crops or tall grass may have reflection coefficients of about -0.3 for small angles of reflection [17].

The distribution of losses in the immediate vicinity of trees does not follow normal probability law but is more accurately represented by Rayleigh's law, which is the distribution of the sum of a large number of equal vectors having random phases.

11.2.3f Effects of the Lower Atmosphere (Troposphere)

Radio waves propagating through the lower atmosphere, or troposphere, are subject to absorption, scattering, and bending. Absorption is negligible in the VHF-UHF frequency range but becomes significant at frequencies above 10 GHz. The index of refraction of the atmosphere, n , is slightly greater than 1 and varies with temperature, pressure, and water vapor pressure, and therefore with height, climate, and local meteorological conditions. An exponential model showing a decrease with height to 37 to 43 mi (60 to 70 kin) is generally accepted [18, 19]. For this model, variation of n is approximately linear for the first kilometer above the surface in which most of the effect on radio waves traveling horizontally occurs. For average conditions, the effect of the atmosphere can be included in the expression of earth diffraction around the smooth earth without discarding the useful concept of straight-line propagation by multiplying the actual earth's radius by k to obtain an effective earth's radius, where

$$k = \frac{1}{1 + a(dn/dh)} \quad (11.2.24)$$

11-42 Broadcast Transmission Systems

Where:

a = the actual radius of the earth

dn/dh = the rate of change of the refractive index with height

Through the use of average annual values of the refractive index gradient, k is found to be $4/3$ for temperate climates.

Stratification and Ducts

As a result of climatological and weather processes such as *subsidence*, *advection*, and surface heating and radiative cooling, the lower atmosphere tends to be stratified in layers with contrasting refractivity gradients [20]. For convenience in evaluating the effect of this stratification, *radio refractivity* N is defined as $N = (n - 1) \times 10^6$ and can be derived from:

$$N = 77.6 \frac{P}{T} + 3.73 \times 10^5 \frac{e}{T^2} \quad (11.2.25)$$

Where:

P = atmospheric pressure, mbar

T = absolute temperature, K

e = water vapor pressure, mbar

When the gradient of N is equal to -39 N -units per kilometer, normal propagation takes place, corresponding to the effective earth's radius ka , where $k = 4/3$.

When dN/dh is less than -39 N -units per kilometer, *subrefraction* occurs and the radio wave is bent strongly downward.

When dN/dh is less than -157 N -units per kilometer, the radio energy may be bent downward sufficiently to be reflected from the earth, after which the ray is again bent toward the earth, and so on. The radio energy thus is trapped in a duct or waveguide. The wave also may be trapped between two elevated layers, in which case energy is not lost at the ground reflection points and even greater enhancement occurs. Radio waves thus trapped or *ducted* can produce fields exceeding those for free-space propagation because the spread of energy in the vertical direction is eliminated as opposed to the free-space case, where the energy spreads out in two directions orthogonal to the direction of propagation. Ducting is responsible for abnormally high fields beyond the radio horizon. These enhanced fields occur for significant periods of time on overwater paths in areas where meteorological conditions are favorable. Such conditions exist for significant periods of time and over significant horizontal extent in the coastal areas of southern California and around the Gulf of Mexico. Over land, the effect is less pronounced because surface features of the earth tend to limit the horizontal dimension of ducting layers [20].

Tropospheric Scatter

The most consistent long-term mode of propagation beyond the radio horizon is that of scattering by small-scale fluctuations in the refractive index resulting from turbulence. Energy is scattered from multitudinous irregularities in the common volume which consists of that portion of troposphere visible to both transmitting and receiving site. There are some empirical data that show a correlation between the variations in the field beyond the horizon and ΔN , the difference

between the reflectivity on the ground and at a height of 1 km [21]. Procedures have been developed for calculating scatter fields for beyond-the-horizon radio relay systems as a function of frequency and distance [22, 23]. These procedures, however, require detailed knowledge of path configuration and climate.

The effect of scatter propagation is incorporated in the statistical evaluation of propagation (considered previously in this chapter), where the attenuation of fields beyond the diffraction zone is based on empirical data and shows a linear decrease with distance of approximately 0.2 dB/mi (0.1 dB/km) for the VHF–UHF frequency band.

11.2.3g Atmospheric Fading

Variations in the received field strengths around the median values are caused by changes in atmospheric conditions. Field strengths tend to be higher in summer than in winter, and higher at night than during the day, for paths over land beyond the line of sight. As a first approximation, the distribution of long-term variations in field strength in decibels follows a normal probability law.

Measurements indicate that the fading range reaches a maximum somewhat beyond the horizon and then decreases slowly with distance out to several hundred miles. Also, the fading range at the distance of maximum fading increases with frequency, while at the greater distances where the fading range decreases, the range is also less dependent on frequency. Thus, the slope of the graph N must be adjusted for both distance and frequency. This behavior does not lend itself to treatment as a function of the earth's radius factor k , since calculations based on the same range of k produce families of curves in which the fading range increases systematically with increasing distance and with increasing frequency.

Effects of the Upper Atmosphere (Ionosphere)

Four principal recognized layers or regions in the ionosphere are the E layer, the $F1$ layer, the $F2$ layer (centered at heights of about 100, 200, and 300 km, respectively), and the D region, which is less clearly defined but lies below the E layer. These *regular* layers are produced by radiation from the sun, so that the ion density—and hence the frequency of the radio waves that can be reflected thereby—is higher during the day than at night. The characteristics of the layers are different for different geographic locations and the geographic effects are not the same for all layers. The characteristics also differ with the seasons and with the intensity of the sun's radiation, as evidenced by the sunspot numbers, and the differences are generally more pronounced upon the $F2$ than upon the $F1$ and E layers. There are also certain random effects that are associated with solar and magnetic disturbances. Other effects that occur at or just below the E layer have been established as being caused by meteors [24].

The greatest potential for television interference by way of the ionosphere is from *sporadic E ionization*, which consists of occasional patches of intense ionization occurring 62 to 75 mi (100 to 120 km) above the earth's surface and apparently formed by the interaction of winds in the neutral atmosphere with the earth's magnetic field. Sporadic E ionization can reflect VHF signals back to earth at levels capable of causing interference to analog television reception for periods lasting from 1 h or more, and in some cases totaling more than 100 h per year. In the U.S., VHF sporadic E propagation occurs a greater percentage of the time in the southern half of the country and during the May to August period [25].

11-44 Broadcast Transmission Systems

11.2.4 References

1. Bullington, K.: "Radio Propagation at Frequencies above 30 Mc," *Proc. IRE*, pg. 1122, October 1947.
2. Fink, D. G., (ed.): *Television Engineering Handbook*, McGraw-Hill, New York, N.Y., 1957.
3. Eckersley, T. L.: "Ultra-Short-Wave Refraction and Diffraction," *J. Inst. Elec. Engrs.*, pg. 286, March 1937.
4. Norton, K. A.: "Ground Wave Intensity over a Finitely Conducting Spherical Earth," *Proc. IRE*, pg. 622, December 1941.
5. Norton, K. A.: "The Propagation of Radio Waves over a Finitely Conducting Spherical Earth," *Phil. Mag.*, June 1938.
6. van der Pol, Balth, and H. Bremmer: "The Diffraction of Electromagnetic Waves from an Electrical Point Source Round a Finitely Conducting Sphere, with Applications to Radiotelegraphy and to Theory of the Rainbow," pt. 1, *Phil. Mag.*, July, 1937; pt. 2, *Phil. Mag.*, November 1937.
7. Burrows, C. R., and M. C. Gray: "The Effect of the Earth's Curvature on Groundwave Propagation," *Proc. IRE*, pg. 16, January 1941.
8. "The Propagation of Radio Waves through the Standard Atmosphere," Summary Technical Report of the Committee on Propagation, vol. 3, National Defense Research Council, Washington, D.C., 1946, published by Academic Press, New York, N.Y.
9. "Radio Wave Propagation," Summary Technical Report of the Committee on Propagation of the National Defense Research Committee, Academic Press, New York, N.Y., 1949.
10. de Lisle, E. W.: "Computations of VHF and UHF Propagation for Radio Relay Applications," RCA, Report by International Division, New York, N.Y.
11. Selvidge, H.: "Diffraction Measurements at Ultra High Frequencies," *Proc. IRE*, pg. 10, January 1941.
12. McPetrie, J. S., and L. H. Ford: "An Experimental Investigation on the Propagation of Radio Waves over Bare Ridges in the Wavelength Range 10 cm to 10 m," *J. Inst. Elec. Engrs.*, pt. 3, vol. 93, pg. 527, 1946.
13. Megaw, E. C. S.: "Some Effects of Obstacles on the Propagation of Very Short Radio Waves," *J. Inst. Elec. Engrs.*, pt. 3, vol. 95, no. 34, pg. 97, March 1948.
14. Dickson, F. H., J. J. Egli, J. W. Herbstreit, and G. S. Wickizer: "Large Reductions of VHF Transmission Loss and Fading by the Presence of a Mountain Obstacle in Beyond-Line-of-Sight Paths," *Proc. IRE*, vol. 41, no. 8, pg. 96, August 1953.
15. Bullington, K.: "Radio Propagation Variations at VHF and UHF," *Proc. IRE*, pg. 27, January 1950.
16. "Report of the Ad Hoc Committee, Federal Communications Commission," vol. 1, May 1949; vol. 2, July 1950.
17. Epstein, J., and D. Peterson: "An Experimental Study of Wave Propagation at 850 Mc," *Proc. IRE*, pg. 595, May 1953.

18. "Documents of the XVth Plenary Assembly," CCIR Report 563, vol. 5, Geneva, 1982.
19. Bean, B. R., and E. J. Dutton: "Radio Meteorology," National Bureau of Standards Monograph 92, March 1, 1966.
20. Dougherty, H. T., and E. J. Dutton: "The Role of Elevated Ducting for Radio Service and Interference Fields," NTIA Report 81-69, March 1981.
21. "Documents of the XVth Plenary Assembly," CCIR Report 881, vol. 5, Geneva, 1982.
22. "Documents of the XVth Plenary Assembly," CCIR Report 238, vol. 5, Geneva, 1982.
23. Longley, A. G., and P. L. Rice: "Prediction of Tropospheric Radio Transmission over Irregular Terrain—A Computer Method," ESSA (Environmental Science Services Administration), U.S. Dept. of Commerce, Report ERL (Environment Research Laboratories) 79-ITS 67, July 1968.
24. National Bureau of Standards Circular 462, "Ionospheric Radio Propagation," June 1948.
25. Smith, E. E., and E. W. Davis: "Wind-induced Ions Thwart TV Reception," *IEEE Spectrum*, pp. 52—55, February 1981.

Chapter
11.3
Radio Broadcasting

Jerry C. Whitaker, Editor-in-Chief

11.3.1 Introduction

Standard broadcasting refers to the transmission of voice and music received by the general public in the 535–1705 kHz frequency band. Amplitude modulation is used to provide service ranging from that needed for small communities to higher powered broadcast stations needed for larger regional areas. The *primary service area* is defined as the area in which the ground or surface-wave signal is not subject to objectionable interference or objectionable fading. The *secondary service area* refers to an area serviced by skywaves and not subject to objectionable interference. *Intermittent service area* refers to an area receiving service from either a surface wave or a skywave but beyond the primary service area and subject to some interference and fading.

11.3.2 Standard Broadcasting

The carrier frequencies for standard broadcasting in the U.S. (referred to internationally as medium wave broadcasting) are designated in the Federal Communications Commission (FCC) Rules and Regulations. A total of 117 carrier frequencies are allocated from 540 to 1700 kHz in 10-kHz intervals. Each carrier frequency is required by the FCC rules to deviate no more than ± 20 Hz from the allocated frequency to minimize heterodyning from two or more interfering stations.

11.3.2a Modulation Techniques

Double-sideband full-carrier modulation, commonly called *amplitude modulation* (AM), is used in standard broadcasting for sound transmission. Basically, the amplitude of an analog radio frequency (RF) carrier is controlled by an analog audio frequency (AF) modulating signal. The resulting RF waveform consists of a carrier wave plus two additional signals that are the result of the modulation process: an upper-sideband signal, which is equal in frequency to the carrier plus the audio modulating frequency, and a lower-sideband signal, which equals the carrier frequency minus the audio modulating frequency.

11-48 Broadcast Transmission Systems

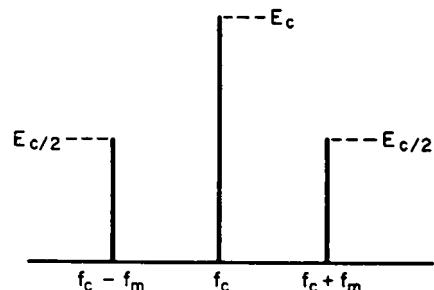


Figure 11.3.1 Double-sideband amplitude modulation.

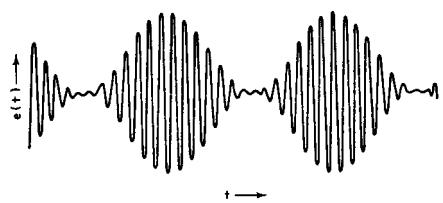


Figure 11.3.2 An amplitude-modulated waveform.

Modulation of the AM Carrier Wave

Figure 11.3.1 shows a fully amplitude modulated signal. When the peak value of the modulated carrier wave (produced by a positive excursion of the audio modulating signal) is twice the value of the unmodulated carrier, and the minimum value of the RF carrier reaches zero (produced by a negative excursion of the modulating signal), 100 percent modulation occurs. If the transmitter is well designed, with sufficient headroom, it is possible to exceed 100 percent positive modulation with an asymmetrical waveform, but the negative modulation may never exceed 100 percent (the RF carrier at zero or completely cut off) without causing distortion of the demodulated signal. The power contained in the audio sidebands created in the process of modulating the carrier 100 percent is equal to 25 percent of the carrier power in both the upper and lower sidebands. As a result, the overall power transmitted during periods of 100 percent modulation is 150 percent of the unmodulated carrier power. (See Figure 11.3.2.)

The overall bandwidth is equal to twice the highest audio modulating frequency. Although full fidelity is possible with amplitude modulation, the FCC requires standard broadcast stations to limit the fidelity, thus restricting occupied bandwidth of the transmitted signal. Typical modulation frequencies for voice and music range from 50 Hz to 10 kHz. Each channel is generally thought of as 10 kHz in width, and thus the frequency band is designated from 535 to 1705 kHz; however, when the modulation frequency exceeds 5 kHz, the radio frequency bandwidth of the channel exceeds 10 kHz and adjacent channel interference may occur.

Stations may also utilize advanced modulation techniques to enhance the transmitted information. *Quadrature modulation* of the AM carrier is utilized a large number of AM stations to transmit stereophonic programming (separate left and right audio channels). The FCC standard for AM stereo modulation does not cause interference in conventional envelope detector receivers. In addition, the system complies with FCC occupied bandwidth limitations, and all international agreements. New, more advanced modulation techniques have successfully been tested allowing for transmission of stereo sound to listeners, free of the static bursts caused by impulse noise. Such systems allow AM stations to transmit digital programming simultaneously, and on

the same channel, with their analog programming. The result is an analog component that is compatible with existing receivers, and a separate digital component offering high-quality stereo audio.

11.3.2b Channel and Station Classifications

In standard broadcast (AM), stations in the U.S. are classified by the FCC according to their operating power, protection from interference, and hours of operation. A Class A station operates with 10 to 50 kW of power servicing a large area with primary, secondary, and intermittent coverage and is protected from interference both day and night. These stations are called *clear channel stations* because the channel is cleared of nighttime interference over a major portion of the country. Class B stations operate full time with transmitter powers of 0.25 to 50 kW and are designed to render primary service only over a principal center of population and contiguous rural area. Whereas nearly all Class A stations operate with 50 kW, most Class B stations must restrict their power to 5 kW or less to avoid interfering with other stations. Class B stations operating in the 1605 to 1705 kHz band are restricted to a power level of 10 kW daytime and 1 kW nighttime. Class C stations operate on six designated channels (1230, 1240, 1340, 1400, 1450, and 1490) with a maximum power of 1 kW or less full time and render primarily local service to smaller communities. Class D stations operate on Class A or B frequencies with Class B transmitter powers during daytime, but nighttime operation—if permitted at all—must be at low power (less than 0.25 kW) with no protection from interference.

When a station is already limited by interference from other stations to a contour of higher value than that normally protected for its class, this higher-value contour is the established protection standard for the station.

Although Class A stations cover large areas at night (approximately a 1000 km radius), the nighttime coverage of Class B, C, and D stations is limited by interference from other stations, electrical devices, and atmospheric conditions to a relatively small area. Class C stations, for example, have an interference-free nighttime coverage radius of approximately 8 to 16 km. As a result there may be large differences in the area that the station covers daytime vs. nighttime. With thousands of AM stations licensed for operation by the FCC, interference, both day and night, is a factor that significantly limits the service stations can provide. In the absence of interference, a daytime signal strength of 2 mV/m is required for reception in populated towns and cities, whereas a signal of 0.5 mV/m is generally acceptable in rural areas without large amounts of man-made interference present. Secondary nighttime service is provided in areas receiving a 0.5-mV/m signal 50 percent or more of the time without objectionable interference. However, it should be noted that these limits apply to new stations and modifications to existing stations. Nearly every station on the air was allocated prior to the implementation of these rules with interference criteria that were less restrictive.

11.3.2c Field Strength

The field strength produced by a standard broadcast station is a key factor in determining the primary and secondary service areas and interference limitations of possible future radio stations. The FCC and others have conducted significant testing to determine what levels of interference are acceptable from adjacent and co-channel stations and have used that data in determining how close stations can be located to one another. If the transmitted power, antenna radiation charac-

11-50 Broadcast Transmission Systems

teristics, and ground conductivity are known, the extent of coverage (or interference) in a given direction for a particular station can be calculated with a high degree of accuracy. These calculations form the basis of the FCC's station allocation system and are based on field intensities with the unit of measure as volts per meter (the voltage induced in an antenna 1 m in length).

11.3.2d Propagation

One of the major factors in the determination of field strength is the propagation characteristic, described as the change in electric field intensity with an increase in distance from the broadcast station antenna. This variation depends on a number of factors including frequency, distance, surface dielectric constant, surface loss tangent, polarization, local topography, and time of day. Generally speaking, surface-wave propagation occurs over shorter ranges both during day and night periods. Skywave propagation in the AM broadcast band permits longer ranges and occurs during night periods, and thus some stations must either reduce power or cease to operate at night to avoid causing interference. Skywave propagation is much less predictable than surface-wave propagation, because it necessarily involves some fading and less predictable field intensities and is most appropriately described in terms of statistics or the percentage of time a particular field strength level is found.

11.3.2e Transmitters

Standard AM broadcast transmitters range in power output from 5 W up to 50 kW units. Modern transmitters utilize low-voltage, high-current metal-oxide-semiconductor field-effect transistor (MOSFET) devices to generate the RF power. However, in a high-power transmitter, it may take hundreds of transistors to generate the rated power. These transistors are combined into modules, the outputs of which are combined to produce the output signal. This design approach has the added benefit that in the event of a module failure, the transmitter continues to operate, but at slightly reduced power.

High-Level AM Modulation

High-level anode modulation is the oldest and simplest way of generating a high power AM signal. In this system, the modulating signal is amplified and combined with the dc supply source to the anode of the final RF amplifier stage. The RF amplifier is normally operated class C. The final stage of the modulator usually consists of a pair of tubes operating class B. A basic modulator of this type is shown in Figure 11.3.3.

The RF signal is normally generated in a low-level transistorized oscillator. It is then amplified by one or more solid-state or vacuum-tube stages to provide final RF drive at the appropriate frequency to the grid of the final class C amplifier. The audio input is applied to an intermediate power amplifier (usually solid state) and used to drive two class B (or class AB) push-pull output stages. The final amplifiers provide the necessary modulating power to drive the final RF stage. For 100 percent modulation, this modulating power is 50 percent of the actual carrier power.

The modulation transformer shown in Figure 11.3.3 does not usually carry the dc supply current for the final RF amplifier. The modulation reactor and capacitor shown provide a means to combine the audio signal voltage from the modulator with the dc supply to the final RF amplifier. This arrangement eliminates the necessity of having dc current flow through the secondary of the modulation transformer, which would result in magnetic losses and saturation effects. In

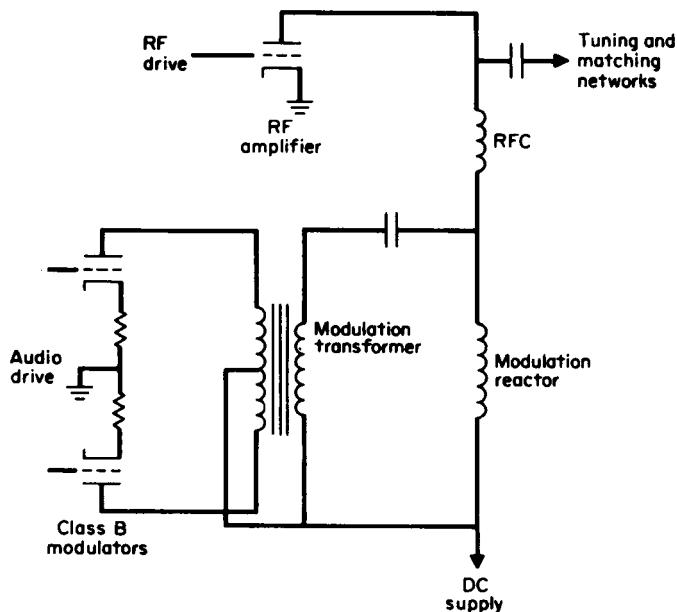


Figure 11.3.3 Basic configuration for high-level amplitude modulation in the standard broadcast band.

some transmitter designs, the modulation reactor is eliminated from the system, thanks to improvements in transformer technology.

The RF amplifier normally operates class C with grid current drawn during positive peaks of the cycle. Typical stage efficiency is 75 to 83 percent.

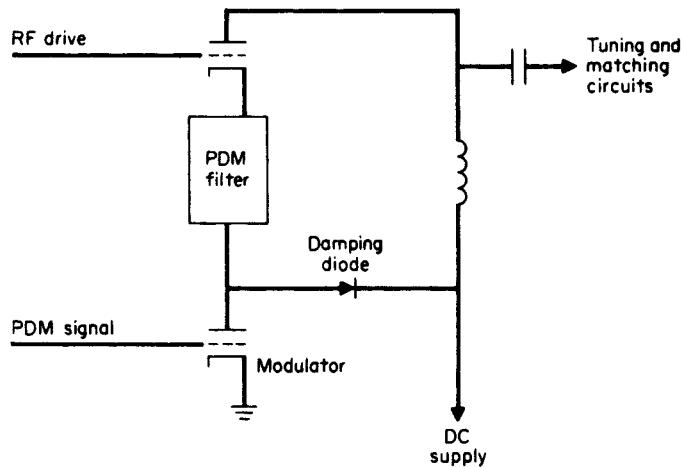
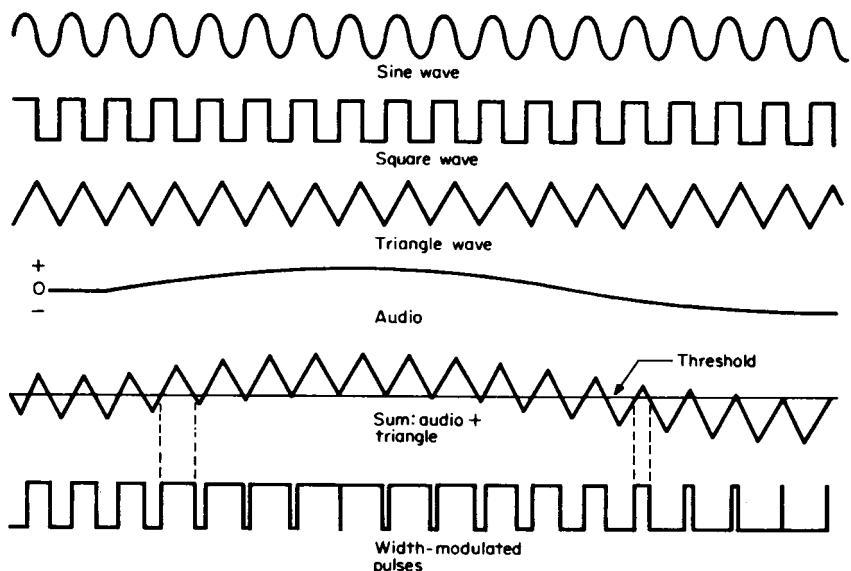
This type of system was popular in AM broadcasting for many years, primarily because of its simplicity. The primary drawback is low overall system efficiency. The class B modulator tubes cannot operate with greater than 50 percent efficiency. Still, with inexpensive electricity, this was not considered to be a significant problem. As energy costs increased, however, more efficient methods of generating high-power AM signals were developed. Increased efficiency normally came at the expense of added technical complexity.

Pulse-Width Modulation

Pulse-width (also known as pulse-duration) modulation is one of the most popular systems developed for modern vacuum-tube AM transmitters. Figure 11.3.4 shows a scheme for pulse-width modulation identified as PDM (as patented by the Harris Corporation). The PDM system works by utilizing a square wave switching system, illustrated in Fig. 11.3.5.

The PDM process begins with a signal generator (Figure 11.3.6). A 75-kHz sine wave is produced by an oscillator and used to drive a square wave generator, resulting in a simple 75-kHz square wave. The square wave is then integrated, resulting in a triangular waveform, which is mixed with the input audio in a summing circuit. The result is a triangular waveform that effec-

11-52 Broadcast Transmission Systems

**Figure 11.3.4** Basic configuration of the PDM system.**Figure 11.3.5** Principle waveforms of the PDM system.

tively rides on the incoming audio (as shown in Figure 11.3.5). This composite signal is then applied to a threshold amplifier, which functions as a switch that is turned on whenever the value of the input signal exceeds a certain limit. The result is a string of pulses in which the width of pulse is proportional to the period of time the triangular waveform exceeds the threshold. The

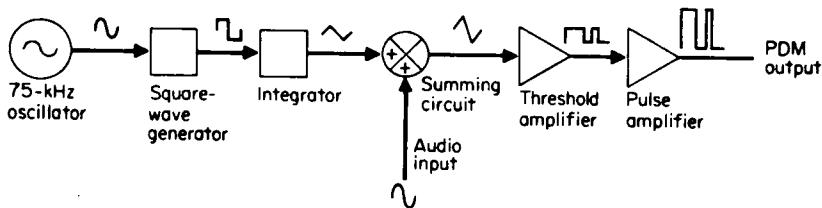


Figure 11.3.6 Simplified block diagram of the PDM system

pulse output is applied to an amplifier to obtain the necessary power to drive subsequent stages. A filter eliminates whatever transients exist after the switching process is complete.

The PDM scheme is—in effect—a digital modulation system with the audio information being sampled at a 75-kHz rate. The width of the pulses contain all the audio information. The pulse-width-modulated signal is applied to a switch or modulator tube. The tube is simply turned on, to a fully saturated state, or off in accordance with the instantaneous value of the pulse. When the pulse goes positive, the modulator tube is turned on and the voltage across the tube drops to a minimum. When the pulse returns to its minimum value, the modulator tube turns off.

This PDM signal becomes the power supply to the final RF amplifier tube. When the modulator is switched on, the final amplifier will have current flow and RF will be generated. When the switch or modulator tube goes off, the final amplifier current will cease. This system, in effect, causes the final amplifier to operate in a highly efficient class D switching mode. A dc offset voltage to the summing amplifier of Figure 11.3.6 is used to set the carrier (no modulation) level of the transmitter.

A high degree of third-harmonic energy will exist at the output of the final amplifier because of the switching-mode operation. This energy is eliminated by a third-harmonic trap. The result is a stable amplifier system that normally operates in excess of 90 percent efficiency. The power consumed by the modulator and its driver is usually a fraction of a full class B amplifier stage.

The damping diode shown in Figure 11.3.4 is used to prevent potentially damaging transient overvoltages during the switching process. When the switching tube turns off the supply current during a period when the final amplifier is conducting, the high current through the inductors contained in the PDM filters could cause a large transient voltage to be generated. The energy in the PDM filter is returned to the power supply by the damping diode. If no alternative route were established, the energy would return by arcing through the modulator tube itself.

The pulse-width-modulation system makes it possible to completely eliminate audio frequency transformers in the transmitter. The result is wide frequency response and low distortion. It should be noted that variations on this amplifier and modulation scheme have been used by other manufacturers for both standard broadcast and short-wave service.

Solid-State Transmitters

Solid-state transmitters make use of various schemes employing pulse modulation techniques. One system is shown in simplified form in Figure 11.3.7a. This method of generating a modulated RF signal employs a solid-state switch to swing back and forth between two voltage levels

11-54 Broadcast Transmission Systems

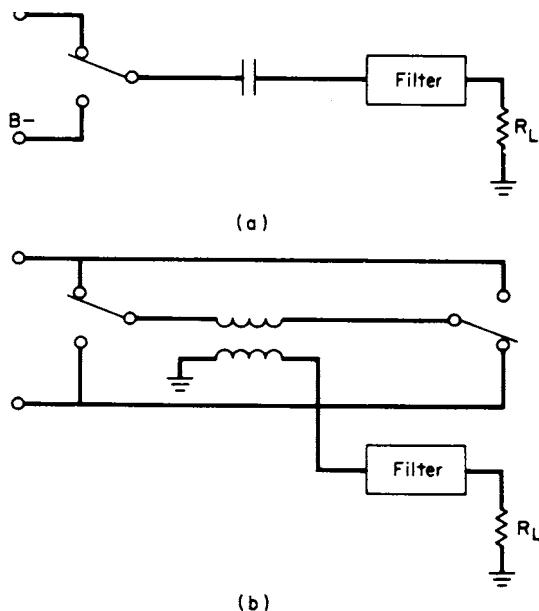


Figure 11.3.7 Solid-state switching RF amplifier: (a) basic system, (b) push-pull configuration.

at the carrier frequency. The result is a square-wave signal that is filtered to eliminate all components except the fundamental frequency itself. A push-pull version of the circuit is shown in Figure 11.3.7b.

Figure 11.3.8 illustrates a class D switching system utilizing bipolar transistors that permits the generation of a modulated carrier. Modern solid-state transmitters use field-effect-transistor (FET) power devices as RF amplifiers. Basically, the dc supply to the RF amplifier stages is switched on and off by an electronic switch in series with a filter. Operating in the class D mode results in a composite signal similar to that generated by the vacuum-tube class D amplifier in the PDM system (at much higher power). The individual solid-state amplifiers are typically combined through a toroidal filter/combiner. The result is a group of low-powered amplifiers operating in parallel and combined to generate the required energy.

Transmitters well above 50 kW have been constructed using this design philosophy. They have proved to be reliable and efficient. The parallel design provides users with automatic redundancy and has ushered in the era of “graceful degradation” failures. In almost all solid-state transmitters, the failure of a single power device will reduce the overall power output or modulation capability of the system but will not take the transmitter off the air. This feature allows repair of defective modules or components at times convenient to the service engineer. The negative effects on system performance are usually negligible.

The audio performance of current technology solid-state transmitters is usually better than vacuum-tube designs of comparable power levels. Frequency response is typically within ± 1 dB from 50 Hz to 10 kHz. Distortion is usually less than 1 percent at 95 percent modulation.

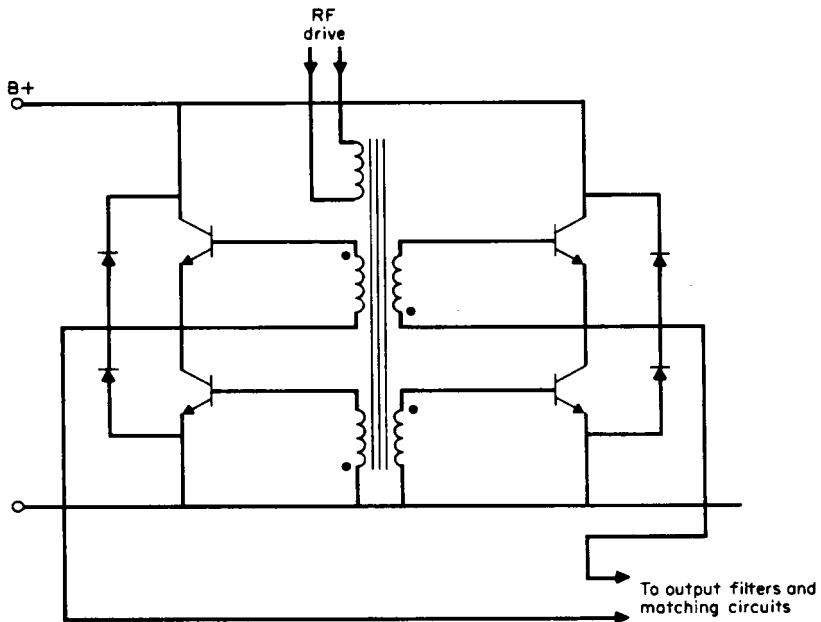


Figure 11.3.8 Bipolar transistor RF stage operating in a class D switching mode for AM service.

11.3.3 FM Broadcasting

Frequency modulation utilizes the audio modulating signal to vary the frequency of the RF carrier. The greater the amplitude of the modulating frequency, the greater the frequency deviation from the center carrier frequency. The rate of the frequency variation is a direct function of the frequency of the audio modulating signal. In FM modulation, multiple pairs of sidebands are produced. The actual number of sidebands that make up the modulated wave is determined by the *modulation index* (MI) of the system. The modulation index is a function of the frequency deviation of the system and the applied modulating signal.

As the MI increases there are more sidebands produced. As the modulating frequency increases for a given maximum deviation, there will be a smaller number of sidebands spaced at wider intervals. Unlike amplitude modulation, which has a percentage of modulation directly proportional to the carrier power, the percentage of modulation in FM is generally referenced to the maximum allowable occupied bandwidth set by regulation. For example, FM broadcast stations are required to restrict frequency deviation to ± 75 kHz from the main carrier. This is referred to as 100 percent modulation for FM broadcast stations.

To determine the frequency spectrum of a transmitted FM waveform, it is necessary to compute a Fourier series or Fourier expansion to show the actual signal components involved. This work is difficult for a waveform of this type, as the integrals that must be performed in the Fourier expansion or Fourier series are difficult to solve. The actual result is that the integral produces a particular class of solution that is identified as the *Bessel function*.

11-56 Broadcast Transmission Systems

Supporting mathematics will show that an FM signal using the modulation indices that occur in a broadcast system will have a multitude of sidebands. To achieve zero distortion, all sidebands would have to be transmitted, received, and demodulated. However, in reality, the $\pm 75\text{-kHz}$ maximum deviation for FM broadcasting allows the transmission and reception of audio with negligible distortion.

The power emitted by an FM transmitter is virtually constant, regardless of the modulating signal. Additional noise or distortion of the amplitude of the waveform has virtually no effect on the quality of the received audio. Thus, FM transmitters may utilize Class C type amplifiers, which cause amplitude distortion but are inherently more efficient than Class A or Class B amplifiers. In addition, atmospheric and man-made noise has little affect, since the receiver clips all amplitude variations off the signal prior to demodulation.

11.3.3a Frequency Modulation

The monophonic system of FM broadcasting was developed to allow sound transmission of voice and music for reception by the general public for audio frequencies from 50 to 15,000 Hz, all to be contained within a $\pm 75\text{-kHz}$ RF bandwidth. This technique provided higher fidelity reception than was available with standard broadcast AM, along with less received noise and interference. FM broadcasting in the U.S. is allocated the 88 to 108 MHz frequency band.

Pre-emphasis is employed in an FM broadcast transmitter to improve the received signal-to-noise ratio. The pre-emphasis upper-frequency limit is based on a time constant of 75 μs , as required by the FCC. Audio frequencies from 50 to 2120 Hz are transmitted with normal FM; whereas audio frequencies from 2120 Hz to 15 kHz are emphasized with a larger modulation index. There is significant signal-to-noise improvement at the receiver, which is equipped with a matching *de-emphasis* circuit.

11.3.3b FM Broadcast Modulation Techniques

FM stereo was developed in 1961 to provide transmission capability for a left- and right-stereo audio signal. Stereophonic transmission is accomplished by adding the left- and right-channel stereo information together in the baseband signal. In addition, a left-minus-right channel is added and frequency multiplexed on a subcarrier of 38 kHz using double sideband suppressed carrier (DSSC) modulation. An unmodulated 19-kHz subcarrier is derived from the 38-kHz subcarrier to provide a synchronous demodulation reference for the stereophonic receiver. In the receiver, a synchronous detector at 38 kHz recovers the left-minus-right channel information, which is then combined with the left-plus-right channel information in sum and difference combiners to produce the original left-channel and right-channel signals. The system works well and, because the baseband mono signal is a combination of both left and right audio, provides full compatibility with monophonic FM receivers.

Modulation Circuits

Early FM transmitters used reactance modulators that operated at a low frequency. The output of the modulator was then multiplied to reach the desired output frequency. This approach was acceptable for monaural FM transmission but not for modern stereo systems or other applications that utilize subcarriers on the FM broadcast signal. Modern FM systems utilize *direct modulation*. That is, the frequency modulation occurs in a modulated oscillator that operates on a

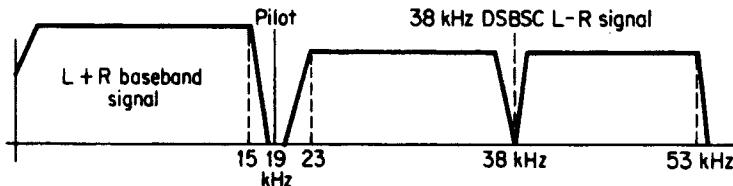


Figure 11.3.9 Composite stereo FM signal.

center frequency equal to the desired transmitter output frequency. In stereo broadcast systems, a composite FM signal is applied to the FM modulator. The basic parameters of this composite signal are shown in Figure 11.3.9.

Various techniques have been developed to generate the direct-FM signal. One of the most popular uses a variable-capacity diode as the reactive element in the oscillator. The modulating signal is applied to the diode, which causes the capacitance of the device to vary as a function of the magnitude of the modulating signal. Variations in the capacitance cause the frequency of the oscillator to vary. Again, the magnitude of the frequency shift is proportional to the amplitude of the modulating signal, and the rate of frequency shift is equal to the frequency of the modulating signal.

The direct-FM modulator is one element of an FM transmitter exciter, which generates the composite FM waveform. A block diagram of a complete FM exciter is shown in Figure 11.3.10. Audio inputs of various types (stereo left and right signals, plus subcarrier programming, if used) are buffered, filtered, and preemphasized before being summed to feed the modulated oscillator. It should be noted that the oscillator is not normally coupled directly to a crystal but is a free-running oscillator adjusted as near as practical to the carrier frequency of the transmitter. The final operating frequency is carefully maintained by an automatic frequency control system employing a *phase-locked loop* (PLL) tied to a reference crystal oscillator or frequency synthesizer.

A solid-state class C amplifier follows the modulated oscillator and increases the operating power of the FM signal to 20–30 W. One or more subsequent amplifiers in the transmitter raise the signal power to several hundred watts for application to the final power amplifier (PA) stage. Nearly all current high-power FM transmitters utilize solid-state amplifiers up to the final RF stage, which is generally a vacuum tube for operating powers of 15 kW and above. All stages operate in the class C mode.

11.3.3c Subsidiary Communications Authorization (SCA)

In addition to its main channel and 38-kHz subcarrier stereo broadcasting, an FM station may utilize additional subcarriers for private commercial use. These SCAs may be used in a variety of ways, such as paging, data transmission, specialized foreign language programs, radio reading services, utility load management, and background music. An FM stereo station may utilize multiplexed subcarriers within the range of 53–99 kHz with up to 20 percent total SCA modulation of the main carrier using any form of modulation, including techniques for the transmission of digital information. The only requirement is that the station does not exceed its occupied bandwidth limitations or cause interference to itself.

11-58 Broadcast Transmission Systems

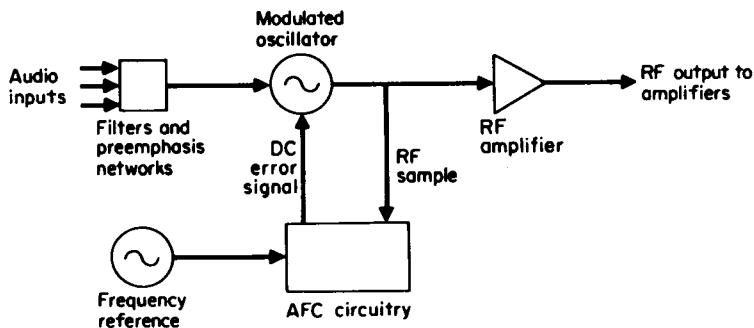


Figure 11.3.10 Block diagram of an FM exciter.

A number of advanced modulation techniques have successfully been applied to transmission of a high-speed digital data stream. These digital audio broadcasting (DAB) systems utilize in-band on channel (IBOC) and in-band adjacent channel (IBAC) techniques to provide high-quality digital audio without interfering with the current analog FM broadcast signals.

11.3.3d Frequency Allocations

The 100 carrier frequencies for FM broadcast range from 88.1 to 107.9 MHz and are equally spaced every 200 kHz. The channels from 88.1 to 91.9 MHz are reserved for educational and noncommercial broadcasting, and those from 92.1 to 107.9 MHz for commercial broadcasting. Each channel has a 200-kHz bandwidth. The maximum frequency swing under normal conditions is ± 75 kHz. Stations operating with an SCA may, under certain conditions, exceed this level, but in no case may exceed a frequency swing of ± 82.5 kHz. The carrier center frequency is required to be maintained within ± 2000 Hz. The frequencies used for FM broadcasting limit the coverage to essentially line-of-sight distances. As a result, consistent FM coverage is limited to a maximum receivable range of a few hundred kilometers depending on the antenna height above average terrain (HAAT) and *effective radiated power* (ERP). The actual coverage area for a given station can be reliably predicted after the power and the antenna height are known. Either increasing the power or raising the antenna will increase the coverage area.

In FM broadcast, stations are classified by the FCC according to their maximum allowable ERP and the transmitting antenna HAAT in their service area. Class A stations provide primary service to a radius of about 28 km with 6000 W of ERP at a maximum HAAT of 100 m. The most powerful class, Class C, operates with maximums of 100,000 W of ERP and heights up to 600 m with a primary coverage radius of over 92 km. All classes may operate at antenna heights above those specified but must reduce the ERP accordingly. Stations may not exceed the maximum power specified, even if the antenna height is reduced. The classification of the station determines the allowable distance to other cochannel and adjacent channel stations.

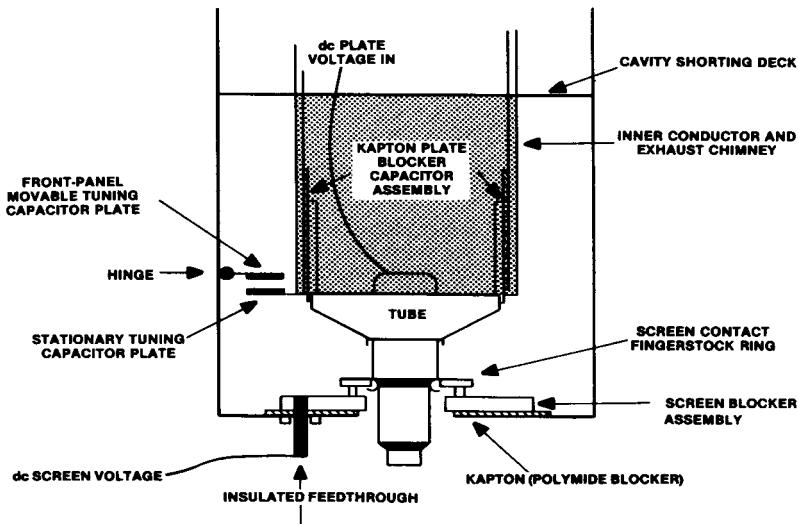


Figure 11.3.11 Physical layout of a common type of 1/4-wave PA cavity for FM broadcast service.

11.3.3e Transmitters

FM broadcast transmitters typically range in power output from 10 W to 50 kW and have performance standards similar to those specified above for AM broadcast transmitters. Digital technology is being used extensively in the processing and exciter stages, allowing for precision modulation control, whereas solid-state devices are predominately used for RF drivers and amplifiers up to several kilowatts. High-power transmitters still rely on tubes in the final stages for the generation of RF power, but manufacturers are developing new devices and technologies that have made high-power solid-state transmitters cost effective and reliable.

FM Power Amplifiers

Nearly all high-power FM transmitters manufactured today employ cavity designs. The 1/4-wavelength cavity is the most common. The design is simple and straightforward. A number of variations can be found in different transmitters, but the underlying theory of operation is the same.

The goal of any cavity amplifier is to simulate a resonant tank circuit at the operating frequency and provide a means to couple the energy in the cavity to the transmission line. Because of the operating frequencies involved (88 to 108 MHz), the elements of the “tank” take on unfamiliar forms.

A typical 1/4-wave cavity is shown in Figure 11.3.11. The plate of the tube connects directly to the inner section (tube) of the plate-blocking capacitor. The blocking capacitor can be formed in one of several ways. In at least one design, it is made by wrapping the outside surface of the inner tube conductor with multiple layers of 8-in-wide and 0.005-in-thick polyimide (Kapton) film. The exhaust chimney-inner conductor forms the other element of the blocking capacitor. The cavity walls form the outer conductor of the 1/4-wave transmission line circuit. The dc plate

11-60 Broadcast Transmission Systems

voltage is applied to the PA tube by a cable routed inside the exhaust chimney and inner tube conductor.

In the design shown in Figure 11.3.11, the screen-contact fingerstock ring mounts on a metal plate that is insulated from the grounded-cavity deck by a Kapton blocker. This hardware makes up the screen-blocker assembly. The dc screen voltage feeds to the fingerstock ring from underneath the cavity deck through an insulated feed through.

Some transmitters that employ the 1/4-wave cavity design use a grounded screen configuration in which the screen contact fingerstock ring is connected directly to the grounded cavity deck. The PA cathode then operates at below ground potential (in other words, at a negative voltage), establishing the required screen voltage for the tube.

The cavity design shown in Figure 11.3.11 is set up to be slightly shorter than a full 1/4-wavelength at the operating frequency. This makes the load inductive and resonates the tube's output capacity. Thus, the physically foreshortened shorted transmission line is resonated and electrically lengthened to 1/4-wavelength.

Tuning the Cavity

Coarse tuning of the cavity is accomplished by adjusting the cavity length. The top of the cavity (the cavity shorting deck) is fastened by screws or clamps and can be raised or lowered to set the length of the assembly for the particular operating frequency.

Fine-tuning is accomplished by a variable-capacity plate-tuning control built into the cavity. In the example, one plate of this capacitor, the stationary plate, is fastened to the inner conductor just above the plate-blocking capacitor. The movable tuning plate is fastened to the cavity box, the outer conductor, and is mechanically linked to the front-panel tuning control. This capacity shunts the inner conductor to the outer conductor and can vary the electrical length and resonant frequency of the cavity.

11.3.4 Transmitter Performance and Maintenance

Assessing the performance of an AM or FM transmitter has been greatly simplified by recent advancements in RF system design. Most transmitters can be checked by measuring the audio performance of the overall system. If the audio tests indicate good performance, chances are good that all the RF parameters also meet specifications.

Both AM and FM broadcasting have certain limitations that prevent them from ever being a completely transparent medium. Top on the list for FM is multipath distortion. In many locations, some degree of multipath is unavoidable. Running a close second is the practical audio bandwidth limitation of the FM stereo multiplex system. The theoretical limit is 19 kHz; however, real-world filter designs result in a high-end passband of between 15 and 17 kHz.

For AM radio, the primary limitation is restricted bandwidth. Modern transmitters are capable of faithfully transmitting wideband audio, but most receivers are not capable of reproducing it. The problem is not just poor-quality AM radios but rather the high-noise environment of the AM band. To reduce adjacent channel interference and atmospheric and artificial noise, receiver designers have restricted the IF and audio bandwidth of their products. The result is a typical AM receiver in which the audio response falls off rapidly beyond 3 kHz.

11.3.4a Key System Measurements

The procedures for measuring AM and FNI transmitter performance vary widely depending on the type of equipment being used. Some generalizations, however, can be made with respect to equipment performance measurements that apply to most systems.

Frequency response is the actual deviation from a constant amplitude across a given span of frequencies. Research has shown that frequency-response accuracy over the audio passband does make an audible difference. Researchers exploring subtle differences in audio amplifier design have found that errors as small as 0.2 dB can be heard. Therefore, flat response (strict adherence to the 75- μ s preemphasis curve) is an important goal in FM broadcasting. Typical performance targets for an FM station are ± 1 dB, 50 Hz to 15 kHz. Typical targets for an AM station are ± 1 dB, 50 Hz to 10 kHz.

Total harmonic distortion (THD) is the creation by a nonlinear device of spurious signals harmonically related to the applied audio waveform. Research has shown that although THD levels greater than 1 percent are detectable during sinewave testing, listeners will tolerate somewhat higher levels of THD on musical material. It must be noted that depending upon the setup of the station's audio processor, distortion targets of 1 percent or less may be impossible to meet with the processor on-line. Because of the highly competitive nature of broadcasting today, audio processors cannot always be adjusted to provide for the purest reproduction of the incoming signal.

Audio processing aside, most FM transmitters will yield THD readings well below 1 percent at frequencies up to 7.5 kHz. AM transmitters can typically produce distortion figures below 1.5 percent at frequencies up to 5 kHz at 95 percent modulation.

The THD test is sensitive to the noise floor of the system under test. If the system has a signal-to-noise ratio of 60 dB, the distortion analyzer's best possible reading will be greater than 0.1 percent ($60 \text{ dB} = 0.001 = 0.1 \text{ percent}$).

Intermodulation distortion (IMD) is the creation by a nonlinear device of spurious signals not harmonically related to the audio waveform. These distortion components are sum-and-difference (*beat notes*) mixing products that research has shown are more objectionable to listeners than even-harmonic distortion products. The IMD measurement is relatively impervious to the noise floor of the system under test. IMD performance targets for AM and FM transmitters are the same as the TH D targets mentioned previously.

Signal-to-noise ratio (S/N) is the amplitude difference, expressed in decibels, between a reference level audio signal and the system's residual noise and hum. In many cases, system noise is the most difficult parameter to bring under control. An FM performance target of -70 dB per stereo channel reflects state-of-the-art, exciter-transmitter performance. Most AM transmitters are capable of -60 dB S/N for monaural operation, and -55 dB per channel for stereo operation.

Separation is a specialized definition for signal crosstalk between the left and right channels of a stereo system. The separation test is performed by feeding a test tone into one channel while measuring leakage into the other channel (whose input is terminated with a $600\text{-}\Omega$ wirewound resistor, or other appropriate value). Typical performance targets for an FM station are -35 dB from 50 to 400 Hz, and -40 dB from 400 Hz to 15 kHz. Targets for a stereo AM station are -20 dB from 50 to 40 Hz and -25 dB from 400 Hz to 10 kHz.

Synchronous AM in FM Systems

The output spectrum of an FM transmitter contains many sideband frequency components, theoretically an infinite number. The practical considerations of transmitter design and frequency

11-62 Broadcast Transmission Systems

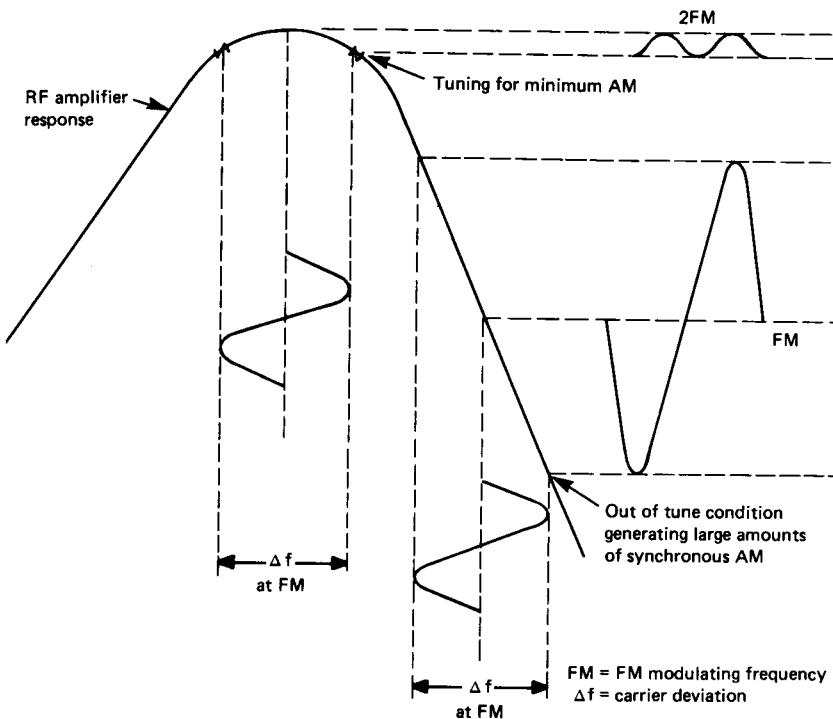


Figure 11.3.12 Synchronous AM is generated by one or more narrow-band stages in an FM transmitter. The amount of AM produced is a function of the flatness of the composite RF system.

allocation make it necessary to restrict the bandwidth of all FM broadcast signals. Bandwidth restriction brings with it the undesirable side-effects of phase shifts through the transmission chain, the generation of synchronous AM components, and distortion in the demodulated output of a receiver.

In most medium- and high-power FM transmitters, the primary out-of-band filtering is performed in the output cavity of the final stage. Previous stages in the transmitter (the exciter and IPA) are designed to be broadband, or at least more broadband than the PA.

As the bandwidth of an FM transmission system is reduced, synchronous amplitude modulation increases for a given carrier deviation (modulation). Synchronous AM is generated as tuned circuits with finite bandwidth are swept by the frequency of modulation. The amount of synchronous AM generated is dependent on tuning, which determines (to a large extent) the bandwidth of the system. Figure 11.3.12 illustrates how synchronous AM is generated through modulation of the FM carrier.

Incidental Phase Modulation

Before the advent of AM stereo, few AM broadcast engineers understood the term *incidental phase modulation* (IPM), let alone concerned themselves with its effect on AM transmitters.

IPM is defined as phase modulation produced by an AM transmitter as a result of amplitude modulation. In other words, as an AM transmitter develops the amplitude-modulated signal, it also produces a phase modulated, or PM, version of the audio as well.

In theory, IPM is of little consequence for monophonic AM because FM and PM do not affect the carrier amplitude. An envelope detector will, in theory at least, ignore IPM. This is true to a point; however, stations attempting to broadcast AM stereo must pay close attention to IPM. Any undesired carrier phase modulation, such as IPM, will have an effect on stereo performance.

As a general rule, because IPM is a direct result of the modulation process, it can be generated in any stage that is influenced by modulation. The most common cause of IPM in plate-modulated and pulse-modulated transmitters is improper neutralization of the final RF amplifier. Adjusting the transmitter for minimum IPM is an accurate way of achieving proper neutralization. The reverse is not always true, however, because some neutralization methods will not necessarily result in the lowest amount of IPM.

Improper tuning of the IPA stage is the second most common cause of IPM. As modulation changes the driver loading to the PA grid, the driver output also may change. The circuits that feed the driver stage usually are isolated enough from the PA that they do not produce IPM. An exception can be found when the power supply for the IPA is influenced by modulation. Such a problem could be caused by a loss of capacitance in the high-voltage power supply itself.

11.3.5 Bibliography

- Bordonaro, Dominic: "Reducing IPM in AM Transmitters," *Broadcast Engineering*, Intertec Publishing, Overland Park, Kan. May 1988.
- Crutchfield, E. B. (ed): *NAB Engineering Handbook*, 8th ed., National Association of Broadcasters, Washington, D.C., 1992.

Chapter

11.4

Radio STL Systems

Jerry C. Whitaker, Editor-in-Chief

11.4.1 Introduction

One of the major concerns in the design and operation of a radio broadcasting facility is the means by which the program audio from the studio is conveyed to the transmitter site. As illustrated in Figure 11.4.1, this link represents an important element in the overall reliability of the transmission chain. Furthermore, as digital technology continues to move into daily radio station operation, the *studio to transmitter link* (STL) must become as transparent as possible. An inferior link will impose an unacceptable limit on overall audio quality. The requirements for reliability and transparent program relay have led to the development of new STL systems based on digital technology.

Changes in FCC broadcast ownership rules and the popularity of *local marketing agreements* (LMAs) have reshaped radio broadcasting. The need for high-quality audio programming is one outgrowth of new competition and new alliances. STL systems are an important component of these audio improvement efforts. Furthermore, increasing numbers of stations are using *intercity relay* (ICR) facilities to share programming. Unfortunately, in many areas of the United States, the demand for 950 MHz STL channels has far outstripped the available supply. The Part 74 bandwidth allocations for STL systems, therefore, necessitate highly-efficient designs to meet the needs of radio broadcasters today.

Market demand for STL systems is not limited to North America. The commercialization of radio broadcasting in Europe and elsewhere has led to increased use of radio links to relay programming from one studio to the next, and from the studio to the transmitter site. In some areas, repeated use of the same frequency places stringent demands on system design.

11.4.1a Radio vs. Landline Links

The arguments over which approach—radio STL or landline (telco)—is the better way to convey program audio from the studio to the transmitter is as old as radio broadcasting itself. Now, however, a new element has been added to the equation that makes the landline option considerably more attractive. That new element is digital technology.

For years, broadcasters were limited in their choice of STL equipment to two main options: 950 MHz radio, and equalized analog telephone lines. Today, a third approach, T1 digital trans-

11-66 Broadcast Transmission Systems

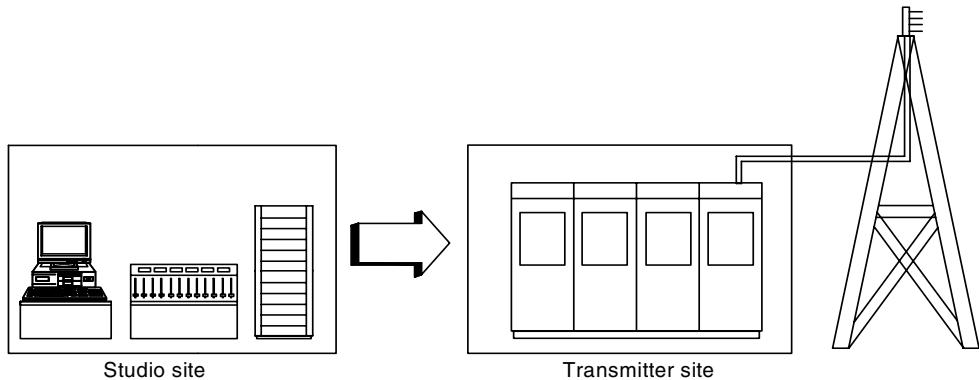


Figure 11.4.1 The basic function of a studio-to-transmitter link. (After [2].)

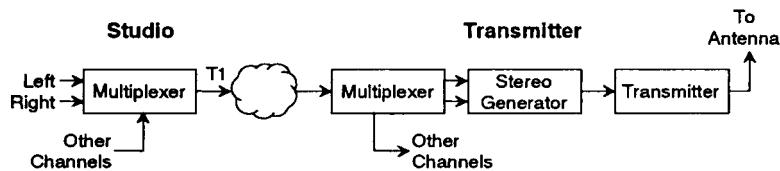


Figure 11.4.2 Discrete channel T1 digital audio STL system. (After [1].)

mission, is rapidly becoming the system of choice for many radio stations [1]. The decision to use T1 may be based on several factors, including:

- Necessity. The station has no line-of-sight to the transmitter, or suitable frequencies are unavailable.
- Sound quality. A digital STL can sound better than even the best analog systems.
- Cost. A single leased T1 line can cost less than multiple leased analog lines.

Whatever the reasons, a broadcaster who decides to use T1 transmission must then choose what type of system to implement. Digital T1 STL systems can be designed either to transmit discrete left and right channel stereo, as shown in Figure 11.4.2, or to transmit a composite stereo signal, illustrated in Figure 11.4.3.

One of the advantages to a broadcaster in using a T1 digital STL is the ability to multiplex several signals together. SCAs, SAP channels, transmitter remote control signals, data, and voice can all be combined with the broadcast audio signal for transmission on a single T1 circuit. The duplex nature of T1 permits use of the same system for both STL and *transmitter-to-studio* (TSL) functions as well. Table 11.4.1 compares T1 bandwidth for various circuit implementations.

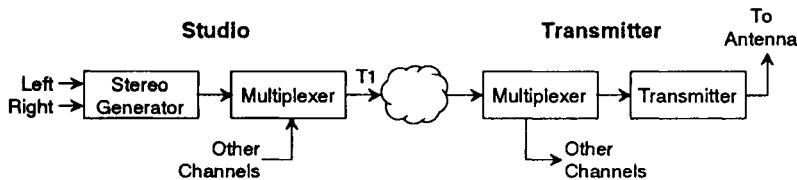


Figure 11.4.3 Composite T1 digital audio STL system. (After [1].)

Table 11.4.1 Digital Audio Sampling Size vs. T1 Bandwidth Usage (After [1].)

System Type	Samples/s (for 15 kHz stereo)	Sample Size	Digital Transmission Rate	Number of T1 Time Slots Used for Broadcast Audio	Number of T1 Time Slots Available for Other Channels
Discrete	$32,000 \times 2$	16-bit	1.152 Mbits/s	18	6
Composite	112,000	16-bit	1.792 Mbits/s	Not usable	--
		15-bit	1.680 Mbits/s	Not usable	--
		14-bit	1.568 Mbits/s	Not usable	--
		13-bit	1.456 Mbits/s	23	1
		12-bit	1.344 Mbits/s	21	3

11.4.2 STL System Configuration

The frequency range used by aural STL systems in the U.S. is 944.5 to 951.5 MHz. Frequencies ranging from 200 to 940 MHz are used in other parts of the world. STL systems are also operated above these traditional bands, however, such links often face difficulties associated with increased susceptibility to signal loss as a result of precipitation, foliage, and fog, which limits usage to short path lengths. Frequency modulation is used for analog STL systems.

The vast majority of radio STL systems in operation today carry a composite stereo baseband signal from the studio to the transmitter. An alternative approach, and one that may be used for monaural AM stations, involves the use of one or two narrowband STL systems. The two approaches are compared in Figure 11.4.4.

The monaural STL typically has an audio bandwidth of up to 15 kHz for program signals, and usually can accommodate a single FM subcarrier at approximately 39 kHz. For stereo or dual mono installations, a common antenna can be used at the transmitter and at the receiver. At the transmit site, a combiner/isolator adds the signals and prevents the generation of intermodulation products. At the receive site, a low-noise antenna splitter feeds the individual receivers.

The composite STL has sufficient bandwidth (220 kHz) to accommodate the output of a stereo FM generator and usually several separate subcarriers. A low-pass filter generally is used at the output of the receiver to separate the signals to be broadcast over-the-air from the non-broadcast (or closed-circuit) signals. A typical configuration is illustrated in Figure 11.4.5. The com-

11-68 Broadcast Transmission Systems

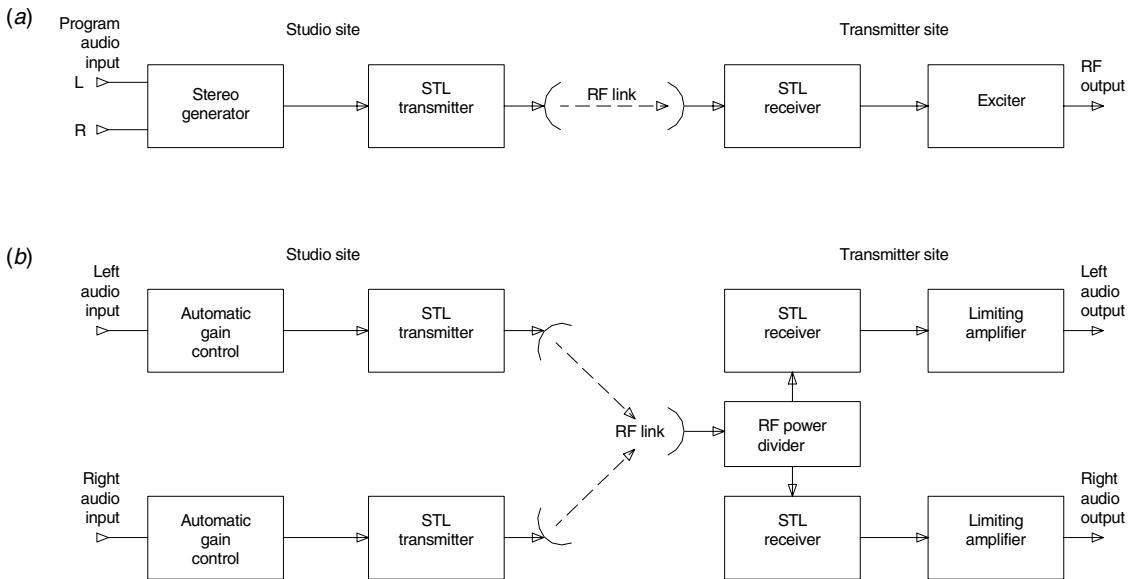


Figure 11.4.4 Comparison of STL systems: (a) composite transmitter-receiver system, (b) dual monaural transmitter-receiver system. (After [2].)

posite STL provides superior stereo performance compared to dual monaural radio links in several respects, including:

- Elimination of interchannel phase and amplitude errors that can arise in a dual channel system.
- Elimination of audio headroom considerations because the STL input signal has already been passed through the station's audio processing system and the stereo generator, which are located at the studio.

For FM applications, the composite signal from the STL receiver is fed directly into the FM transmitter exciter. An alternate method involves the use of *Reciter* [2], which functions as an STL receiver and FM exciter but with an IF interface rather than a baseband interface. The elimination of one stage of demodulation and remodulation reduces noise and distortion in the transmitted FM signals.

Because AM stereo excitors require discrete right and left channel audio inputs, a complementary stereo decoder is used to recover the left and right channel audio if a composite STL (fed by a stereo generator) is used.

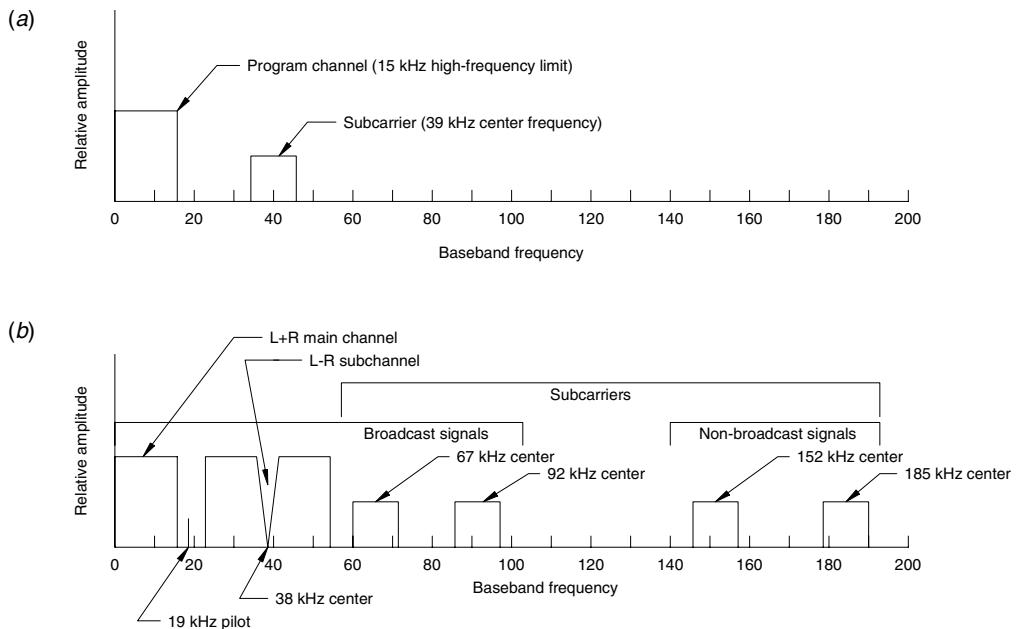


Figure 11.4.5 Baseband spectrum of STL systems: (a) monaural, (b) composite. (After [2].)

11.4.2a Digital STL vs. Analog STL Link

Digital signal encoding of audio frequency information provides a number of advantages over conventional analog systems. The principal strengths of digital coding for transmission and storage vs. analog methods include the following:

- Greater immunity to noise and interference in the transmission path. Differences of 40 dB SNR vs. 76 dB are common.
- Elimination of transmission-path-dependent distortion mechanisms, such as harmonic distortion, intermodulation distortion, and crosstalk.
- Efficient use of baseband and RF spectrum.
- Efficient and predictable regeneration of the digital signal.
- Easy and effective encryption for security and coding purposes.
- Uniformity in transmitting audio and data signals.

The last item merits additional consideration. Intense development has been underway for a number of years to develop efficient and cost-effective digital coding and compression systems for audio frequency signals. The results of this work can be seen in everything from personal computer sound boards to recordable MP3-type minidisks. This technology, obviously, has many spinoff effects that can be used to advantage by radio broadcasters. Advanced digital devices and

11-70 Broadcast Transmission Systems

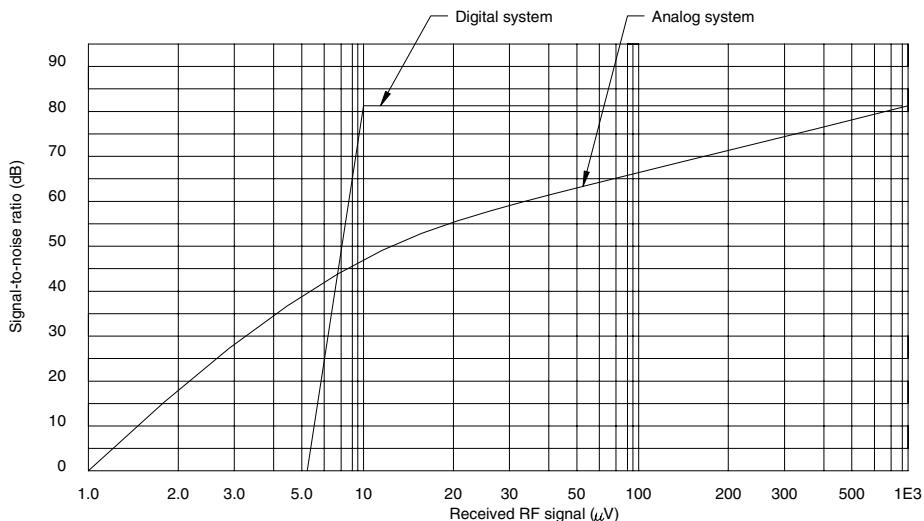


Figure 11.4.6 The benefits of digital vs. analog STL systems in terms of S/N and received RF level. (After [2].)

algorithms are now being adapted for use in a wide variety of broadcast products, most notably studio storage and mixing systems, and STL systems.

The use of digital coding and modulation in an STL system provides the user a number of benefits in terms of performance and path reliability. By its nature, digital coding offers a more robust link. As long as the received signal strength and carrier-to-noise exceed a defined minimum value, the *bit error rate* remains relatively constant, and the decoded baseband signal is an exact duplicate of the input signal at the studio. (See Figure 11.4.6.) With adequate error correction, signals can be received error-free. With a digital STL, multiple links are practical because signal degradations are not additive, as in an analog system.

It is fair to point out that digitization of the input audio signal always brings with it a measure of degradation (quantization errors), but with the high sampling rates typically used for professional audio applications, such degradation is minor and almost always completely inaudible.

The process of quantization is illustrated in Figure 11.4.7. The sampling rate and quality of the sampling circuit determine, in large part, the overall quality of the digital system. A properly operating transmission channel can be assumed to provide error-free throughput. This being the case, the digital signal can be regenerated at the receiving point as an exact duplicate of the input waveform, quantization errors excepted. Figure 11.4.8 shows a general representation of a digital communications channel. In the case of a radio link, such as an STL, the transmission medium is analog in nature (FM). The circuits used to excite the FM modulator, however, are essentially identical to those used for an all-digital link, such as fiber optic cable.

The functions of the encoder and decoder, shown in Figure 11.4.8a, usually are formed into a single device, or set of devices (a chip set), known as a *codec* (*coding and decoding* device). At the transmission end, the codec provides the necessary filtering to band-limit the analog signal to avoid aliasing, thereby preventing *analog-to-digital* (A/D) conversion errors. At the receiver, the

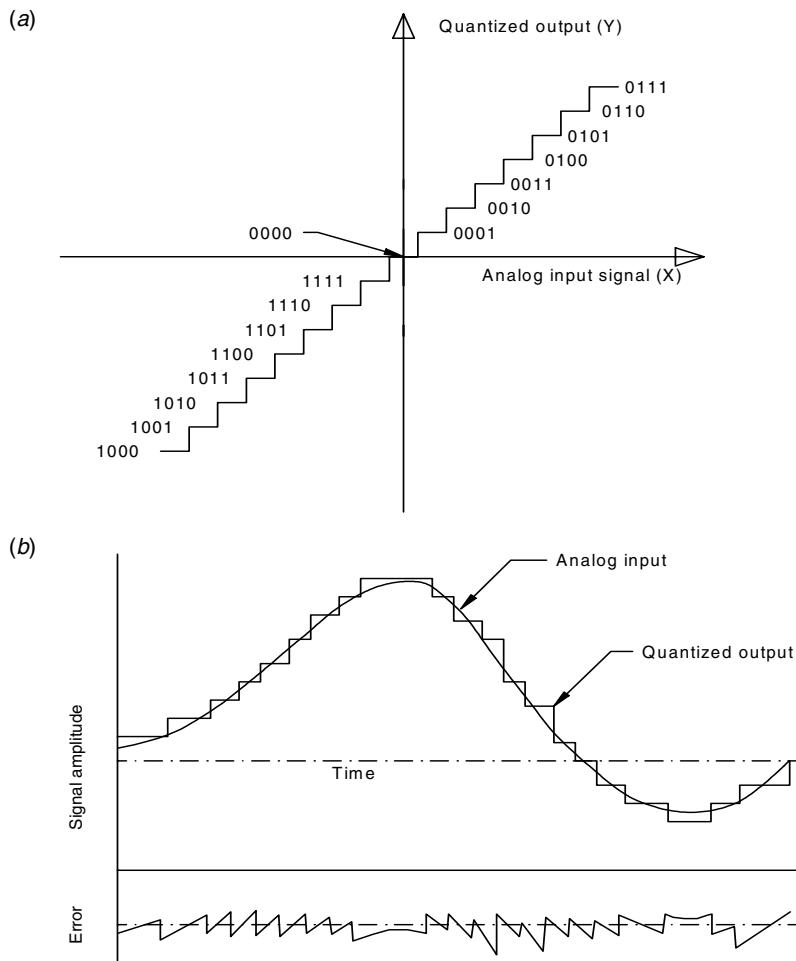


Figure 11.4.7 Quantization of an input signal: (a) quantization steps, (b) quantization error signal. (After [2].)

codec performs the reciprocal *digital-to-analog* (D/A) conversion and interpolates (smooths) the resulting analog waveform.

Digital STL Performance

The benefits of a digital STL for radio broadcasters can best be appreciated by comparing the performance of a digital system and an analog system. A digital STL typically permits broadcasters to extend the fade margin of an existing analog link by 20 dB or more. Furthermore, audio *signal-to-noise* (S/N) improvements of at least 10 dB can be expected for a given RF signal strength. Alternatively, for the same S/N, the maximum possible path distance of a given com-

11-72 Broadcast Transmission Systems

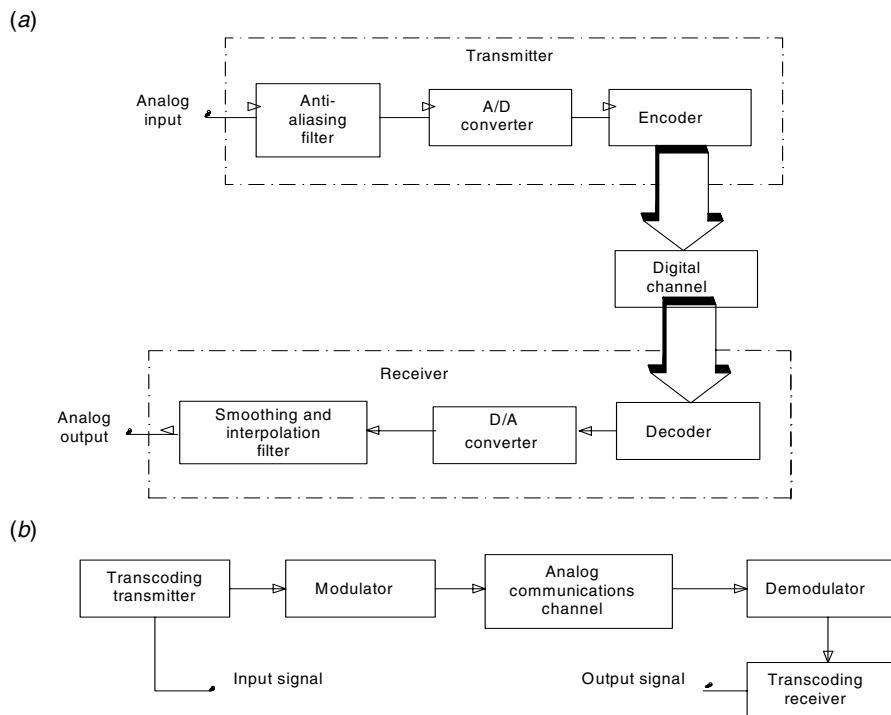


Figure 8 Digital transmission system: (a) coding/decoding functions, (b) overall communications link. (After [2].)

posite STL transmitter and receiver can be extended. These features could, in some cases, make the difference between a one-hop system and a two-hop system.

The spectrum-efficiency of a digital STL is of great importance today in highly-congested markets. The system may, for example, be capable of relaying four program channels and two voice-grade channels. The use of digital coding also makes the signals more tolerant of co-channel interference than a comparable analog STL.

Coding System

Several approaches may be used to digitize or encode the input audio signals. The complexity of the method used is a function of the availability of processing power and encoder memory, and of the resulting delay incurred during the encoding/decoding process. For a real-time function such as an STL, significant encoding/decoding delay is unacceptable. *Pulse code modulation* (PCM) is a common scheme that meets the requirements for speed and accuracy. In the PCM process, the sampled analog values of the input waveform are coded into unique and discrete values. This quantization may be uniform, as illustrated in Figure 11.4.7, or nonuniform. With nonuniform quantization, compression at the coder and subsequent expansion at the decoder is performed. By using larger quantization steps for high energy signals and smaller steps for low energy sig-

nals, efficient use is made of the data bits, while maintaining a specified signal-to-quantization noise level. This process is known as *companding* (compression and expansion).

PCM encoding, in a simple real-time system, provides a high-speed string of discrete digital values that represent the input audio waveform. Each value is independent of all previous samples. No encoder memory is required. This approach, while simple and fast, is not particularly efficient insofar as the transmission channel is concerned. There are many redundancies in any given input signal. By eliminating the redundancies, and taking advantage of the *masking* effects of human hearing, greater transmission efficiency can be realized. Viewed from another perspective, for a given radio transmission bandwidth, more information can be transferred by using a compression system that removes nonessential data bits.

11.4.2b Analog Composite STL Transmitter and Receiver Characteristics

As mentioned previously, the composite aural STL is the workhorse of the radio industry. A number of hardware variations can be found in transmitter and receiver systems, but all are designed to accomplish basically the same goal: relay of a baseband signal from the studio site to the transmitter. Figure 11.4.9 illustrates the basic STL configurations. A block diagram of a representative composite STL is given in Figure 11.4.10.

The RF carrier is generated by a *voltage-controlled oscillator* (VCO) that is phase-locked to a modulating VCO operating at an intermediate frequency of approximately 70 MHz. Direct composite FM modulation takes place at the IF VCO, which is also phase-locked to an oven-controlled crystal oscillator that serves as the master timebase for the transmitter. Because frequency multiplication is not used, the RF carrier is an exact reproduction of the IF VCO. Therefore, a number of potential compromises resulting from frequency-multiplication and post-heterodyne filtering are eliminated, including:

- Degradation of S/N
- Generation of spurious signals
- Degraded stereo performance

IF modulation provides excellent overall stereo performance, as documented in Table 11.4.2.

In this design, the operating frequency of the transmitter is programmable by DIP switches in 12.5 kHz steps, so that frequency changes can be made in the field. The composite baseband signal accommodates the full stereo baseband, including two SCA and/or MUX channels. These features permit all processing equipment for analog operation to be located at the studio. An illustration of the composite baseband of such a system is shown in Figure 11.4.11.

A block diagram of the companion composite STL receiver is shown in Figure 11.4.12. Like the transmitter, the receiver is user-programmable in 12.5 kHz steps through the use of internal DIP switches. The front-end uses cascaded high-*Q* cavity filters and *surface acoustic wave* (SAW) IF filters to provide high selectivity and phase linearity. Triple conversion IF is used to feed a pulse-counting discriminator for linear baseband demodulation.

11.4.2c Components of an STL System

Any STL installation is only as good as the hardware used to interconnect the link. All components, from the transmitter to the output connectors, must be carefully chosen and properly

11-74 Broadcast Transmission Systems

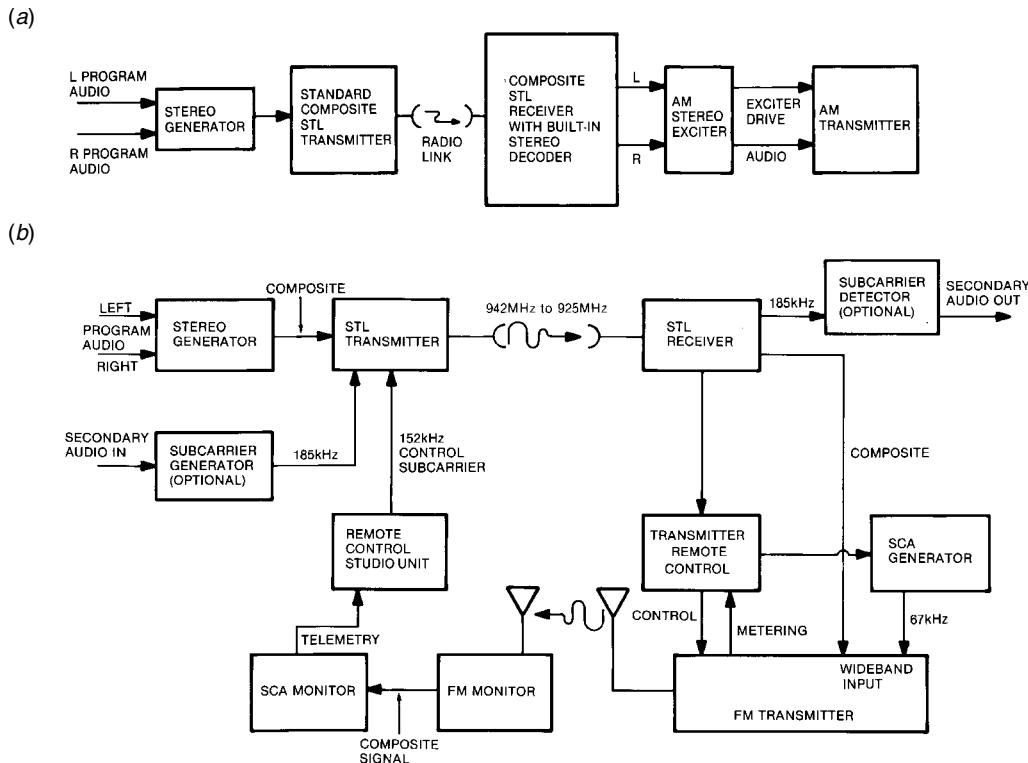


Figure 11.4.9 Analog composite STL links: (a) AM radio application, (b) FM radio application. (After [2].)

installed. A well-designed system will provide years of trouble-free service. A poorly-designed system will cause problems on a regular basis.

Transmitter and Receiver Requirements

When choosing an STL transmitter, it will be necessary to select a unit that will deliver sufficient power to overcome the losses determined by path gain/loss calculations. Figure 11.4.13 provides a starting point for system planning. In cases where additional power is needed, a separate power amplifier can be used. For convenience in manipulating figures, the transmitter power output should be converted to gain in decibels above a 1 mW reference (dBm). (See Table 11.4.3.)

When choosing an STL receiver, specifications should be carefully analyzed, particularly receiver sensitivity. This figure, necessary in STL path calculations, is usually specified as a signal level required for a specified S/N. This value should be converted to dBm. (See Table 11.4.4.) For example, a receiver may require 100 μ V for 60 dB S/N. This is equivalent to -66.9 dBm. In receiver design, sensitivity, S/N, selectivity, and the method of demodulation are determining factors of receiver quality. The use of SAW filters provide sharper selectivity and more

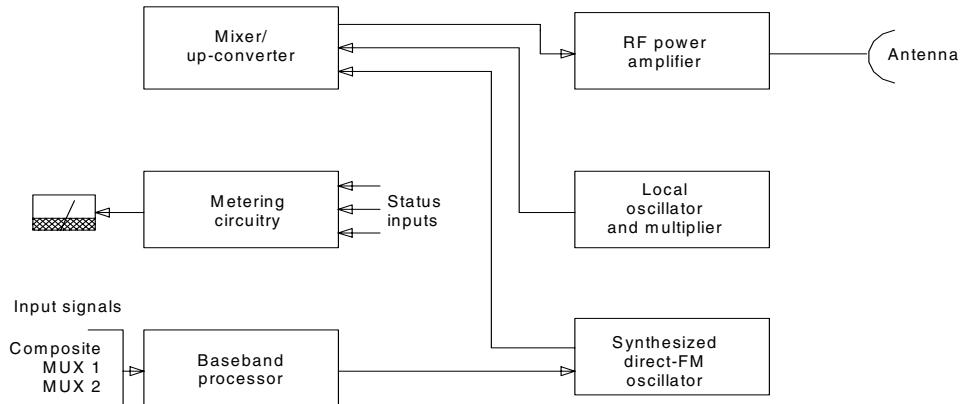


Figure 11.4.10 Block diagram of a composite STL transmitter. (After [2].)

Table 11.4.2 Specifications for a Typical Composite STL System (After [2].)

Parameter	Specification
Power output	6 to 8 W
Frequency stability	$\pm 0.0002\%$, 0 to 50°C
Spurious emissions	60 dB below maximum carrier power
Baseband frequency response	± 0.1 dB or less, 30 Hz to 75 kHz
Stereo separation	Greater than 55 dB at 1 kHz
Total harmonic distortion	0.02%, 75 μs deemphasis
S/N	85 dB below ± 75 kHz deviation, 75 μs deemphasis
Nonlinear crosstalk	50 dB or less
Subchannel-to-main crosstalk	60 dB or less

linear phase response. These attributes yield better stereo separation and lower distortion. A pulse-counting discriminator also provides low distortion and accurate demodulation of the received signal. Phase-linear lowpass filtering is critical for best stereo separation.

A low-noise RF preamplifier may be added to the system when the received signal level is low. For best performance, the preamplifier should be mounted directly at the receive antenna. Care must be taken, however, to prevent overloading the receiver front-end by unwanted, and often strong, interfering signals.

In areas of frequency congestion, narrowband receivers are important in preventing interference from other transmitters. STL manufacturers have responded to the needs of broadcasters in congested RF environments by providing suitable narrowband STL systems. Such receivers typically incorporate bandpass cavity filters, helical resonators, or mechanical coaxial filters. SAW filters and ceramic filters in IF stages also may be included to improve selectivity.

11-76 Broadcast Transmission Systems

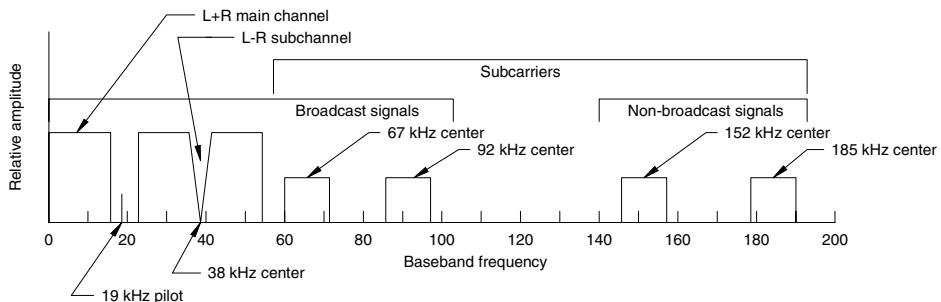


Figure 11.4.11 The baseband of a composite STL transmitter. (After [2].)

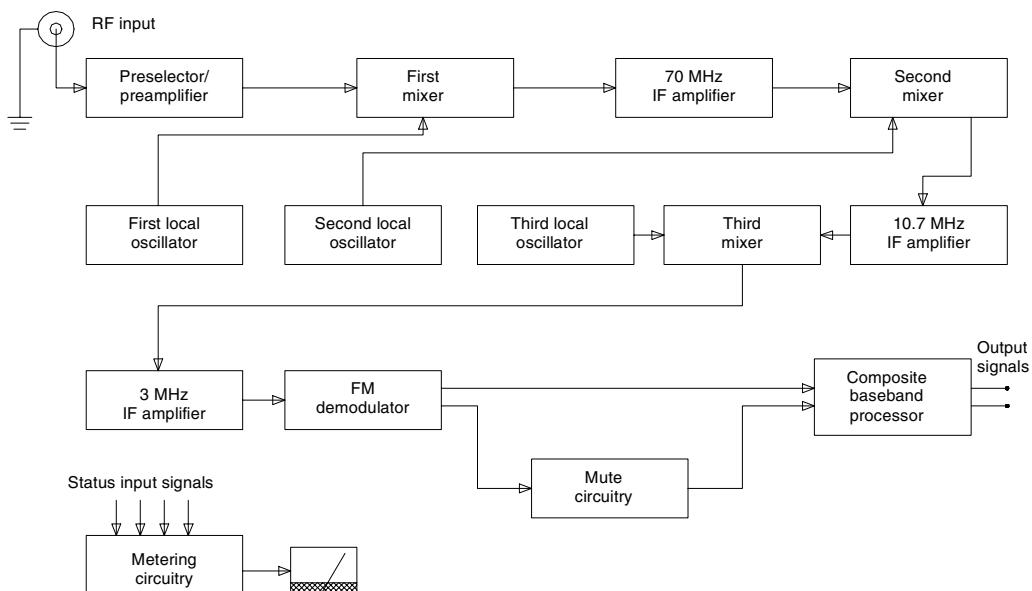


Figure 11.4.12 Block diagram of a composite STL receiver. (After [2].)

Choose a monaural, composite, or digital STL as needed. Compare performance specifications and carefully analyze the long-term benefits of each system. Examine what the STL will need to handle two or five years from now.

Example: $H_T = 60$ feet. $H_R = 500$ feet. Radio path length = 41.5 miles

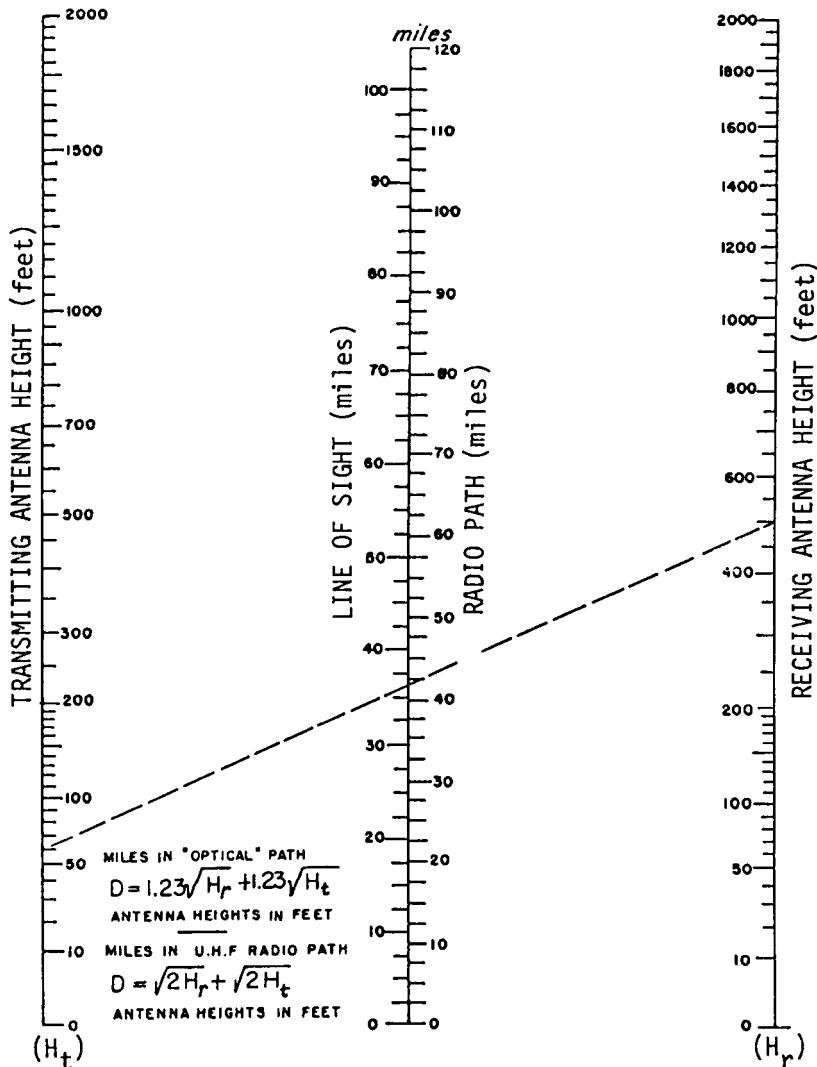


Figure 11.4.13 Nomogram to quickly determine if a path is viable. (After [2].)

Transmission Lines

Figure 11.4.14 shows the primary hardware elements required for an aural STL. Transmission

11-78 Broadcast Transmission Systems

Table 11.4. 3 Conversion of Watts to dBm (50 Ω Impedance)

Power in Watts	dBm	Power in Watts	dBm
5.0	37.0	18.0	42.6
5.5	37.42	19.0	42.8
6.0	37.79	20.0	43.0
6.5	38.12	25.0	43.9
7.0	38.46	30.0	44.8
7.5	38.75	35.0	45.4
8.0	39.0	40.0	46.0
8.5	39.3	45.0	46.5
9.0	39.55	50.0	47.0
9.5	39.78	55.0	47.4
10.0	40.0	60.0	47.8
10.5	40.2	65.0	48.1
11.0	40.4	70.0	48.5
12.0	40.8	75.0	48.8
13.0	41.1	80.0	49.0
14.0	41.5	85.0	49.3
15.0	41.8	90.0	49.5
16.0	42.0	95.0	49.9
17.0	42.3	100.0	50

line sections, connections, and strain-relief provisions are important for long-term reliability. The main criteria in the selection of transmission line include the following:

- Amount of signal attenuation
- Physical parameters (dielectric material and size)
- Purchase and installation cost

In general, the larger the diameter of the transmission line, the lower the loss, and the greater the cost of the line. Loss is also affected by the type of dielectric material used. The most common types of dielectric are air and foam. Air dielectric cable typically requires pressurization and is, therefore, seldom used for 950 MHz installations. For the purpose of gain/loss calculations, cable loss for a particular installation can be determined from the transmission line manufacturer's specification table, given in decibels per unit of measure. Attenuation for common types of coaxial line are given in Table 11.4.5.

Other electrical and mechanical specifications (mainly cable size and impedance) must be compatible with the transmitter, receiver, and antennas to be used. Connector loss must also be considered. It is important to minimize each potential source of signal loss when assembling the system. There are no "minor details" in the installation of an STL.

Table 11.4.4 Conversion of Microvolts to dBm (50 Ω Impedance)

Microvolts	dBm	Microvolts	dBm
0.10	-127	40	-74.9
0.12	-125.25	45	-73.9
0.14	-124	50	-72.9
0.16	-122.9	60	-71.25
0.18	-121.9	70	-70
0.20	-120.9	80	-68.9
0.25	-119	90	-67.9
0.30	-117.25	100	-66.9
0.35	-116	120	-65.25
0.40	-115	140	-64
0.60	-111.25	160	-62.9
0.70	-110	180	-61.9
0.80	-108.9	200	-60.9
0.90	-107.9	250	-59
1.0	-106.9	300	-57.5
1.2	-105.25	350	-56
1.4	-104	400	-54.9
1.6	-102.9	450	-53.9
1.8	-101.9	500	-52.9
2.0	-100.9	600	-51.25
2.5	-99	700	-50
3.0	-97.5	800	-49
3.5	-96	900	-48
4.0	-95	1000	-46.9
4.5	-93.9	1200	-45.25
5.0	-92.9	1400	-44
6.0	-91.25	1600	-42.9
7.0	-90	1800	-41.9
8.0	-88.9	2000	-40.9
9.0	-87.9	2500	-39
10	-86.9	3000	-37.25
1.1	-86	3500	-36
12	-85.25	4000	-34.9
14	-84	4500	-33.9
16	-82.9	5000	-33
18	-81.9	6000	-31.25
20	-80.9	7000	-30
25	-79	8000	-28.9
30	-77.25	9000	-27.9
35	-76	10,000	-26.9

11-80 Broadcast Transmission Systems

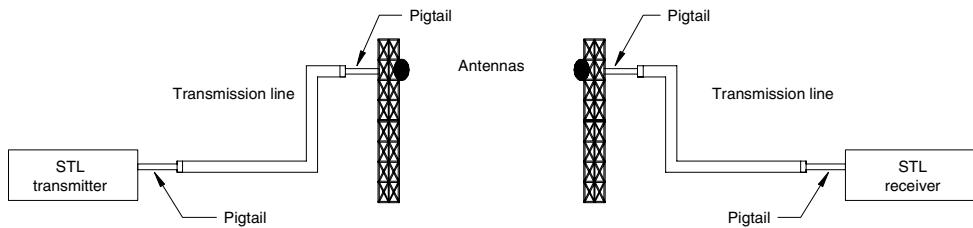


Figure 11.4.14 Principle physical components of an STL system. (After [2].)

Table 11.4.5 Typical Coaxial Transmission Line Loss at 950 MHz

Cable Type	Loss (dB/100 Feet)
RG-8/U	8.5
RG-218/U	3.8
1/2-in foam dielectric	3.0
7/8-in foam dielectric	2.0
1/2-in low-loss foam dielectric	2.4
7/8-in low-loss foam dielectric	1.4
1-5/8-in foam dielectric	1.4
1/2-in air dielectric	2.7
7/8-in air dielectric	1.4
1-5/8-in air dielectric	0.7

Stain relief must be provided on each end of the cable run (at the transmitter and at the receiver). So-called pigtail or jumper cables are commonly used for this purpose. They permit movement without straining cable and chassis connections. Because the pigtails commonly are terminated with *N-type* male connectors on both ends, the main transmission line must be configured with female *N-type* connectors on both ends if a pair of pigtails are used.

Antenna System

At the frequencies commonly used for STL operation, antennas can readily focus the signal into a narrow beam. This focusing, required for point-to-point communications, provides high gain at both the transmitter and the receiver. Several types of antennas are available, including parabolic and parabolic section. Parabolic antennas are available in solid or grid styles, while parabolic section antennas usually are grid-type in design. Antenna models differ in a number of respects, including:

- Gain (directly proportional to size)
- Operating frequency range

- Polarization (most antennas can be set for either horizontal or vertical polarization using universal mounting hardware kits)
- Beamwidth
- Front-to-back ratio
- Windloading
- Structural strength

Environmental factors may also need to be considered so that suitable antenna accessories, such as deicers and/or radomes, are specified.

Antenna gain is specified in decibels referenced to an isotropic antenna (dBi) or decibels referenced to a dipole antenna (dBd). Conversion from dBd to dBi is as follows:

$$\text{dBi} = 2.2 + \text{dBd} \quad (11.4.1)$$

For path analysis calculations of system gains and losses, dBi is used.

Figure 11.4.15 plots the response of a common type of STL antenna. Note the high directivity provided by the device and the difference in gain between operation in the 450 MHz band as opposed to the 950 MHz band. Table 11.4.6 lists the typical gain provided by antennas of different physical size.

Mounting Structures

Each antenna requires a mounting structure, such as the side of a roof or building, or an existing or custom-built tower. The structure must be sufficiently high to mount the antenna at the minimum elevation determined by STL path calculations. The structure must also be capable of supporting the weight of the antenna and transmission line, as well as withstanding any forces that affect the antenna tower, such as wind and the weight of snow and ice. Be certain to take into account the mounting hardware used with the antenna, which typically accommodates pipe up to a diameter of 3.5 in. Provisions must be made to secure the transmission line. The structure may also be subject to local building and zoning codes. Strict adherence to these codes is mandatory. Adequate physical clearance must be considered to allow for proper horizontal and vertical alignment during installation.

Hardware Considerations

Depending on the complexity of the STL system, additional hardware may be required. Where more than one transmitter or receiver is used with a single antenna, a combiner, splitter, or duplexer will be needed. These items contribute additional loss and must be accounted for in path calculations.

Certain installations may require the installation of an external power amplifier in cases where the output of a standard STL transmitter is insufficient for a particular path. In path calculations, the power output of the external power amplifier (converted to dBm) is substituted for the output of the STL transmitter. In general practice, most engineers choose to use an external power amplifier only as a last resort. Higher gain antennas usually are a more attractive alternative.

11-82 Broadcast Transmission Systems

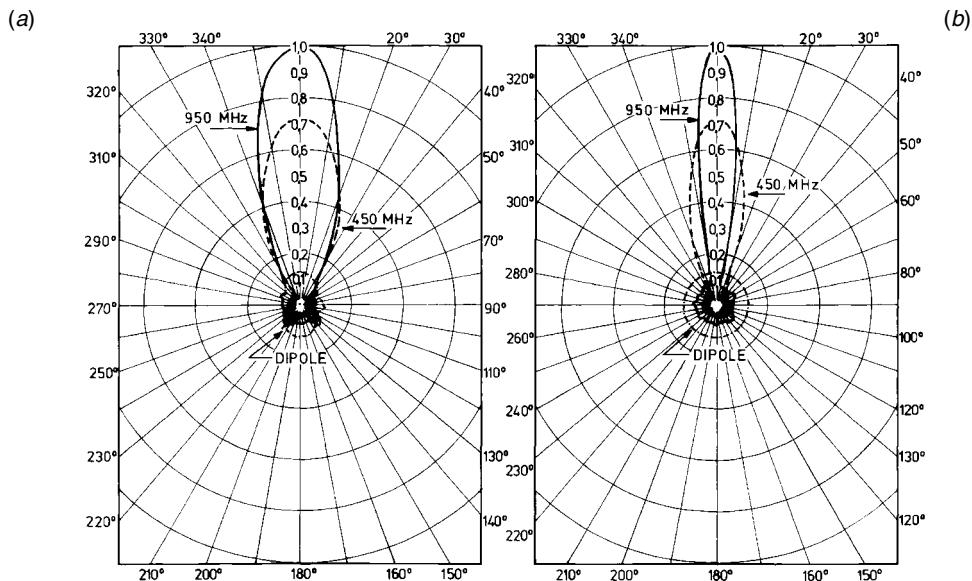


Figure 11.4.15 Radiating plots for a parabolic section antenna at 450 MHz and 950 MHz: (a) horizontal response, (b) vertical response. (After [2].)

Table 11.4.6 Typical Isotropic Gain for Full Parabolic Antennas at 950 MHz

Antenna Diameter	Isotropic Gain
4 feet	18.9 dBi
5 feet	21.0 dBi
6 feet	22.0 dBi
8 feet	25.0 dBi
10 feet	27.0 dBi

11.4.3 System Planning

The ultimate goal in selecting STL equipment is to choose a combination of STL transmitter, transmission line, antennas, and STL receiver that will give adequate quieting (S/N) when transmitting over the path between the studio and the station transmitter. Allowance in the form of *fade margin* must also be made for the uncertainties that the path imposes on the received signal.

The STL transmitter supplies the operating power that drives the system. The transmit and receive antennas add gain to the link. All remaining components add loss. By far the greatest loss results from the path itself. Passive components of the system, including the transmission line, connectors, and pigtail cables, also add loss. All gains and losses must be taken into account to

ensure that the RF level at the receiver input is adequate. Losses should be minimized wherever possible.

11.4.3a Frequency Selection

Frequency selection would be a relatively simple procedure were it not for other broadcasters, who also may be using STL frequencies, and perhaps sharing the same transmitter site. In practice, the choice of operating frequency is governed by the availability of unused STL channels in the area, and is further dictated by the need to avoid interference with other STL users. The local frequency coordinator (under the auspices of the Society of Broadcast Engineers in most areas of the U.S.) usually can provide information on available frequencies and potential frequency conflicts.

STL engineering can be divided into three broad categories:

- Frequency selection
- Path layout
- Path gain/loss calculations

Path layout concerns the factors affecting the physical location of the STL transmitter and receiver, as well as the intended path. The goal in path layout is to ensure a correctly-oriented and unobstructed route for unhampered propagation of the radio wave. Path gain/loss calculations involve analyzing RF power levels from the transmitter output to the receiver input so that an adequate receive level is provided.

Spectrum Considerations

In view of the serious spectrum congestion problems that exist today in many areas of the U.S., an STL system should be designed to be as spectrum-efficient as possible and—equally important—to be as immune to undesired transmissions as possible. Even if the system will be operated in an area that currently does not have a spectrum congestion problem, there is no guarantee that such a situation will not surface in the near future. In any event, a well-engineered system is also a spectrum-efficient system.

The first rule of spectrum-efficiency is to use only the *effective radiated power* (ERP) necessary to do the job. There is no justification for putting 15 W into the air when 5 W will provide the required receiver quieting and fade margin.

A simple and sometimes effective spectrum coordination tool is cross-polarization. Two stations on nearby frequencies may achieve as much as 25 dB RF isolation through the use of different polarizations of transmit antennas, matched by like polarization at their respective receive antennas. Cross-polarization results in varying degrees of success, depending upon the frequency of operation and the surrounding terrain.

11.4.3b Path Layout

Careful path engineering should be performed prior to any licensing work to determine if the proposed locations of the STL transmitter and receiver will be able to achieve the desired results. There is much more to path engineering than simply pointing the transmitting and receiving antennas at each other and turning the equipment on. Antennas cannot always be located in the

11-84 Broadcast Transmission Systems

best possible geographic points because of space availability problems, excessive construction or site rental costs, or local/federal licensing difficulties. In such cases, the required path is not the ideal path and the link will have to be engineered around these fixed points.

Planning for any STL system should begin with an accurate, detailed U.S. Geological Survey (USGS) map covering the proposed path. Note should be made of any natural obstructions (such as mountains, hills, or vegetation) or man-made obstructions (such as buildings, water tanks, or transmitting towers) in the path.

Line of Sight

Because microwave frequencies are used for STL systems, the signal path is theoretically limited to the line-of-sight between the studio and transmitter locations. In reality, the *radio horizon* is frequently situated beyond the visual horizon. This is the result of the gradual decrease in the refractive index of the atmosphere with increasing altitude above the earth. This effect bends radio waves downward. The degree of bending is characterized by the *K factor*, which is the ratio of the effective earth radius to the true earth radius. A typical value for *K* is 4/3, or 1.33, valid over 90 percent of the time in most parts of the world. For long paths, consult a map showing the minimum *K* factor occurring in the specific area so that proper antenna heights can be planned. Figure 11.4.16 plots an example path on 4/3 earth graph paper.

STL Site Selection

Usually, there is little choice for the STL transmitter and receiver locations, because most STL systems will be required to provide a path between already determined studio and transmitter points. If, however, a new studio location is being planned, the possible STL path is an important consideration. Figure 11.4.17a illustrates a poor path. Generally speaking, a good path is literally line-of-sight, with no obstructions to block the signal, and no other terrestrial or atmospheric conditions that would compromise the path (Figure 11.4.17b).

Terrain Considerations

One of the major tasks required to engineer an STL system is the path analysis between the STL transmitter at the studio and the STL receiver location. To determine what constitutes a clear path, the concept of Fresnel zones for optical theory is applied to radio waves. Most of the electromagnetic energy at a receiving point is concentrated in an elliptical volume that is a function of the distance between the transmit and receive points and the wavelength. The energy outside this volume either cancels or reinforces the energy within the volume, depending on whether the distance that the energy travels to the receive point is longer by an even or odd number of 1/4-wavelengths. Even distances result in radio wave cancellations; odd distances result in radio wave reinforcement. (See Figure 11.4.18.) The radius of the first Fresnel zone, which defines the boundary of the elliptical volume, is given by the following:

$$F_1 = 72.1 \sqrt{\frac{d_1 d_2}{fD}} \quad (11.4.2)$$

Where:

F_1 = first Fresnel zone radius in feet

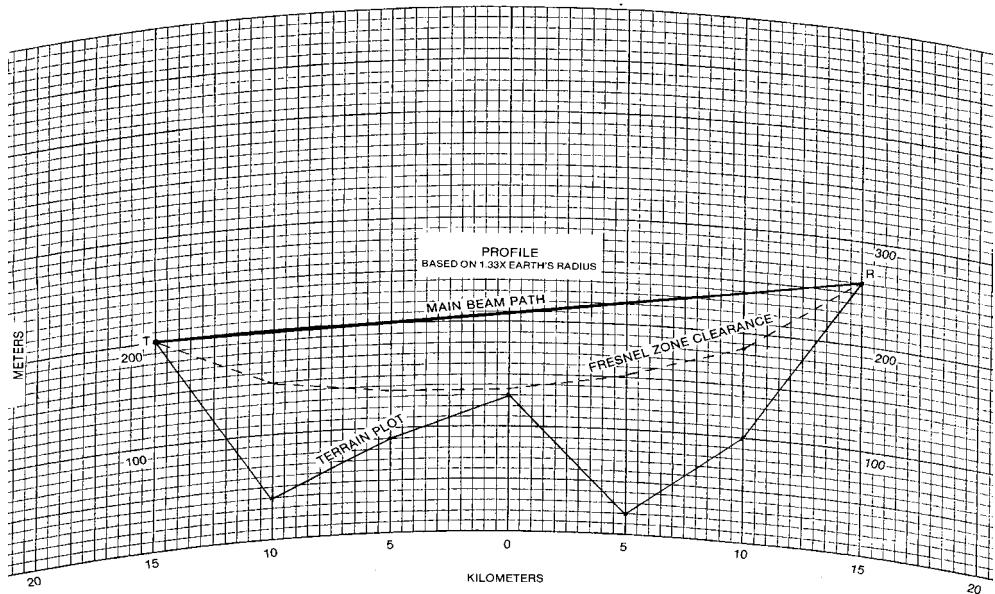


Figure 11.4.16 An aural STL path profile drawn on true-earth radius graph paper. (After [2].)

d_1 = distance from the transmitting antenna to the obstruction in miles

D = total path length in miles

$d_2 = D - d_1$ in miles

f = frequency in GHz

H = distance from the top of the obstruction to the radio path

For reliable operation, obstructions should not project into the area thus defined. Empirical studies, however, have shown that performance is substantially the same as long as H is greater than $0.6 F_1$.

The first step in evaluating the path is to make a subjective check of the planned route. First, determine that a reasonable possibility of success exists. Next, draw a best-case route for the path. If obstructions are found, consider accomplishing the path by using a repeater system, or by shifting the location of the STL transmit antenna. Although a detailed path analysis may not be required in cases where line-of-sight is clearly present and the distance is less than about 10 miles, it is still good engineering practice to review some of the basic elements of the evaluation.

Obtain an accurate map showing the topography between the STL transmitter and receiver locations. After determining the transmitter and receiver sites on the map, connect them with a straight line showing the proposed path. After the path has been drawn, a protractor can be used to determine the direction (azimuth) of the path in degrees from true North. This data will later assist in antenna setup, and is necessary for filling out the appropriate FCC application.

Using the scale of miles on the map, measure the distance between the transmitter and receiver sites. Determine the altitude of the proposed transmit antenna location from the contour

11-86 Broadcast Transmission Systems

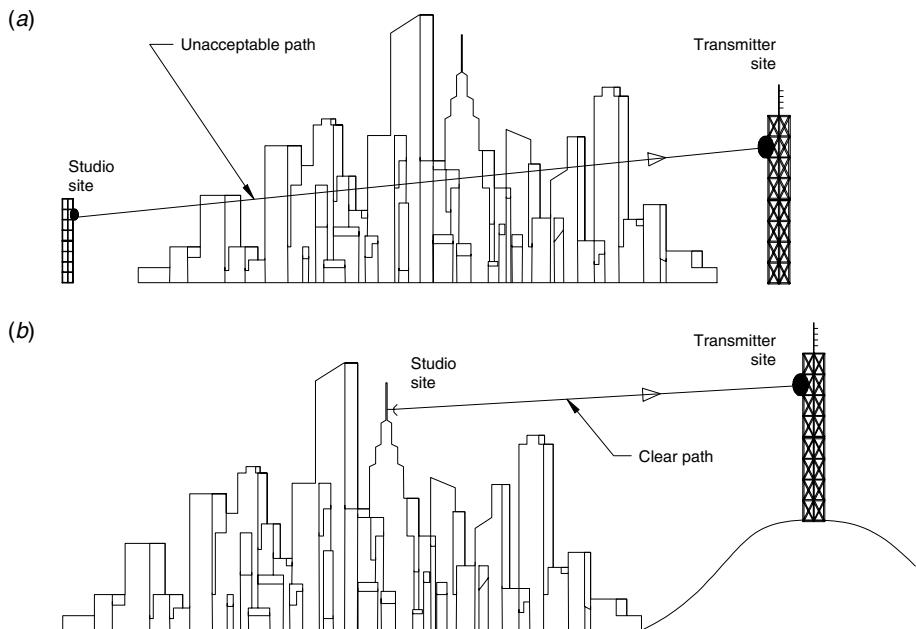


Figure 11.4.17 Path considerations in planning an STL: (a) unacceptable path because of obstructions, (b) usable path. (After [2].)

lines on the map, and add to that the height of the building or other structure on which the antenna will be mounted. Make a similar determination for the receive antenna location. Adjust the heights to ensure that all obstructions are cleared. Depending on the path length and the height of the antennas, it may be necessary to take the curvature of the earth into account, and use earth radius graph paper to plot a cross-section of the path.

Study the map to see what terrain features are between the path points. Prior to making any other evaluations, conduct a visual field survey. Check for any structures or features not listed on the map. Anticipate any possible new construction or tree growth that may cause problems in the future.

The terrain from the transmitting antenna to the receiving antenna must be examined not only for obstructions, but for reflection possibilities as well. A large body of water will usually cause problems for an STL system. If the water is an even number of Fresnel zones from the direct path, signal attenuation will likely occur at the receiver. Temperature changes and tidal conditions will also have an effect. Likewise, thick vegetation or forested areas can be reflective to RF signals when wet, creating a similar (but not so troublesome) problem. Generally, the solution to reflection conditions is to change either the transmitting or receiving antenna height. In extreme cases, a diversity reception system may also be used.

Path Reliability

The long-term reliability of an STL path is determined in large part by the initial engineering work done before the system is installed. Figure 11.4.19 charts gains and losses through a typical

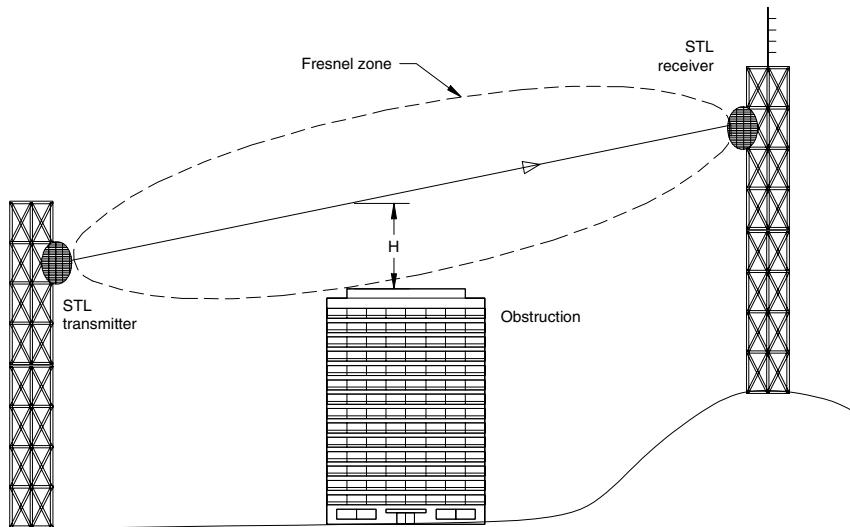


Figure 11.4.18 Fresnel zone clearance for an STL path. (After [2].)

system. The most important factors are free space loss and allowance for fade margin. Thereafter, effects such as diffraction, reflection, refraction, absorption, scattering, and terrain loss must be considered.

A gain and loss balance sheet should be computed to determine the fade margin of the planned STL system. An adequate fade margin is vital to reliable performance because a link that is operating on the edge of the minimum acceptable receiver quieting will encounter problems later down the road. Normal component aging in the receiver or transmitter can cause a loss in received signal level and, thus, degrade system performance. Atmospheric conditions, such as severe weather in the area or ice on the transmitting or receiving antennas, can also cause sharp fading and even a complete loss of signal if an adequate fade margin above minimum receiver quieting is not provided. The STL fade margin can be computed using the following equations:

$$G_s = G_t + G_{ta} + G_{ra} \quad (11.4.3)$$

Where:

G_s = total system gain in decibels

G_t = transmitter power output in dBm

G_{ta} = transmit antenna gain in dBi

G_{ra} = receive antenna gain in dBi

The values for G_{ta} and G_{ra} are gathered from the antenna manufacturer's literature.

The value for G_t is given by the following:

$$G_t = 30 + 10 \log P_o \quad (11.4.4)$$

11-88 Broadcast Transmission Systems

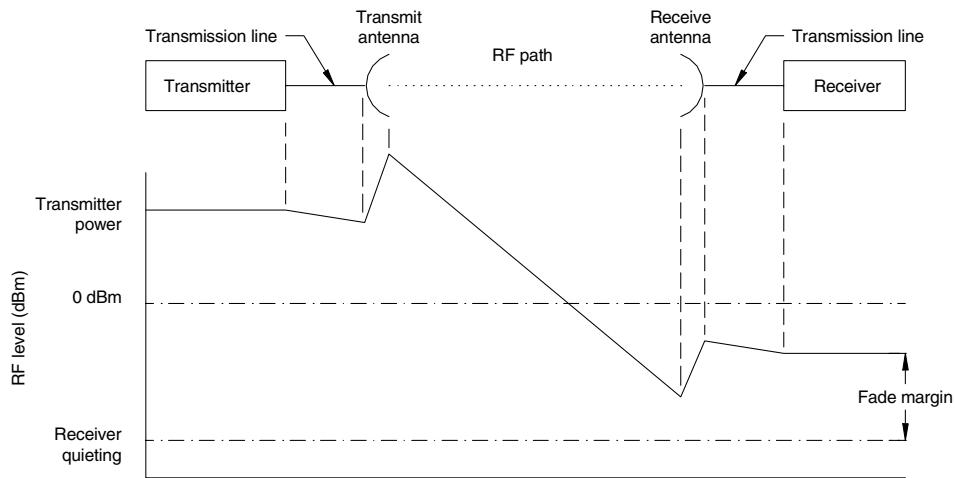


Figure 11.4.19 Path analysis plotting gains vs. losses for an STL. (After [2].)

Where:

G_t = transmitter power output in dBm

P_o = transmitter power output in watts

The total system losses are then computed:

$$L_s = L_p + L_l + L_c + L_m \quad (11.4.5)$$

Where:

L_s = total system losses in decibels

L_p = path loss in dB

L_l = transmission line loss in dB

L_c = connector losses in dB

L_m = miscellaneous losses in dB

The values for L_t and L_c can be determined from manufacturer's literature. Figure 11.4.20 shows typical loss values for 1/2-in foam-filled transmission line. A reasonable value for connector loss with components normally used in 1/2-in coax installations is 0.5 dB. The value for L_p can be determined by using the following formula:

$$L_p = 36.6 + 20 \log F + 20 \log D \quad (11.4.6)$$

Where:

L_p = free space attenuation loss between two isotropic radiators (in dB)

F = frequency of operation in megahertz

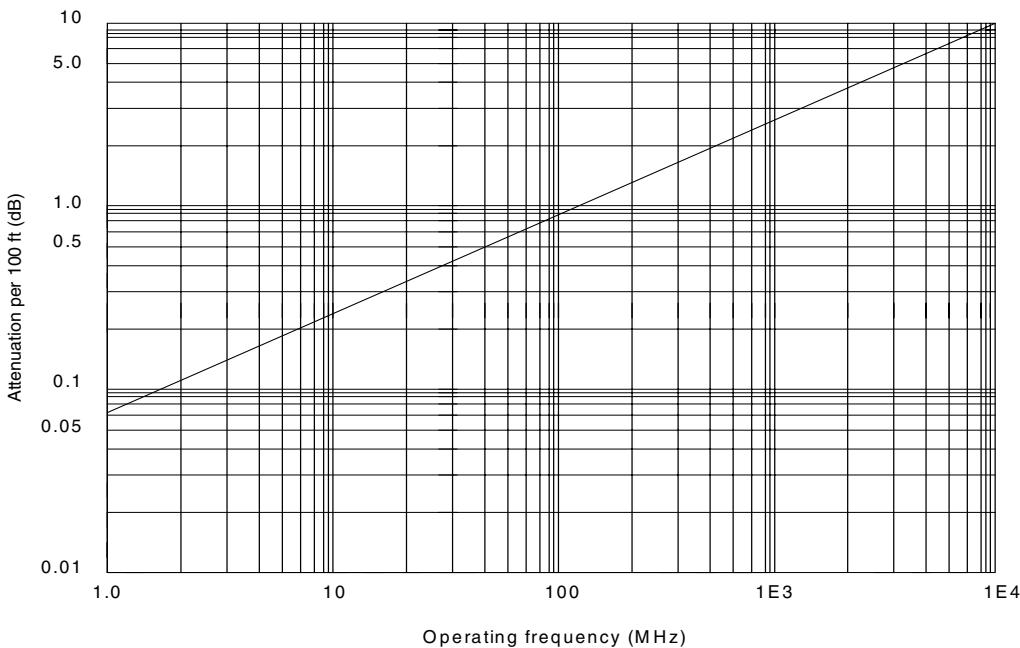


Figure 11.4.20 Loss vs. frequency for 1/2-inch foam dielectric transmission line. (After [2].)

D = distance between the antennas in statute miles

Free space loss can also be found using a table of approximate values, as given in Table 11.4.7.

With the foregoing information, the fade margin can be calculated:

$$M_f = G_s - L_s - R_m \quad (11.4.7)$$

Where:

M_f = fade margin (in dB)

G_s = total system gain (dB)

L_s = total system losses (dB)

R_m = minimum signal strength required for the target S/N in dBm (a negative number)

G_s and L_s are determined by the equations given previously. R_m (receiver sensitivity) is determined from the receiver manufacturer's specifications. If the manufacturer gives a receiver sensitivity figure in microvolts, the following formula can be used to convert to dBm:

$$R_m = 20 \log \left(\frac{V_r \times 10^{-6}}{0.7746} \right) \quad (11.4.8)$$

11-90 Broadcast Transmission Systems

Table 11.4.7 Path Attenuation at 950 MHz

Distance in Miles	Loss in dB	Distance in Miles	Loss in dB
1	-96.2	16	-120.2
2	-102.2	17	-120.7
3	-105.7	18	-121.3
4	-108.2	19	-121.7
5	-110.1	20	-122.2
6	-111.7	21	-122.6
7	-113.1	22	-123.0
8	-114.2	23	-123.4
9	-115.3	24	-123.8
10	-116.2	25	-124.1
11	-116.9	26	-124.5
12	-117.7	27	-124.8
13	-118.4	28	-125.1
14	-119.1	29	-125.4
15	-119.7	30	-125.7

Where:

R_m = minimum required signal strength (in dBm)

V_r = receiver sensitivity (in μ V)

In order to predict accurately the performance of the STL radio link, the value of R_m must be determined carefully. For maximum system performance and reliability, the fade margin determination should be made based upon the signal level required to provide the minimum acceptable receiver S/N performance. Longer paths require greater margins for fade.

The primary cause of signal fade in an STL system below 1.0 GHz is changes in the refractive indexes of the atmosphere along the signal path. These fluctuations affect the path of reflected or refracted signals differently from the direct, line-of-sight signal. When the interfering signals combine with the direct signal, the level at the receiver increases or decreases depending upon the degree of reinforcement or cancellation. Because atmospheric conditions are seldom stable, some fade margin is always necessary.

Another cause of signal fade is *earth bulge* (or *inverse beam*) fading, where the overall refractive index in the vicinity of the signal path decreases, thus hindering the full signal from reaching the receive antenna. Again, allowance for signal fade will minimize degradation. Precipitation is another potential cause of signal fading, although it is not generally considered significant at frequencies below 1.0 GHz.

Fade margin also can be determined approximately from Table 11.4.8. The relationship between system reliability and fade margin is detailed in Table 11.4.9.

Table 11.4.8 Recommended Fade Margin as a Function of Path Length

Path Length	Fade Margin
5 mi	5 dB
10 mi	7 dB
15 mi	15 dB
20 mi	22 dB
25 mi	27 dB
30 mi	30 dB

11.4.3c Dealing with Problem Paths

The STL systems described thus far have been simple installations with a transmitter at the studio and a receiver at the broadcast transmitter site. There are cases, however, where a simple, single-hop STL is not be practical. Reasons include the following:

- The distance is too great to be reasonably covered by a single-hop system.
- A direct path cannot be used because of obstructions of some type.
- An unusually large fade margin is required for the application.

Figure 11.4.21 shows three instances where a two-hop system is required. A two-hop repeater uses two transmitters and two receivers, with one pair located at the intermediate site. As shown in Figure 11.4.21c, interference avoidance is another practical application of a multi-hop system.

With a conventional active repeater, the received signal is demodulated down to the baseband at the intermediate site before being retransmitted. This approach is shown in Figure 11.4.22. Unfortunately, this method results in inevitable degradation of the signal. A better approach, an IF repeater, is shown in Figure 11.4.23. Instead of being demodulated to the baseband level, the received signal is brought down to the IF level before being injected into the repeater transmitter.

Hot-Standby Redundant Systems

The importance of continuous, uninterrupted service from an STL system cannot be overstated. Modern STL equipment is stable and reliable. Failures rarely occur. However, a station may elect to install a hot-standby redundant system to ensure zero downtime from the STL. Figure 11.4.24 shows a block diagram of a common hot-standby system. The installation includes a second transmitter and a second receiver. An automatic changeover unit is included for switching purposes. The changeover system monitors each transmitter and receiver pair to sense critical parameters. If a failure occurs, the changeover controller switches from the faulty transmitter to the standby transmitter, or from the faulty receiver to the standby receiver. The changeover units work independently.

This system can be simplified to include only hot-standby provisions at the transmitter site, or at the receiver site. If this approach is taken, a case can be made for either backup system. One argument states that the transmitter is more likely to fail because it contains high-power stages, which typically are more prone to problems than low-power circuits. On the other hand, the

11-92 Broadcast Transmission Systems

Table 11.4.9 Relationship Between Fade Margin, Reliability, and Outage Time for Rayleigh Distributed Paths

Fade Margin (dB)	Path Reliability/ Availability	Outage Hours per Year	Outage Minutes per Month	Outage Seconds per Day
10	90.4837	834.20	4170.98	8222.05
20	99.0050	87.22	436.12	859.69
21	99.2088	69-35	346.77	683.58
22	99.3710	55.14	275.68	543.43
23	99.5001	43.82	219.12	431.94
24	99.6027	34.83	174.14	343.28
25	99.6843	27.68	138.38	272.79
26	99.7491	21.99	109.96	216.75
27	99.9007	17.47	87.37	172.22
28	99.8416	13.88	69.41	136.83
29	99.8742	11.03	55.14	108.70
30	99.9000	8.76	43.81	86.36
31	99.9206	6.96	34.80	68.60
32	99.9369	5.53	27.65	54.50
33	99.9499	4.39	21.96	43.29
34	99.9602	3.49	17.45	34.39
35	99.9684	2.77	13.86	27.32
36	99.9749	2.20	11.01	21.70
37	99.9800	1.75	8.74	17.24
38	99.9842	1.39	6.95	13.69
39	99.9874	1.10	5.52	10.88
40	99.9900	0.88	4.38	8.64
41	99.9921	0.70	3.48	6.86
42	99.9937	0.55	2.77	5.45
43	99.9950	0.44	2.20	4.33
44	99.9960	0.35	1.74	3.44
45	99.9968	0.28	1.39	2.73
50	99.9990	0.09	0.44	0.86
55	99.9997	0.03	0.14	0.27
60	99.9999	0.01	0.04	0.09

transmitter is almost always more accessible for repair—being at the studio site—than the receiver.

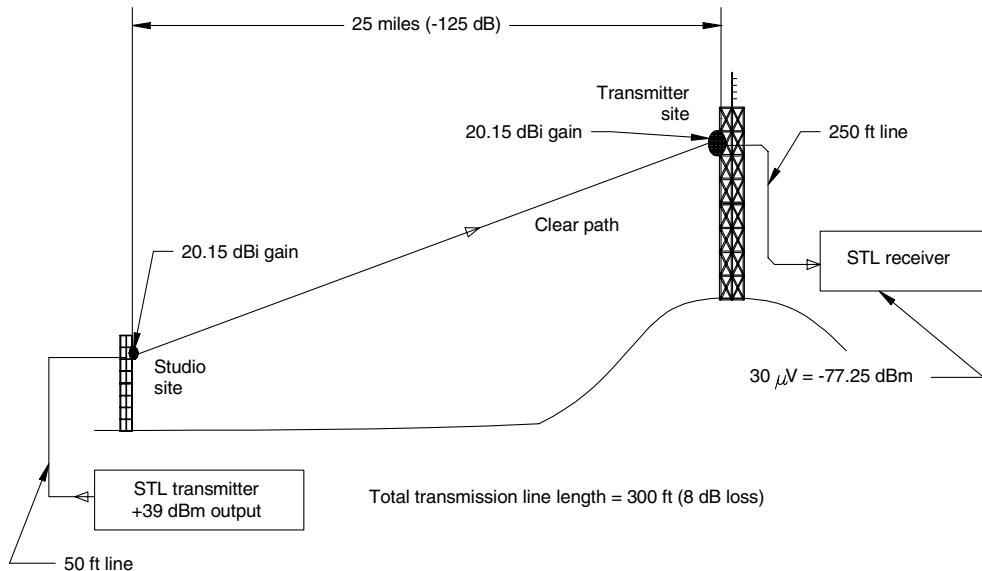


Figure 11.4.21 Applications requiring the use of a multi-hop STL system: (a) excessively long path, (b) path obstructions, (c) interference concerns. (After [2].)

11.4.3d Pre-Installation Checkout

The best planned STL system, well engineered and comprised of the finest equipment, will not perform properly if incorrectly installed. Quality control steps taken before installation actually begins is an important part of the installation process. Delivery schedules for STL hardware vary widely. After all the equipment is ordered, components will arrive, typically, over a period of some weeks. Good engineering practice dictates that each component is checked or inspected as it arrives. This process is preferable to waiting for all of the components to arrive on site because any problems discovered at that point will delay the project.

The STL transmitter can be operated into a dummy load to confirm proper operation. With an ideal load, the transmitter front panel readings should correspond closely with the final test sheet supplied by the manufacturer. Modulation can be applied to the transmitter to ensure proper operation.

The STL receiver is more difficult to test effectively with the instruments typically available at a radio station. The engineer can, however, bench test the receiver by operating the transmitter into a dummy load and attaching a short wire to the receiver antenna input. Be sure to use the proper connector on the receiver to avoid possible damage to the center pin. Maintain sufficient separation of the transmitter and receiver to prevent overloading the receiver front-end. Arranged in this manner, the STL system can be checked from end-to-end with a signal generator (or program material) on the bench.

For a digital STL, the transmitter, receiver, and modem/multiplexer can be tested in a similar way. Bench tests before installation usually resolve hidden or unexpected problems associated with equipment operation and/or interfacing. Functionality of the transmitter and receiver should

11-94 Broadcast Transmission Systems

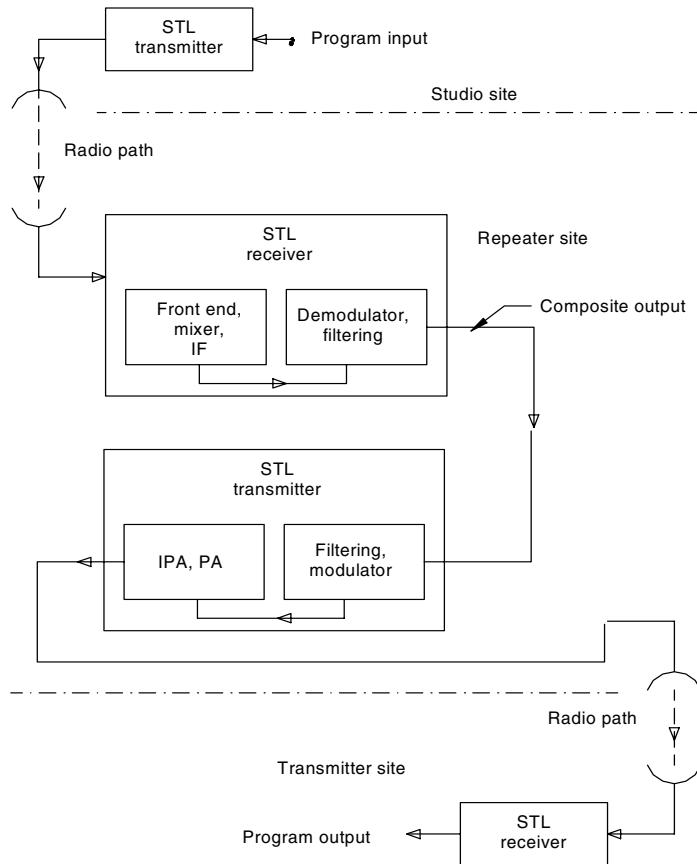


Figure 11.4.22 Repeater link using conventional demodulation and remodulation of the composite signal. (After [2].)

be checked before adding the encoder to the system. In this manner, any problems detected can be readily pinpointed.

An STL using one or more subcarriers for program relay, closed-circuit communications, or remote control functions can be bench tested in a similar manner. Having all hardware on one bench and easily accessible makes adjustment of levels and other parameters much easier than when the units are separated by many miles.

Nearly all STL systems checked in this manner pass specification testing with no problem. Still, there may be instances where a unit was damaged during shipment. It is far easier to solve such problems before the hardware is installed.

Antennas should be given a close visual inspection. Many antenna models used for STL work are shipped in a partially disassembled state. Final assembly of the antennas should be completed before the installation process is begun. Pay particular attention to mounting brackets; make sure

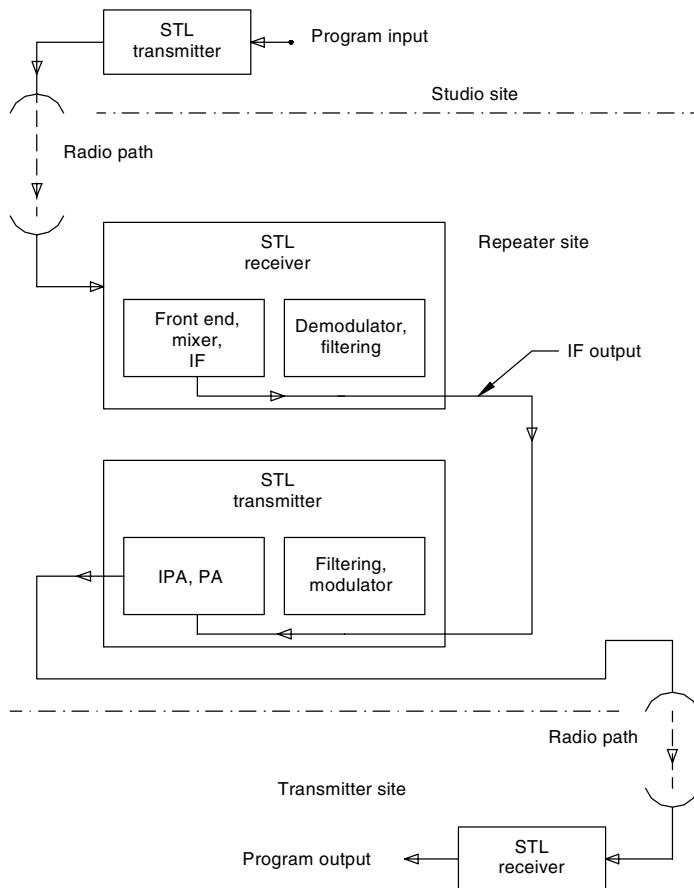


Figure 11.4.23 Repeater link using IF transfer of the composite signal. (After [2].)

the hardware will mate properly with the tower or structure on which the antennas will be mounted. Check carefully for missing parts.

The transmission line and connectors require no pre-installation quality control, however, it is suggested that as many connectors as possible be installed in the shop, rather than at the site. Type N connectors require patience and skill to install. If possible, have the cable supplier attach the proper connectors to both ends of the cable reel. In this way, only two connectors will need to be installed in the field. Always place the manufacturer-installed connector ends at the antennas. This provides the best assurance of trouble free operation in an exposed environment. Consider ordering a couple extra connectors just in case a part is lost or damaged during construction. The engineer may also want to use one connector as a test case to become more familiar with the required installation technique.

11-96 Broadcast Transmission Systems

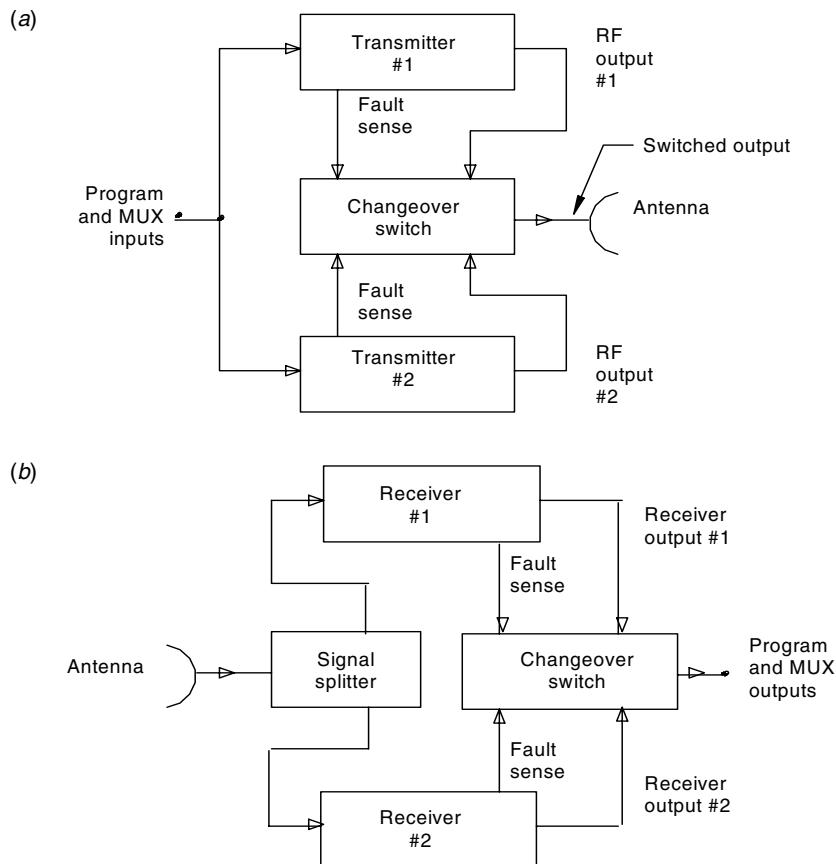


Figure 11.4.24 Equipment configuration for a hot-standby system: (a) studio site, (b) broadcast transmitter site. (After [2].)

The test equipment required for pre-installation checkout is usually readily available in the radio station shop. Basic items include the following:

- A high-quality $50\ \Omega$ dummy load capable of dissipating approximately 25 W.
- An in-line RF power output meter capable of reading forward and reverse power at 1.0 GHz.
- Audio frequency signal generator.
- Audio frequency distortion analyzer.
- Frequency counter accurate to 1.0 GHz.

If possible, a spectrum analyzer also should be available. Although most radio stations do not own a spectrum analyzer, engineers in a given market sometimes rent one to be shared among the stations for a period of several weeks.

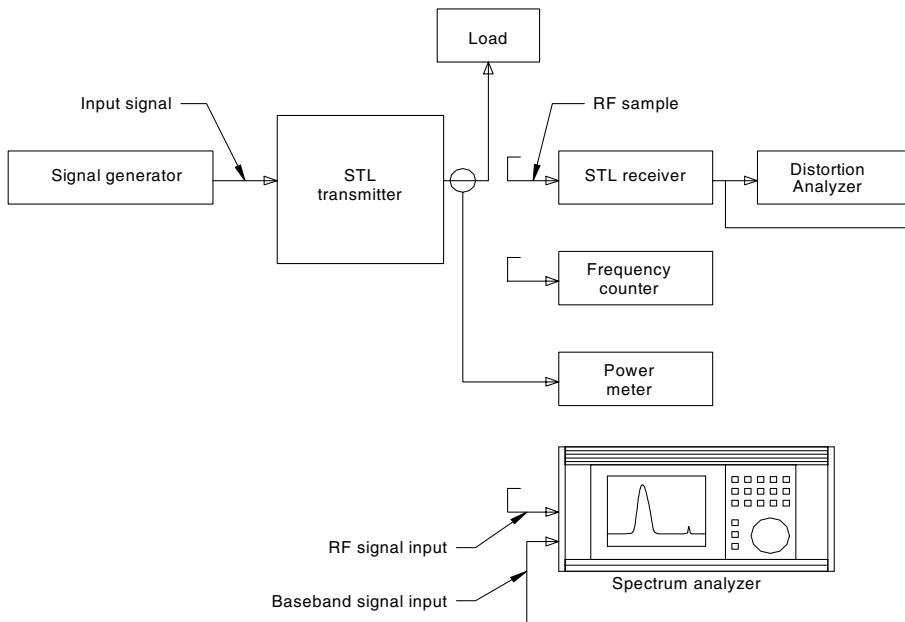


Figure 11.4.25 Equipment setup for bench testing an STL system prior to installation. (After [2].)

Figure 11.4.25 shows a typical bench test setup for an STL system.

Planning Cable Runs

Routing of the transmission line from the transmitter to the antenna is an often overlooked element of STL installation. Confirm that there is a practical route for the coax from the transmitter to the rooftop or tower location. Double-check the length of the transmission line path at both the studio site and the station transmitter site. If the actual run is longer than predicted during fade margin calculations, run the numbers again to be certain that a problem will not develop during system commissioning. Confirm also that the pigtail cables to be used are sufficiently long and have the proper connector terminations.

11.4.3e Installation

The locations commonly used for broadcast transmitter sites are seldom ideal from an environmental standpoint. The site may be difficult to reach during certain parts of the year, very hot during the summer, and very cold during the winter. For these reasons, rugged equipment should be chosen and properly installed. Temperature extremes can cause problems for frequency-determining elements, as well as accessories such as cavity filters, preselectors, and preamplifiers. Environmental control at the broadcast transmitter site, therefore, is highly recommended.

The STL transmitter and receiver should be mounted in an equipment rack in a protected location adjacent to the stereo generator at the studio site, and adjacent to the exciter at the

11-98 Broadcast Transmission Systems

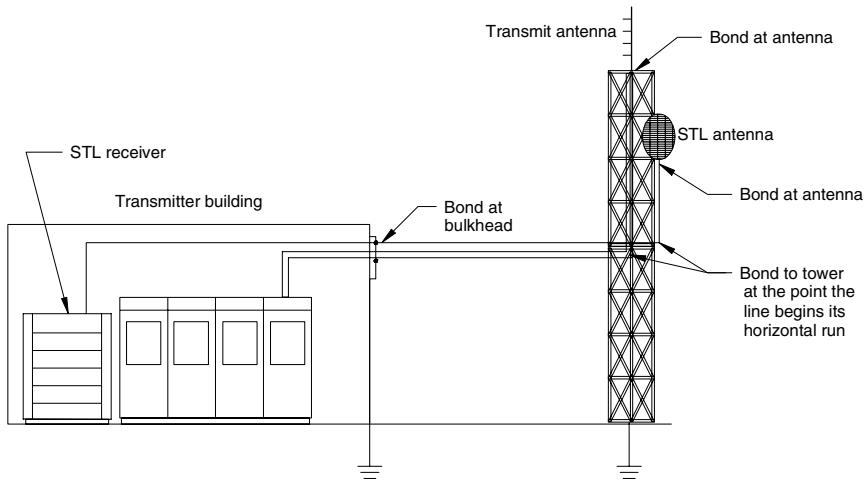


Figure 11.4.26 Grounding practice for an STL transmission line on a tower. (After [2].)

broadcast transmitter site. Keep all cable runs as short and direct as possible. Excessive cable lengths between the stereo generator and the STL transmitter, or between the STL receiver and the exciter, can degrade stereo separation and frequency response. Follow good grounding practices at all times.

The antenna presents probably the greatest installation challenge. Because of its directional nature, the antenna must be properly oriented. Compass bearings are desirable along with whatever visual sightings are practical. The received signal strength at the broadcast transmitter site can be used as an indication of proper orientation of the STL receive antenna. Ensure also that the antenna is set to the proper angle relative to the horizon. Because the STL antennas may be located high on a tower or building, a professional tower crew may be required to mount the devices. Make certain that all cables on the tower are securely fastened to prevent stress and wear. Seal the external connections with sealant made for that purpose and completely wrap the connection joints with tape.

While it is desirable to mount the receive antenna high on a tower to ensure a good path, it is also good engineering practice to keep the antenna as far as possible from radiating elements on the structure, such as the station's main broadcast antenna. Other potential sources of RF problems include 2-way radio, cellular radio, and TSL systems. If the STL receive antenna is located in a strong RF field, the front-end may be overloaded or *desensitized*. Placing the STL antenna close to other antennas on the structure also can detune the antennas, degrading the performance of both systems.

One of the most common problems encountered during installation is damage to the transmission line. Lines must not be twisted or bent beyond their minimum bending radius. To retain its characteristic impedance, the line must be treated with care. A line damaged by carelessness can cause poor performance of the system.

The transmission line must be properly grounded. As illustrated in Figure 11.4.26, grounding typically is performed at the point where the line leaves the tower and at the point where it enters

the building. This procedure, obviously, applies only to grounded tower systems. Grounding kits are available for this purpose.

In situations where the STL antenna is to be mounted on an AM antenna tower, which typically is insulated from ground, an *isocoupler* will need to be installed where the line begins its run up the tower. The isocoupler will pass the STL signal efficiently, while providing a high impedance to ground for the tower at the AM frequency. When this is done, the base impedance of the tower should be expected to change slightly.

11.4.3f Checking System Operation

After all of the components of the STL system have been installed, checkout can begin. Start by powering-up the STL transmitter. Confirm that all meter readings are within normal parameters. If any problems are detected, shut the system down immediately and troubleshoot the failure.

After proper operation has been confirmed, adjust the modulation levels into the STL transmitter according to the instructions given with the unit. If a digital modem/multiplexer is used, make the necessary adjustments on that unit.

Let the transmitter run for a period of time—several hours or so—and check the operating parameters on a regular basis to be certain that no problems develop. Pay particular attention to VSWR and internal high-temperature alarms. If abnormal operation is suspected, consult the manufacturer's equipment manual.

After checkout of the STL transmitter is complete, power up the STL receiver and confirm correct parameters on the front-panel meter. Check and adjust program levels as needed. If a digital modem/multiplexer is used in the installation, check the status monitors on the unit. If displayed, observe the bit error rate monitor to be sure the link is operating properly.

Using an accurate modulation monitor at the broadcast transmitter site, adjust the modulation level of the transmitter. Check operation of the STL system carefully for an hour or so before committing the system to the air.

After the link is operating properly, run a complete proof-of-performance on the broadcast chain. Document the measured parameters so they can be used as a benchmark for later reference.

11.4.3g Troubleshooting

The most common problem associated with STL commissioning is high VSWR. If the indicated VSWR is outside normal limits, the transmission line, a connector, or an antenna usually is at fault. A dummy load may be used to determine where the fault lies. Begin at the first pigtail and substitute the dummy load for the transmission line. At each point the load is substituted for the line and antenna, check the displayed VSWR on the transmitter. It is necessary, of course, to power-down the transmitter while the dummy load is moved from one point to the next.

11.4.3h Operating the System

Regular performance tests should be made on the STL system. Routine inspections allow the engineer to spot potential problems that could cause a system failure if left unattended.

The STL transmitter and receiver should provide trouble-free performance for many years. If problems develop, the likely trouble areas will be the antennas and transmission lines—the com-

11-100 Broadcast Transmission Systems

ponents exposed to the elements. The best check of system performance at the studio is the STL transmitter multimeter. A periodic check of the meter will help to spot oncoming problems. Gradually changing readings can provide clues that will help prevent a failure. Many users will wish to monitor forward power continuously. Reverse power also should be checked regularly to monitor the condition of the line and antennas, as discussed previously. Compare the test data taken after system commissioning with the operating parameters observed during routine checks. They should closely agree.

Operation of the receiver is best checked with its built-in multimeter. The RF level indication should be carefully noted during initial installation, and subsequent observations should be compared with this reference. Remember there will be some change in signal strength readings because of weather conditions and temperature variations. Consider the fade margin conditions used in the path analysis calculations when making judgments about the observed readings. Be aware of unusual variations resulting from temperature inversions, if they are common in the area. If trouble is experienced with the receiver, the possibility of interference from another STL or another service should not be overlooked. A spectrum analyzer is invaluable for such troubleshooting.

For digital STL systems, error rate monitors are provided to assess overall performance. As long as the status indicators on the front panel indicate proper operation, the system usually can be considered to be transparent to the program material.

The most definitive overall check of the system will be the audio proof of performance. First attention should be given to the noise measurement. If this is not satisfactory, it will be impossible to achieve meaningful distortion measurements, because noise will be indicated as distortion by the analyzer. For dual monaural systems, the engineer will need to carefully check left and right channel balance.

Any broadcast system is only as strong as its weakest component. Before placing blame for poor performance on the STL, start at the beginning of the broadcast chain and follow the signal step-by-step to the point where it deteriorates.

11.4.4 References

1. Rollins, William W., and Robert L. Band: "T1 Digital STL: Discrete vs. Composite Transmission," *NAB 1996 Broadcast Engineering Conference Proceedings*, National Association of Broadcasters, Washington, D.C., pp. 356–359, 1996
2. Whitaker, Jerry C., (ed.): *A Primer: Digital Aural Studio to Transmitter Links*, TFT, Santa Clara, CA, 1994.

11.4.5 Bibliography

Hauptstuek, Jim: "Interconnecting the Digital Chain," *NAB 1996 Broadcast Engineering Conference Proceedings*, National Association of Broadcasters, Washington, D.C., pp. 360–358, 1996.

- McClanahan, M. E.: "Aural Broadcast Auxiliary Links," in *NAB Engineering Handbook*, 8th ed., E. B. Crutchfield (ed.), National Association of Broadcasters, Washington, D.C, pp. 671–678, 1992.
- Parker, Darryl: "TFT DMM92 Meets STL Requirements," *Radio World*, Falls Church, VA, October 21, 1992.
- Salek, Stanley: "Analysis of FM Booster System Configurations," *Proceedings of the 1992 NAB Broadcast Engineering Conference*, National Association of Broadcasters, Washington, DC, April 1992.
- Whitaker, Jerry C., and Skip. Pizzi: "Radio Electronic News Gathering and Field Production," in *NAB Engineering Handbook*, 8th ed., E. B. Crutchfield (ed.), National Association of Broadcasters, Washington, D.C, pp. 1051–1072, 1992.

Chapter
11.5
Transmitting Antennas

Jerry C. Whitaker, Editor-in-Chief

11.5.1 Introduction

Transmission is accomplished by the emission of coherent electromagnetic waves in free space from one or more radiating elements that are excited by RF currents. Although, by definition, the radiated energy is composed of mutually dependent magnetic and electric vector fields, it is conventional practice to measure and specify radiation characteristics in terms of the electric field only.

The purpose of an antenna is to radiate efficiently the power supplied to it by the transmitter. A simple antenna, consisting of a single vertical element over a ground plane can do this job quite well at low-to-medium frequencies. Antenna systems may also be required to concentrate the radiated power in a given direction and minimize radiation in the direction of other stations sharing the same or adjacent frequencies. To achieve such directionality may require a complicated antenna system that incorporates a number of individual elements or towers and matching networks.

As the operating frequency increases into VHF and above, the short wavelengths permit the design of specialized antennas that offer high directivity and gain.

11.5.2 Operating Characteristics

Wavelength is the distance traveled by one cycle of a radiated electric signal. The frequency of the signal is the number of cycles per second. It follows that the frequency is inversely proportional to the wavelength. Both wavelength and frequency are related to the speed of light. Conversion between the two parameters can be accomplished with the formula

$$c = f \times w$$

Where:

c = speed of light

f = operating frequency

w = wavelength

11-104 Broadcast Transmission Systems

The velocity of electric signals in air is essentially the same as that of light in free space (2.9983×10^{10} cm/s).

The *electrical length* of a radiating element is the most basic parameter of an antenna.

$$H = \frac{H_t F_o}{2733}$$

Where:

H = length of the radiating element in electrical degrees

H_t = length of the radiating element in feet

F_o = frequency of operation in kHz

When the radiating element is measured in meters

$$H = \frac{H_t F_o}{833.23}$$

The *radiation resistance* of an antenna is defined by the equation

$$R = \frac{P}{I^2}$$

Where:

R = radiation resistance

P = power delivered to the antenna

I = driving current at the antenna base

11.5.2a Antenna Bandwidth

Bandwidth is a general classification of the frequencies over which an antenna is effective. This parameter requires specification of acceptable tolerances relating to the uniformity of response over the intended operating band.

Strictly speaking, antenna bandwidth is the difference in frequency between two points at which the power output of the transmitter drops to one-half the midrange value. These points are called the *half-power points*. A half-power point is equal to a VSWR of 5.83:1, or the point at which the voltage response drops to 0.7071 of the midrange value. In most communications systems a VSWR of less than 1.2:1 within the occupied bandwidth of the radiated signal is preferable.

Antenna bandwidth depends upon the radiating element impedance and the rate at which the reactance of the antenna changes with frequency. Bandwidth and RF coupling go hand in hand regardless of the method used to excite the antenna. All elements between the transmitter output circuit and the antenna must be analyzed, first by themselves, then as part of the overall system bandwidth. In any transmission system, the *composite bandwidth*, not just the bandwidths of individual components, is of primary concern.

11.5.2b Polarization

Polarization is the angle of the radiated electric field vector in the direction of maximum radiation. Antennas may be designed to provide horizontal, vertical, or circular polarization. Horizontal or vertical polarization is determined by the orientation of the radiating element with respect to earth. If the plane of the radiated field is parallel to the ground, the signal is said to be horizontally polarized. If it is at right angles to the ground, it is said to be vertically polarized. When the receiving antenna is located in the same plane as the transmitting antenna, the received signal strength will be maximum.

Circular polarization (CP) of the transmitted signal results when equal electrical fields in the vertical and horizontal planes of radiation are out-of-phase by 90° and are rotating a full 360° in one wavelength of the operating frequency. The rotation can be clockwise or counterclockwise, depending on the antenna design. This continuously rotating field gives CP good signal penetration capabilities because it can be received efficiently by an antenna of any random orientation. Figure 11.5.1 illustrates the principles of circular polarization.

11.5.2c Antenna Beamwidth

Beamwidth in the plane of the antenna is the angular width of the directivity pattern where the power level of the received signal is down by 50 percent (3 dB) from the maximum signal in the desired direction of reception.

11.5.2d Antenna Gain

Directivity and gain are measures of how well energy is concentrated in a given direction. Directivity is the ratio of power density in a specified direction to the power density that would be produced if the energy were radiated isotropically. The reference can be linearly or circularly polarized. Directivity is usually given in dBi (decibels above isotropic).

Gain is the field intensity produced in a given direction by a fixed input power to the antenna, referenced to a dipole. Gain is frequently used as a figure of merit. Gain is closely related to directivity, which in turn is dependent upon the radiation pattern. High values of gain are usually obtained with a reduction in beamwidth.

An antenna is typically configured to exhibit “gain” by narrowing the beamwidth of the radiated signal to concentrate energy toward the intended coverage area. The actual amount of energy being radiated is the same with a unity gain antenna or a high gain antenna, but the useful energy (commonly referred to as the *effective radiated power*, or ERP) can be increased significantly.

Electrical *beam tilt* can also be designed into a high gain antenna. A conventional antenna typically radiates more than half of its energy above the horizon. This energy is lost for practical purposes in most applications. Electrical beam tilt, caused by delaying the RF current to the lower elements of a multi-element antenna, can be used to provide more useful power to the service area.

Pattern optimization is another method that may be used to maximize radiation to the intended service area. The characteristics of the transmitting antenna are affected, sometimes greatly, by the presence of the supporting tower, if side-mounted, or by nearby tall obstructions (such as another transmitting tower) if top-mounted. Antenna manufacturers use various methods to reduce pattern distortions. These generally involve changing the orientation of the radiators with respect to the tower, and adding parasitic elements.

11-106 Broadcast Transmission Systems

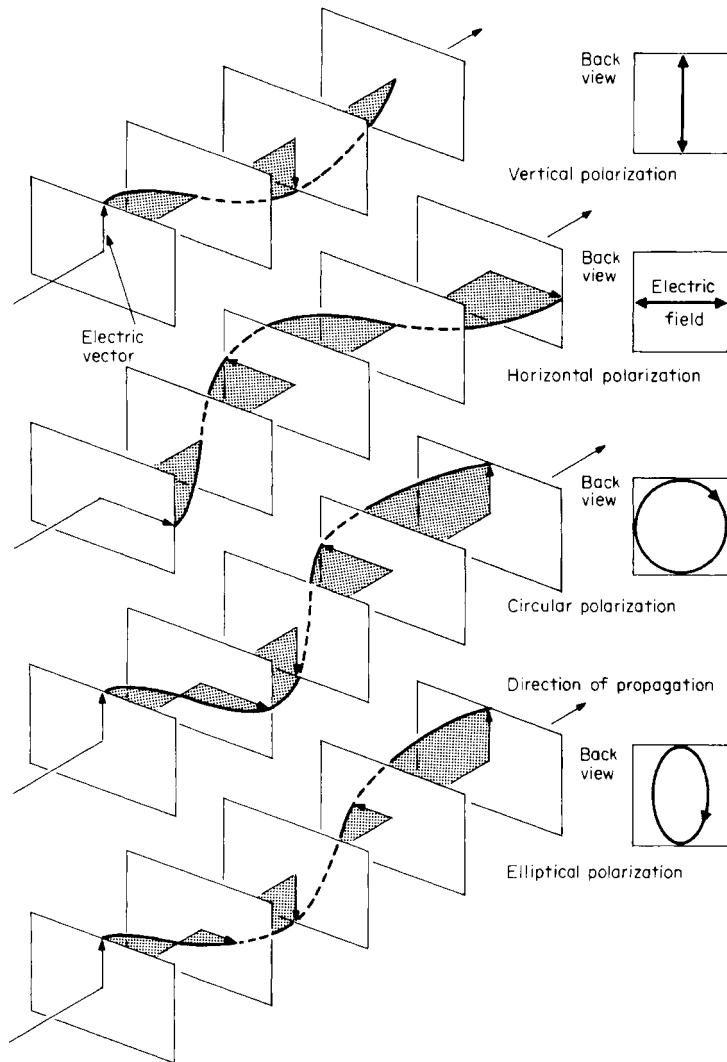


Figure 11.5.1 Polarization of the electric field of a transmitted wave.

11.5.2e Space Regions

Insofar as the transmitting antenna is concerned, space is divided into three regions:

- *Reactive near-field region.* This region is the area of space immediately surrounding the antenna in which the reactive components predominate. The size of the region varies, depending on the antenna design. For most antennas, the reactive near-field region extends 2 wavelengths or less from the radiating elements.

- *Radiating near-field region.* This region is characterized by the predictable distribution of the radiating field. In the near-field region, the relative angular distribution of the field is dependent on the distance from the antenna.
- *Far-field region.* This region is characterized by the independence of the relative angular distribution of the field with varying distance. The pattern is essentially independent of distance.

11.5.2f Impedance Matching

Most practical antennas require some form of impedance matching between the transmission line and the radiating elements. The implementation of a matching network can take on many forms, depending upon the operating frequency and output power.

The *negative sign* convention is generally used in impedance matching analysis. That is, if a network delays or retards a signal by θ degrees, the phase shift across the network is said to be minus θ degrees. For example, a 1/4-wavelength of transmission line, if properly terminated, has a phase shift of -90° . Thus a *lagging* or low-pass network has a negative phase shift, and a *leading* or high-pass network has a positive phase shift. There are three basic network types that can be used for impedance matching: L, pi, and tee.

L Network

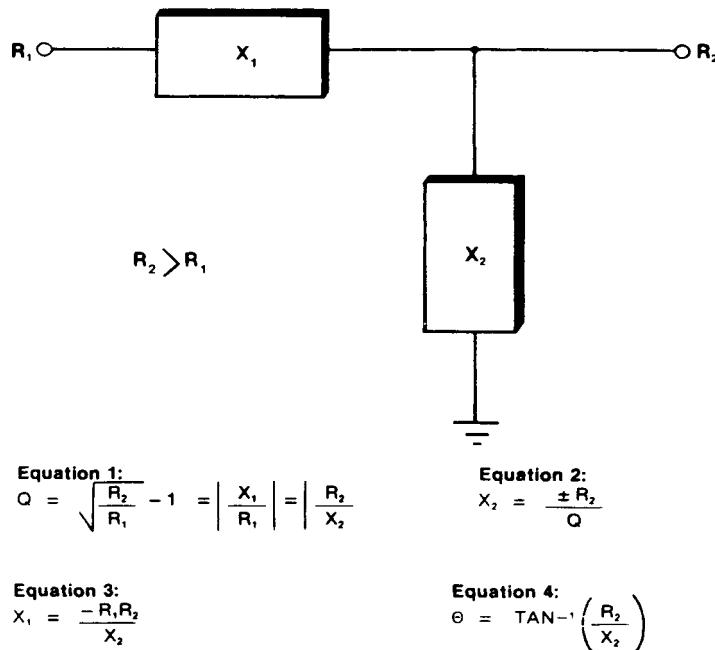
The L network is shown in Figure 11.5.2. The loaded Q of the network is determined from Equation (1). Equation (2) defines the shunt leg reactance, which is negative (capacitive) when θ is negative, and positive (inductive) when θ is positive. The series leg reactance is found using Equation (3), the phase shift via Equation (4), and the currents and voltages via Ohm's Law. Note that R_2 (the resistance on the shunt leg side of the L network) must always be greater than R_1 . An L network cannot be used to match equal resistances, or to adjust phase independently of resistance.

Tee Network

The tee network is shown in Figure 11.5.3. This configuration can be used to match unequal resistances. The tee network has the added feature that phase shift is independent of the resistance transformation ratio. A tee network can be considered simply as two L networks back-to-back. Note that there are two loaded Q s associated with a tee network: an input Q and output Q . In order to gauge the bandwidth of the tee network, the lower value of Q must be ignored. Note that the Q of a tee network increases with increasing phase shift.

Equations (5) through (14) describe the tee network. It is a simple matter to find the input and output currents via Ohm's Law, and the shunt leg current can be found via the Cosine Law, Equation (12). Note that this current increases with increasing phase shift. Equation (13) describes the mid-point resistance of a tee network, which is always higher than R_1 or R_2 . Equation (14) is useful when designing a *phantom tee network*; that is, where X_2 is made up only of the antenna reactance, and there is no physical component in place of X_2 . Keep in mind that a tee network is considered as having a lagging or negative phase shift when the shunt leg is capacitive (X_3 negative), and vice versa. The input and output arms can go either negative or positive, depending on the resistance transformation ratio and desired phase shift.

11-108 Broadcast Transmission Systems



Where: R_1 = L network input resistance (ohms)

R_2 = L network output resistance (ohms)

X_1 = Series leg reactance (ohms)

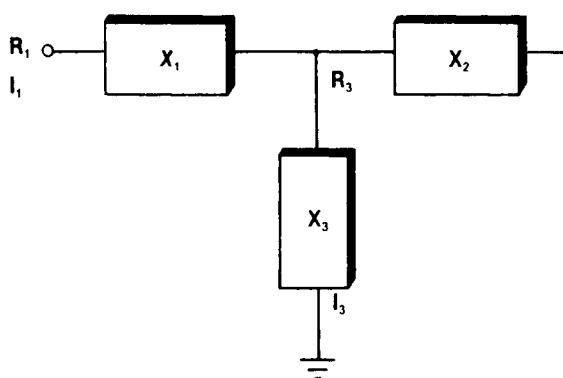
X_2 = Shunt leg reactance (ohms)

Q = Loaded Q of the L network

Figure 11.5.2 L network parameters.

Pi Network

The pi network is shown in Figure 11.5.4. It can also be considered as two L networks back-to-back and, therefore, the same comments regarding overall loaded Q apply. Note that susceptances have been used in Equations (15) through (19) instead of reactances in order to simplify calculations. The same conventions regarding tee network currents apply to pi network voltages, Equations (20, 21, and 22). The mid-point resistance of a pi network is always less than R_1 or R_2 . A pi network is considered as having a negative or lagging phase shift when Y_3 is positive, and vice versa.



Where:
 R_1 = Tee network input resistance (ohms)
 R_2 = Tee network output resistance (ohms)
 I_1 = Tee network input current (amps)
 I_2 = Tee network output current (amps)
 I_3 = Shunt element current (amps)
 X_1 = Network input element reactance (ohms)
 X_2 = Network output element reactance (ohms)
 X_3 = Network shunt element reactance (ohms)
 P = Input power (watts)
 Q_1 = Input loaded Q
 Q_2 = Output loaded Q
 R_3 = Midpoint resistance of the network (ohms)

$$\text{Equation 5: } X_3 = \frac{\sqrt{R_1 R_2}}{\sin(\theta)}$$

$$\text{Equation 6: } X_1 = \frac{R_1}{\tan(\theta)} - X_3$$

$$\text{Equation 7: } X_2 = \frac{R_2}{\tan(\theta)} - X_3$$

$$\text{Equation 8: } Q_1 = \left| \frac{X_1}{R_1} \right|$$

$$\text{Equation 9: } Q_2 = \left| \frac{X_2}{R_2} \right|$$

$$\text{Equation 10: } I_1 = \sqrt{\frac{P}{R_1}}$$

$$\text{Equation 11: } I_2 = \sqrt{\frac{P}{R_2}}$$

$$\text{Equation 12: } I_3 = \sqrt{I_1^2 + I_2^2 - 2(I_1)(I_2)\cos(\theta)}$$

$$\text{Equation 13: } R_3 = (Q_1^2 + 1)R_2$$

$$\text{Equation 14: } \theta = \tan^{-1} \left(\frac{X_1}{R_1} \right) \pm \tan^{-1} \left(\frac{X_2}{R_2} \right)$$

Figure 11.5.3 T network parameters.

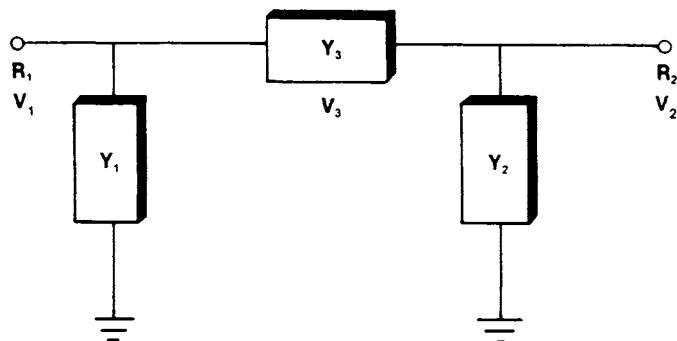
Line Stretcher

A *line stretcher* makes a transmission line look longer or shorter in order to produce sideband impedance symmetry at the transmitter PA (see Figure 11.5.5). This is done to reduce audio distortion in an envelope detector—the kind of detector that most monophonic AM receivers employ. Symmetry is defined as equal sideband resistances, and equal—but opposite sign—sideband reactances.

There are two possible points of symmetry, each 90° from the other. One produces sideband resistances greater than the carrier resistance, and the other produces the opposite effect. One side will create a pre-emphasis effect, and the other a de-emphasis effect.

Depending on the Q of the transmitter output network, one point of symmetry may yield lower sideband VSWR at the PA than the other. This results from the Q of the output network opposing the Q of the antenna in one direction, but aiding the antenna Q in the other direction.

11-110 Broadcast Transmission Systems

**Equation 15:**

$$Y_3 = \frac{1}{-\sin(\theta)\sqrt{R_1 R_2}}$$

Equation 16:

$$Y_1 = \frac{\tan(\theta)}{R_1 - Y_3}$$

Equation 17:

$$Y_2 = \frac{\tan(\theta)}{R_2 - Y_3}$$

Equation 18:

$$Q_1 = |R_1 Y_1|$$

Equation 19:

$$Q_2 = |R_2 Y_2|$$

Equation 20:

$$V_1 = \sqrt{R_1 P}$$

Equation 21:

$$V_2 = \sqrt{R_2 P}$$

Equation 22:

$$V_3 = \sqrt{V_1^2 + V_2^2 - 2(V_1)(V_2) \cos(\theta)}$$

Equation 23:

$$R_3 = \frac{Q_2^2 + 1}{R_2}$$

Where: R_1 = Pi network input resistance (ohms)

R_2 = Output resistance (ohms)

V_1 = Input voltage (volts)

V_2 = Output voltage (volts)

V_3 = Voltage across series element (volts)

P = Power input to pi network (watts)

Y_1 = Input shunt element susceptance (mhos)

Y_2 = Output shunt element susceptance (mhos)

Y_3 = Series element susceptance (mhos)

Q_1 = Input loaded Q

Q_2 = Output loaded Q

Figure 11.5.4 Pi network parameters.

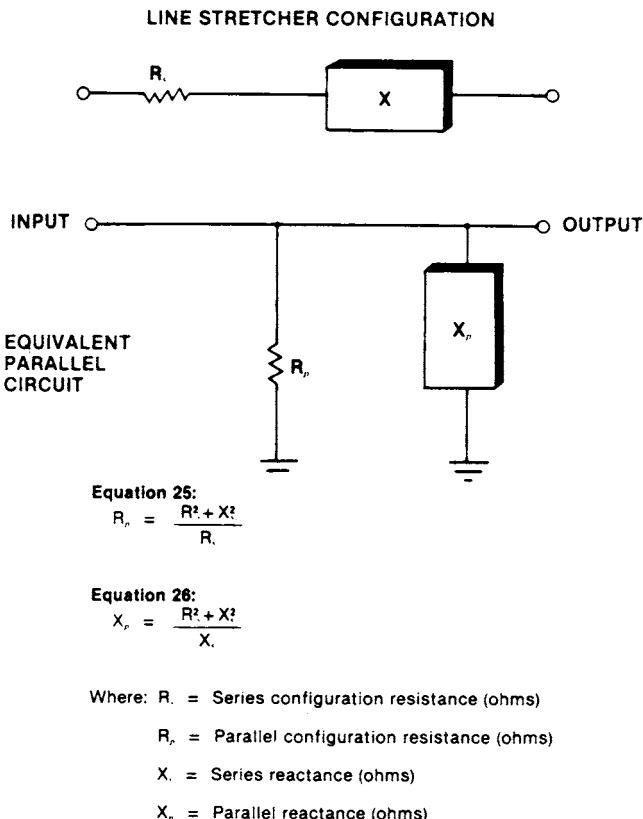


Figure 11.5.5 Line stretcher configuration.

11.5.3 Antenna Types

The *dipole antenna* is simplest of all antennas, and the building block of most other designs. The dipole consists of two in-line rods or wires with a total length equal to 1/2-wave at the operating frequency. Figure 11.5.6 shows the typical configuration, with two 1/4-wave elements connected to a transmission line. The radiation resistance of a dipole is on the order of 73Ω . The bandwidth of the antenna can be increased by increasing the diameter of the elements, or by using cones or cylinders rather than wires or rods, as shown in the figure. Such modifications also increase the impedance of the antenna.

The dipole can be straight (in-line) or bent into a V-shape. The impedance of the V-dipole is a function of the V angle. Changing the angle effectively tunes the antenna. The vertical radiation pattern of the V-dipole antenna is similar to the straight dipole for angles of 120° or less.

A *folded dipole* may be fashioned as shown in Figure 11.5.7. Such a configuration results in increased bandwidth and impedance. Impedance can be further increased by using rods of different diameter and by varying the spacing of the elements. The 1/4-wave dipole elements con-

11-112 Broadcast Transmission Systems

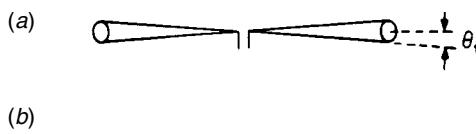


Figure 11.5.6 Half-wave dipole antenna: (a) conical dipole; (b) conventional dipole.

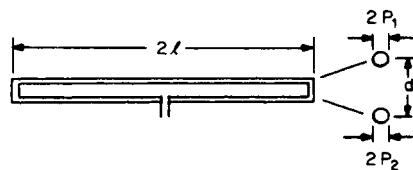


Figure 11.5.7 Folded dipole antenna.

nected to the closely-coupled 1/2-wave element act as a matching stub between the transmission line and the single-piece 1/2-wave element. This broadbands the folded dipole antenna by a factor of 2.

A corner-reflector antenna may be formed as shown in Figure 11.5.8. A ground plane or flat reflecting sheet is placed at a distance of 1/16- to 1/4-wavelengths behind the dipole. Gain in the forward direction can be increased by a factor of 2 with this type of design.

11.5.3a Quarter-Wave Monopole

A conductor placed above a ground plane forms an image in the ground plane such that the resulting pattern is a composite of the *real antenna* and the *image antenna* (see Figure 11.5.9). The operating impedance is 1/2 of the impedance of the antenna and its image when fed as a physical antenna in free space. An example will help illustrate this concept. A 1/4-wave monopole mounted on an infinite ground plane has an impedance equal to 1/2 the free-space impedance of a 1/2-wave dipole. It follows, then, that the theoretical characteristic resistance of a 1/4-wave monopole with an infinite ground plane is 37Ω .

For a real-world antenna, an infinite ground plane is neither possible nor required. An antenna mounted on a ground plane that is 2 to 3 times the operating wavelength has about the same impedance as a similar antenna mounted on an infinite ground plane.

11.5.3b Log-Periodic Antenna

The log-periodic antenna can take-on a number of forms. Typical designs include the *conical log spiral*, *log-periodic V*, and *log-periodic dipole*. The most common of these antennas is the log-periodic dipole. The antenna can be fed either by using alternating connections to a balanced line, or by a coaxial line running through one of the feeders from front-to-back. In theory, the log-periodic antenna may be designed to operate over many octaves. In practice, however, the upper frequency is limited by the precision required in constructing the small elements, feed lines, and support structure of the antenna.

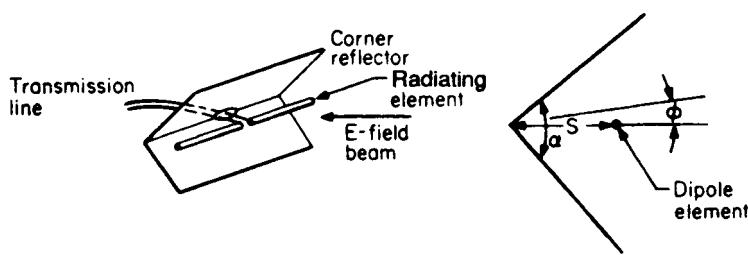


Figure 11.5.8 Corner-reflector antenna.

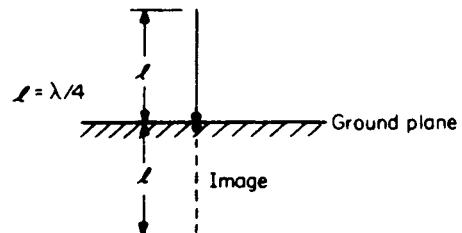


Figure 11.5.9 Vertical monopole mounted above a ground plane.

11.5.3c Yagi-Uda Antenna

The *Yagi-Uda* is an *end-fire array* consisting typically of a single driven dipole with a *reflector dipole* behind the driven element, and one or more *parasitic director elements* in front (Figure 11.5.10). Common designs use from one to 7 director elements. As the number of elements is increased, directivity increases. Bandwidth, however, decreases as the number of elements is increased. Arrays of more than 4 director elements are typically narrowband.

The radiating element is 1/2-wavelength at the center of the band covered. The single reflector element is slightly longer, and the director elements are slightly shorter, all spaced approximately 1/4-wavelength from each other.

Table 11.5.1 demonstrates how the number of elements determines the gain and beamwidth of a Yagi-Uda antenna.

11.5.3d Waveguide Antenna

The waveguide antenna consists of a dominant-mode-fed waveguide opening onto a conducting ground plane. Designs may be based on rectangular, circular, or coaxial waveguide (also called an *annular slot*). The slot antenna is simplicity itself. A number of holes of a given dimension are placed at intervals along a section of waveguide. The radiation characteristics of the antenna are determined by the size, location, and orientation of the slots. The antenna offers optimum reliability because there are no discrete elements, except for the waveguide section itself.

11-114 Broadcast Transmission Systems

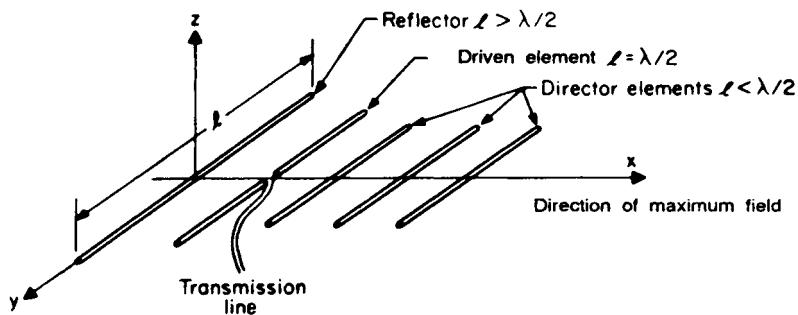


Figure 11.5.10 The Yagi-Uda array.

Table 11.5.1 Typical Characteristics of Single-Channel Yagi-Uda Arrays

Number of Elements	Gain, dB	Beamwidth, Degrees
2	3 to 4	65
3	6 to 8	55
4	7 to 10	50
5	9 to 11	45
9	12 to 14	37
15	14 to 16	30

11.5.3e Horn Antenna

The horn antenna can be considered a natural extension of the dominant-mode waveguide feeding the horn in a manner similar to the wire antenna, which is a natural extension of the two-wire transmission line. The most common types of horns are the *E-plane sectoral*, *H-plane sectoral*, *pyramidal horn* (formed by expanding the walls of the $\text{TE}_{0,1}$ -mode-fed rectangular waveguide), and the *conical horn* (formed by expanding the walls of the $\text{TE}_{1,1}$ -mode-fed circular waveguide). The H-plane sectoral horn is shown in Figure 1.5.11. Dielectric-loaded waveguides and horns offer improved pattern performance over unloaded horns. Ridged and tapered horn designs improve the bandwidth characteristics. Horn antennas are available in single and dual polarized configurations.

11.5.3f Reflector Antenna

The reflector antenna is formed by mounting a radiating feed antenna above a reflecting ground plane. The most basic form of reflector is the loop or dipole spaced over a finite ground plane. This concept is the basis for the parabolic or *spherical reflector antenna*. The *parabolic reflector antenna* can be fed directly or through the use of a subreflector in the focal region of the parabola. In this approach, the subreflector is illuminated from the parabolic surface. The chief disad-

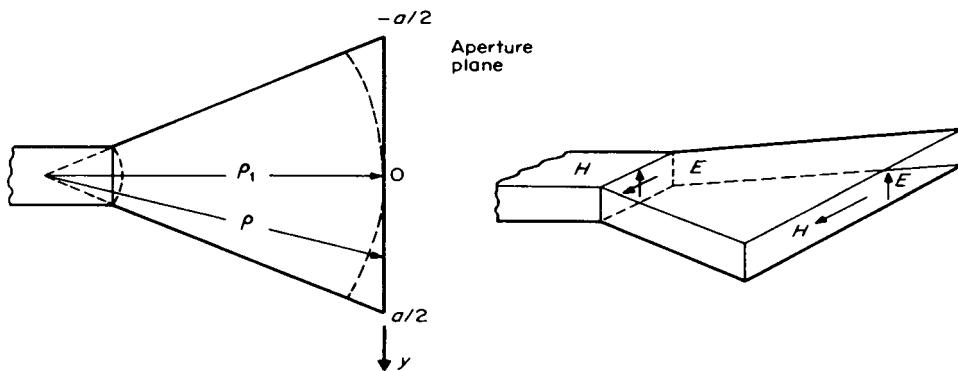


Figure 11.5.11 Geometry of an *H*-plane sectoral horn.

vantage of this design is the aperture blockage of the subreflector, which restricts its use to large aperture antennas. The operation of a parabolic or spherical reflector antenna is typically described using physical optics.

Parabolic-reflector antennas are usually illuminated by a flared-horn antenna with a flare angle of less than 18° . A rectangular horn with a flare angle less than 18° has approximately the same aperture field as the dominant-mode rectangular waveguide feeding the horn.

11.5.3g Spiral Antenna

The bandwidth limitations of an antenna are based on the natural change in the critical dimensions of the radiating elements caused by differing wavelengths. The spiral antenna overcomes this limitation because the radiating elements are specified only in angles. A two-arm equiangular spiral is shown in Figure 11.5.12. This common design gives wideband performance. Circular polarization is inherent in the antenna. Rotation of the pattern corresponds to the direction of the spiral arms. The gain of a spiral antenna is typically slightly higher than a dipole.

The basic spiral antenna radiates on both sides of the arms. Unidirectional radiation is achieved through the addition of a reflector or cavity.

11.5.3h Array Antenna

The term “array antenna” covers a wide variety of physical structures. The most common configuration is the *planar array antenna*, which consists of a number of radiating elements regularly spaced on a rectangular or triangular lattice. The *linear array antenna*, where the radiating elements are placed in a single line, is also common. The pattern of the array is the product of the element pattern and the array configuration. Large array antennas may consist of 20 or more radiating elements.

Correct phasing of the radiating elements is the key to operation of the system. The radiating pattern of the structure, including direction, can be controlled through proper adjustment of the relative phase of the elements.

11-116 Broadcast Transmission Systems

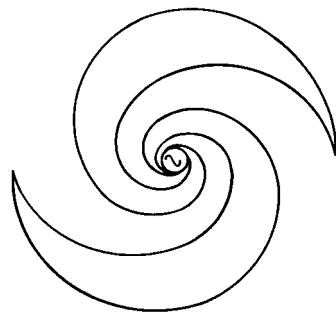


Figure 11.5.12 Two-arm equiangular spiral antenna.

11.5.4 AM Broadcast Antenna Systems

Vertical polarization of the transmitted signal is used for AM broadcast stations because of its superior groundwave propagation, and because of the simple antenna designs that it affords. The FCC and licensing authorities in other countries have established classifications of AM stations with specified power levels and hours of operation. Protection requirements set forth by the FCC specify that some AM stations (in the U.S.) reduce their transmitter power at sunset, and return to full power at sunrise. This method of operation is based on the propagation characteristics of AM band frequencies. AM signals propagate further at nighttime than during the day.

The different day/night operating powers are designed to provide each AM station with a specified coverage area that is free from interference. Theory rarely translates directly into practice insofar as coverage is concerned, however, because of the increased interference that all AM stations suffer at nighttime.

The tower structure visible at any AM radio station transmitter site is only half of the antenna system. The second element is a buried ground system. Current on a tower does not simply disappear, rather it returns to earth through the capacitance between the earth and the tower. Ground losses are greatly reduced if the tower has a radial copper ground system. A typical single tower ground system is made up of 120 radial ground wires each 140 electrical degrees long (at the operating frequency) equally spaced out from the tower base. This is often augmented with an additional 120 interspersed radials 50 ft long.

11.5.4a Directional AM Antenna Design

When a non-directional antenna with a given power does not radiate enough energy to serve the station's primary service area, or radiates too much energy toward other radio stations on the same or adjacent frequencies, it is necessary to employ a directional antenna system. Rules set out by the FCC and regulatory agencies in other countries specify the protection requirements to be provided by various classes of stations, for both daytime and nighttime hours. These limits tend to define the shape and size of the most desirable antenna pattern.

A directional antenna functions by carefully controlling the amplitude and phase of the RF currents fed to each tower in the system. The directional pattern is a function of the number and spacing of the towers (vertical radiators), and the relative phase and magnitude of their currents. The number of towers in a directional AM array can range from two to six or even more in a

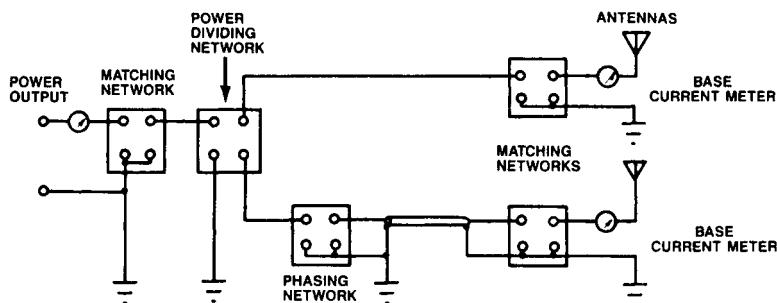


Figure 11.5.13 Block diagram of an AM directional antenna feeder system for a two tower array.

complex system. One tower is defined as the *reference tower*. The amplitude and phase of the other towers are measured relative to this reference.

A complex network of power splitting, phasing, and antenna coupling elements is required to make a directional system work. Figure 11.5.13 shows a block diagram of a basic two tower array. A power divider network controls the relative current amplitude in each tower. A phasing network provides control of the phase of each tower current, relative to the reference tower. Matching networks at the base of each tower couple the transmission line impedance to the base operating impedance of the radiating towers.

In practice, the system shown in the figure would not consist of individual elements. Instead, the matching network, power dividing network, and phasing network are usually combined into a single unit, referred to as the *phasor*.

Antenna Pattern Design

The pattern of any AM directional antenna system (array) is determined by a number of factors, including:

- Electrical parameters (phase relationship and current ratio for each tower)
- Height of each tower
- Position of each tower with respect to the other towers (particularly with respect to the reference tower)

A *directional array* consists of two or more towers arranged in a specific manner on a plot of land. Figure 11.5.14 shows a typical 3-tower array, and the pattern such an array would produce. This is an *in-line array*, meaning that all the elements (towers) are in line with one another. Notice that the *major lobe* is centered on the same line as the line of towers, and that the *pattern nulls (minima)* are positioned symmetrically about the line of towers, protecting co-channel stations A and B at true bearings of 315° and 45° , respectively.

Figure 11.5.15 shows the same array, except that it has been rotated by 10° . Notice that the pattern shape is not changed, but the position of the major lobe and the nulls follow the line of towers. Also notice that the nulls no longer are pointed at the stations to be protected. Figures 11.5.16 and 11.5.17 show that the position of the nulls can be changed by varying the electrical

11-118 Broadcast Transmission Systems

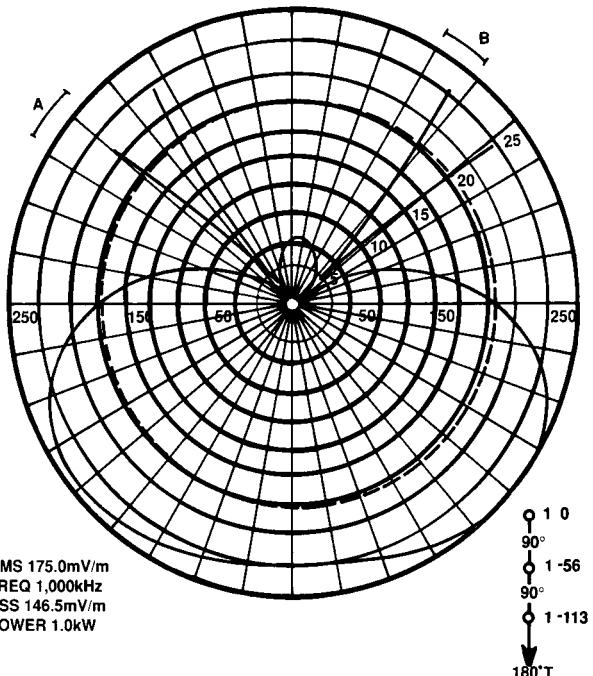


Figure 11.5.14 Radiation pattern generated with a 3-tower in-line directional array using the electrical parameters and orientation shown.

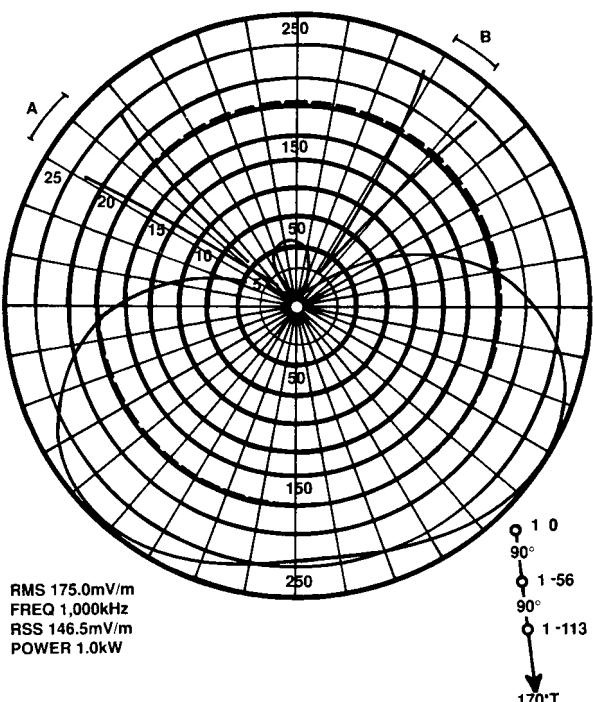


Figure 11.5.15 Radiation pattern produced when the array of Figure 11.5.14 is rotated to a new orientation.

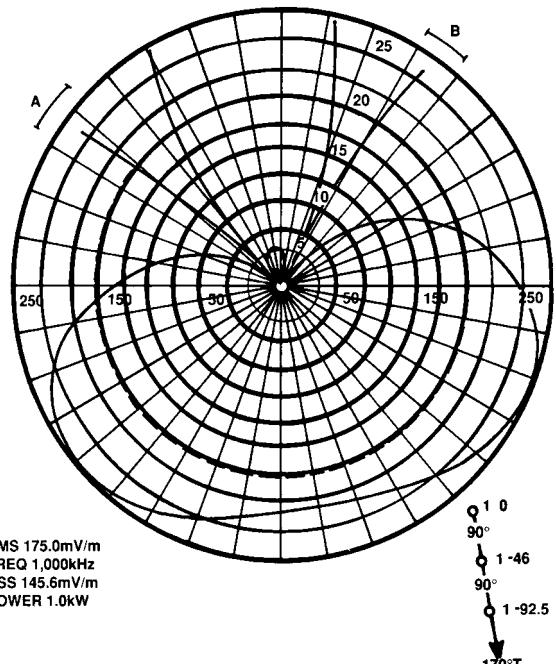


Figure 11.5.16 Directional pattern generated with the orientation shown in Figure 11.5.15, but with different electrical parameters.

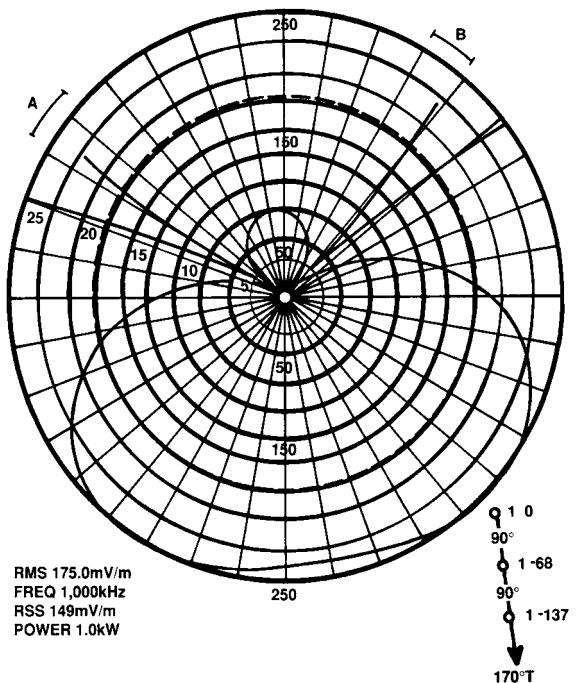


Figure 11.5.17 Directional pattern generated by the tower orientation shown in Figure 11.5.15, but with a new set of electrical parameters.

11-120 Broadcast Transmission Systems

parameters so that one or the other can be pointed in the required direction, but not both. Also, when this is attempted, the size and shape of the small *back lobe* is changed, as is the shape of the major lobe, especially the radiation on the line of towers.

If this directional antenna system were constructed on a gigantic turntable, the pattern could be rotated without affecting the shape. But, to accomplish the required protections and to have the major lobe(s) oriented in the right direction, there is only one correct position. In most cases, the position of the towers will be specified with respect to a single reference tower. The location of the other towers will be given in the form of a distance and bearing from that reference. Occasionally, a reference point, usually the center of the array, will be used for a geographic coordinate point.

11.5.4b Bearing

The bearing or azimuth of the towers from the reference tower or point is specified clockwise in degrees from *true north*. The distinction between true and *magnetic north* is vital. The magnetic North Pole is not at the true or geographic North Pole. (In fact, it is in the vicinity of 74° north, 101° west, in the islands of northern Canada.) The difference between magnetic and true bearings is called variation or *magnetic declination*. Declination, a term generally used by surveyors, varies for different locations; it is not a constant. The earth's magnetic field is subject to a number of changes in intensity and direction. These changes take place over daily, yearly, and long-term (or *secular*) periods. The secular changes result in a relatively constant increase or decrease in declination over a period of many years.

Antenna Monitoring System

Monitoring the operation of an AM directional antenna basically involves measuring the power into the system, the relative value of currents into the towers, their phase relationships, and the levels of radiated signal at certain monitoring points some distance from the antenna. Figure 11.5.18 shows a block diagram of a typical monitoring system for a three tower array. For systems with additional towers, the basic layout is extended by adding more pickup elements, sample lines, and a monitor with additional inputs.

Phase/Current Sample Loop

Two types of phase/current sample pickup elements are commonly used: the *sample loop* and *toroidal current transformer* (TCT). The sample loop consists of a single turn unshielded loop of rigid construction, with a fixed gap at the open end for connection of the sample line. The device must be mounted on the tower near the point of maximum current. The loop can be used on towers of both uniform and non-uniform cross section. It must operate at tower potential, except for towers of less than 130° electrical height, where the loop can be operated at ground potential.

When the sample loop is operated at tower potential, the coax from the loop to the base of the tower is also at tower potential. In order to bring the sample line across the base of the tower, a sample line isolation coil is used.

A shielded toroidal current transformer can also be used as the phase/current pickup element. Such devices offer several advantages over the sample loop including greater stability and

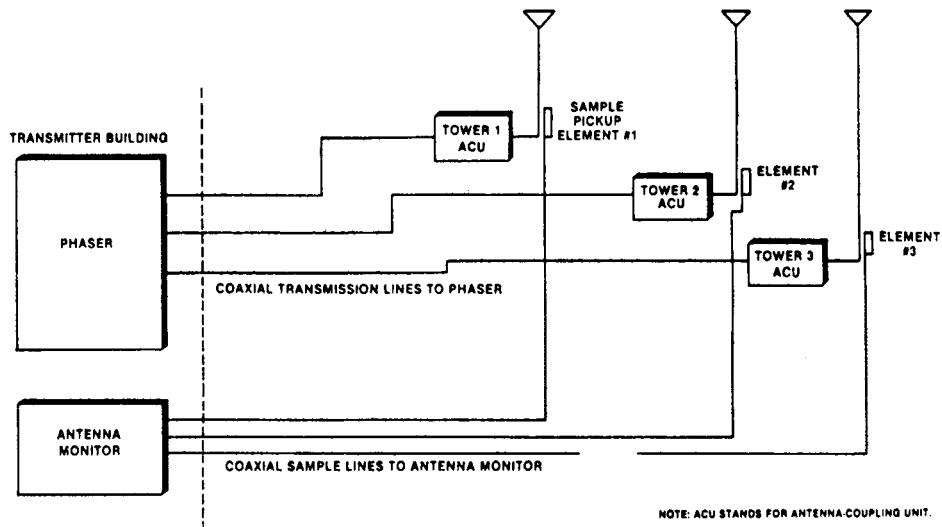


Figure 11.5.18 A typical 3-tower directional antenna monitoring system.

reliability. Because they are located inside the tuning unit cabinet or house, TCTs are protected from wind, rain, ice, and vandalism.

Unlike the rigid, fixed sample loop, toroidal current transformers are available in several sensitivities, ranging from 0.25–1.0 V per ampere of tower current. Tower currents of up to 40 A or more can be handled, providing a more usable range of voltages for the antenna monitor.

Figure 11.5.19 shows the various arrangements that can be used for phase/current sample pickup elements.

Sample lines

The selection and installation of the sampling lines for a directional monitoring system are important factors in the ultimate accuracy of the overall array.

With *critical arrays* (antennas requiring operation within tight limits specified in the station license) all sample lines must be of equal electrical length and installed in such a manner that corresponding lengths of all lines are exposed to equal environmental conditions.

While sample lines can be run above ground on supports (if protected and properly grounded) the most desirable arrangement is direct burial using jacketed cable. Burial of sample line cable is almost a standard practice because proper burial offers good protection against physical damage and a more stable temperature environment.

The Common Point

The power input to a directional antenna is measured at the phasor *common point*. Power is determined by the *direct method*.

11-122 Broadcast Transmission Systems

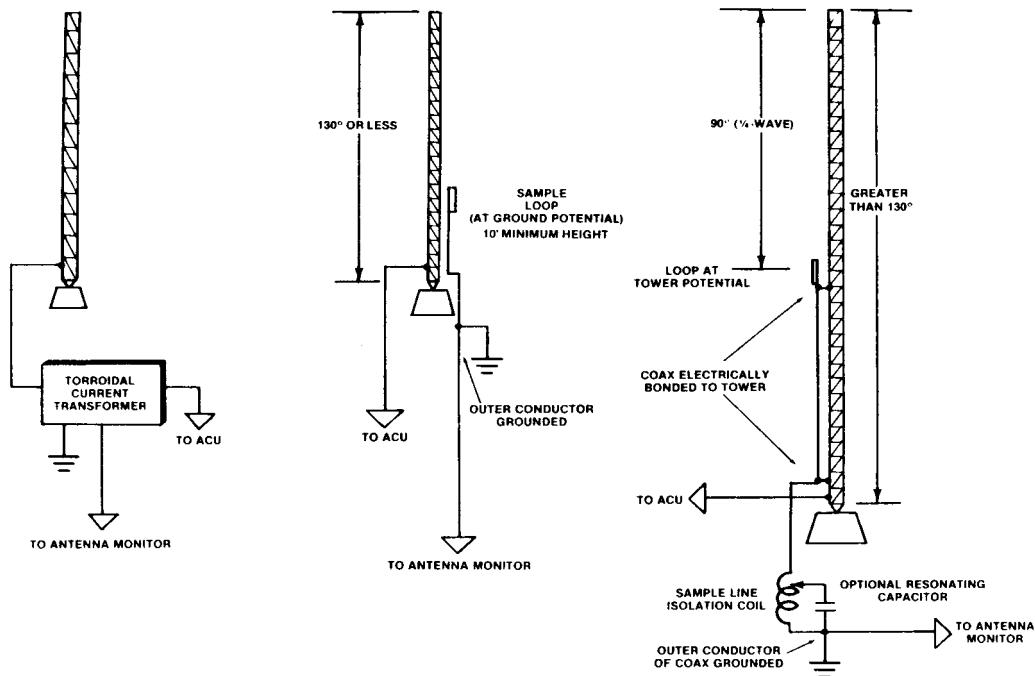


Figure 11.5.19 Three possible circuit configurations for phase sample pickup.

$$P = I^2 R$$

Where:

P = power in W

I = the common point current in A

R = the common point resistance in Ω

Monitor Points

Routine monitoring of a directional antenna involves the measurement of field intensity at certain locations away from the antenna, called *monitor points*. These points are selected and established during the initial tune-up of the antenna system. Measurements at the monitor points should confirm that radiation in prescribed directions does not exceed a value that would cause interference to other stations operating on the same or adjacent frequencies. The field intensity limits at these points are normally specified in the station license. Measurements at the monitor points may be required on a weekly or a monthly basis, depending on several factors and conditions relating to the particular station. If the system is not a critical array, quarterly measurements may be sufficient.

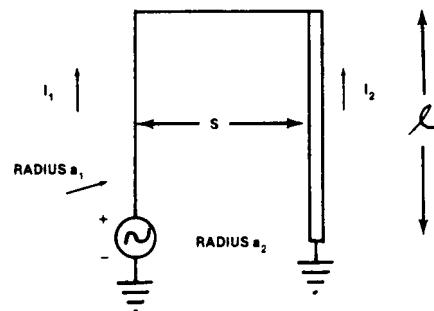


Figure 11.5.20 The folded unipole antenna can be thought of as a 1/2-wave folded dipole antenna perpendicular to the ground and cut in half.

11.5.4c Folded Unipole Antenna

The *folded unipole* antenna consists of a grounded vertical structure with one or more conductors folded back parallel to the side of the structure. It can be visualized as a half-wave folded dipole perpendicular to the ground and cut in half (see Figure 11.5.21). This design makes it possible to provide a wide range of resonant radiation resistances by varying the ratio of the diameter of the folded-back conductors in relation to the tower. Top loading can also be used to broaden the antenna bandwidth. A side view of the folded unipole is shown in Figure 11.5.22.

The folded unipole antenna could be considered a modification of the standard shunt-fed system. Instead of a slant wire that leaves the tower at an approximate 45° angle (as used for shunt-fed systems), the folded unipole antenna has one or more wires attached to the tower at a predetermined height. The wires are supported by standoff insulators and run parallel to the sides of the tower down to the base.

The tower is grounded at the base. The folds, or wires, are joined together at the base and driven through an impedance matching network. Depending upon the tuning requirements of the folded unipole, the wires may be connected to the tower at the top and/or at predetermined levels along the tower with shorting stubs.

The folded unipole can be used on tall (130° or greater) towers. However, if the unipole is not divided into two parts, the overall efficiency (*unattenuated field intensity*) will be considerably lower than the normally expected field for the electrical height of the tower.

11.5.5 FM Broadcast Antenna Systems

The propagation characteristics of VHF FM radio are much different than for MF AM. There is essentially no difference between day and night FM propagation. FM stations have relatively uniform day and night service areas with the same operating power.

A wide variety of antennas is available for use in the FM broadcast band. Nearly all employ circular polarization. Although antenna designs differ from one manufacturer to another, generalizations can be made that apply to most units.

11-124 Broadcast Transmission Systems

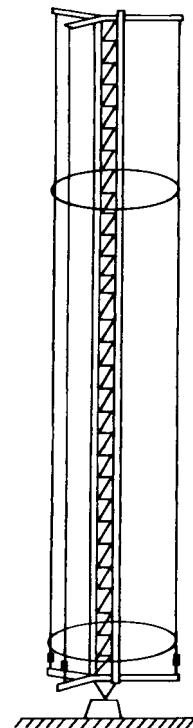


Figure 11.5.21 The folds of the unipole antenna are arranged either near the legs of the tower or near the faces of the tower.

11.5.5a Antenna Types

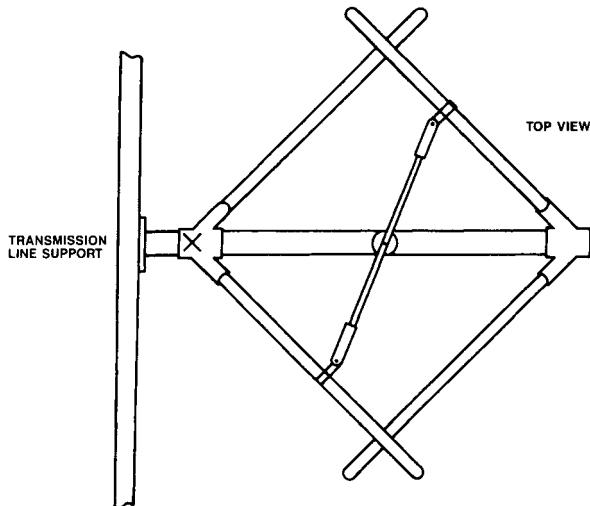
There are three basic classes of FM broadcast transmitting antennas in use today: *ring stub* and *twisted ring*, *shunt-* and *series-fed slanted dipole*, and *multi-arm short helix*. While each design is unique, all have the following items in common:

- The antennas are designed for side mounting to a steel tower or pole.
- Radiating elements are shunted across a common rigid coaxial transmission line.
- Elements are placed along the rigid line every one wavelength.
- Antennas with one to seven bays are fed from the bottom of the coaxial transmission line.
- Antennas with more than seven bays are fed from the center of the array to provide more predictable performance in the field.

Antennas generally include a means of tuning out reactances after the antenna has been installed through the adjustment of variable capacitive or inductive elements at the feed point.

Figure 11.5.23 shows a shunt-fed slanted dipole antenna that consists of two half-wave dipoles offset 90°. The two sets of dipoles are rotated 22.5° (from their normal plane) and are *delta-matched* to provide a 50Ω impedance at the radiator input flange. The lengths of all four dipole arms can be matched to resonance by mechanical adjustment of the end fittings. Shunt-feeding (when properly adjusted) provides equal currents in all four arms.

(a)



(b)

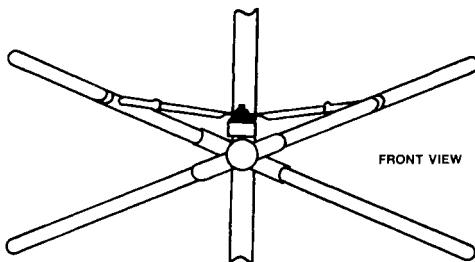


Figure 11.5.22 Mechanical configuration of one bay of a circularly-polarized FM transmitting antenna: (a) top view of the antenna, (b) front view.

Wideband *panel antennas* are a fourth common type of antenna used for FM broadcasting. Panel designs share some of the characteristics listed previously, but are intended primarily for specialized installations in which two or more stations will use the antenna simultaneously. Panel antennas are larger and more complex than other FM antennas, but offer the possibility for shared tower space among several stations and custom coverage patterns that would be difficult or even impossible with more common designs.

The ideal combination of antenna gain and transmitter power for a particular installation involves the analysis of a number of parameters. As shown in Table 11.5.2, a variety of pairings can be made to achieve the same ERP.

11-126 Broadcast Transmission Systems

Table 11.5.2 Various Combinations of Transmitter Power and Antenna Gain that will Produce 100 KW Effective Radiated Power (ERP) for an FM Station

Number of Bays	Antenna Gain	Required Transmitter Power (kW)*
3	1.5888	66.3
4	2.1332	49.3
5	2.7154	38.8
6	3.3028	31.8
7	3.8935	27.0
8	4.4872	23.5
10	5.6800	18.5
12	6.8781	15.3

* Effects of transmission line and related losses not included.

11.5.6 Bibliography

- Benson, B., and Whitaker, J.: *Television and Audio Handbook for Technicians and Engineers*, McGraw-Hill, New York, N.Y., 1989.
- Bingeman, Grant: "AM Tower Impedance Matching," *Broadcast Engineering*, Intertec Publishing, Overland Park, Kan., July 1985.
- Bixby, Jeffrey: "AM DAs—Doing it Right," *Broadcast Engineering*, Intertec Publishing, Overland Park, Kan., February 1984.
- Chick, Elton B.: "Monitoring Directional Antennas," *Broadcast Engineering*, Intertec Publishing, Overland Park, Kan., July 1985.
- Fink, D., and D. Christiansen (eds.): *Electronics Engineer's Handbook*, 3rd ed., McGraw-Hill, New York, N.Y., 1989.
- Jordan, Edward C. (ed.): "Reference Data for Engineers—Radio, Electronics, Computer and Communications," 7th ed., Howard W. Sams, Indianapolis, Ind., 1985.
- Mullaney, John H.: "The Folded Unipole Antenna for AM Broadcast," *Broadcast Engineering*, Intertec Publishing, Overland Park, Kan., January 1960.
- Mullaney, John H.: "The Folded Unipole Antenna," *Broadcast Engineering*, Intertec Publishing, Overland Park, Kan., July 1986.
- Westberg, J. M.: "Effect of 90° Stub on Medium Wave Antennas," *NAB Engineering Handbook*, 7th ed., National Association of Broadcasters, Washington, D.C., 1985.

Chapter
11.6

RF Combiner and Diplexer Systems

Cecil Harrison, Robert A. Surette

Jerry C. Whitaker, Editor-in-Chief

11.6.1 Introduction

The basic purpose of an RF combiner is to add two or more signals to produce an output signal that is a composite of the inputs. The combiner performs this signal addition while providing isolation between inputs. Combiners perform other functions as well, and can be found in a wide variety of RF transmission equipment. Combiners are valuable devices because they permit multiple amplifiers to drive a single load. The isolation provided by the combiner permits tuning adjustments to be made on one amplifier—including turning it on or off—without significantly affecting the operation of the other amplifier. In a typical application, two amplifiers drive the hybrid and provide two output signals:

- A combined output representing the sum of the two input signals, typically directed toward the antenna.
- A difference output representing the difference in amplitude and phase between the two input signals. The difference output typically is directed toward a dummy (reject) load.

For systems in which more than two amplifiers must be combined, two or more combiners are cascaded.

Diplexers are similar in nature to combiners but permit the summing of output signals from two or more amplifiers operating at different frequencies. This allows, for example, the outputs of several transmitters operating on different frequencies to utilize a single broadband antenna.

11.6.2 Passive Filters

A *filter* is a multiport-network designed specifically to respond differently to signals of different frequency [1]. This definition excludes *networks*, which incidentally behave as filters, sometimes to the detriment of their main purpose. Passive filters are constructed exclusively with passive

11-128 Broadcast Transmission Systems

elements (i.e., resistors, inductors, and capacitors). Filters are generally categorized by the following general parameters:

- Type
- Alignment (or class)
- Order

11.6.2a Filter Type

Filters are categorized by type, according to the magnitude of the frequency response, as one of the following [1]:

- *Low-pass* (LP)
- *High-pass* (HP)
- *Band-pass* (BP)
- *Band-stop* (BS).

The terms *band-reject* or *notch* are also used as descriptive of the BS filter. The term *all-pass* is sometimes applied to a filter whose purpose is to alter the phase angle without affecting the magnitude of the frequency response. Ideal and practical interpretations of the types of filters and the associated terminology are summarized in Figure 11.6.1.

In general, the voltage gain of a filter in the *stop band* (or *attenuation band*) is less than $\sqrt{2}/2$ (≈ 0.707) times the maximum voltage gain in the pass band. In logarithmic terms, the gain in the stop band is at least 3.01 dB less than the maximum gain in the pass band. The cutoff (*break* or *corner*) frequency separates the pass band from the stop band. In BP and BS filters, there are two cutoff frequencies, sometimes referred to as the *lower* and *upper* cutoff frequencies. Another expression for the cutoff frequency is *half-power frequency*, because the power delivered to a resistive load at cutoff frequency is one-half the maximum power delivered to the same load in the pass band. For BP and BS filters, the center frequency is the frequency of maximum or minimum response magnitude, respectively, and bandwidth is the difference between the upper and lower cutoff frequencies. *Rolloff* is the transition from pass band to stop band and is specified in gain unit per frequency unit (e.g., gain unit/Hz, dB/decade, dB/octave, etc.)

11.6.2b Filter Alignment

The *alignment* (or *class*) of a filter refers to the shape of the frequency response [1]. Fundamentally, filter alignment is determined by the coefficients of the filter network transfer function, so there are an indefinite number of filter alignments, some of which may not be realizable. The more common alignments are:

- Butterworth
- Chebyshev
- Bessel
- Inverse Chebyshev
- Elliptic (or Cauer)

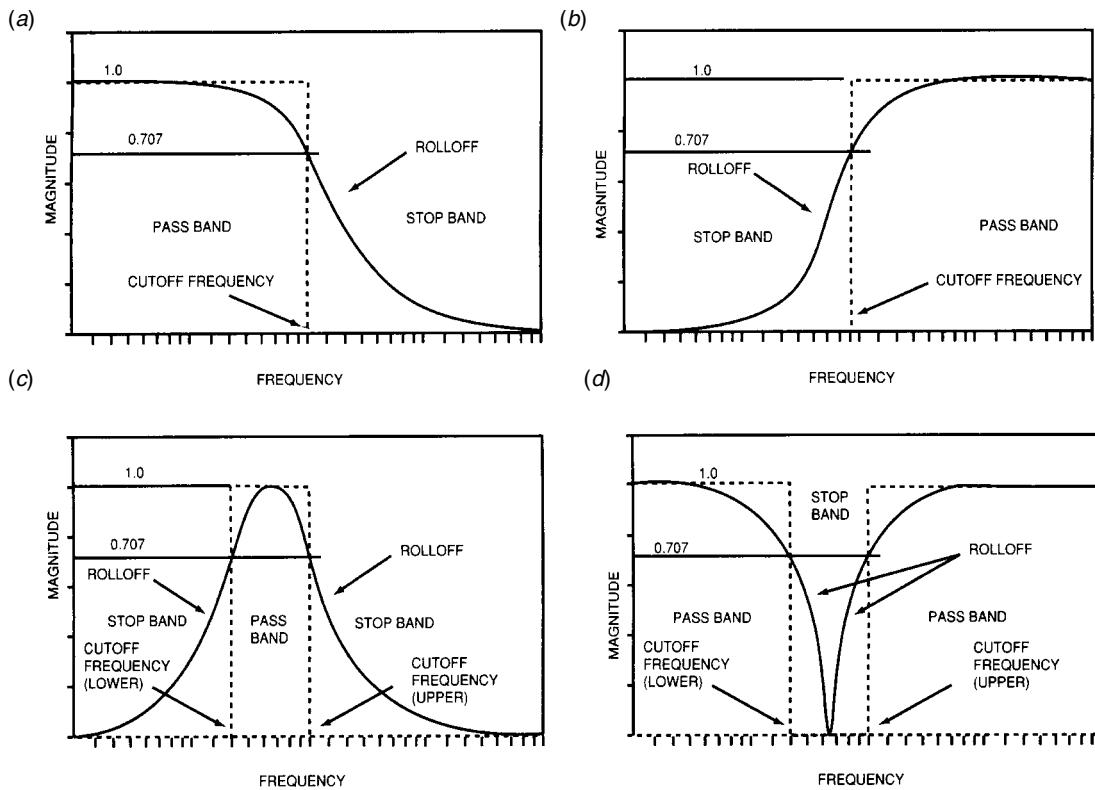


Figure 11.6.1 Filter characteristics by type: (a) low-pass, (b) high-pass, (c) bandpass, (d) band-stop. (From [1]. Used with permission.)

Each filter alignment has a frequency response with a characteristic shape, which provides some particular advantage. (See Figure 11.6.2.) Filters with Butterworth, Chebyshev, or Bessel alignment are called *all-pole filters* because their low-pass transfer functions have no zeros. Table 11.6.1 summarizes the characteristics of the standard filter alignments.

11.6.2c Filter Order

The *order* of a filter is equal to the number of poles in the filter network transfer function [1]. For a lossless *LC* filter with resistive (nonreactive) termination, the number of reactive elements (inductors or capacitors) required to realize a LP or HP filter is equal to the order of the filter. Twice the number of reactive elements are required to realize a BP or a BS filter of the same order. In general, the order of a filter determines the slope of the rolloff—the higher the order, the steeper the rolloff. At frequencies greater than approximately one octave above cutoff (i.e., $f \gg 2f_c$), the rolloff for all-pole filters is $20n$ dB/decade (or approximately $6n$ dB/octave), where

11-130 Broadcast Transmission Systems

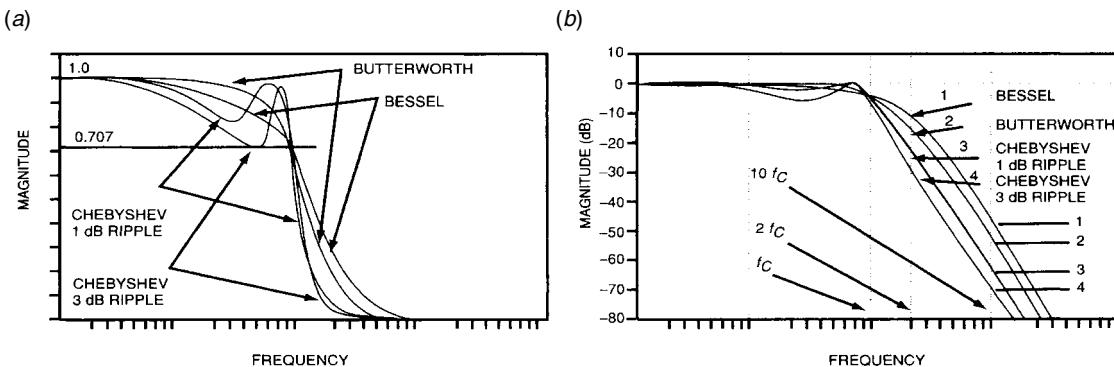


Figure 11.6.2 Filter characteristics by alignment, third-order, all-pole filters: (a) magnitude, (b) magnitude in decibels. (From [1]. Used with permission.)

Table 11.6.1 Summary of Standard Filter Alignments (After [1].)

Alignment	Pass Band Description	Stop Band Description	Comments
Butterworth	Monotonic	Monotonic	All-pole; maximally flat
Chebyshev	Rippled	Monotonic	All-pole
Bessel	Monotonic	Monotonic	All-pole; constant phase shift
Inverse Chebyshev	Monotonic	Rippled	
Elliptic (or Cauer)	Rippled	Rippled	

n is the order of the filter (Figure 11.6.3). In the vicinity of f_c , both filter alignment and filter order determine rolloff.

11.6.3 Four-Port Hybrid Combiner

A hybrid combiner (coupler) is a reciprocal four-port device that can be used for either splitting or combining RF energy over a wide range of frequencies. An exploded view of a typical 3 dB 90° hybrid is illustrated in Figure 11.6.4. The device consists of two identical parallel transmission lines coupled over a distance of approximately one-quarter wavelength and enclosed within a single outer conductor. Ports at the same end of the coupler are in phase, and ports at the opposite end of the coupler are in quadrature (90° phase shift) with respect to each other.

The phase shift between the two inputs or outputs is always 90° and is virtually independent of frequency. If the coupler is being used to combine two signals into one output, these two signals must be fed to the hybrid in phase quadrature. When the coupler is used as a power splitter, the division is equal (half-power between the two ports). The hybrid presents a constant impedance to match each source.

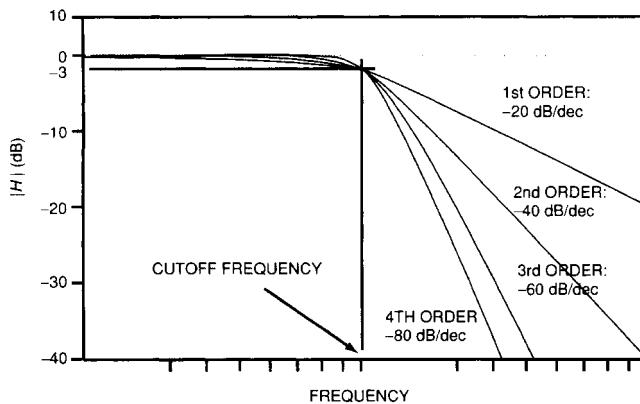


Figure 11.6.3 The effects of filter order on rolloff (Butterworth alignment). (From [1]. Used with permission.)

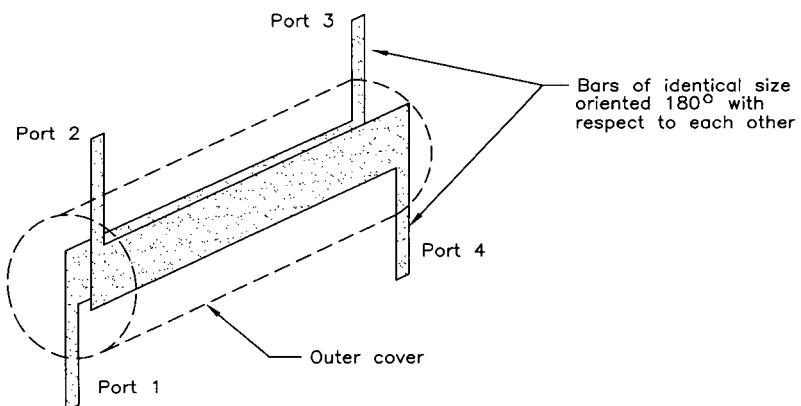


Figure 11.6.4 Physical model of a 90° hybrid combiner.

Operation of the combiner can best be understood through observation of the device in a practical application. Figure 11.6.5 shows a four-port hybrid combiner used to add the outputs of two transmitters to feed a single load. The combiner accepts one RF source and splits it equally into two parts. One part arrives at output port *C* with 0° phase (no phase delay; it is the *reference phase*). The other part is delayed by 90° at port *D*. A second RF source connected to input port *B*, but with a phase delay of 90°, also will split in two, but the signal arriving at port *C* now will be in phase with source 1, and the signal arriving at port *D* will cancel, as shown in the figure.

Output port *C*, the summing point of the hybrid, is connected to the load. Output port *D* is connected to a resistive load to absorb any residual power resulting from slight differences in

11-132 Broadcast Transmission Systems

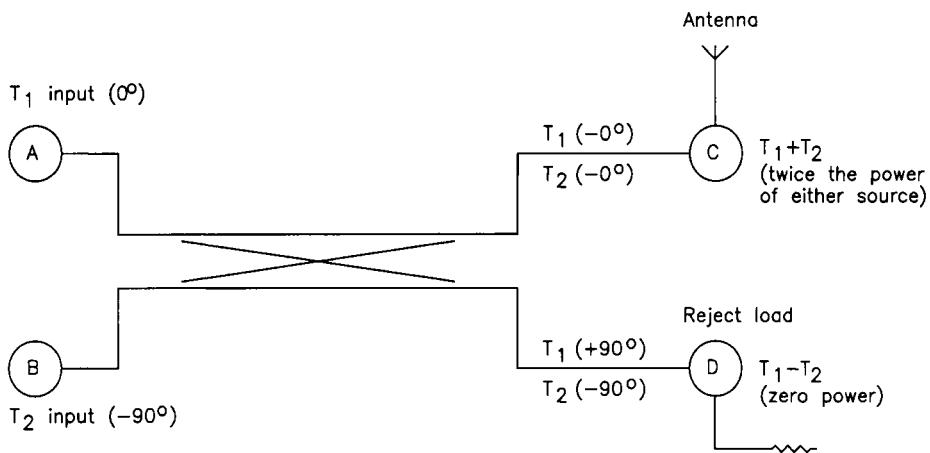


Figure 11.6.5 Operating principles of a hybrid combiner. This circuit is used to add two identical signals at inputs A and B.

amplitude and/or phase between the two input sources. If one of the RF inputs fails, half of the remaining transmitter output will be absorbed by the resistive load at port *D*.

The four-port hybrid works only when the two signals being mixed are identical in frequency and amplitude, and when their relative phase is 90° .

Operation of the hybrid can best be described by a *scattering matrix* in which vectors are used to show how the device operates. Such a matrix is shown in Table 11.6.2. In a 3 dB hybrid, two signals are fed to the inputs. An input signal at port 1 with 0° phase will arrive in phase at port 3, and at port 4 with a 90° lag (-90°) referenced to port 1. If the signal at port 2 already contains a 90° lag (-90° referenced to port 1), both input signals will combine in phase at port 4. The signal from port 2 also experiences another 90° change in the hybrid as it reaches port 3. Therefore, the signals from ports 1 and 2 cancel each other at port 3.

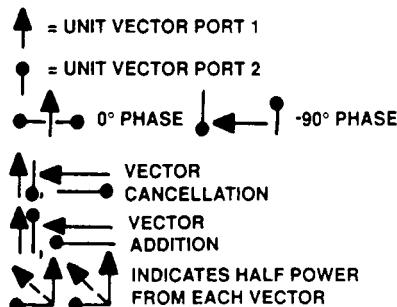
If the signal arriving at port 2 leads by 90° (mode 1 in the table), the combined power from ports 1 and 2 appears at port 4. If the two input signals are matched in phase (mode 4), the output ports (3 and 4) contain one-half of the power from each of the inputs.

If one of the inputs is removed, which would occur in a transmitter failure, only one hybrid input receives power (mode 5). Each output port then would receive one-half the input power of the remaining transmitter, as shown.

The input ports present a predictable load to each amplifier with a VSWR that is lower than the VSWR at the output port of the combiner. This characteristic results from the action of the difference port, typically connected to a dummy load. Reflected power coming into the output port will be directed to the reject load, and only a portion will be fed back to the amplifiers. Figure 11.6.6 illustrates the effect of output port VSWR on input port VSWR, and on the isolation between ports.

Table 11.6.2 Single 90° Hybrid System Operating Modes

MODE	INPUT		SCHEMATIC	OUTPUT	
	1	2		3	4
1	$P_1 / 0^\circ$	$P_2 / -90^\circ$		0	$P_1 + P_2$
2	$P_1 / 0^\circ$	$P_2 / 90^\circ$		$P_1 + P_2$	0
3	$P_1 / 0^\circ$	$P_2 / 0^\circ$		$P_{1/2} + P_{2/2}$	$P_{1/2} + P_{2/2}$
4	$P_1 / 0^\circ$	$P_2 = 0$		$P_{1/2}$	$P_{1/2}$
5	$P_1 = 0$	$P_2 / 0^\circ$		$P_{2/2}$	$P_{2/2}$



As noted previously, if the two inputs from the separate amplifiers are not equal in amplitude and not exactly in phase quadrature, some power will be dissipated in the difference port reject load. Figure 11.6.7 plots the effect of power imbalance, and Figure 11.6.8 plots the effects of

11-134 Broadcast Transmission Systems

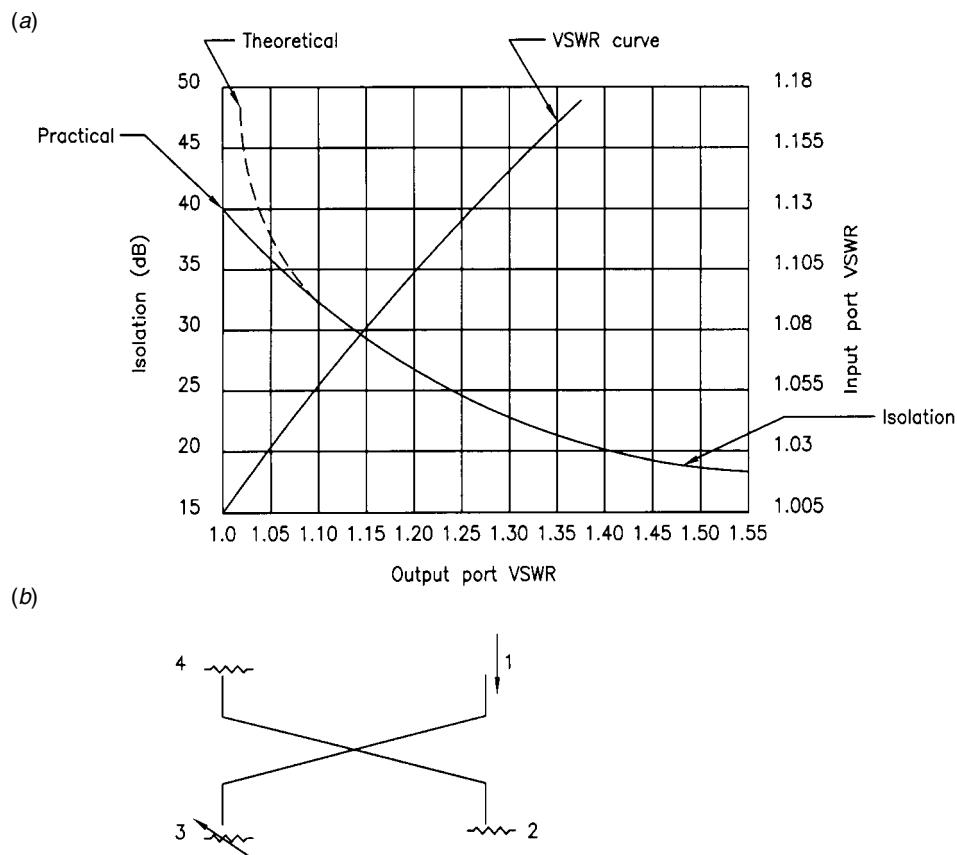


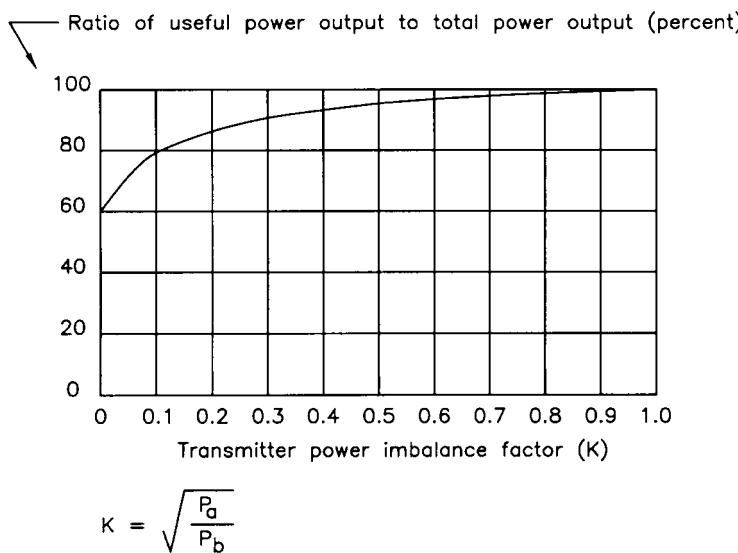
Figure 11.6.6 The effects of load VSWR on input VSWR and isolation: (a) respective curves, (b) coupler schematic.

phase imbalance. The power lost in the reject load can be reduced to a negligible value by trimming the amplitude and/or phase of one (or both) amplifiers.

11.6.3a Microwave Combiners

Hybrid combiners typically are used in microwave amplifiers to add the output energy of individual power modules to provide the necessary output from an RF generator. Quadrature hybrids effect a VSWR-canceling phenomenon that results in well-matched power amplifier inputs and outputs that can be broadbanded with proper selection of hybrid tees. Several hybrid configurations are possible, including the following:

- Split-tee



$$K = \sqrt{\frac{P_a}{P_b}}$$

P_a = Output power of transmitter 1 (lower power system)

P_b = Output power of transmitter 2

Figure 11.6.7 The effects of power imbalance at the inputs of a hybrid coupler.

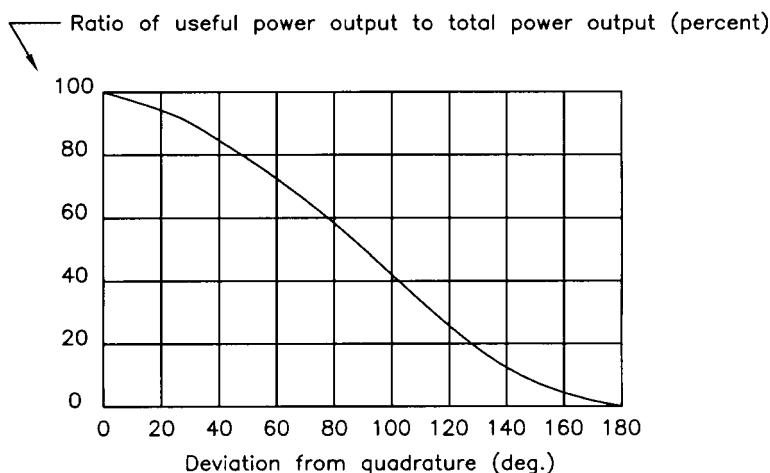


Figure 11.6.8 Phase sensitivity of a hybrid coupler.

11-136 Broadcast Transmission Systems

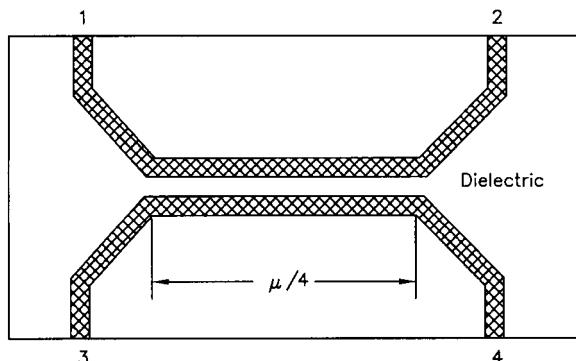


Figure 11.6.9 Reverse-coupled 1/4-wave hybrid combiner.

- Branch-line
- Magic-tee
- Backward-wave

Key design parameters include coupling bandwidth, isolation, and ease of fabrication. The equal-amplitude quadrature-phase reverse-coupled TEM 1/4-wave hybrid is particularly attractive because of its bandwidth and amenability to various physical implementations. Such a device is illustrated in Figure 11.6.9.

11.6.3b Hot Switching Combiners

Switching RF is nothing new. Typically, the process involves coaxial switches, coupled with the necessary logic to ensure that the “switch” takes place with no RF energy on the contacts. This process usually takes the system off-line for a few seconds while the switch is completed. Through the use of hybrid combiners, however, it is possible to redirect RF signals without turning the carrier off. This process is referred to as *hot switching*. Figure 11.6.10 illustrates two of the most common switching functions (SPST and DPDT) available from hot switchers.

The unique phase-related properties of an RF hybrid make it possible to use the device as a switch. The input signals to the hybrid in Figure 11.6.11a are equally powered but differ in phase by 90°. This phase difference results in the combined signals being routed to the output terminal at port 4. If the relative phase between the two input signals is changed by 180°, the summed output then appears on port 3, as shown in Figure 11.6.11b. The 3 dB hybrid combiner, thus, functions as a switch.

This configuration permits the switching of two RF generators to either of two loads. Remember, however, that the switch takes place when the phase difference between the two inputs is 90°. To perform the switch in a useful way requires adding a high-power phase shifter to one input leg of the hybrid. The addition of the phase shifter permits the full power to be combined and switched to either output. This configuration of hybrid and phase shifter, however, will not permit switching a main or standby transmitter to a main or auxiliary load (DPDT function). To accomplish this additional switch, a second hybrid and phase shifter must be added, as shown in Figure 11.6.12. This configuration then can perform the following switching functions:

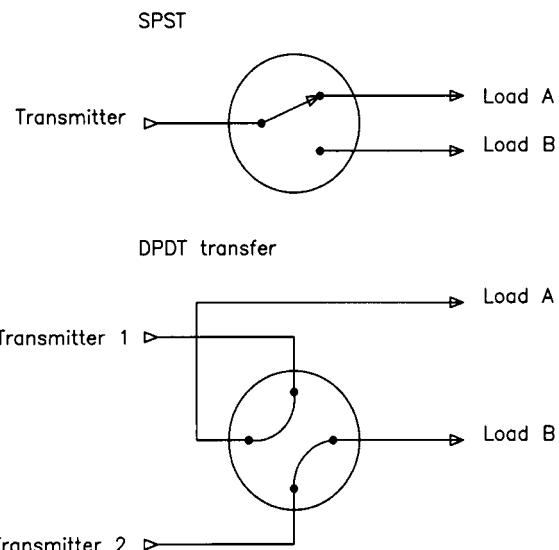


Figure 11.6.10 Common RF switching configurations.

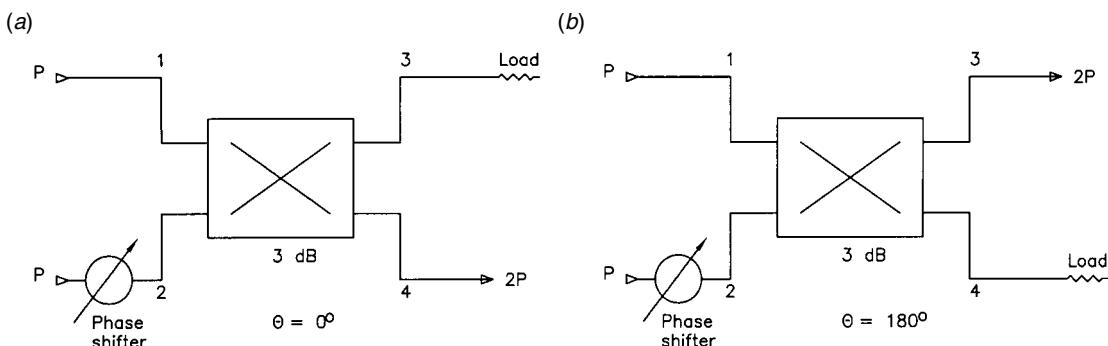


Figure 11.6.11 Hybrid switching configurations: (a) phase set so that the combined energy is delivered to port 4, (b) phase set so that the combined energy is delivered to port 3.

- RF source 1 routed to output *B*
- RF source 2 routed to output *A*
- RF source 1 routed to output *A*
- RF source 2 routed to output *B*

11-138 Broadcast Transmission Systems

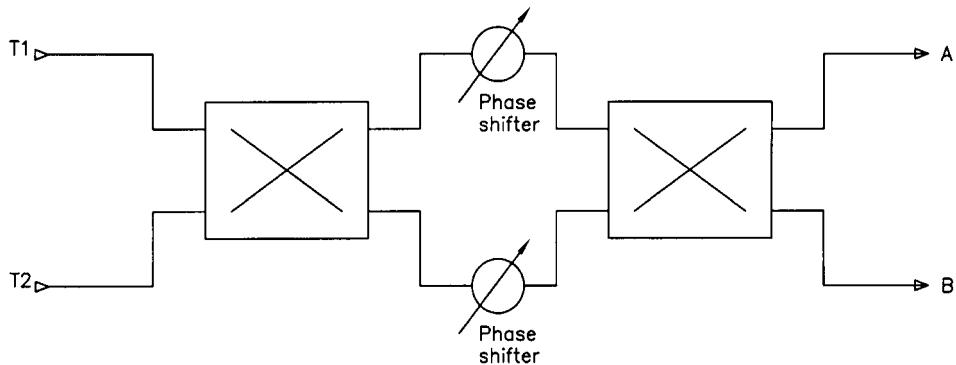


Figure 11.6.12 Additional switching and combining functions enabled by adding a second hybrid and another phase shifter to a hot switching combiner.

The key element in developing such a switch is a high-power phase shifter that does not exhibit reflection characteristics. In this application, the phase shifter allows the line between the hybrids to be electrically lengthened or shortened. The ability to adjust the relative phase between the two input signals to the second hybrid provides the needed control to switch the input signal between the two output ports.

If a continuous analog phase shifter is used, the transfer switch shown in Figure 11.6.12 also can act as a hot switchless combiner where RF generators 1 and 2 can be combined and fed to either output *A* or *B*. The switching or combining functions are accomplished by changing the physical position of the phase shifter.

Note that it does not matter whether the phase shifter is in one or both legs of the system. It is the phase difference ($\theta_1 - \theta_2$) between the two input legs of the second hybrid that is important. With 2-phase shifters, dual drives are required. However, the phase shifter needs only two positions. In a 1-phase shifter design, only a single drive is required, but the phase shifter must have four fixed operating positions.

11.6.4 High-Power Isolators

The high-power ferrite isolator offers the ability to stabilize impedance, isolate the RF generator from load discontinuities, eliminate reflections from the load, and absorb harmonic and intermodulation products. The isolator also can be used to switch between an antenna or load under full power, or to combine two or more generators into a common load.

Isolators commonly are used in microwave transmitters at low power to protect the output stage from reflections. Until recently, however, the insertion loss of the ferrite made use of isolators impractical at high-power levels (25 kW and above). Ferrite isolators are now available that can handle 500 kW or more of forward power with less than 0.1 dB of forward power loss.

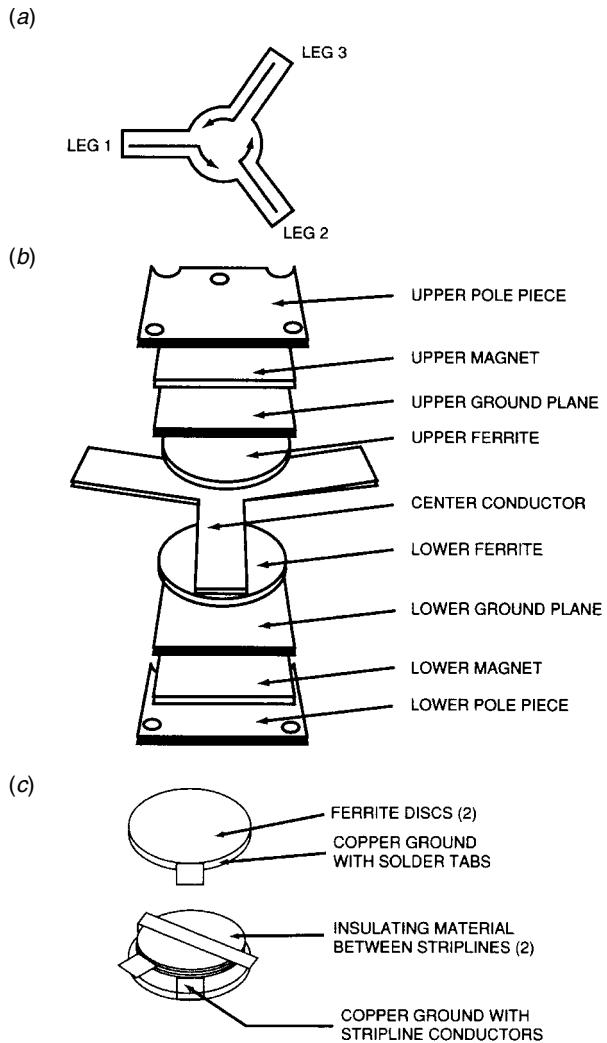


Figure 11.6.13 Basic characteristics of a circulator: (a) operational schematic, (b) distributed constant circulator, (c) lump constant circulator. (From [2]. Used with permission.)

11.6.4a Theory of Operation

High-power isolators are three-port versions of a family of devices known as *circulators*. The circulator derives its name from the fact that a signal applied to one of the input ports can travel in only one direction, as shown in Figure 11.6.13. The input port is isolated from the output port. A signal entering port 1 appears only at port 2; it does not appear at port 3 unless reflected from port 2. An important benefit of this one-way power transfer is that the input VSWR at port 1 is dependent only on the VSWR of the load placed at port 3. In most applications, this load is a resistive (dummy) load that presents a perfect load to the transmitter.

11-140 Broadcast Transmission Systems

The unidirectional property of the isolator results from magnetization of a ferrite alloy inside the device. Through correct polarization of the magnetic field of the ferrite, RF energy will travel through the element in only one direction (port 1 to 2, port 2 to 3, and port 3 to 1). Reversing the polarity of the magnetic field makes it possible for RF flow in the opposite direction. Recent developments in ferrite technology have resulted in high isolation with low insertion loss.

In the basic design, the ferrite is placed in the center of a Y-junction of three transmission lines, either waveguide or coax. Sections of the material are bonded together to form a thin cylinder perpendicular to the electric field. Even though the insertion loss is low, the resulting power dissipated in the cylinder can be as high as 2 percent of the forward power. Special provisions must be made for heat removal. It is efficient heat-removal capability that makes high-power operation possible.

The insertion loss of the ferrite must be kept low so that minimal heat is dissipated. Values of ferrite loss on the order of 0.05 dB have been produced. This equates to an efficiency of 98.9 percent. Additional losses from the transmission line and matching structure contribute slightly to loss. The overall loss is typically less than 0.1 dB, or 98 percent efficiency. The ferrite element in a high-power system is usually water-cooled in a closed-loop path that uses an external radiator.

The two basic circulator implementations are shown in Figures 11.6.13a and 11.6.13b. These designs consist of Y-shaped conductors sandwiched between magnetized ferrite discs [2]. The final shape, dimensions, and type of material varies according to frequency of operation, power handling requirements, and the method of coupling. The *distributed constant circulator* is the older design; it is a broad-band device, not quite as efficient in terms of insertion loss and leg-to-leg isolation, and considerably more expensive to produce. It is useful, however, in applications where broad-band isolation is required. More commonly today is the *lump constant circulator*, a less expensive and more efficient, but narrow-band, design.

At least one filter is always installed directly after an isolator, because the ferrite material of the isolator generates harmonic signals. If an ordinary band-pass or band-reject filter is not to be used, a harmonic filter will be needed.

11.6.4b Applications

The high-power isolator permits a transmitter to operate with high performance and reliability despite a load that is less than optimum. The problems presented by ice formations on a transmitting antenna provide a convenient example. Ice buildup will detune an antenna, resulting in reflections back to the transmitter and high VSWR. If the VSWR is severe enough, transmitter power will have to be reduced to keep the system on the air. An isolator, however, permits continued operation with no degradation in signal quality. Power output is affected only to the extent of the reflected energy, which is dissipated in the resistive load.

A high-power isolator also can be used to provide a stable impedance for devices that are sensitive to load variations, such as klystrons. This allows the device to be tuned for optimum performance regardless of the stability of the RF components located after the isolator. Figure 11.6.14 shows the output of a wideband (6 MHz) klystron operating into a resistive load, and into an antenna system. The power loss is the result of an impedance difference. The periodicity of the ripple shown in the trace is a function of the distance of the reflections from the source.

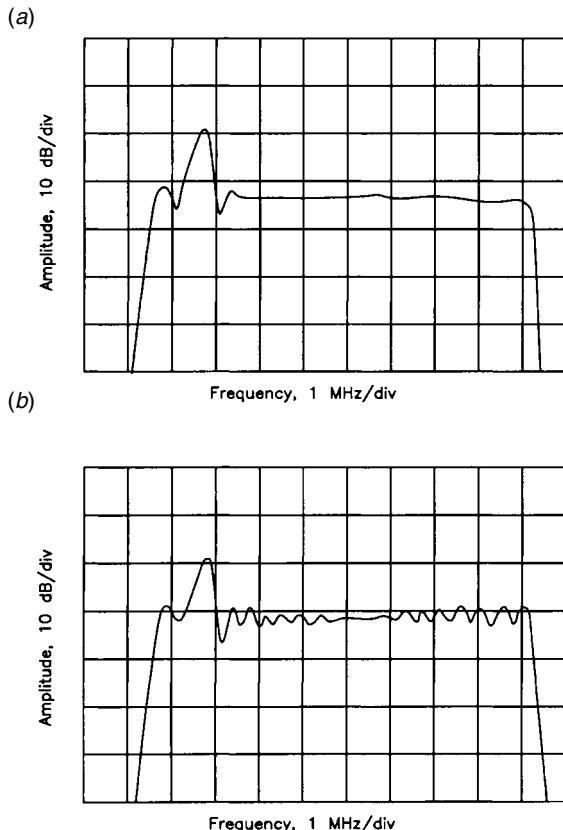


Figure 11.6.14 Output of a klystron operating into different loads through a high-power isolator: (a) resistive load, (b) an antenna system.

Hot Switch

The circulator can be made to perform a switching function if a short circuit is placed at the output port. Under this condition, all input power will be reflected back into the third port. The use of a high-power stub on port 2, therefore, permits redirecting the output of an RF generator to port 3.

At odd 1/4-wave positions, the stub appears as a high impedance and has no effect on the output port. At even 1/4-wave positions, the stub appears as a short circuit. Switching between the antenna and a test load, for example, can be accomplished by moving the shorting element 1/4 wavelength.

Diplexer

An isolator can be configured to combine the aural and visual outputs of a TV transmitter into a single output for the antenna. The approach is shown in Figure 11.6.15. A single notch cavity at the aural frequency is placed on the visual transmitter output (circulator input), and the aural signal is added (as shown). The aural signal will be routed to the antenna in the same manner as it is reflected (because of the hybrid action) in a conventional diplexer.

11-142 Broadcast Transmission Systems

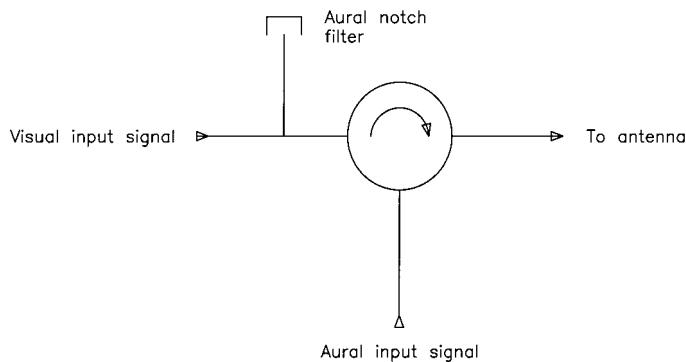


Figure 11.6.15 Use of a circulator as a diplexer in TV applications.

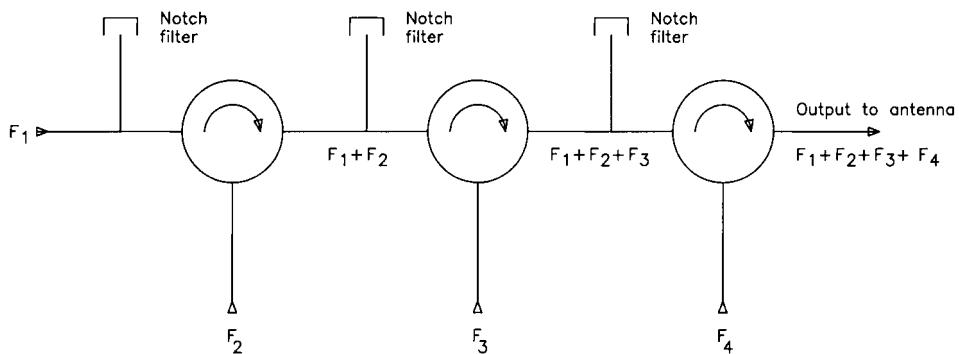


Figure 11.6.16 Use of a bank of circulators in a multiplexer application.

Multiplexer

A multiplexer can be formed by cascading multiple circulators, as illustrated in Figure 11.6.16. Filters must be added, as shown. The primary drawback of this approach is the increased power dissipation that occurs in circulators nearest the antenna.

11.6.5 References

1. Harrison, Cecil: "Passive Filters," in *The Electronics Handbook*, Jerry C. Whitaker (ed.), CRC Press, Boca Raton, Fla., pp. 279–290, 1996.

2. Surette, Robert A.: "Combiners and Combining Networks," in *The Electronics Handbook*, Jerry C. Whitaker (ed.), CRC Press, Boca Raton, Fla., pp. 1368–1381, 1996.

11.6.6 Bibliography

- DeComier, Bill: "Inside FM Multiplexer Systems," *Broadcast Engineering*, Intertec Publishing, Overland Park, Kan., May 1988.
- Heymans, Dennis: "Hot Switches and Combiners," *Broadcast Engineering*, Overland Park, Kan., December 1987.
- Stenberg, James T.: "Using Super Power Isolators in the Broadcast Plant," *Proceedings of the Broadcast Engineering Conference*, Society of Broadcast Engineers, Indianapolis, IN, 1988.
- Vaughan, T., and E. Pivit: "High Power Isolator for UHF Television," *Proceedings of the NAB Engineering Conference*, National Association of Broadcasters, Washington, D.C., 1989.

Section**12****Radio Receivers**

The development of radio transmission and reception must be considered one of the major technical achievements of the twentieth century. The impact of voice broadcasts to the public, whether by commercial stations or government-run organizations, has expanded the horizons of everyday citizens in virtually every country on earth. It is hard to overestimate the power of radio broadcasting.

Technology has dramatically reshaped the transmission side of AM and FM broadcasting. Profound changes have also occurred in receiver technology. Up until 1960, radio broadcasting was basically a stationary medium. The receivers of that time were physically large and heavy, and required 120 V ac power to drive them. The so-called portable radios of the day relied on bulky batteries that offered only a limited amount of listening time. Automobile radios incorporated vibrator-choppers to simulate ac current. All the receivers available for commercial use during the 1940s and 1950s used vacuum tubes exclusively.

The first technical breakthrough for radio broadcasting was the transistor, available commercially at reasonable prices during the early 1960s. The transistor brought with it greatly reduced physical size and weight, and even more importantly, it eliminated the necessity of ac line current to power the radio. The first truly portable AM radios began to appear during the early 1960s, with AM-FM versions following by the middle of the decade.

Many of the early receiver designs were marginal from a performance stand-point. The really good receivers were still built around vacuum tubes. As designers learned more about transistors, and as better transistors became available, tube-based receivers began to disappear. By 1970, *transistorized* radios, as they were called, commanded the consumer market.

The integrated circuit (IC) was the second technical breakthrough in consumer receiver design. This advance, more than anything else, made high-quality portable radios possible. It also accelerated the change in listening habits from AM to FM. IC-based receivers allowed designers to put more sophisticated circuitry into a smaller package, permitting consumers to enjoy the benefits of FM broadcasting without the penalties of the more complicated receiver required for FM stereo.

The move to smaller, lighter, more power-efficient radios has led to fundamental changes in the way radios are built and serviced. In the days of vacuum-tube and transistor-based receivers, the designer would build a radio out of individual stages that interconnected to provide a working unit. The stages for a typical radio included:

12-2 Section Twelve

- RF amplifier
- Local oscillator
- Intermediate frequency (IF) amplifier
- Detector and audio preamplifier

Today, however, large-scale integration (LSI) or even very large scale integration (VLSI) techniques have permitted virtually all the active circuits of an AM-FM radio to be placed on a single IC. Advanced circuitry has also permitted radio designers to incorporate all-electronic tuning, eliminating troublesome and sometimes expensive mechanical components. Electronically tuned radios (ETRs) have made features such as "station scan" and "station seek" possible. Some attempts were made to incorporate scan and seek features in mechanically tuned radios, but the results were never very successful.

The result of LSI-based receiver design has been twofold. First, radios based on advanced chip technology are much easier to build and are, therefore, usually less expensive to consumers. Second, such radios are not serviceable. Most consumers today would not bother to have a broken radio repaired. They would simply buy a new one and throw the old one away.

Still, however, it is important to know what makes a radio work. Although radios being built with LSI and VLSI technology do not lend themselves to stage-by-stage troubleshooting as earlier radios did, it is important to know how each part of the system functions to make a working unit. Regardless of the sophistication of a VLSI-based receiver, the basic principles of operation are the same as a radio built of discrete stages.

In This Section:

Chapter 12.1: Receiver Characteristics	12-7
Introduction	12-7
Practical Receivers	12-8
The Receiving System	12-9
Gain, Sensitivity, and Noise Figure	12-10
NF	12-10
Minimum Detectable Signal	12-12
Selectivity	12-13
Dynamic Range	12-14
Desensitization	12-16
AM Cross Modulation	12-16
IM	12-17
Error Vector Magnitude	12-21
Gain Control	12-21
Digital Receiver Characteristics	12-25
BER Testing	12-25
Transmission and Reception Quality	12-26
Bibliography	12-27
Chapter 12.2: The Radio Channel	12-29
Introduction	12-29
Channel Impulse Response	12-30

Doppler Effect	12-34
Transfer Function	12-37
Time Response of Channel Impulse Response & Transfer Function	12-39
References	12-40
Chapter 12.3: AM and FM Receivers	12-41
Introduction	12-41
Superheterodyne Receiver	12-41
Radio Wave Propagation	12-43
AM Band Propagation	12-43
FM Band Propagation	12-43
Radio Receivers	12-44
Antenna Systems	12-45
Antenna Coupling Network	12-45
Whip Antenna	12-45
Loop Antenna	12-46
Filter Types	12-47
LC Filter	12-48
Quartz Crystal Filter	12-48
Monolithic Quartz Filter	12-48
Ceramic Filter	12-49
RF Amplifier and AGC	12-49
The AGC Loop	12-50
Mixer	12-51
Passive Mixer	12-52
Active Mixer	12-53
Local Oscillator	12-54
PLL Synthesizer	12-55
Frequency Divider	12-56
Variable-Frequency Oscillator	12-56
Diode Switching	12-57
Crystal-Controlled Oscillator	12-59
AM-FM Demodulation	12-59
AM Demodulation	12-62
FM Demodulation	12-64
Amplitude Limiter	12-66
Stereo Systems	12-68
FM Stereo	12-68
Generating the Stereo Signal	12-69
Decoding the Stereo Signal	12-70
AM Stereo	12-71
Decoding the C-QUAM Signal	12-71
References	12-73
Bibliography	12-73
Chapter 12.4: Stereo Television	12-75
Introduction	12-75
Audio Chain	12-75

12-4 Section Twelve

Stereo Generator	12-75
Subchannel Filters	12-78
Audio System Specifications and Performance Objectives	12-78
Frequency Response	12-80
Signal-to-Noise Ratio (SNR)	12-80
Total Harmonic Distortion (THD)	12-80
Intermodulation Distortion (IMD)	12-80
Separation	12-81
Headroom	12-81
Channel-to-Channel Amplitude and Phase Tracking	12-81
Stereo Generator Specifications and Performance Objectives	12-81
Separation	12-81
Linear Crosstalk	12-82
Nonlinear Crosstalk	12-82
Spurious Suppression	12-82
Deviation Calibration	12-83
Composite STL Specifications and Performance Objectives	12-83
Signal-to-Noise Ratio	12-83
Distortion	12-83
Amplitude Stability	12-84
Aural Exciter Specifications and Performance Objectives	12-84
Signal-to-Noise Ratio	12-84
Amplitude Stability	12-84
References	12-84

Reference Documents for this Section

- Amos, S. W.: "FM Detectors," *Wireless World*, vol. 87, no. 1540, pg. 77, January 1981.
- Benson, K. Blair, and Jerry C. Whitaker: *Television and Audio Handbook for Engineers and Technicians*, McGraw-Hill, New York, N.Y., 1990.
- Engelson, M., and J. Herbert: "Effective Characterization of CDMA Signals," *Microwave Journal*, pg. 90, January 1995.
- Howald, R.: "Understand the Mathematics of Phase Noise," *Microwaves & RF*, pg. 97, December 1993.
- Johnson, J. B.: "Thermal Agitation of Electricity in Conduction," *Phys. Rev.*, vol. 32, pg. 97, July 1928.
- Nyquist, H.: "Thermal Agitation of Electrical Charge in Conductors," *Phys. Rev.*, vol. 32, pg. 110, July 1928.
- Pleasant, D.: "Practical Simulation of Bit Error Rates," *Applied Microwave and Wireless*, pg. 65, Spring 1995.
- Rohde, Ulrich L.: *Digital PLL Frequency Synthesizers*, Prentice-Hall, Englewood Cliffs, N.J., 1983.
- Rohde, Ulrich L.: "Key Components of Modern Receiver Design—Part 1," *QST*, pg. 29, May 1994.

- Rohde, Ulrich L. Rohde and David P. Newkirk: *RF/Microwave Circuit Design for Wireless Applications*, John Wiley & Sons, New York, N.Y., 2000.
- Rohde, Ulrich L., and Jerry C. Whitaker: *Communications Receivers*, 3rd ed., McGraw-Hill, New York, N.Y., 2000.
- “Standards Testing: Bit Error Rate,” application note 3SW-8136-2, Tektronix, Beaverton, OR, July 1993.
- Using Vector Modulation Analysis in the Integration, Troubleshooting and Design of Digital RF Communications Systems*, Product Note HP89400-8, Hewlett-Packard, Palo Alto, Calif., 1994.
- Watson, R.: “Receiver Dynamic Range; Pt. 1, Guidelines for Receiver Analysis,” *Microwaves & RF*, vol. 25, pg. 113, December 1986.
- “Waveform Analysis: Noise and Jitter,” application note 3SW8142-2, Tektronix, Beaverton, OR, March 1993.
- Wilson, E.: “Evaluate the Distortion of Modular Cascades,” *Microwaves*, vol. 20, March 1981.
- Whitaker, Jerry C. (ed.): *NAB Engineering Handbook*, 9th ed., National Association of Broadcasters, Washington, D.C., 1999.

Chapter

12.1

Receiver Characteristics

Ulrich L. Rohde

Jerry C. Whitaker, Editor-in-Chief

12.1.1 Introduction¹

The superheterodyne receiver makes use of the heterodyne principle of mixing an incoming signal with a signal generated by a local oscillator (LO) in a nonlinear element (Figure 12.1.1). However, rather than synchronizing the frequencies, the superheterodyne receiver uses a LO frequency offset by a fixed intermediate frequency (IF) from the desired signal. Because a nonlinear device generates identical difference frequencies if the signal frequency is either above or below the LO frequency (and also a number of other spurious responses), it is necessary to provide sufficient filtering prior to the mixing circuit so that this undesired signal response (and others) is substantially suppressed. The frequency of the undesired signal is referred to as an *image frequency*, and a signal at this frequency is referred to as an *image*. The image frequency is separated from the desired signal frequency by a difference equal to twice the IF. The preselection filtering required at the signal frequency is much broader than if the filtering of adjacent channel signals were required. The channel filtering is accomplished at IF. This is a decided advantage when the receiver must cover a wide frequency band, because it is much more difficult to maintain constant bandwidth in a tunable filter than in a fixed one. Also, for receiving different signal types, the bandwidth can be changed with relative ease at a fixed frequency by switching filters of different bandwidths. Because the IF at which channel selectivity is provided is often lower than the signal band frequencies, it may be easier to provide selectivity at IF, even if wide-band RF tuning is not required.

1. This chapter is based on: Rohde, Ulrich L., and Jerry C. Whitaker: *Communications Receivers: Principles and Design*, 3rd ed., McGraw-Hill, New York, N.Y., 2000. Used with permission.

12-8 Broadcast Receiver Systems

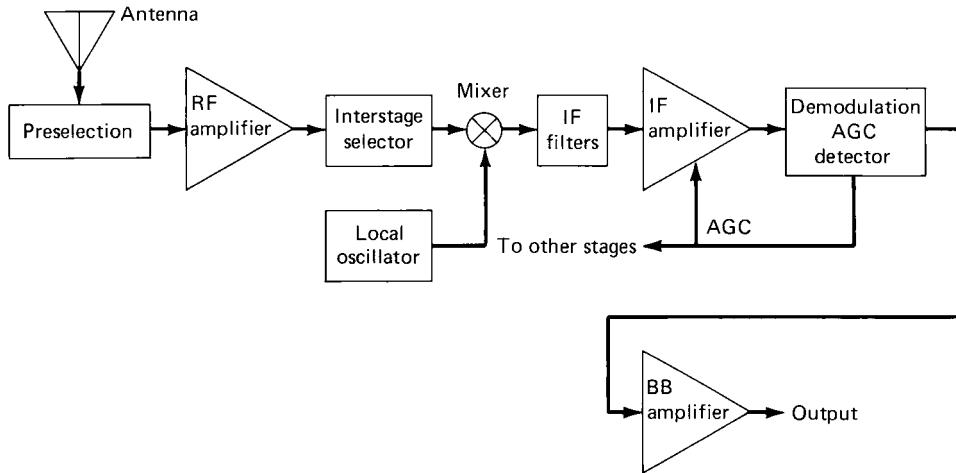


Figure 12.1.1 Block diagram of a superheterodyne receiver.

12.1.1a Practical Receivers

Because of the nature of active electronic devices, it is generally easier to provide high stable gain in a fixed-frequency amplifier than in a tunable one, and gain is generally more economical at lower frequencies. Thus, although the superheterodyne receiver does introduce a problem of spurious responses not present in the other receiver types, its advantages are such that it has replaced other types except for special applications.

Referring again to Figure 12.1.1, the signal is fed from the antenna to a preselector filter and amplifier. The input circuit is aimed at matching the antenna to the first amplifying device so as to achieve the best sensitivity while providing sufficient selectivity to reduce the probability of overload from strong undesired signals in the first amplifier. Losses from the antenna coupling circuit and preselection filters decrease the sensitivity. Because sufficient selectivity must be provided against the image and other principal spurious responses prior to the mixing circuit, the preselection filtering is often broken into two or more parts with intervening amplifiers to minimize the effects of the filter loss on the *noise factor* (NF). The LO provides a strong stable signal at the proper frequency to convert the signal frequency to IF. This conversion occurs in the mixer. (This element has also been referred to as the *first detector*, converter, or *frequency changer*.)

The output from the mixer is applied to the IF amplifier, which amplifies the signal to a suitable power level for the demodulator. This circuit derives from the IF signal the modulation signal, which may be amplified by the baseband amplifier to the level required for output. Normally, the output of an audio amplifier is fed to a headphone or loudspeaker at the radio, or coupled to a transmission line for remote use. A video signal requires development of sweep, intensity, and usually color signals from the amplified video demodulation prior to display. In other cases, the output may be supplied to a data demodulator to produce digital data signals from the baseband signal. The data demodulator may be part of the receiver or may be provided separately as part of a data modem. The data modem may also be fed directly from the receiver at IF. Data demodulation is typically accomplished using digital processing circuits rather than ana-

log demodulators and amplifiers. In this case, the IF amplifier must be designed to provide the appropriate level to an A/D converter so that digital processing can be carried out. Additional IF filtering, data demodulation, and error control coding can all be performed by digital circuits or a microprocessor, either in the receiver or as part of an external modem.

An alternative to IF sampling and A/D conversion is the conversion of the signal to baseband in two separate *coherent demodulators* driven by quadrature LO signals at the IF. The two outputs are then sampled at the appropriate rate for the baseband by two A/D converters or a single multiplexed A/D converter, providing the in-phase and quadrature samples of the baseband signal. Once digitized, these components can be processed digitally to provide filtering, frequency changing, phase and timing recovery, data demodulation, and error control.

12.1.2 The Receiving System

The first essential function of any radio receiver is to effect the transfer of energy picked up by the antenna to the receiver itself through the input circuits. Maximum energy is transferred if the impedance of the input circuit matches that of the antenna (inverse reactance and same resistance) throughout the frequency band of the desired signal. This is not always feasible, and the best energy transfer is not essential in all cases. A receiver may also be connected with other receivers through a *hybrid* or *active multicoupler* to a single antenna. Such arrangements are sometimes very sensitive to mismatches.

There are at least three antenna matching problems in a receiver. The first and, in many cases, most crucial problem is that the receiver may be used from time to time with different antennas whose impedances the potential users cannot specify fully. Second, antennas may be used in mobile applications or in locations subject to changing foliage, buildings, or waves at sea, so that the impedance—even if measured accurately at one time—is subject to change from time to time. Third, at some frequencies, the problems of matching antennas are severely limited by available components, and the losses in a matching network may prove greater than for a simpler lower-loss network with poorer match.

When antenna matching is important over a wide band, it may be necessary to design a network that can be tuned mechanically or electrically under either manual or automatic control in response to a performance measure in the system. In older receivers with a wide tuning range, it was common to have a mechanically tuned preselector that could be adjusted by hand and was generally connected directly to the *variable-frequency oscillator* (VFO) used for the first conversion. At times a trimmer was used in the first circuit to compensate for small antenna mismatches. Thus, tuning of the circuit could be modified to match the effects of the expected antenna impedance range. Modern wide tuning receivers often use one-half-octave switchable filters in the preselector, which may be harder to match, but are much easier to control by computer. Similarly, the first oscillator is generally a microprocessor-controlled synthesizer.

Often the problem of antenna matching design is solved by the user specification that defines one or more “dummy antenna” impedances to be used with a signal generator to test the performance of the receiver for different receiver input circuits. In this case, the user’s system is designed to allow for the mismatch losses in performance that result from the use of actual antennas. When it is necessary to measure receiver input impedance accurately, it is best accomplished through a network analyzer.

12-10 Broadcast Receiver Systems

A number of other receiver input considerations may occur in certain cases. The input circuits may be balanced or unbalanced, or may need to be connectable either way. The input circuits may require grounding, isolation from ground, or either connection. The circuits may need protection from high-voltage discharges or from impulses. They may need to handle, without destruction, high-power nearby signals, both tuned to the receiver frequency and off-tune. Thus, the input circuit can—at times—present significant design challenges.

12.1.2a Gain, Sensitivity, and Noise Figure

Any given receiver is usually required to receive and process a wide range of signal powers, but in most cases it is important that they be capable of receiving distant signals whose power has been attenuated billions of times during transmission. The extent to which such signals can be received usefully is determined by the noise levels received by the antenna (natural and man-made) and those generated within the receiver itself. It is also necessary that the receiver produce a level of output power suitable for the application. Generally the ratio of the output power of a device to the input power is known as the *gain*. The design of a receiver includes gain distribution among the various stages so as to provide adequate receiver gain and an optimum compromise among the other operating characteristics.

While there are limits to the amount of gain that can be achieved practically at one frequency because of feedback, modern receivers need not be gain-limited. When the gain is sufficiently high, the weakest signal power that may be processed satisfactorily is *noise-limited*. This signal level is referred to as the *sensitivity* of the system at a particular time and varies depending on the external noise level. It is possible in some systems for the external noise to fall sufficiently so that the system sensitivity is established by the internal noise of the receiver. A receiver's sensitivity is one of its most important characteristics. There are no universal standards for its measurement, although standards have been adopted for specific applications and by specific user groups. Figure 12.1.2 shows a block diagram of the test setup and the typical steps involved in determining receiver sensitivity.

NF

Sensitivity measures depend upon specific signal characteristics. The NF measures the effects of inherent receiver noise in a different manner. Essentially it compares the total receiver noise with the noise that would be present if the receiver generated no noise. This ratio is sometimes called the noise factor F , and when expressed in dB, the NF. F is also defined equivalently as the ratio of the S/N of the receiver output to the S/N of the source. The source generally used to test receivers is a signal generator at local room temperature. An antenna, which receives not only signals but noises from the atmosphere, the galaxy, and man-made sources, is unsuitable to provide a measure of receiver NF. However, the NF required of the receiver from a system viewpoint depends on the expected S/N from the antenna. The effects of external noise are sometimes expressed as an equivalent antenna NF.

For the receiver, we are concerned with internal noise sources. Passive devices such as conductors generate noise as a result of the continuous thermal motion of the free electrons. This type of noise is referred to generally as *thermal noise*, and is sometimes called *Johnson noise* after the person who first demonstrated it. Using the statistical theory of thermodynamics, Nyquist showed that the mean-square thermal noise voltage generated by any impedance between two frequencies f_1 and f_2 can be expressed as

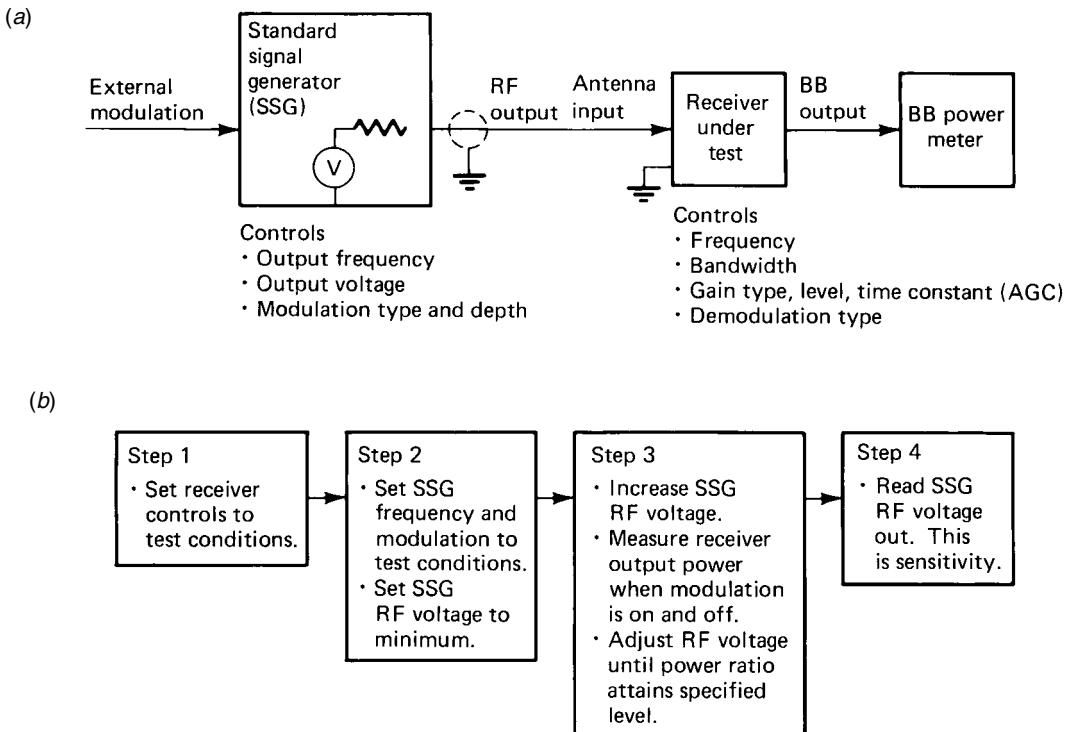


Figure 12.1.2 Receiver sensitivity measurement: (a) test setup, (b) procedure.

$$\overline{V_n}^2 = 4 k t \int_{f_1}^{f_2} R(f) d f \quad (12.1.1)$$

where $R(f)$ is the resistive component of the impedance.

Magnetic substances also produce noise, depending upon the residual magnetization and the applied dc and RF voltages. This is referred to as the *Barkhausen effect*, or *Barkhausen noise*. The greatest source of receiver noise, however, is generally that generated in semiconductors. Like the older thermionic tubes, transistors and diodes also produce characteristic noise. *Shot noise* resulting from the fluctuations in the carrier flow in semiconductor devices produces wide-band noise, similar to thermal noise. Low-frequency noise or $1/f$ noise, also called *flicker effect*, is roughly inversely proportional to frequency and is similar to the contact noise in contact resistors. All of these noise sources contribute to the “excess noise” of the receiver, which causes the NF to exceed 0 dB.

The NF is often measured in a setup similar to that of Figure 12.1.2, using a specially designed and calibrated white-noise generator as the input. The receiver is tuned to the correct frequency and bandwidth, and the output power meter is driven from a linear demodulator or the

12-12 Broadcast Receiver Systems

final IF amplifier. The signal generator is set to produce no output, and the output power is observed. The generator output is then increased until the output has risen 3 dB. The setting on the generator is the NF in decibels.

The NF of an amplifier can also be calculated as the ratio of input to output S/N, per the equation

$$NF = 10 \log \left[\frac{(S/N)_1}{(S/N)_0} \right] \quad (12.1.2)$$

where NF is the noise figure in dB and $(S/N)_1$ and $(S/N)_0$ are the amplifier input and output SNR, respectively.

The NF for a noiseless amplifier or lossless passive device is 0 dB; it is always positive for nonideal devices. The NF of a lossy passive device is numerically equal to the device insertion loss. If the input of a nonideal amplifier of gain G (dB) and noise figure NF (dB) were connected to a matched resistor, the amplifier output noise power P_{No} (dB) would be

$$P_{No} = 10 \log (kT) + 10 \log (B) + G + NF \quad (12.1.3)$$

where k is Boltzmann's constant ($\text{mW}/\text{°K}$), T is the resistor temperature in °K , and B is the noise bandwidth in Hz.

When amplifiers are cascaded, the noise power rises toward the output as noise from succeeding stages is added to the system. Under the assumption that noise powers add noncoherently, the noise figure NF_T of a cascade consisting of two stages of numerical gain A_1 and A_2 and noise factor N_1 and N_2 , is given by Friis' equation

$$NF_T = 10 \log \left[\frac{N_1 + (N_2 - 1)}{A_1} \right] \quad (12.1.4)$$

where the noise factor is $N = 10^{(NF/10)}$ and the numerical gain is $A = 10^{(G/10)}$. The system NF, therefore, is largely determined by the first stage NF when A_1 is large enough to make $(N_2 - 1)/A_1$ much smaller than N_1 .

Minimum Detectable Signal

Another measure of sensitivity is the *minimum detectable signal* (MDS). The measurement procedure is similar to the NF measurement except that a sinusoidal signal generator replaces the noise generator to produce the doubling of output power over noise power alone. This signal

power, just equal to the noise power, is defined as the MDS. Because receiver noise varies with bandwidth, so does the MDS, which can be expressed as

$$MDS = k T B_n F \quad (12.1.5)$$

In dBm, $MDS = -174 + 10\log(B_n + NF)$, where B_n is the noise bandwidth of the receiver. (0 dbm = decibels referenced to 1mW.)

The available thermal noise power per hertz is -174 dBm at 290°K (63°F), an arbitrary reference temperature near standard room temperatures. When any two of the quantities in the expression are known, the third may be calculated. As in the case of NF measurements, care is required in measuring MDS, because a large portion of the power being measured is noise, which produces MDS' typical fluctuations.

12.1.3 Selectivity

Selectivity is the property of a receiver that allows it to separate a signal or signals at one frequency from those at all other frequencies. At least two characteristics must be considered simultaneously in establishing the required selectivity of a receiver. The selective circuits must be sufficiently sharp to suppress the interference from adjacent channels and spurious responses. On the other hand, they must be broad enough to pass the highest sideband frequencies with acceptable distortion in amplitude and phase. Each class of signals to be received may require different selectivity to handle the signal sidebands adequately while rejecting interfering transmissions having different channel assignment spacings. However, each class of signal requires about the same selectivity throughout all the frequency bands allocated to that class of service. Older receivers sometimes required broader selectivity at their higher frequencies to compensate for greater oscillator drift. This requirement has been greatly reduced by the introduction of synthesizers for control of LOs and economical high-accuracy and high-stability crystal standards for the reference frequency oscillator. Consequently, except at frequencies above VHF, or in applications where adequate power is not available for temperature-controlled ovens, only the accuracy and stability of the selective circuits themselves may require selectivity allowances today.

Quantitatively the definition of selectivity is the bandwidth for which a test signal x decibels stronger than the minimum acceptable signal at a nominal frequency is reduced to the level of that signal. This measurement is relatively simple for a single selective filter or single-frequency amplifier, and a selectivity curve can be drawn showing the band offset both above and below the nominal frequency as the selected attenuation level is varied. Ranges of 80 to 100 dB of attenuation can be measured readily, and higher ranges—if required—can be achieved with special care. A test setup similar to Figure 12.1.2 may be employed with the receiver replaced by the selective element under test. Proper care must be taken to achieve proper input and output impedance termination for the particular unit under test. The power output meter need only be sufficiently sensitive, have uniform response over the test bandwidth, and have monotonic response so that the same output level is achieved at each point on the curve. A typical IF selectivity curve is shown in Figure 12.1.3.

12-14 Broadcast Receiver Systems

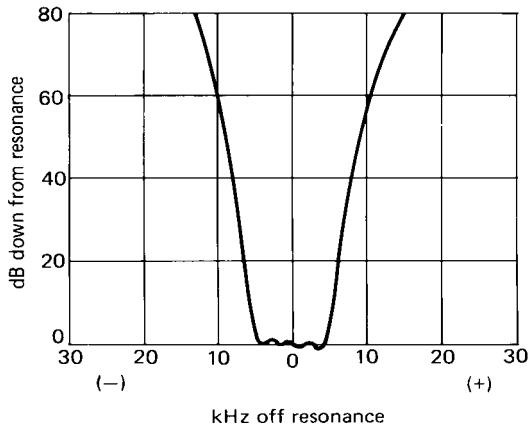


Figure 12.1.3 Example of an IF selectivity curve.

The measurement of overall receiver selectivity, using the test setup of Figure 12.1.2, presents some difficulties. The total selectivity of the receiving system is divided among RF, IF, and baseband selective elements. There are numerous amplifiers and frequency converters, and at least one demodulator intervening between input and output. Hence, there is a high probability of nonlinearities in the nonpassive components affecting the resulting selectivity curves. Some of the effects that occur include overload, modulation distortion, spurious signals, and spurious responses. If there is an AGC, it must be disabled so that it cannot change the amplifier gain in response to the changing signal levels in various stages of the receiver. If there is only an AM or FM demodulator for use in the measurement, distortions occur because of the varying attenuation and phase shift of the circuits across the sidebands.

When measuring complete receiver selectivity, with either a modulated or nonmodulated signal, it is wise to use an output power meter calibrated in decibels. The measurement proceeds as described previously. However, if any unusual changes in attenuation or slope are observed, the generator level may be increased in calibrated steps; it should be noted whether the output changes decibel for decibel. If not, what is being observed at this point is not the selectivity curve, but one of the many nonlinearities or responses of the system.

12.1.4 Dynamic Range

The term *dynamic range*, especially in advertising literature, has been used to mean a variety of things. We must be especially careful in using a common definition when comparing this characteristic of receivers. In some cases, the term has been used to indicate the ratio in decibels between the strongest and weakest signals that a receiver could handle with acceptable noise or distortion. This is the ratio between the signal that is so strong that it causes maximum tolerable distortion and the one that is so weak that it has the minimum acceptable S/N. This measure is of limited value in assessing performance in the normal signal environment where the desired sig-

nal may have a range of values, but is surrounded by a dense group of other signals ranging from very weak to very strong. The selective circuits of a receiver can provide protection from many of these signals. However, the stronger ones, because of the nonlinearity of the active devices necessary to provide amplification and frequency conversion, can degrade performance substantially. In modern parlance, dynamic range refers to the ratio of the level of a strong out-of-band signal that in some manner degrades signal performance of the receiver to a very weak signal. The most common weak signal considered is the MDS, and differing strong degrading signal levels may be used. It is, therefore, important to know which definition is meant when evaluating the meaning of the term *dynamic range*.

If the foregoing discussion of dynamic range seems vague, it is because there is not one characteristic but several that is encompassed by the term. Each may have a different numeric value. A receiver is a complex device with many active stages separated by different degrees of selectivity. The response of a receiver to multiple signals of different levels is extremely complex, and the results do not always agree with simple theory. However, such theory provides useful comparative measures. If we think of an amplifier or mixer as a device whose output voltage is a function of the input voltage, we may expand the output voltage in a power series of the input voltage

$$V_o = \sum a_n V_i^n \quad (12.1.6)$$

where a_1 is the voltage amplification of the device and the higher-order a_n cause distortion.

Because the desired signal and the undesired interference are generally narrow-band signals, we may represent V_i as a sum of sinusoids of different amplitudes and frequencies. Generally $(A_1 \sin 2\pi f_1 t + A_2 \sin 2\pi f_2 t)^n$, as a result of trigonometric identities, produces a number of components with different frequencies, $mf_1 \pm (n-m)f_2$, with m taking on all values from 0 to n . These intermodulation (IM) products may have the same frequency as the desired signal for appropriate choices of f_1 and f_2 . When n is even, the minimum difference between the two frequencies for this to happen is the desired frequency itself. This type of *even* IM interference can be reduced substantially by using selective filters.

When n is odd, however, the minimum difference can be very small. Because m and $n-m$ can differ by unity, and each can be close to the signal frequency, if the adjacent interferer is ξf from the desired signal, the second need be only $(2\xi f)/(n-1)$ further away for the product to fall at the desired frequency. Thus, *odd-order* IM products can be caused by strong signals only a few channels removed from the desired signal. Selective filtering capable of reducing such signals substantially is not available in most superheterodyne receivers prior to the final IF. Consequently, odd-order IM products generally limit the dynamic range significantly.

Other effects of odd-order distortion are *desensitization* and *cross modulation*. For the case where n is odd, the presence of the desired signal and a strong interfering signal results in a product of the desired signal with an even order of the interfering signal. One of the resulting components of an even power of a sinusoid is a constant, so the desired signal is multiplied by that constant and an even power of the interferer's signal strength. If the interferer is sufficiently strong, the resulting product will subtract from the desired signal product from the first power term, reducing the effective gain of the device. This is referred to as *desensitization*. If the interferer is amplitude-modulated, the desired signal component will also be amplitude-modulated by

12-16 Broadcast Receiver Systems

the distorted modulation of the interferer. This is known as *cross modulation* of the desired signal by the interfering signal.

This discussion provides a simple theory that can be applied in considering strong signal effects. However, the receiver is far more complicated than the single device, and strong signal performance of single devices by these techniques can become rapidly intractable as higher-order terms must be considered. Another mechanism also limits the dynamic range. LO noise sidebands at low levels can extend substantially from the oscillator frequency. A sufficiently strong off-tune signal can beat with these noise sidebands in a mixer, producing additional noise in the desired signal band. Other characteristics that affect the dynamic range are spurious signals and responses, and blocking.

The effects described here occur in receivers, and tests to measure them are essential to determining the dynamic range. Most of these measurements involving the dynamic range require more than one signal input. They are conducted using two or three signal generators in a test setup such as that indicated in Figure 12.1.4.

12.1.4a Desensitization

Desensitization measurements are related to the 1-dB compression point and general linearity of the receiver. Two signal generators are used in the setup of Figure 12.1.4. The controls of the receiver under test are set as specified, usually to one of the narrower bandwidths and with MGC set as in sensitivity measurements so as to avoid effects of the AGC system. The signal in the operating channel is modulated and set to a specified level, usually to produce an output S/N or SINAD measurement of a particular level, for example, 13 dB. The interfering signal is moved off the operating frequency by a predetermined amount so that it does not affect the S/N measurement because of beat notes and is then increased in level until the S/N measurement is reduced by a specified amount, such as 3 dB. More complete information can be obtained by varying the frequency offset and plotting a desensitization selectivity curve. In some cases, limits for this curve are specified. The curve may be carried to a level of input where spurious responses, reciprocal mixing, or other effects prevent an unambiguous measurement. Measurements to 120 dB above sensitivity level can often be achieved.

If the degradation level at which desensitization is measured is set to 1-dB, and the desensitizing signal is well within the passband of the preselector filters, the desensitization level corresponds to the 1 dB *gain compression* (GC), which is experienced by the system up to the first mixer. (See the subsequent discussion of intermodulation and intercept points.) A gain compression (or *blocking*) dynamic range can be defined by comparing the input signal level at 1-dB GC to the MDS, i. e., dynamic range (dB) equals the GC (input dBm) minus the MDS (input dBm). This is sometimes referred to as the *single-tone dynamic range*, because only a single interfering signal is needed to produce GC.

12.1.4b AM Cross Modulation

Although many saturation effects in receivers have been called cross modulation, SSB and FM are not cross-modulated in the same sense as described previously. Cross modulation occurs in AM and VSB signals by a strong modulated signal amplitude-modulating a weak signal through the inherent nonlinearities of the receiver. Cross modulation typically occurs in a band allocated for AM use and requires a much higher interfering signal level than for the generation of IM

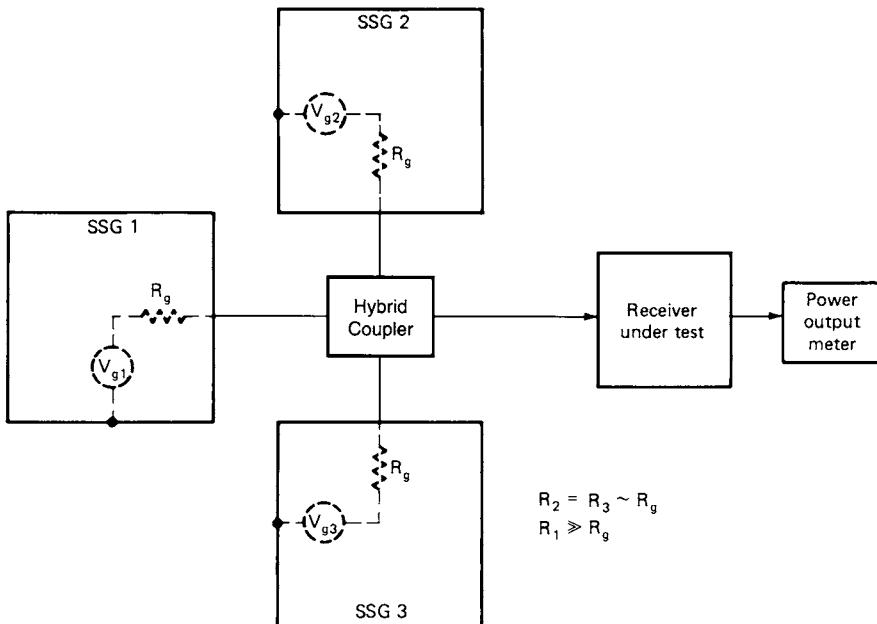


Figure 12.1.4 Test setup for measuring the dynamic range properties of a receiver.

products. The typical measurement setup is similar to that for overload measurements, except that the interfering signal is amplitude-modulated, usually at a high level, such as 90 percent. The modulation is at a different frequency than that for the operating channel (if it is modulated), and a band-pass filter is used in the output to assure that the transferred modulation is being measured. The out-of-channel interfering signal is increased in level until the desired signal has a specified level of output at the cross modulation frequency, for example, the equivalent of 10 percent modulation of the desired carrier. One or more specific offsets may be specified for the measurement, or a cross-modulation selectivity curve may be taken by measuring carrier level versus frequency offset to cause the specified degree of cross modulation.

In analog television systems, cross modulation can result in a ghost of an out-of-channel modulation being visible on the operating channel. The so-called three-tone test for television signals is a form of cross-modulation test.

12.1.4c IM

As described in previous sections, IM produces sum and difference frequency products of many orders that manifest themselves as interference. The measurement of the IM distortion performance is one of the most important tests for a communications receiver. No matter how sensitive a receiver may be, if it has poor immunity to strong signals, it will be of little use. Tests for even-order products determine the effectiveness of filtering prior to the channel filter, while odd-order

12-18 Broadcast Receiver Systems

products are negligibly affected by those filters. For this reason, odd-order products are generally much more troublesome than even-order products and are tested for more frequently. The second- and third-order products are generally the strongest and are the ones most frequently tested. A two-signal generator test set is required for testing, depending on the details of the specified test.

For IM measurements, the controls of the receiver under test are set to the specified bandwidths, operating frequency, and other settings as appropriate, and the gain control is set on manual (or AGC disabled). One signal generator is set on the operating frequency, modulated and adjusted to provide a specified S/N (that for sensitivity, for example). The modulation is disabled, and the output level of this signal is measured. This must be done using the IF output, the SSB output with the signal generator offset by a convenient audio frequency, or with the BFO on and offset. Alternatively, the dc level at the AM demodulator can be measured, if accessible. The signal generator is then turned off. It may be left off during the remainder of the test or retuned and used to produce one of the interfering signals.

For second-order IM testing, two signal generators are now set to two frequencies differing from each other by the operating frequency. These frequencies can be equally above and below the carrier frequency at the start, and shifted on successive tests to assure that the preselection filters do not have any weak regions. The signal with the frequency nearest to the operating frequency must be separated far enough to assure adequate channel filter attenuation of the signal (several channels). For third-order IM testing, the frequencies are selected in accordance with the formula given previously so that the one further from the operating frequency is twice as far as the one nearer to the operating frequency. For example, the nearer interferer might be three channels from the desired frequency and the further one, six channels in the same direction.

In either case, the voltage levels of the two interfering signal generators are set equal and are gradually increased until an output equal to the original channel output is measured in the channel. One of several performance requirements may be specified. If the original level is the sensitivity level, the ratio of the interfering generator level to the sensitivity level may have a specified minimum. Alternatively, for any original level, an interfering generator level may be specified that must not produce an output greater than the original level. Finally, an *intercept point* (IP) may be specified.

The IP for the n th order of intermodulation occurs because the product is a result of the interfering signal voltages being raised to the n th power. With equal voltages, as in the test, the resultant output level of the product increases as

$$V_{dn} = c_n V^n \quad (12.1.7)$$

where c_n is a proportionality constant and V is the common input level of the two signals. Because the output from a single signal input V at the operating frequency is G_{vB} , there is a theoretical level at which the two outputs would be equal. This value V_{IPn} is the n th IP, measured at the input. It is usually specified in dBm. In practice the IPs are not reached because as the amplifiers approach saturation, the voltage at each measured frequency becomes a combination of components from various orders of n . Figure 12.1.5 indicates the input-output power relationships in second- and third-order IPs.

In Equation 12.1.7 we note that at the IP

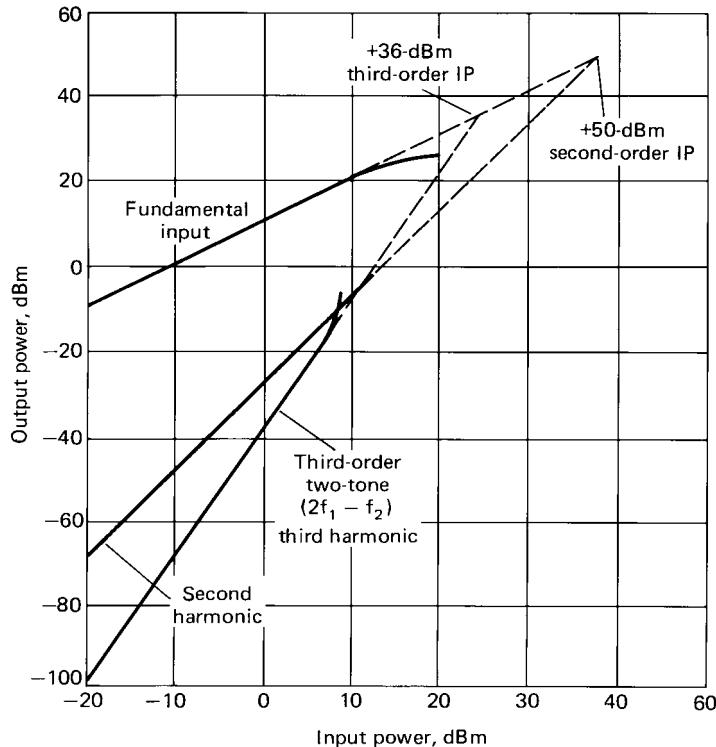


Figure 12.1.5 Input/output power relationships for second- and third-order intercept points.

$$V_{dn} = c_n (V_{IPn})^n \text{ and } (V_{IPn})_{out} = G_v (V_{IPn})_{in} \quad (12.1.8)$$

This leads to

$$c_n = G_v (V_{IPn})^{1-n} \text{ and } V_{dn} = G_v V \left[\frac{V}{V_{IPn}} \right]^{1-n} \quad (12.1.9)$$

The ratio of signal to distortion becomes $(V_{IPn}) / V^{n-1}$. In decibels it becomes

12-20 Broadcast Receiver Systems

$$R_{dn} = 20 \log \left[\frac{V}{V_{dn}} \right] = (n-1) [20 \log V_{IP_n} - 20 \log V] \quad (12.1.10)$$

If the intercept level is expressed in dBm rather than voltage, then the output power represented by V must be similarly expressed.

The IM products we have been discussing originate in the active devices of the receiver, so that the various voltages or power levels are naturally measured at the device output. The IP is thus naturally referred to the device output and is so specified in most data sheets. In the foregoing discussion, we have referred the IP to the voltage level at the device input. If the input power is required, we subtract from the output intercept level in decibels, the amplifier power gain or loss. The relationship between input and output voltage at the IP is given in Equation 12.1.8. Reference of the IP to the device input is somewhat unnatural but is technically useful because the receiver system designer must deal with the IP generation in all stages and needs to know at what antenna signal level the receiver will produce the maximum tolerable IM products.

Consider the input power (in each signal) that produces an output IM product equal to the MDS. The ratio of this power to the MDS may be called the *third-order IM dynamic range*. It also is sometimes referred to as the *two-tone dynamic range*. Expressing Equation 12.1.10 in terms of input power and input IP measured in dBm, we have

$$R_{dn} = (n-1) [IP_{n(in)} - P_{(in)}] \quad (12.1.11)$$

When we substitute MDS for the distortion and $MDS + DR$ for $P_{(in)}$ we obtain

$$\begin{aligned} DR &= (n-1) [IP_{n(in)} - MDS - DR], \\ nDR &= (n-1) [IP_{n(in)} - MDS] \end{aligned} \quad (12.1.12)$$

When n is 3, we find the relationship

$$DR = \frac{2 [IP_{3(in)} - MDS]}{3} \quad (12.1.13)$$

A dynamic range could presumably be defined for other orders of IM, but it is not common to do so. From the three different definitions of dynamic range described in this section, it should be clear why it is important to be careful when comparing receiver specifications for this characteristic.

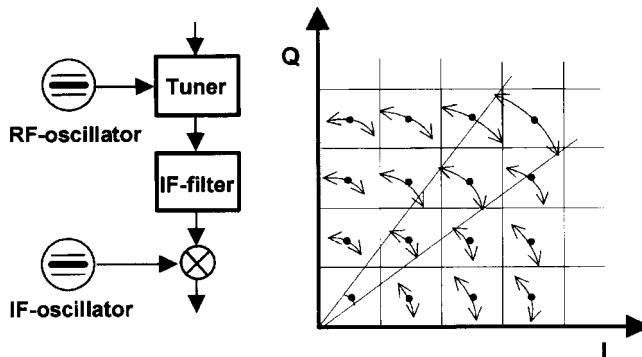


Figure 12.1.6 Phase noise is critical to digitally modulated communication systems because of the modulation errors it can introduce. *Inter-symbol interference (ISI)*, accompanied by a rise in BER, results when state values become so badly error-blurred that they fall into the regions of adjacent states. This drawing depicts only the results of phase errors introduced by phase noise; in actual systems, thermal noise, AM-to-PM conversion, differential group delay, propagation, and other factors may also contribute to the spreading of state amplitude and phase values. (Courtesy of Rohde & Schwarz.)

12.1.4d Error Vector Magnitude

As shown in Figure 12.1.6, particular sources of amplitude and/or phase error can shift the values of digital emission data states toward decision boundaries, resulting in increased BER because of intersymbol interference. Figures 12.1.7, 12.1.8, and 12.1.9 show three additional sources of such errors.

A figure of merit known as *error vector magnitude* (EVM) has been developed as sensitive indicator of the presence and severity of such errors. The error vector magnitude of an emission is the magnitude of the phasor difference as a function of time between an ideal reference signal and the measured transmitted signal after its timing, amplitude, frequency, phase, and dc offset have been modified by circuitry and/or propagation. Figure 12.1.10 illustrates the EVM concept.

12.1.5 Gain Control

Receivers must often be capable of handling a signal range of 100 dB or more. Most amplifiers remain linear over only a much smaller range. The later amplifiers in a receiver, which must provide the demodulator with about 1 V on weak signals, would need the capability to handle thousands of volts for strong signals without some form of gain control. Consequently, receivers customarily provide means for changing the gain of the RF or IF amplifiers, or both.

For applications where the received signal is expected to remain always within narrow limits, some form of manually selectable control can be used, which may be set on installation and seldom adjusted. There are few such applications. Most receivers, however, even when an operator is available, must receive signals that vary by tens of decibels over periods of fractions of sec-

12-22 Broadcast Receiver Systems

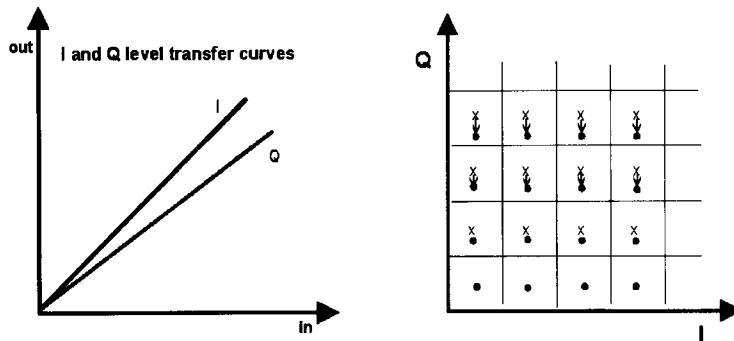


Figure 12.1.7 Effect of gain imbalance between *I* and *Q* channels on data signal phase constella-tion. (Courtesy of Rohde & Schwarz.)

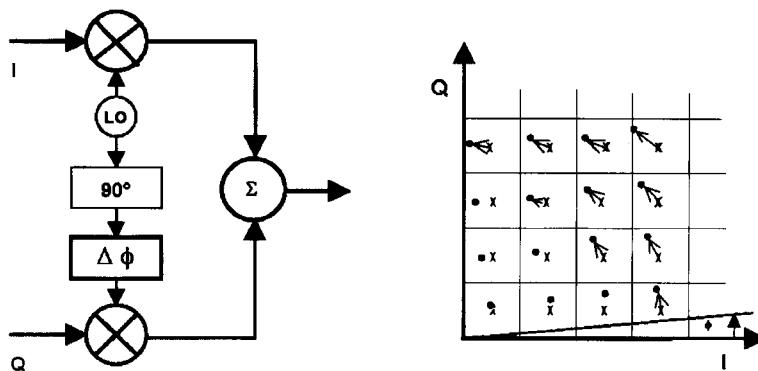


Figure 12.1.8 Effect of quadrature offset on data signal phase constellation. (Courtesy of Rohde & Schwarz.)

onds to minutes. The level also changes when the frequency is reset to receive other signals that may vary over similar ranges but with substantially different average levels. Consequently, an AGC is very desirable.

Some angle modulation receivers provide gain control by using amplifiers that limit on strong signals. Because the information is in the angle of the carrier, the resulting amplitude distortion is of little consequence. Receivers that must preserve AM or maintain very low angle modulation distortion use amplifiers that can be varied in gain by an external control voltage. In some cases, this has been accomplished by varying the operating points of the amplifying devices, but most modern systems separate solid-state circuits or switched passive elements to obtain variable attenuation between amplifier stages with minimum distortion. For manual control, provision can be made to let an operator set the control voltage for these variable attenuators. For automatic control, the output level from the IF amplifiers or the demodulator is monitored by the AGC cir-

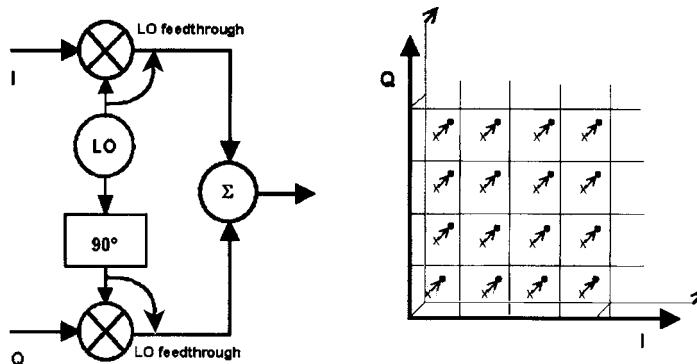


Figure 12.1.9 Effect of LO-feedthrough-based *I*/*Q* offset on data signal phase constellation. (*Courtesy of Rohde & Schwarz.*)

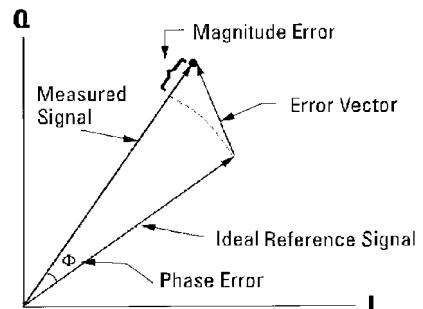


Figure 12.1.10 The concept of error vector magnitude. (*Courtesy of Hewlett-Packard.*)

cuit and a low-pass negative-feedback voltage is derived to maintain a relatively constant signal level.

A block diagram of a dual AGC loop system is illustrated in Figure 12.1.11. One loop is driven by first IF energy that is band-limited, and the other loop is driven by second IF energy that is band-limited by optional second IF filters. The first loop controls a PIN diode pi attenuator ahead of the first mixer. The second loop controls the second IF amplifier stages. In this design, a microprocessor adjusts the time constants of both loops so that time delays introduced by the filters do not cause AGC oscillation.

A number of tests of gain control characteristics are customarily required. MGC may be designed to control gain continuously or in steps. It is important that the steps be small enough that operators do not detect large jumps as they adjust the gain. Because the gain must be controlled over a very wide range, the MGC is easiest to use if it tends to cause a logarithmic variation. Usually, the testing of the MGC is confined to establishing that a specified range of gain control exists and measuring the degree of decibel linearity versus control actuation.

12-24 Broadcast Receiver Systems

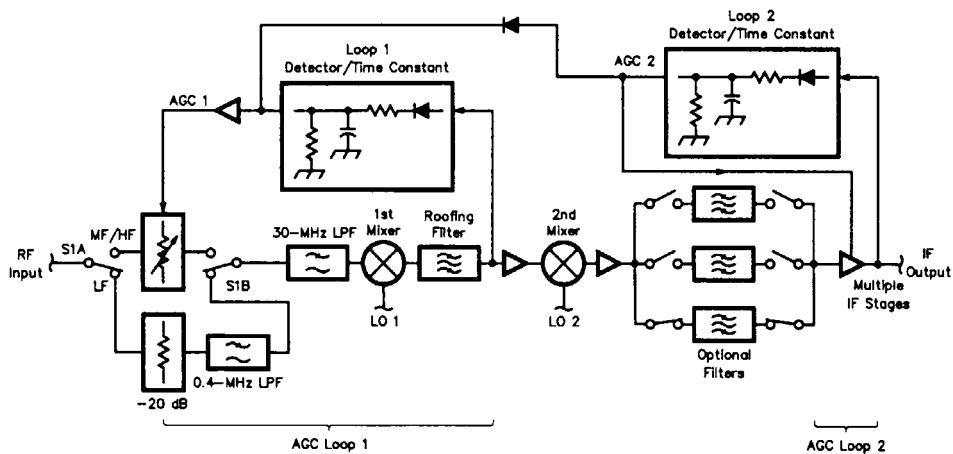


Figure 12.1.11 Block diagram of a dual loop AGC system for a communications receiver.

The principal AGC characteristics of importance are the steady-state control range and output-input curve, and the attack and decay times. In some systems, a variety of time constants are provided for the AGC to allow for different modulation types. For AM voice modulation, the radiated carrier is constant and the lowest sidebands are usually several hundred hertz removed from the carrier. At the receiver, the carrier component can be separated from the demodulated wave by a low-pass filter and can serve as the AGC control voltage. The response time of the filter, which is often just an RC network, need only be fast enough to respond to the fading rate of the medium, which is a maximum of 5 or 10 dB/s in most AM applications. A response time of 0.1 to 0.2 s is required for such a fading rate. For the more common slower rates, responses up to a second or more can be used.

To test for the AGC range and input-output curve, a single signal generator is used (as in Figure 18.8.2) in the AM mode with the receiver's AGC actuated. The signal generator is set to several hundred microvolts, and the baseband output level is adjusted to a convenient level for output power measurement. The signal generator is then tuned to its minimum level and the output level is noted. The signal is gradually increased in amplitude, and the output level is measured for each input level, up to a maximum specified level, such as 2 V. Figure 12.1.12 shows some typical AGC curves. In most cases, there will be a low-input region where the signal output, rising out of the noise, varies linearly with the input. At some point, the output curve bends over and begins to rise very slowly. At some high level, the output may drop off because of saturation effects in some of the amplifiers. The point at which the linear relationship ends is the AGC *threshold* of the AGC action. The point at which the output starts to decrease, if within a specified range, is considered the upper end of the AGC control range. The difference between these two input levels is the AGC *control range*. If the curve remains monotonic to the maximum input test level, that level is considered the upper limit of the range. A measure of AGC effectiveness is the increase in output from a specified lower input voltage level to an upper input voltage

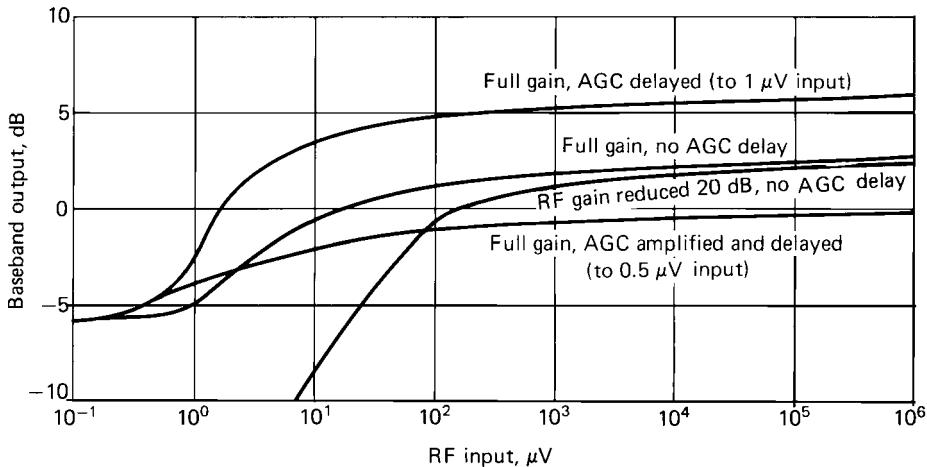


Figure 12.1.12 Representative input-output AGC curves.

level. For example, a good design might have an AGC with a threshold below 1 μV that is monotonic to a level of 1 V and has the 3 dB increase in output between 1 μV and 0.1 V.

The foregoing applies to a purely analog system. The advantage of a DSP-based system include that AGC is handled internally. The receiver still requires a dual-loop AGC of which the input stage AGC will remain analog.

12.1.6 Digital Receiver Characteristics

The foregoing has not exhausted the analog receiver characteristics that may be of interest but has reviewed some of the more significant ones. For example, in FM sets, the *capture ratio* is important. Clearly, an area of increasing interest is the characterization of systems utilizing digital modulation techniques. Because a digital radio system is a hybrid A/D device, many of the test procedures outlined previously for analog receivers are useful and important in characterizing a digital radio system. Additional tests, primarily involving the analysis of bit error rates (BER), must also be run to properly identify any weak points in a receiver design.

12.1.6a BER Testing

The primary method for testing the quality of transmission over a high speed digital communications link is the BER, defined as the number of bit errors divided by the number of bits transmitted. The BER is also used to qualify the sensitivity and noise characteristics of a receiver. The major contributor to BER is *jitter*, which results from noise in the system. This noise causes the output comparator to vary the time of its transition relative to the data clock. If the transition time changes too much, an error will result.

12-26 Broadcast Receiver Systems

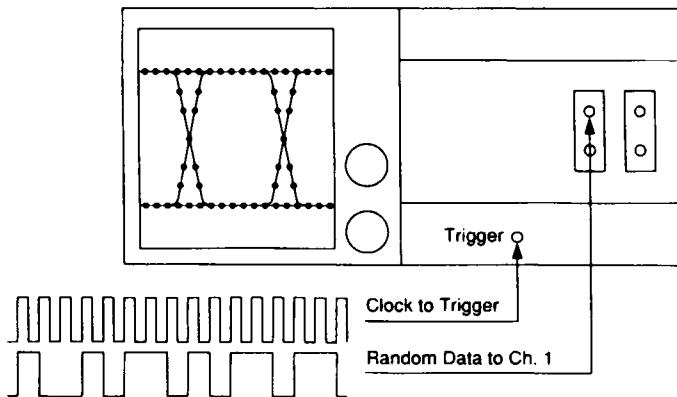


Figure 12.1.13 Test setup for eye pattern measurement.

Using a signal analyzer specifically designed for BER testing, jitter can be displayed directly, or the BER is simply tabulated by the analyzer. The format of the testing signal is determined by the application to which the digital radio system is applied. A variety of formats and protocols are used. For end-to-end testing, the analyzer feeds a reference RF signal generator whose characteristics are known, and the output signal is applied to the receiver antenna input.

The noise and jitter on a data waveform provides vital information about the quality of the signal. A typical setup for capturing an *eye pattern* is shown in Figure 12.1.13. Eye patterns are the traditional method of displaying high-speed digital data (Figure 12.1.14). Some communications signal analyzers augment this information with a built-in statistical database, which allows additional analysis, including automated noise and jitter measurements on random data. Sophisticated software can also analyze the form of the distribution, providing mean, rms, and standard deviation results.

Proper receiver design involves identifying the areas of the system that are likely to cause problems. LO phase noise is one such area. Phase noise can seriously impair the performance of a digital receiver. Excessive phase noise can increase the BER, especially in systems using phase-based modulation schemes, such as binary PSK and quadrature PSK. For a given statistical phase-error characteristic, BER is degraded according to the percentage of time that the phase error causes the signal position in signal space to cross a decision boundary.

12.1.6b Transmission and Reception Quality

Testing of digital circuits deviates from the typical analog measurements, and yet the analog measurements are still necessary and related. In particular, because of the Doppler effect and the use of digital rather than analog signals, where the phase information is significant, the designer ends up using coding schemes for error-correction—specifically, *forward error correction* (FEC). The S/N as we know it from analog circuits now determines the BER, and its tolerable values depend on the type of modulation used. The actual bit error rate depends on the type of filtering, coding, modulation, and demodulation.

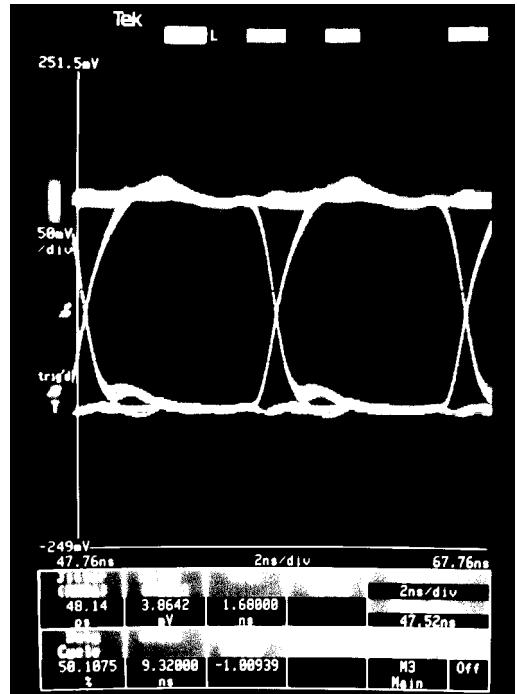


Figure 12.1.14 Eye pattern display of BER measurement. (Courtesy of Tektronix.)

The *adjacent-channel power ratio* (ACPR), a factor involving the transmitter of a second station, is another problem that receivers must deal with. Given the fact that a transmitter handling digital modulation delivers its power in pulses, its transmissions may affect adjacent channels by producing transient spurious signals similar to what we call *spatter* in analog SSB systems. This is a function of the linearity of the transmitter system all the way out to the antenna, and forces most designers to resort to less-efficient Class A operation. As possible alternatives, some researchers have designed systems using Class D or E modulation.

It is not uncommon to do many linear measurements, and then by using correlation equations, translate these measured results into their digital equivalents. Therefore, the robustness of the signal as a function of antenna signal at the receiver site, constant or known phase relationships, and high adjacent power ratios will provide good transfer characteristics.

12.1.7 Bibliography

- Engelson, M., and J. Herbert: "Effective Characterization of CDMA Signals," *Microwave Journal*, pg. 90, January 1995.
- Howald, R.: "Understand the Mathematics of Phase Noise," *Microwaves & RF*, pg. 97, December 1993.
- Johnson, J. B.: "Thermal Agitation of Electricity in Conduction," *Phys. Rev.*, vol. 32, pg. 97, July 1928.

12-28 Broadcast Receiver Systems

- Nyquist, H.: "Thermal Agitation of Electrical Charge in Conductors," *Phys. Rev.*, vol. 32, pg. 110, July 1928.
- Pleasant, D.: "Practical Simulation of Bit Error Rates," *Applied Microwave and Wireless*, pg. 65, Spring 1995.
- Rohde, Ulrich L.: "Key Components of Modern Receiver Design—Part 1," *QST*, pg. 29, May 1994.
- Rohde, Ulrich L. Rohde and David P. Newkirk: *RF/Microwave Circuit Design for Wireless Applications*, John Wiley & Sons, New York, N.Y., 2000.
- "Standards Testing: Bit Error Rate," application note 3SW-8136-2, Tektronix, Beaverton, OR, July 1993.
- Using Vector Modulation Analysis in the Integration, Troubleshooting and Design of Digital RF Communications Systems*, Product Note HP89400-8, Hewlett-Packard, Palo Alto, Calif., 1994.
- Watson, R.: "Receiver Dynamic Range; Pt. 1, Guidelines for Receiver Analysis," *Microwaves & RF*, vol. 25, pg. 113, December 1986.
- "Waveform Analysis: Noise and Jitter," application note 3SW8142-2, Tektronix, Beaverton, OR, March 1993.
- Wilson, E.: "Evaluate the Distortion of Modular Cascades," *Microwaves*, vol. 20, March 1981.

Chapter
12.2
The Radio Channel

Ulrich L. Rohde

Jerry C. Whitaker, Editor-in-Chief

12.2.1 Introduction¹

The transmission of information from a fixed station to one or more mobile stations is considerably influenced by the characteristics of the radio channel [1]. The RF signal arrives at the receiving antenna not only on the direct path but is normally reflected by natural and artificial obstacles in its way. Consequently, the signal arrives at the receiver several times in the form of echoes that are superimposed on the direct signal, as illustrated in Figure 12.2.1. This superposition may be an advantage as the energy received in this case is greater than in single-path reception. However, this characteristic may be a disadvantage when the different waves cancel each other under unfavorable phase conditions. In conventional car radio reception this effect is known as *fading*. It is particularly annoying when the vehicle stops in an area where the field strength is reduced because of fading (for example, at traffic lights). Additional difficulties arise when digital signals are transmitted. If strong echo signals (compared to the directly received signal) arrive at the receiver with a delay in the order of a symbol period or more, time-adjacent symbols interfere with each other. In addition, the receive frequency may be falsified at high vehicle speeds because of the Doppler effect so that the receiver may have problems estimating the instantaneous phase in the case of angle-modulated carriers. Both effects lead to a high symbol error rate even if the field strength is sufficiently high.

Radio broadcasting systems using conventional frequency modulation are not seriously affected by these interfering effects in most cases. If an analog system is replaced by a digital one that is expected to offer advantages over the previous system, the designer must ensure that the expected advantages—for example, improved audio S/N and the possibility to offer supplementary services to the subscriber—are not achieved at the expense of reception in hilly terrain or at high vehicle speeds because of extreme fading. For this reason, a modulation method combined

1. This chapter is based on: Rohde, Ulrich L., and Jerry C. Whitaker: *Communications Receivers: Principles and Design*, 3rd ed., McGraw-Hill, New York, N.Y., 2000. Used with permission.

12-30 Broadcast Receiver Systems

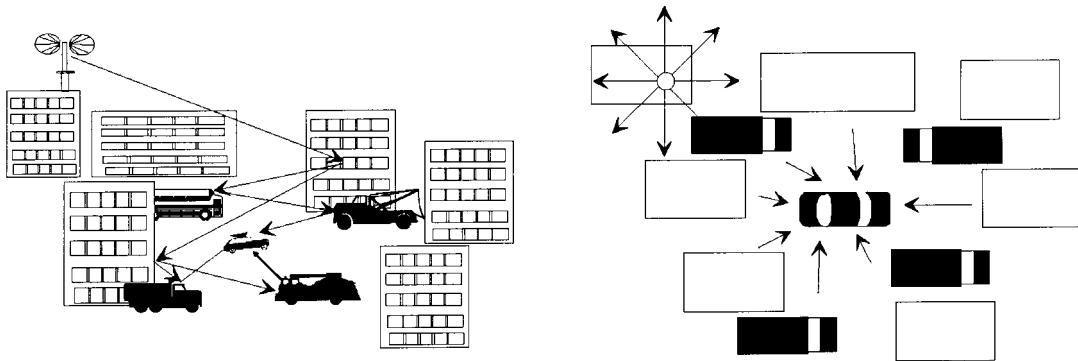


Figure 12.2.1 Mobile receiver affected by fading. (Courtesy of Rohde & Schwarz.)

with suitable error protection must be found for mobile reception in a typical radio channel that is immune to fading, echo, and Doppler effects.

With a view to this design objective, more detailed information on the radio channel is required. The channel can be described by means of a model. In the worst case, which may be the case for reception in urban areas, it can be assumed that the mobile receives the signal on several indirect paths but not on a direct one. The signals are reflected, for example, by large buildings; the resulting signal delays are relatively long. In the vicinity of the receiver, these paths are split up into a great number of subpaths; the delays of these signals are relatively short. These signals may again be reflected by buildings but also by other vehicles or natural obstacles such as trees. Assuming the subpaths being statistically independent of each other, the superimposed signals at the antenna input cause considerable time- and position-dependent field-strength variations with an amplitude obeying the Rayleigh distribution (Figures 12.2.2 and 12.2.3). If a direct path is received in addition to the reflected ones, the distribution changes to the Rice distribution and finally, when the direct path becomes dominant, the distribution follows the Gaussian distribution with the field strength of the direct path being used as the center value.

In a Rayleigh channel the bit error rate increases dramatically compared to the BER in an AWGN (additive white Gaussian noise) channel (Figure 12.2.4).

12.2.1a Channel Impulse Response

The scenario outlined in the previous section can be demonstrated by means of the *channel impulse response* [1]. Assume that a very short pulse of extremely high amplitude—in the ideal case a *Dirac pulse* $\delta(t)$ —is sent by the transmitting antenna at a time $t_0 = 0$. This pulse arrives at the receiving antenna direct and in the form of reflections with different delays τ_i and different amplitudes because of path losses. The impulse response of the radio channel is the sum of all received pulses (Figure 12.2.5). Because the mobile receiver and some of the reflecting objects are moving, the channel impulse response is a function of time and of delays τ_i ; that is, it corresponds to

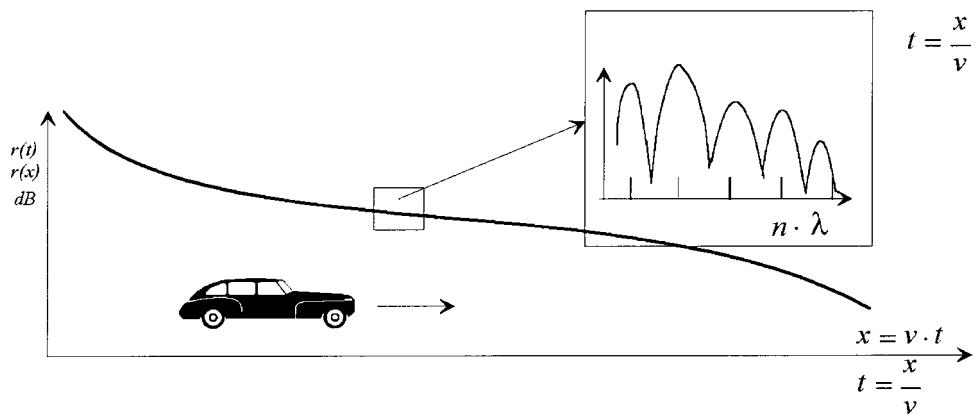


Figure 12.2.2 Receive signal as a function of time or position. (Courtesy of Rohde & Schwarz.)

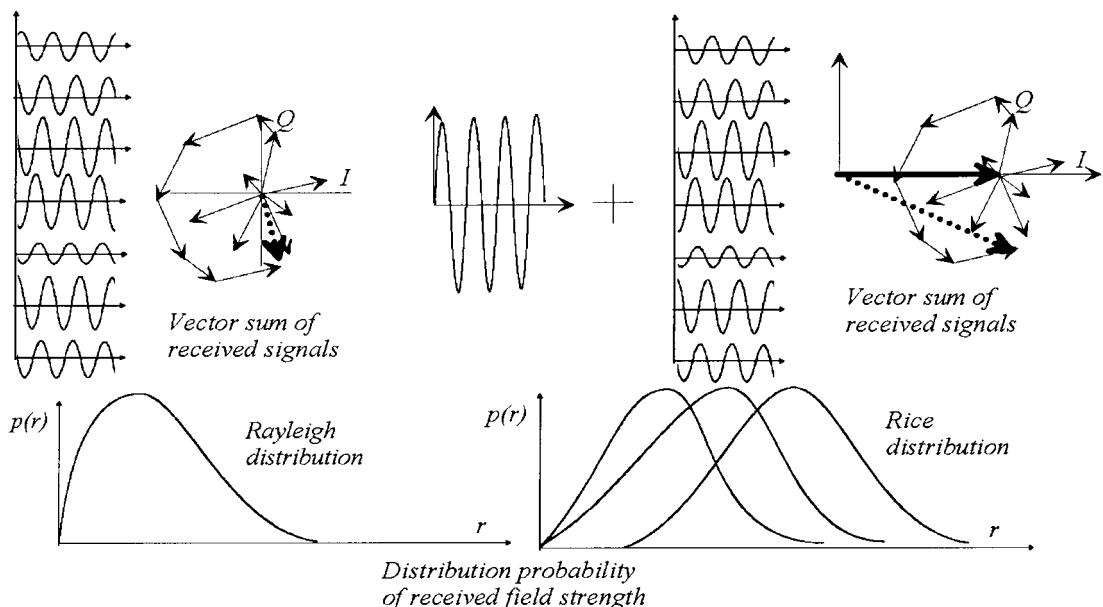


Figure 12.2.3 Rayleigh and Rice distribution. (Courtesy of Rohde & Schwarz.)

$$h(t, \tau) = \sum_N a_i \delta(t - \tau_i) \quad (12.2.1)$$

12-32 Broadcast Receiver Systems

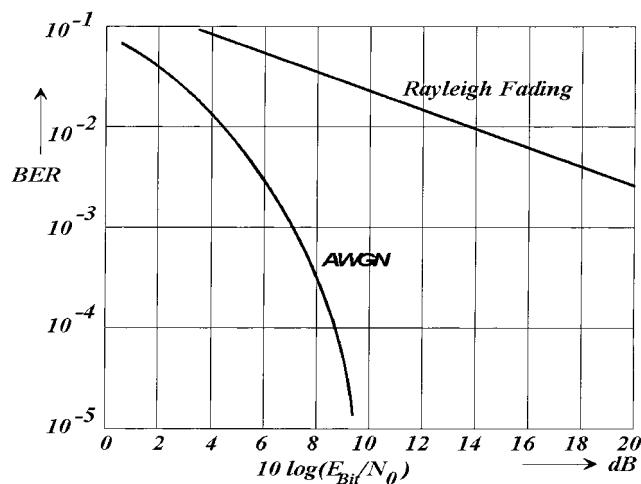


Figure 12.2.4 Bit error rate in a Rayleigh channel. (Courtesy of Rohde & Schwarz.)

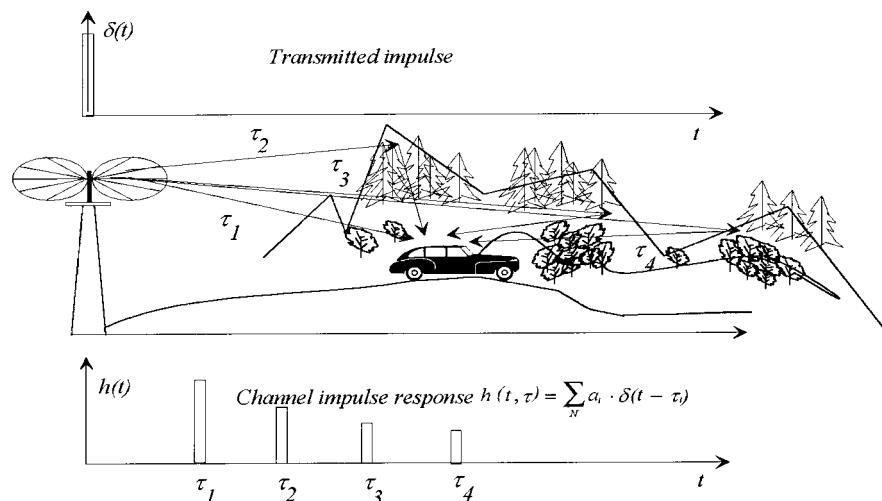


Figure 12.2.5 Channel impulse response. (Courtesy of Rohde & Schwarz)

This shows that delta functions sent at different times t cause different reactions in the radio channel.

In many experimental investigations, different landscape models with typical echo profiles were created. The most important are:

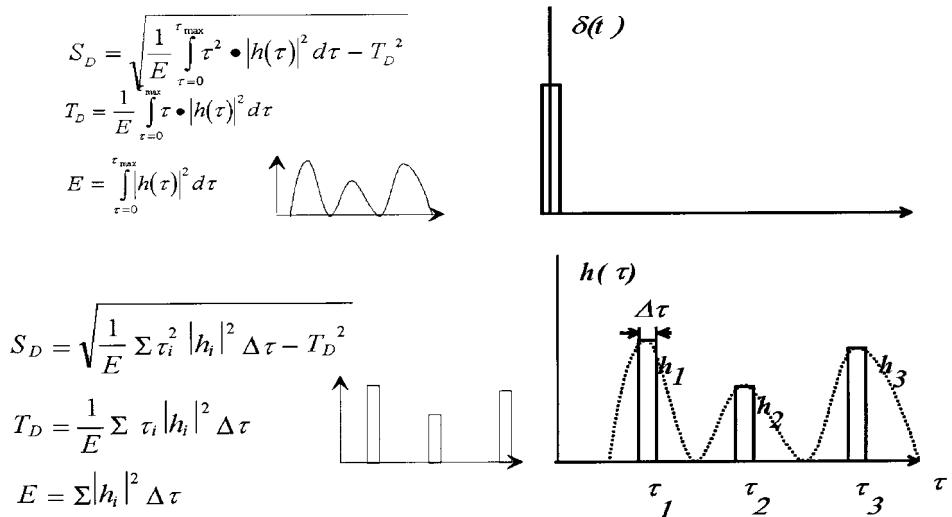


Figure 12.2.6 Calculation of delay spread. (Courtesy of Rohde & Schwarz.)

- Rural area (RA)
- Typical urban area (TU)
- Bad urban area (BA)
- Hilly terrain (HT)

The channel impulse response tells us how the received power is distributed to the individual echoes. A useful parameter, the *delay spread* can be calculated from the channel impulse response, permitting an approximate description of typical landscape models, as illustrated in Figure 12.2.6.

The delay spread also roughly characterizes the modulation parameters, carrier frequency, symbol period, and duration of guard interval, which have to be selected in relation to each other. If the receiver is located in an area with a high delay spread (for example, in hilly terrain), echoes of the symbols sent at different times are superimposed when broadband modulation methods with a short symbol period are used. An adjacent transmitter emitting the same information on the same frequency has the effect of an artificial echo (Figure 12.2.7).

A constructive superposition of echoes is only possible if the symbol period is much greater than the delay spread. The following holds

$$T_s > 10T_d \quad (12.2.2)$$

12-34 Broadcast Receiver Systems

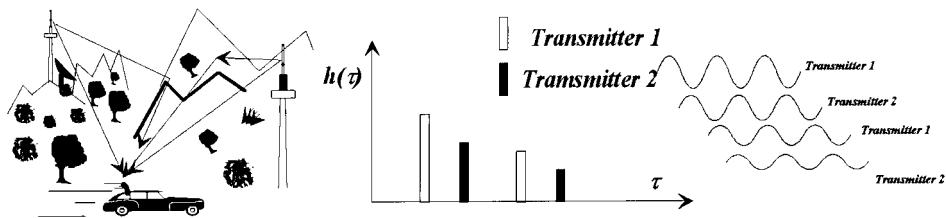


Figure 12.2.7 Artificial and natural echoes in the single-frequency network. (Courtesy of Rohde & Schwarz.)

This has the consequence that relatively narrowband modulation methods have to be used. If this is not possible, *channel equalizing* is required.

For channel equalizing, a continuous estimation of the radio channel is necessary. The estimation is performed with the aid of a periodic transmission of data known to the receiver. In networks, a midamble consisting of 26 bits—the *training sequence*—can be transmitted with every burst. The training sequence corresponds to a characteristic pattern of *I/Q* signals that is held in a memory at the receiver. The baseband signals of every received training sequence are correlated with the stored ones. From this correlation, the channel can be estimated; the properties of the estimated channel will then be fed to the equalizer, as shown in Figure 12.2.8.

The equalizer uses the *Viterbi algorithm* (maximum sequence likelihood estimation) for the estimation of the phases that most likely have been sent at the sampling times. From these phases the information bits are calculated (Figure 12.2.9). A well designed equalizer then will superimpose the energies of the single echoes constructively, so that the results in an area where the echoes are moderately delayed (delay times up to 16 μs at the receiver) are better than in an area with no significant echoes (Figure 12.2.10). The remaining bit errors are eliminated using another Viterbi decoder for the convolutionally encoded data sequences.

The ability of a mobile receiver to work in an hostile environment such as the radio channel with echoes must be proven. The test is performed with the aid of a *fading simulator*, which simulates various scenarios with different delay times and different Doppler profiles. A signal generator produces undistorted *I/Q* modulated RF signals that are downconverted into the baseband. Next, the *I/Q* signals are digitized and split into different channels where they are delayed and attenuated, and where Doppler effects are superimposed. After combination of these distorted signals at the output of the baseband section of the simulator, the signals modulate the RF carrier, which is the test signal for the receiver under test (Figure 12.2.11).

12.2.1b Doppler Effect

Because the mobile receiver and some of the reflecting objects are in motion, the receive frequency is shifted as a result of the Doppler effect [1]. In the case of single-path reception, this shift is calculated as follows

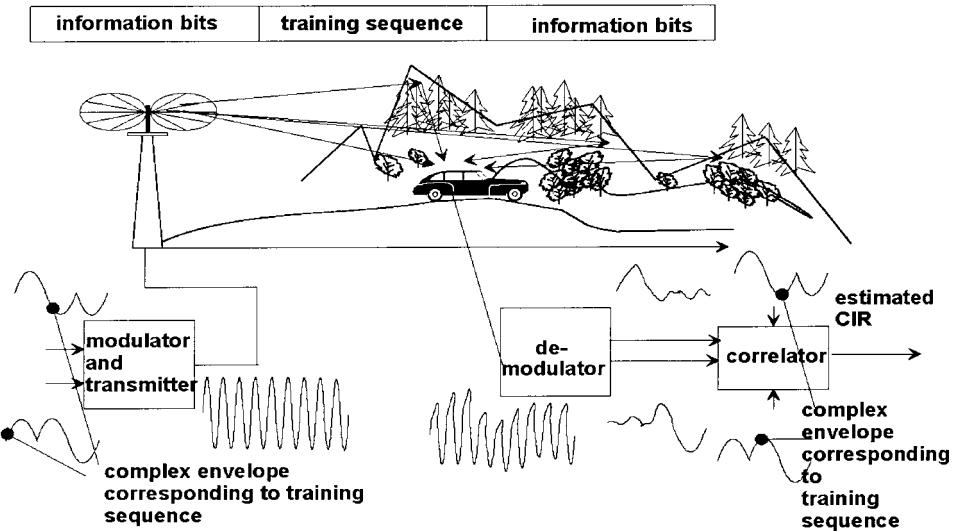


Figure 12.2.8 Channel estimation. (Courtesy of Rohde & Schwarz.)

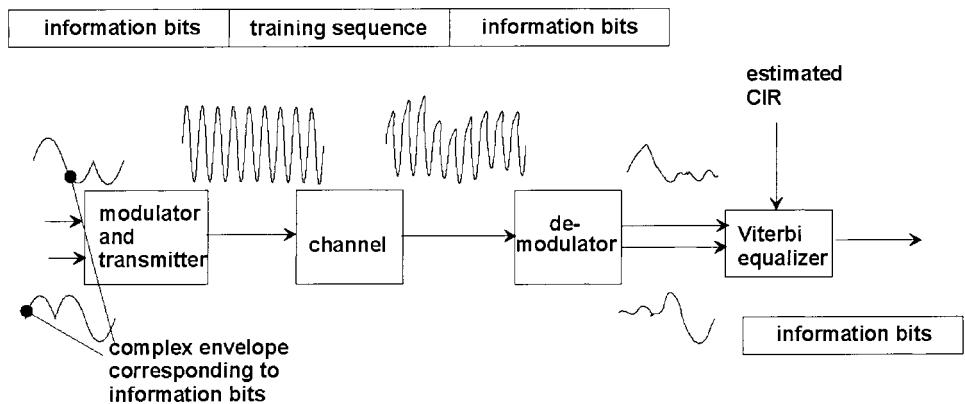


Figure 12.2.9 Channel equalization. (Courtesy of Rohde & Schwarz.)

$$f_d = \frac{v}{c} f_c \cos \alpha \quad (12.2.3)$$

Where:

12-36 Broadcast Receiver Systems

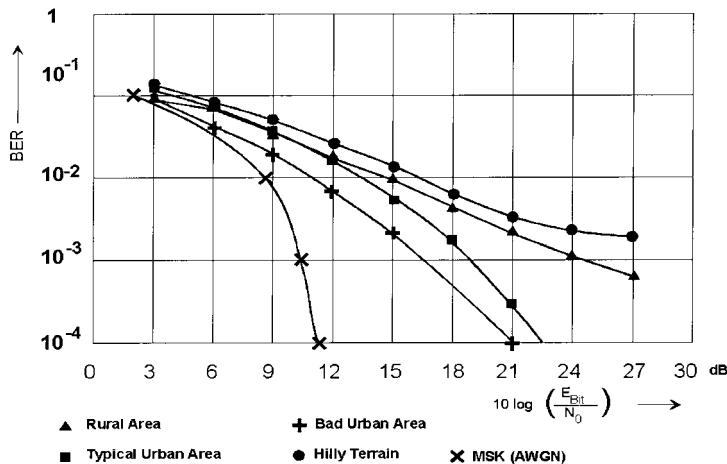


Figure 12.2.10 BERs after the channel equalizer in different areas. (Courtesy of Rohde & Schwarz.)

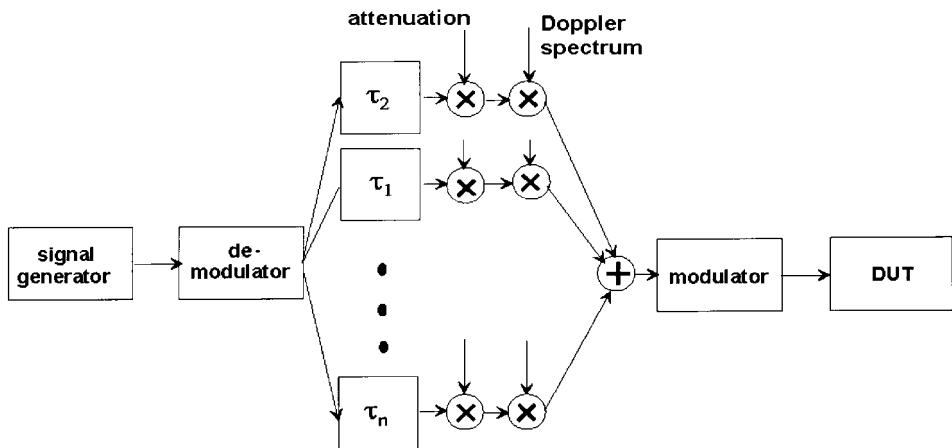


Figure 12.2.11 Fading simulator. (Courtesy of Rohde & Schwarz.)

v = speed of vehicle

c = speed of light

f = carrier frequency

α = angle between v and the line connecting transmitter and receiver

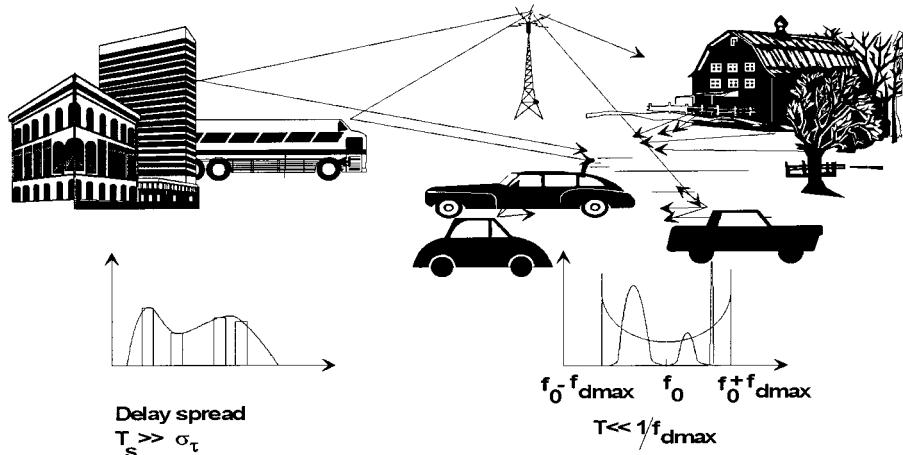


Figure 12.2.12 Doppler spread. (Courtesy of Rohde & Schwarz.)

In the case of multipath reception, the signals on the individual paths arrive at the receiving antenna with different Doppler shifts because of the different angles α_i , and the receive spectrum is spread. Assuming an equal distribution of the angles of incidence, the power density spectrum can be calculated as follows

$$P(f) = \frac{1}{\pi} \frac{1}{\sqrt{f_d^2 - f^2}} \text{ for } |f| < |f_d| \quad (12.2.4)$$

where f_d = maximum Doppler frequency.

Of course, other Doppler spectra are possible in addition to the pure Doppler shift; for example, spectra with a Gaussian distribution using one or several maxima. A Doppler spread can be calculated from the Doppler spectrum analogously to the delay spread shown in Figure 12.2.12.

12.2.1c Transfer Function

The FFT value of the channel impulse response is the transfer function $H(f,t)$ of the radio channel, which is also time-dependent. The transfer function describes the attenuation of frequencies in the transmission channel. When examining the frequency dependence, it will be evident that the influence of the transmission channel on two sine-wave signals of different frequencies becomes greater with increasing frequency difference. This behavior can be adequately described by the *coherence bandwidth*, which is approximately equal to the reciprocal delay spread; that is

12-38 Broadcast Receiver Systems

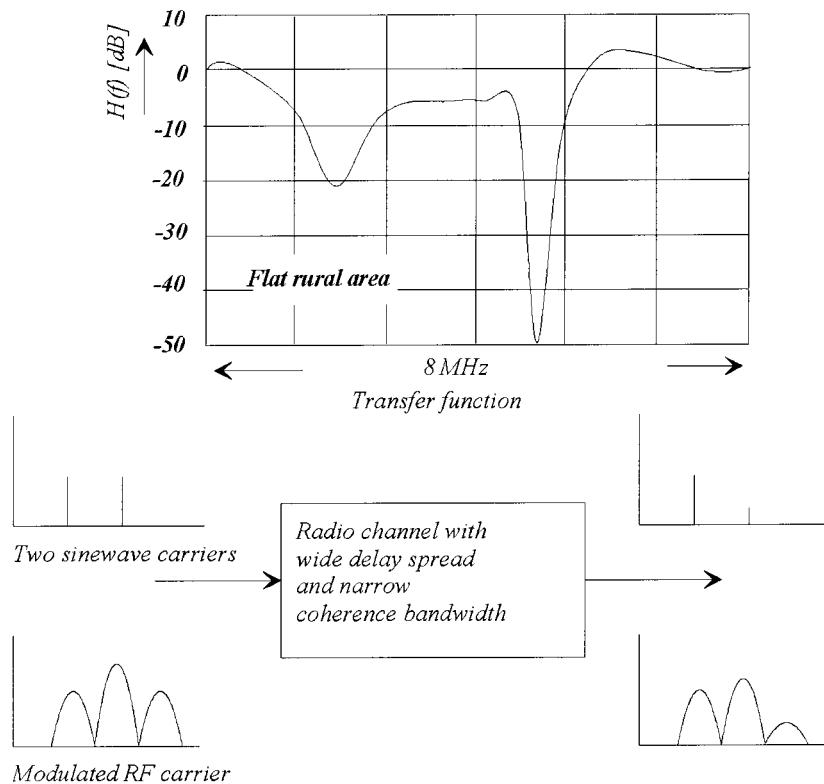


Figure 12.2.13 Effect of transfer function on modulated RF signals. (Courtesy of Rohde & Schwarz.)

$$(\Delta f)_c = \frac{1}{T_d} \quad (12.2.5)$$

If the coherence bandwidth is sufficiently wide and—consequently—the associated delay spread is small, the channel is not frequency-selective. This means that all frequencies are subject to the same fading. If the coherence bandwidth is narrow and the associated delay spread wide, even very close adjacent frequencies are attenuated differently by the channel. The effect on a broadband-modulated carrier with respect to the coherence bandwidth is obvious. The sidebands important for the transmitted information are attenuated to a different degree. The result is a considerable distortion of the receive signal combined with a high bit error rate even if the received field strength is high. This characteristic of the radio channel again speaks for the use of narrowband modulation methods. (See Figure 12.2.13).

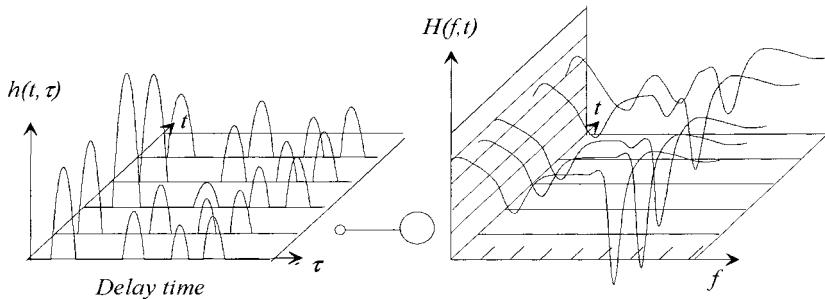


Figure 12.2.14 Channel impulse response and transfer function as a function of time. (Courtesy of Rohde & Schwarz.)

12.2.1d Time Response of Channel Impulse Response and Transfer Function

The time response of the radio channel can be derived from the Doppler spread. It is assumed that the channel rapidly varies at high vehicle speeds. The time variation of the radio channel can be described by a figure, the *coherence time*, which is analogous to the coherence bandwidth. This calculated value is the reciprocal bandwidth of the Doppler spectrum. A wide Doppler spectrum therefore indicates that the channel impulse response and the transfer function vary rapidly with time, as shown in Figure 12.2.14. If the Doppler spread is reduced to a single line, the channel is *time-invariant*. In other words, if the vehicle has stopped or moves at a constant speed in a terrain without reflecting objects, the channel impulse response and the transfer function measured at different times are the same.

The effect on information transmission can be illustrated with a simple example. In the case of MPSK modulation using hard keying, the transmitter holds the carrier phase for a certain period of time; that is, for the symbol period T . In the case of soft keying with low-pass-filtered baseband signals for limiting the modulated RF carrier, the nominal phase is reached at a specific time—the sampling time. In both cases the phase error $\phi_f = f_{dTS}$ is superimposed onto the nominal phase angle, which yields a phase uncertainty of $\Delta\phi = 2\phi_f$ at the receiver. The longer the symbol period, the greater the angle deviation (Figure 12.2.15). Considering this characteristic of the transmission channel, a short symbol period of $T_s \ll (\Delta t)_c$ should be used. However, this requires broadband modulation methods.

Figure 12.2.16 shows the field strength or power arriving at the mobile receiver if the vehicle moves in a Rayleigh distribution channel. Because the phase depends on the vehicle position, the receiver moves through positions of considerably differing field strength at different times (time-dependence of radio channel). In the case of frequency-selective channels, this applies to one frequency only; that is, to a receiver using a narrowband IF filter for narrowband emissions. As Figure 12.2.16 shows, this effect can be reduced by increasing the bandwidth of the emitted signal and consequently the receiver bandwidth.

12-40 Broadcast Receiver Systems

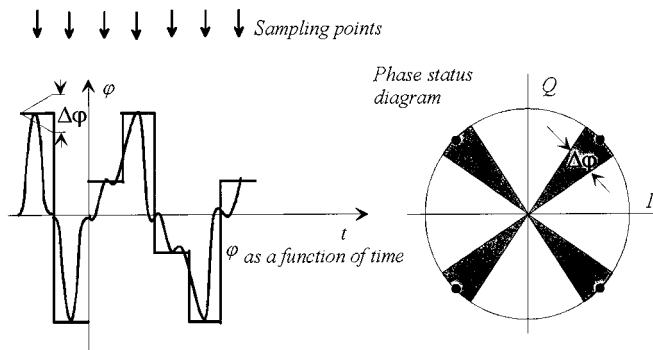


Figure 12.2.15 Phase uncertainty caused by Doppler effect. (Courtesy of Rohde & Schwarz.)

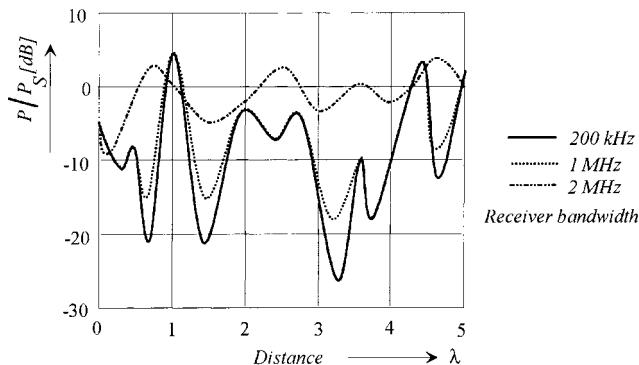


Figure 12.2.16 Effect of bandwidth on fading. (Courtesy of Rohde & Schwarz.)

12.2.2 References

1. Rohde, Ulrich L., and David P. Newkirk: *RF/Microwave Circuit Design for Wireless Applications*, John Wiley & Sons, New York, N.Y., 2000.

Chapter
12.3
AM and FM Receivers

Jerry C. Whitaker, Editor-in-Chief

12.3.1 Introduction¹

A *simplex* transmitter-receiver system is the simplest type of communication system possible. A single source transmits to a single receiver. When the receiver can respond to a transmission from the source and relay voice or other information back to the transmission site, a *duplex* arrangement is realized. A two-way business radio system operating on different transmit and receive frequencies employs a duplex arrangement. AM and FM broadcast systems utilize a *simplex star* configuration in which one transmitter feeds many receivers. For greatest efficiency, the receivers in a simplex star should be made as simple as possible, concentrating technical complexity at the transmit site. Radio broadcasting today follows this basic rule.

AM and FM stations transmit at high power levels to facilitate simpler radio designs. AM stations in the U.S. can operate at up to 50 kW, while FM stations can operate at up to 100 kW. This essentially eliminates the need for a sensitive antenna at the receiver. Stereo AM and stereo FM systems are, likewise, designed to permit the more complex operations to be performed in the encoding stage at the transmitter, rather than in the decoding circuits of the receiver.

12.3.1a Superheterodyne Receiver

Virtually every AM-FM radio manufactured today incorporates the superheterodyne method of reception. This system makes use of the heterodyne principle of mixing an incoming signal with a signal generated by a local oscillator (LO), as illustrated in Figure 12.3.1. The LO is offset by a fixed intermediate frequency (IF) from the desired signal. Because the mixer (a nonlinear device) generates a difference frequency that is identical if the desired signal is either above or below the LO frequency (and also a number of other spurious responses), filtering is required prior to the mixer to suppress the undesired signal. The frequency of the undesired signal is referred to as an *image frequency* and is separated from the desired signal frequency by a difference equal to twice the IF.

1. Portions of this chapter were adapted from Rohde, Ulrich L., and Jerry C. Whitaker: *Communications Receivers*, 3rd ed., McGraw-Hill, New York, N.Y., 2000. Used with permission.

12-42 Radio Receivers

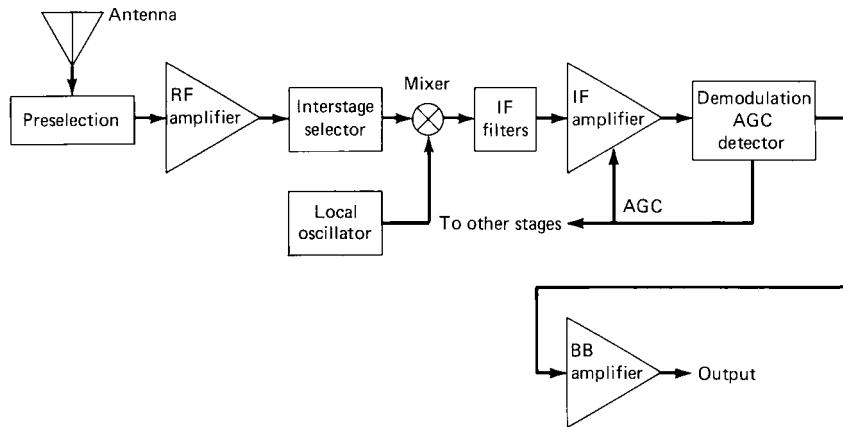


Figure 12.3.1 Simplified block diagram of a superheterodyne receiver. Depending on the sophistication of the radio, some of the stages shown may be combined into a single circuit. (From [1]. Used with permission.)

Channel filtering is accomplished by one or more fixed-frequency filters in the IF. This is a decided advantage when the receiver must cover a wide frequency band, because it is more difficult to maintain constant bandwidth in a tunable filter than in a fixed one. An IF frequency of 455 kHz is used for AM receivers and 10.7 MHz for FM receivers.

As shown in Figure 12.3.1, the input signal is fed from the antenna to a preselector filter and RF amplifier. The input circuit matches the antenna to the first amplifying device to achieve the best sensitivity. It also provides the necessary selectivity to reduce the possibility of overload in the first amplifier stage caused by strong, undesired signals. Because sufficient selectivity must be provided to eliminate the image and other spurious signals prior to the mixer, preselection filtering may be broken into two or more parts with intervening amplifiers (in premium receiver designs) to minimize the effects of filter losses on the receiver *noise figure* (NF). The noise figure is a comparison in decibels (dB) of the noise power generated by a receiver with the noise power generated in an equivalent resistor.

The LO provides a strong stable signal at the proper frequency to the mixer for conversion of the desired signal to the IF. (The mixer stage may also be called the *first detector*, *converter*, or *frequency changer*.) The operating frequency of the LO may be controlled by a variable capacitor or varactor diode operating in a phase-locked-loop (PLL) system. Frequency synthesizer PLL designs are the basis for electronically-tuned radios (ETRs).

The output of the mixer is applied to the IF amplifier, which raises the desired signal to a suitable power level for the demodulator. The demodulator derives from the IF signal the modulated baseband waveform, which may be amplified by a baseband amplifier before being applied to a decoder (for stereo broadcasts) or audio preamplifier (for monophonic broadcasts).

12.3.2 Radio Wave Propagation

Radio waves are electromagnetic energy that can be attenuated, reflected, refracted, and scattered by changes in the media through which they propagate. Radio waves in free space have electric and magnetic field components that are mutually perpendicular and lie in a plane transverse to the direction of propagation. Radio waves travel at a velocity of about 300,000 km/s.

The received signal field is accompanied by noise generated as the result of natural sources (electrical activities in the atmosphere and space) and/or artificial sources (other radio transmissions and machinery of various types). In addition, the receiver itself is a source of noise. Electrical noise limits the range and performance of radio communications by requiring a sufficient signal strength at the receiver to overcome the undesired noise.

AM Band Propagation

AM broadcasting (530 to 1610 kHz) falls in the middle of the medium-frequency (MF) band, which extends from 300 kHz to 3 MHz. During daylight hours, ground wave propagation is predominate, and the effects of atmospheric noise are minimal. The receiver NF has little effect on overall signal quality unless the antenna system is very inefficient. At night, however, sky wave propagation is significant, permitting reception of signals hundreds or even thousands of miles away. Unfortunately, atmospheric noise is also greatest at night.

Fading at the receiver occurs in portions of a station's coverage area where both the ground wave and sky wave are comparable in strength. Fading can become quite deep during periods when the two waves are nearly equal. The ground wave will always reach the receiver through the most direct route. The sky wave will arrive later in time, out of phase with the ground wave, resulting in attenuation of the composite signal. Fading can also occur as a result of two or more sky waves with different numbers of reflections combining at the receiver.

When fading is caused by two or more waves that interfere as a result of having traveled over paths of different lengths, different frequencies within the transmitted spectrum can be attenuated to various extents. This phenomenon is known as *selective fading*. It can result in severe distortion of the signal.

Noise-free AM reception is particularly difficult at night because of the tremendous amount of noise present on the AM dial caused by sky wave interference from distant stations on the same or adjacent frequencies. The only solution to this problem is to increase the signal strength of the desired station at the antenna of the receiver. This essentially means that the effective service area of an AM station will be restricted during nighttime hours to locations in which a strong local ground wave signal is available.

FM Band Propagation

FM broadcasting (88 to 108 MHz) occupies a portion of the very-high-frequency (VHF) band, which extends from 30 to 300 MHz. For greatest reliability, most VHF communication relies on a line-of-sight path from the transmitter to the receivers. For this reason, FM transmitting antennas are located on high towers, tall buildings, or mountaintop sites. The wavelength at FM frequencies is sufficiently short that FM can penetrate automobiles and steel-frame buildings with a minimum of loss. FM is not affected by nighttime sky wave interference. The propagation properties of FM radio are essentially the same from day to night.

FM is, however, affected by reflections of the transmitted signal that can mix at the receiver and cause fading or distortion. The phenomenon is known as *multipath* (see Figure 12.3.2).

12-44 Radio Receivers

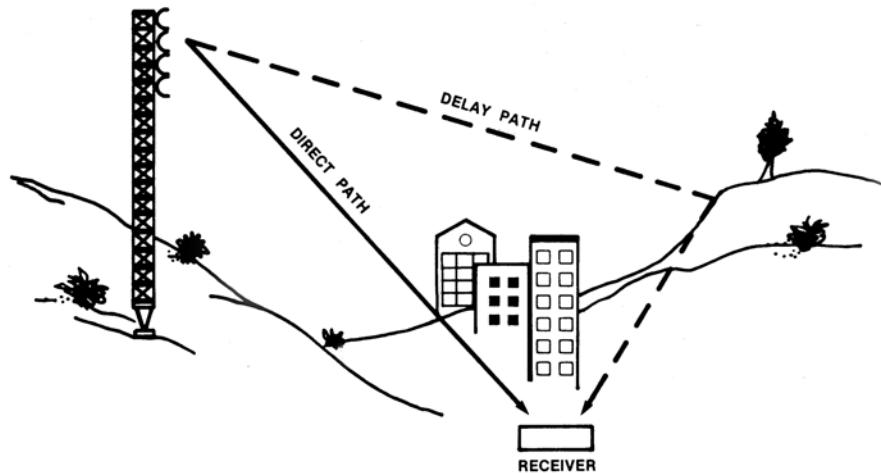


Figure 12.3.2 The causes of multipath in FM broadcasting. Distortion of the demodulated FM signal is related to both the direct-reflected signal ratio and the delay time of the reflected signal. Distortion increases as the signal ratio approaches 1:1, and as the secondary path delay increases.

Reflections can be caused by mountains, steel-frame buildings, vehicles, and other objects. The interference patterns set up as a result of multipath cause the signal strength to vary from one location to another in an apparently random manner. It is often possible, for example, to improve reception simply moving the antenna 2 ft or so (about 1/4 wavelength at FM frequencies).

The noise level at VHF is low compared to the MF band. Artificial noise can produce impulsive interference, but the use of hard limiting in receivers will eliminate most amplitude-related noise. Interference from other FM stations can, however, produce noise at the receiver that usually cannot be stripped off by limiting.

12.3.3 Radio Receivers

The design of an AM or FM receiver is a complicated process involving a huge number of technical, economic, and perceptual characteristics. It is not enough for a receiver to perform well—it must also meet a particular price point and respond to the needs (or wants) of the consumer. Take, for example, the small, portable ("Walkman"-type) AM-FM radios available today. The consumer wants a radio that is small and light enough to clip on a belt or slip into a shirt pocket. The radio does not need to be feature-rich, but it must perform well without any external antenna in a wide variety of working environments. The price point of the radio to consumers must be below \$20.

These specifications place enormous demands on the manufacturer. It is far easier to build a large, heavy, feature-rich radio that uses an outdoor antenna, is powered from the ac line, and can be priced upwards of \$400 than it is to build the radio just described. The design of any AM-FM radio for use by consumers is an exercise in compromise.

Most radios today have been reduced to just a handful of LSI chips, or even a single VLSI device. This dramatic move toward miniaturization has eliminated the traditional stages that technicians are familiar with. Still, each stage exists in one form or another in virtually every radio produced today. The circuits may be hidden on a slab of silicon, but they are there just the same.

12.3.3a Antenna Systems

The antenna must be designed to efficiently capture the transmitted signal and meet the physical limitations of the receiver. This is easier said than done. The ideal antenna for an AM radio would be a tall vertical wire or tower placed over a buried ground system. The FM equivalent would be a high-gain Yagi antenna mounted 10 to 30 ft above the ground. Such antennas are completely impractical for all but a few specialized, fixed installations. For most applications, especially mobile or portable radios, compromises in antenna design must be made.

Antenna efficiency, impedance, bandwidth, and reception pattern are all a function of the antenna dimensions relative to wavelength. A ferrite loop antenna is typically used for fixed and portable AM radios, while a short (less than 3 ft) telescoping vertical antenna is normally used for automobile AM radios. FM antennas usually consist of simply a length of wire (about 18 in) that is tacked up on a nearby wall or run along a baseboard for fixed-location receivers, or a short telescoping antenna for portable radios. Automobile FM radios use the AM auto antenna discussed previously. Shirt-pocket-sized radios simply rely on a length of wire run within the case of the unit.

None of the antennas just discussed are particularly effective in converting radio waves into electric signals. The constraints of space and cost have led manufacturers to sacrifice antenna efficiency for a compact, convenient product. Supplemental external antennas are used only in fixed applications requiring exceptional performance.

Antenna Coupling Network

Whatever form the antenna takes, it must be coupled efficiently to the first RF stage of the receiver. The coupling network must exhibit low loss and adequate bandwidth to cover the band selected (AM or FM). Until recently, it was customary practice to couple the antenna to a tuned circuit connected to the input RF amplifier. Figure 12.3.3 illustrates several common coupling arrangements. The examples differ mainly in the details of coupling. With ETR systems gaining wide acceptance, mechanical tuning is becoming obsolete. Voltage-tuned capacitors (varactors) are increasingly being used in radios today.

Whip Antenna

Nearly all automobile and some portable radios use whip antennas of about 2 to 3 ft in length for both AM and FM reception. Seldom does the mounting surface resemble a plane. The problem of coupling a whip optimally to the first active circuit is a difficult one. The antenna represents a complex set of impedance, capacitance, reactance, and resistive components that vary with the frequency and the surrounding physical structures. Still, given the ability to switch coupling circuitry between AM and FM bands, coupling to the RF amplifier can be satisfactory for reasonable efficiency and bandwidth.

12-46 Radio Receivers

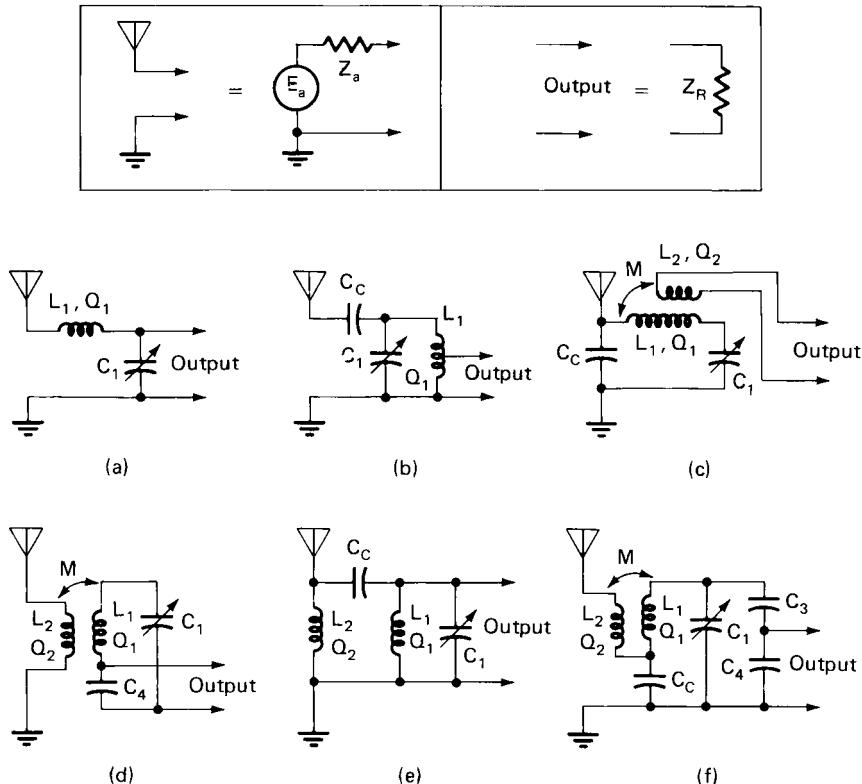


Figure 12.3.3 Typical circuits used for coupling an antenna to a tuned resonant circuit. (From [1]. Used with permission.)

Loop Antenna

The loop antenna has been used in portable AM receivers for many years. Its response differs from the monopole in that when the face of the loop is vertical, it responds to the magnetic field rather than the electric field. Instead of being omnidirectional in azimuth (like a whip), the loop responds to the cosine of the angle between its face and the direction of the desired transmission. This yields the familiar figure-eight pattern, which makes the loop useful for direction finding by providing a sharp null for waves arriving perpendicular to the face.

Loops used for AM broadcast reception incorporate a high-permeability (ferrite) core to facilitate reduced size. Such a loop may be tuned by a capacitance and connected directly to the input device of the receiver. Coupling is usually simple, as illustrated in Figure 12.3.4. If the loop has an inductance lower than that required for proper input impedance, it may be connected in series with an additional inductance for tuning (as shown). If the loop impedance is too high, the receiver input may be tapped down on the circuit.

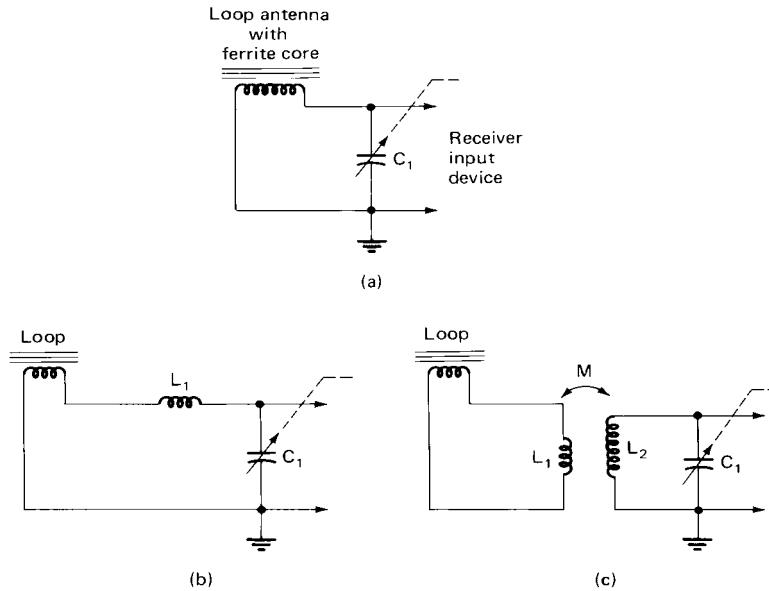


Figure 12.3.4 Examples of coupling circuits used for AM broadcast reception using a loop antenna. (From [1]. Used with permission.)

12.3.3b Filter Types

Filters are perhaps the most basic element of any receiver. If you are old enough to remember “crystal radio” hobbyist sets popular some years ago, you will recall that these childhood toys consisted of little more than a tuned circuit (filter) and detector. Although these sets have pretty much been retired to museums, the principle still holds true.

Filter design has come a long way within the past 10 to 20 years. Filters that once were used only in military applications because of their high cost are now produced by the tens of thousands. With reduced costs have come better designs. Radios today include filters that designers of consumer products could have only dreamed about 20 years ago.

Conventional filter design may be implemented using a number of different types of resonators. The principal available technologies include:

- LC filter
- Electrical resonator
- Quartz crystal filter
- Monolithic quartz filter
- Ceramic filter

The classical approach to radio filtering involved cascading single- or dual-resonator filters separated by amplifier stages. Overall selectivity was provided by this combination of one- or two-pole filters. This approach, however, had two distinct disadvantages: 1) the circuits were dif-

12-48 Radio Receivers

ficult to align properly, and 2) the system was susceptible to IM and overload (even in the early IF stages) from out-of-band signals. The classic approach did have some advantages, though. First, limiting from strong impulse noise would occur in early stages where the broad bandwidth would reduce the noise energy more than after complete selectivity had been achieved. Second, such designs were relatively inexpensive to manufacture.

Modern radios use multiresonator filters inserted as early as possible in the amplification chain to reduce nonlinear distortion, simplify alignment, and permit easy attainment of a variety of selectivity patterns. The simple single- or dual-resonator pairs are now used primarily for impedance matching between stages or to reduce noise between broadband cascaded amplifiers.

LC Filter

Inductor-capacitor resonators are limited to Q values on the order of a few hundred for reasonable sizes. In most cases, designers must be satisfied with rather low Q . The size of the filter depends on the center frequency. Two separate LC filters can easily cover the AM and FM broadcast bands. Skirt selectivity depends on the number of resonators used. Ultimate filter rejection can be made higher than 100 dB with careful design. Filter loss depends on the percentage bandwidth required and the resonator Q . It can be as high as 1 dB per resonator at narrow bandwidths.

This type of filter does not generally suffer from nonlinearities. Frequency stability is limited by the individual components and cannot be expected to achieve much better than 0.1 percent of center frequency under extremes of temperature and aging. Except for front ends that require broad bandwidth filters, *LC* filters have been largely superseded in modern radios by new filter technologies.

Quartz Crystal Filter

While several piezoelectric materials have been used for filter resonators, quartz crystal units have proved to be the most satisfactory. Filters are available in frequencies that can easily cover the AM broadcast band, and the FM band IF frequency (10.7 MHz). Standard filter response shapes are available to simplify the receiver design process and keep costs down. Ultimate filter rejection can be in excess of 100 dB. Input and output impedances are determined by input and output matching networks in the filters. Insertion loss varies from about 1 to 10 dB, depending on filter bandwidth and complexity. Frequency accuracy and temperature stability can be maintained to tight tolerances.

Monolithic Quartz Filter

A monolithic quartz filter is made up of a number of resonators constructed on a single quartz substrate. The principal energy of each resonator is confined primarily to the region between the plated electrodes, with a small amount of energy escaping to provide coupling. Usually these filters are limited to about four resonators. Monolithic quartz filters are available for operation through the FM band with characteristics generally similar to discrete quartz resonator filters, except that the bandwidth is limited to several tenths of a percent. Monolithic quartz filters are also smaller and lighter than discrete resonator filters.

Ceramic Filter

Piezoelectric ceramics can be used to achieve some of the characteristics of a quartz filter, but at a lower cost. Such filters are comparable in size to monolithic quartz filters but are available over a limited center frequency range (100 to 700 kHz). This limits ceramic filters to IF applications for AM reception (455 kHz). The cutoff rate, stability, and accuracy of a ceramic filter are not as good as quartz but are adequate for many applications. Single- and double-resonator structures are available. Multiple-resonator filters use electrical coupling between sections.

12.3.3c RF Amplifier and AGC

Depending on the receiver design, an amplifier stage may be used to increase the voltage of the signal received by the antenna (which may be 1 mV or less) to a level sufficient to drive the mixer. Because of the wide range of signals to which a receiver must respond, the input device must be capable of wide dynamic range. It must also be as linear as possible to minimize the generation of IM products from strong signals at the input. It follows that the number of strong signals should be minimized by restricting receiver bandwidth at as low a gain level as possible. Thus, gain should be low prior to the most narrow bandwidth stage in the receiver (the IF).

At AM frequencies, it is common practice to avoid RF amplification and use the mixer as the input device of the receiver. Bandwidth restriction is handled by filters in the first IF amplifier stage.

When the desired signal is relatively strong, the RF amplifier may raise it to a level that will cause distortion in later stages. Automatic gain control (AGC) is, therefore, provided for most RF amplifiers. AGC circuits are basically low-frequency feedback systems. They are necessary in a receiver to maintain a relatively constant output level when the input signal changes frequently. The successful design of an AGC circuit that will perform satisfactorily under all expected signal conditions is a major challenge.

Gain control is generally distributed over a number of stages. The gain in later IF stages is reduced first, and gain in earlier (RF and first IF) stages is reduced only for signal levels sufficiently high to assure the desired S/N. Variable gain amplifiers are controlled electronically. When attenuators are used in a receiver, they are usually operated electrically by either variable voltages for continuous attenuators or by electric switches (relays or diodes) for fixed or stepped attenuators.

The simplest method of gain control is to design one or more amplifier stages to change gain in response to a control voltage. One common configuration uses a differential pair of common-emitter amplifiers whose emitters are supplied through a separate common-emitter stage. The gain-control voltage is applied to the base of the latter stage, while the signal is applied to one (or both, if balanced) of the bases of the differential pair. This approach has been implemented in a variety of linear integrated circuits. Figure 12.3.5 shows a schematic diagram of one such IC. The gain-control voltage curves are given in Figure 12.3.6.

A PIN diode attenuator can also provide the low-distortion gain control that is especially important prior to the mixer. Figure 12.3.7 shows such a circuit and its control elements. The control curve has approximately linear decibel variation over most of its 60-dB range. The pi-type attenuator circuit provides a good match between terminations over the control range. The minimum useful frequency for a PIN diode attenuator varies inversely with the minority carrier lifetime. For available diodes, the low end of the HF band is near this limit.

12-50 Radio Receivers

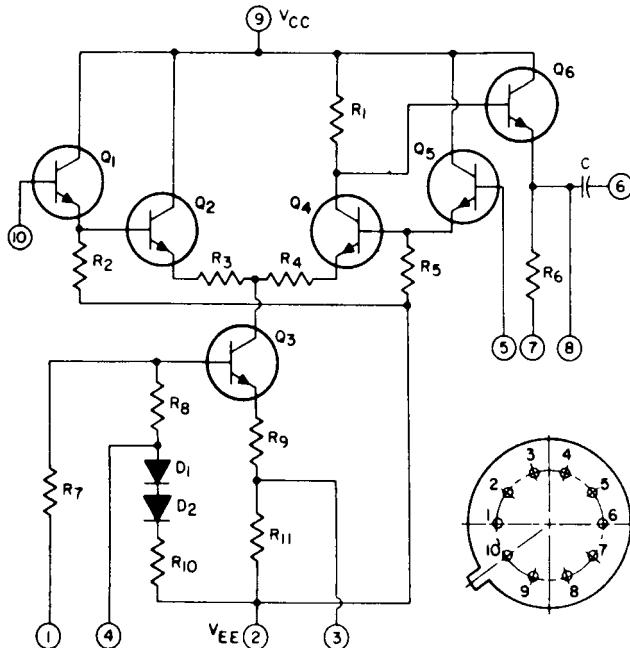


Figure 12.3.5 A gain-controlled RF amplifier integrated circuit suitable for use in AM-FM receivers.

The AGC Loop

The design objective of the AGC loop is to provide a substantially constant signal level to the demodulator despite changes in the input signal level. Figure 12.3.8 shows a simplified block diagram of a receiver with AGC. The input voltage to the RF amplifier stage may range between 1 μ V and 1 V. The envelope of this voltage is detected at the input to the detector stage. This signal is processed to produce the control voltages necessary for the variable-gain device (or devices) in the IF amplifier chain. This control voltage may be increased in level by an amplifier stage, as shown in Figure 12.3.9.

As the AGC tries to maintain constant output voltage with a varying input voltage, the issue of attack and decay times comes into play. The AGC system has a finite delay in its response to a change at the input of the receiver. In practice, it is not desirable for the AGC to have too fast a reaction time. In such a case, any static pulse, ignition noise, or other impulsive interference with a fast rise time would be detected by the AGC and desensitize the receiver for a given "hold time." The loop must be fast enough, though, to respond to signal level variations caused by terrain, artificial structures, and multipath.

12.3.3d Mixer

In the mixer circuit, the RF and LO signals are acted upon by the nonlinear properties of a device (or devices) to produce a third frequency, the IF. At AM frequencies, receivers are usually built

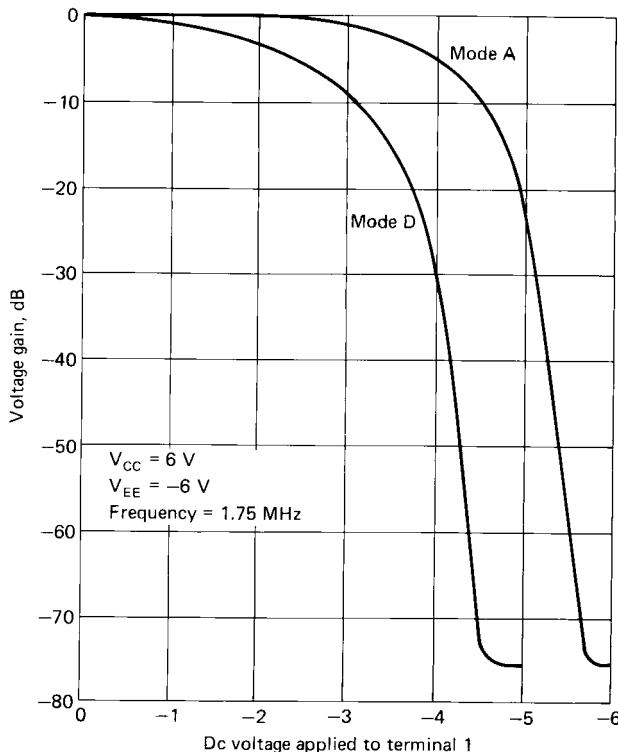


Figure 12.3.6 Gain-control curves of the CA3002 RF amplifier shown in Figure 12.3.5. The “Mode A” and “Mode B” traces refer to different circuit configurations for the device.

without RF preamplifiers; the antenna is fed directly to the mixer stage. In this frequency range, artificial and atmospheric noise is usually greater than the receiver NF. At FM frequencies, an RF amplifier is typically used. The mixer is located in the signal chain prior to the narrow filtering of the first IF.

Ideally, the mixer should accept the RF and LO inputs and produce an output having only one frequency (sum or difference), with signal modulation precisely transferred to this IF. In actual practice, however, mixers produce the desired IF but also many undesired components that must be filtered. Any device with nonlinear transfer characteristics can act as a mixer. For the purposes of this discussion, two classes will be discussed:

- Passive mixers, which use diodes as the mixing elements
- Active mixers, which employ gain devices (e.g., bipolar transistors or FETs)

12-52 Radio Receivers

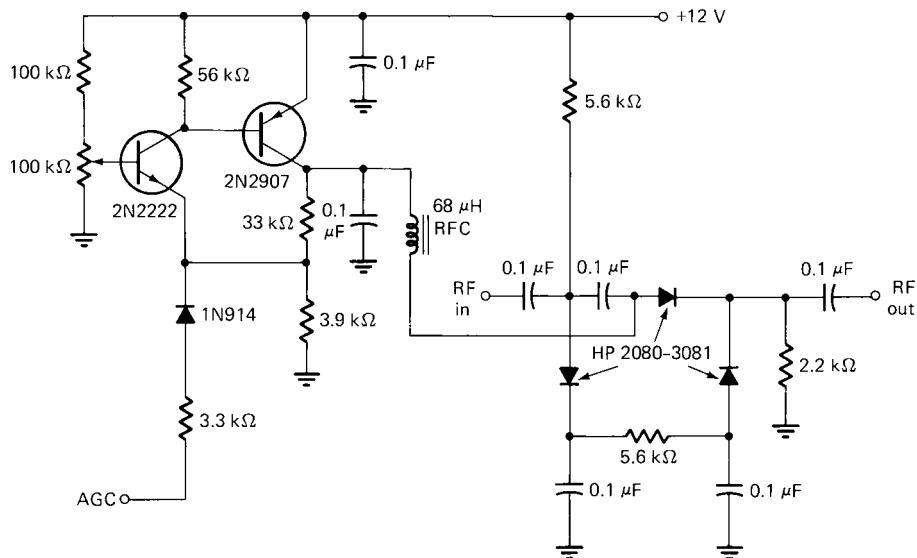


Figure 12.3.7 Schematic of a PIN diode attenuator. (From [1]. Used with permission.)

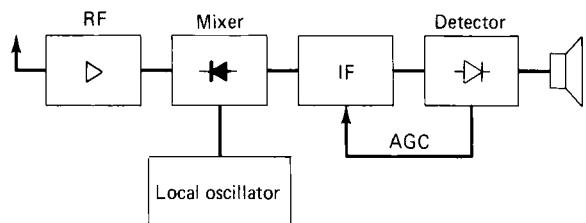


Figure 12.3.8 Simplified block diagram of a receiver with AGC. (From [1]. Used with permission.)

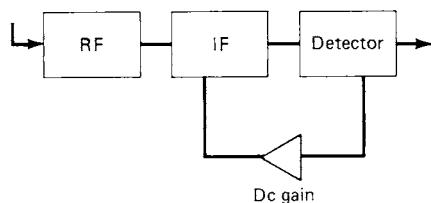


Figure 12.3.9 Block diagram of a receiver with amplified AGC. (From [1]. Used with permission.)

Passive Mixer

Passive mixers have been built using germanium and silicon diodes. The development of *hot carrier diodes*, however, has resulted in a significant improvement in passive mixers.

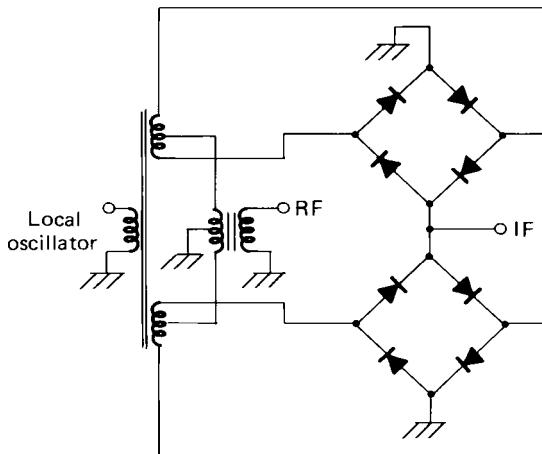


Figure 12.3.10 Schematic of a double-balanced passive mixer. (From [1]. Used with permission.)

A single diode can be used to build a mixer. The performance of such a circuit is poor, though, because the RF and LO frequencies (as well as their harmonics and other odd and even mixing products) all appear at the output. As a result, a large number of spurious components are produced that are difficult to remove. Moreover, there is no isolation of the LO and its harmonics from the input circuit, necessitating the use of an RF amplifier to prevent oscillator radiation from the antenna.

A better approach can be found in the double-balanced mixer, shown in Figure 12.3.10. This circuit, with its balanced diodes and transformers, cancels even harmonics of both RF and LO frequencies and provides isolation among the various ports. For optimum performance, careful matching of the diodes and transformers is necessary. The manufacturing process for hot carrier diodes has provided the tight tolerances that make them substantially better than other diode types available for mixer applications. Passive mixers can be described as low-, medium- or high-level mixers, depending on the diodes used and the number of diodes in the ring.

Active Mixer

The simplest type of active mixer uses a FET or bipolar transistor with the LO and RF signals applied to the gate-source or base-emitter junction. This unbalanced mixer has the same drawbacks as the simple diode mixer and is not used for high-performance receivers.

An improved configuration uses a dual-gate FET or cascode bipolar circuit with the LO and RF signals applied to different gates (bases). The balanced transistor arrangement of Figure 12.3.5 can also be used as a mixer with the LO applied to the base of Q3 and the signal applied to the bases of Q1 and/or Q5.

Active mixers can be implemented using a wide variety of devices and configurations, depending on the specifications and cost structure of the receiver. Figure 12.3.11 shows a push-pull balanced FET mixer. The circuit uses two dual-gate FETs in a push-pull arrangement

12-54 Radio Receivers

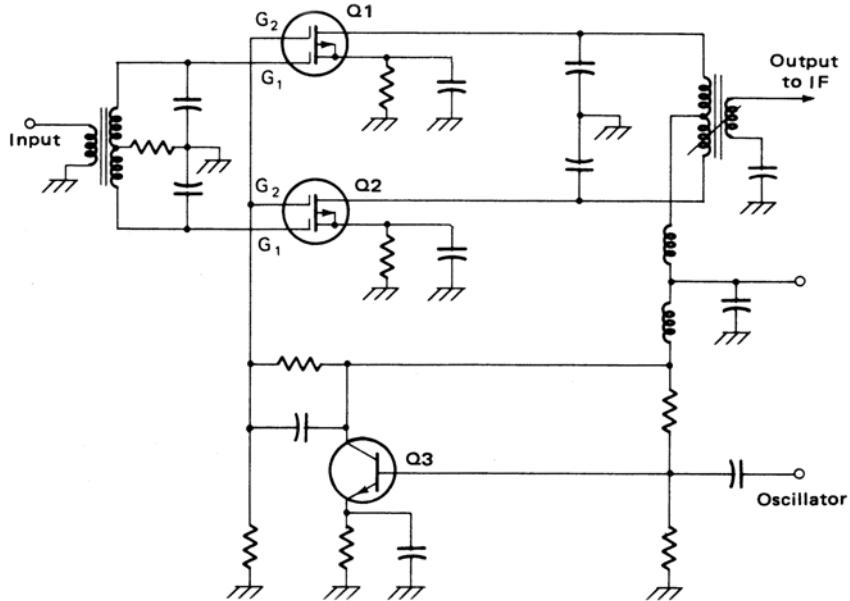


Figure 12.3.11 Schematic diagram of a push-pull dual gate FET balanced mixer.

between the RF input (applied to the first gates) and the IF output. The oscillator is injected in parallel on the second gates.

Active mixers have gain and are sensitive to mismatch conditions. If operated at high levels, the collector or drain voltage can become so high that the base-collector or gate-drain junction can open during a cycle and cause severe distortion. Control of the RF input by filtering out-of-band signals and AGC are important considerations for active mixer designs. Advantages of the active mixer include lower LO drive requirements and the possible elimination of an RF preamplifier stage.

12.3.3e Local Oscillator

Most modern high-performance receivers utilize a frequency-synthesized local oscillator to generate all individual frequencies needed over the required band(s). Synthesized oscillators generally use varactor diodes for the required variable-tuning capability. This arrangement offers a stable frequency source that is easily tunable in precise steps and is easily adapted to preprogrammed stations. The days of station presets based on moving electromechanical elements are, thankfully, gone for good. The LO must meet three basic requirements:

- Spectral purity to simplify IF filtering requirements
- Frequency agility to facilitate rapid station selection

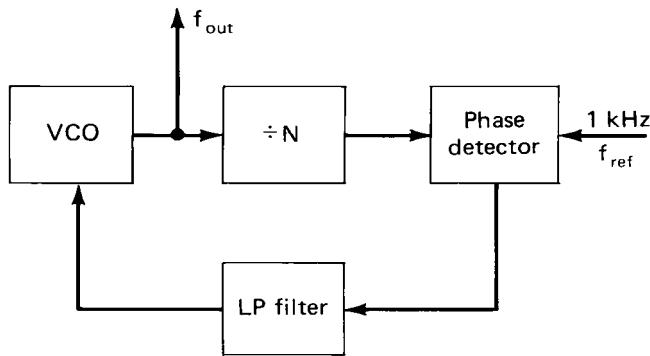


Figure 12.3.12 Block diagram of a single-loop frequency synthesizer.

- Frequency adjustment accuracy to match the center carrier frequencies of AM and FM broadcast stations

Synthesizers may categorized into two basic classes:

- Direct*, in which the LO output is derived from the product of multiple mixing and filtering
- Indirect*, in which the LO output is derived from a phase-locked loop that samples the direct output to reduce spurious signals.

PLL Synthesizer

Most AM-FM receivers incorporating a frequency synthesized LO use a single-loop digital PLL of the type shown in Figure 12.3.12. When describing frequency synthesizers mathematically, a linearized model is generally used. Because most effects occurring in the phase detector are highly nonlinear, however, only the so-called piecewise linear treatment allows adequate approximation. The PLL is nonlinear because the phase detector is nonlinear. However, it can be accurately approximated by a linear model when the loop is in lock.

Assume that the voltage-controlled oscillator (VCO) of Figure 12.3.12 is tunable over a range of 88 to 108 MHz. The output is divided to the reference frequency in a programmable divider stage whose output is fed to one of the inputs of the phase-frequency detector and compared with the reference frequency (fed to the other input). The loop filter at the output of the phase detector suppresses the reference frequency components, while also serving as an integrator. The dc control voltage output of the loop filter pulls the VCO until the divided frequency and phase equal those of the reference. A fixed division of the frequency standard oscillator (not shown in Figure 12.3.12) produces the reference frequency of appropriate step size. The operating range of the PLL is determined by the maximum operating frequency of the programmable divider, its division range ratio, and the tuning range of the VCO.

There are various choices of loop filter types and response. Because the VCO by itself is an integrator, a simple RC filter following the phase detector can be used. If the gain of the passive loop is too low to provide adequate drift stability of the output phase (especially if a high division ratio is used), an active amplifier may be used as an integrator. In most frequency synthesizers,

12-56 Radio Receivers

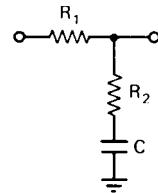


Figure 12.3.13 Schematic diagram of a PLL passive RC filter.

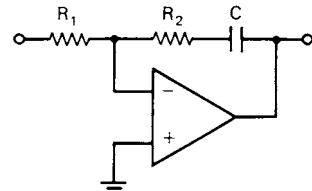


Figure 12.3.14 Schematic diagram of an active filter for a second-order PLL.

an active filter-integrator is preferred to a passive one. Figure 12.3.13 shows a passive RC filter for the second-order loop typically used in PLL synthesizers. Figure 12.3.14 shows an active filter for the second-order loop.

Frequency Divider

Frequency dividers are commonly built using transistor-transistor logic (TTL), complementary MOS (CMOS), and low-power emitter-coupled logic (ECL) IC technologies. Dividers come in two common categories: synchronous counters and asynchronous counters. The frequency range of the CMOS, depending on the process, is limited to 10 to 30 MHz. TTL operates successfully up to 100 MHz in a ripple counter configuration. In a synchronous counter configuration, TTL is limited to perhaps 30 MHz.

Frequency extension is possible through the use of an ECL prescaler, available in variable-ratio and fixed-ratio configurations. The term *prescaling* is generally used in the sense of a predivider that is nonsynchronous with the rest of the chain. Fixed-ratio prescalers are used as ripple counters preceding a synchronous counter. A single-loop synthesizer loses resolution by the amount of prescaling.

Figure 12.3.15 shows a block diagram of the MC12012 (Motorola) variable-ratio dual-modulus presealer. Through external programming, this ECL divider can be made to divide in various ratios. With proper clocking, the device can be considered a synchronous counter. With such a system, it is possible to increase the maximum operating frequency to about 400 MHz without losing resolution.

Variable-Frequency Oscillator

The LO in the receiver must be capable of being turned over a specified frequency range, offset from the desired operating band(s) by the IF. Prior to the advent of the varactor diode and good switching diodes, it was customary to tune an oscillator mechanically using a variable capacitor with an air dielectric, or in some cases by moving a powdered iron core inside a coil to make a variable inductor. Automobile radios typically used the variable-inductor method of tuning the

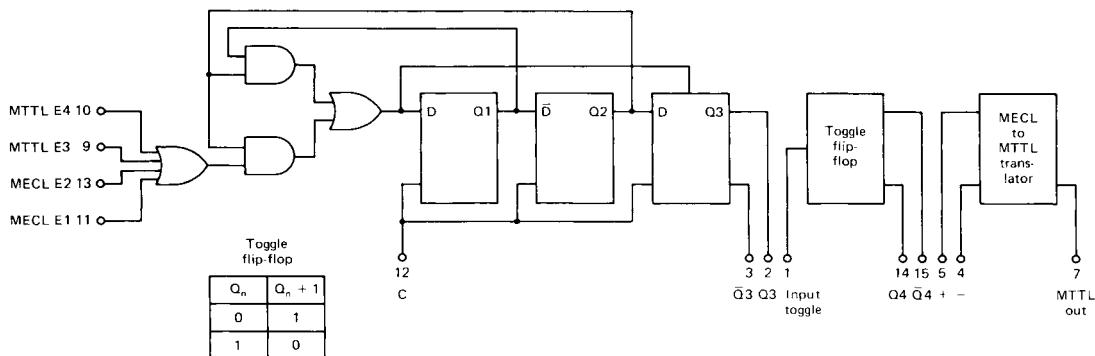


Figure 12.3.15 Block diagram of a divide-by-10/11 dual-modulus prescaler IC, the Motorola MC12012. (Courtesy of Motorola.)

AM broadcast band. Figure 12.3.16 shows the classic VFO circuits commonly used in receivers. Different configurations are used in different applications, depending on the range of tuning and whether the tuning elements are completely independent or have a common element (such as the rotor of a tuning capacitor).

Newer receivers control the oscillator frequency by electronic rather than mechanical means. Tuning is accomplished by a voltage-sensitive capacitor (varactor diode). Oscillators that are tuned by varying the input voltage are referred to as *voltage-controlled oscillators* (VCOs).

The capacitance versus voltage curves of a varactor diode depend on the physical composition of the diode junction. Maximum values range from a few hundred picofarads, and useful capacitance ratios range from about 5 to 15. Figure 12.3.17 shows three typical tuning circuits incorporating varactor diodes. In all cases the voltage is applied through a large value resistor.

Diode Switching

Because diodes have a low resistance when biased in one direction and a very high resistance when biased in the other, they may be used to switch RF circuits. A sufficiently large bias voltage may be applied to keep the diode *on* when it is carrying RF currents, or off when it is subjected to RF voltages. It is important that, in the forward-biased condition, the diode add as little resistance as possible to the circuit and that it be capable of handling the maximum RF current plus the bias current. When the diode is reverse-biased, the breakdown voltage must be higher than the combined bias and RF peak voltage in the circuit. Almost any type of diode can perform switching, but at high frequencies, PIN diodes are especially useful. Figure 12.3.18 shows three examples of diode switching in RF circuits.

The advantage of electronic tuning using varactor diodes is only fully realized when band selection also takes place electronically. Diode switches are preferable to mechanical switches because of their high reliability. Diode switches eliminate the need for a mechanical link between front panel controls and the tuned circuits to be switched.

12-58 Radio Receivers

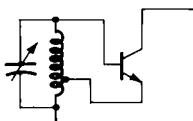
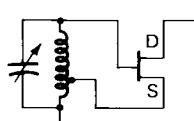
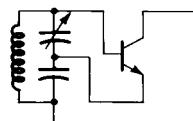
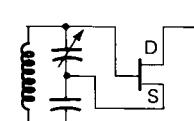
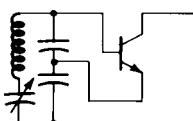
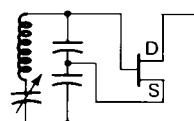
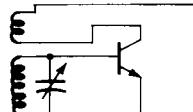
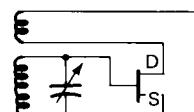
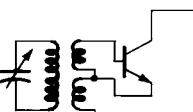
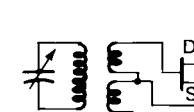
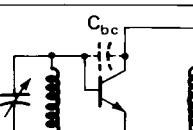
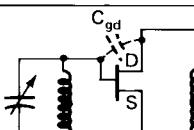
Oscillator type	Bipolar transistor RF circuit	FET RF circuit
Hartley		
Colpitts		
Clapp (GOURIET)		
Transformer feedback		
Meissner		
Tuned input/tuned output		

Figure 12.3.16 Schematic diagrams of common oscillator circuits using vacuum-tube, transistor, and FET active circuits. (From [1]. Used with permission.)

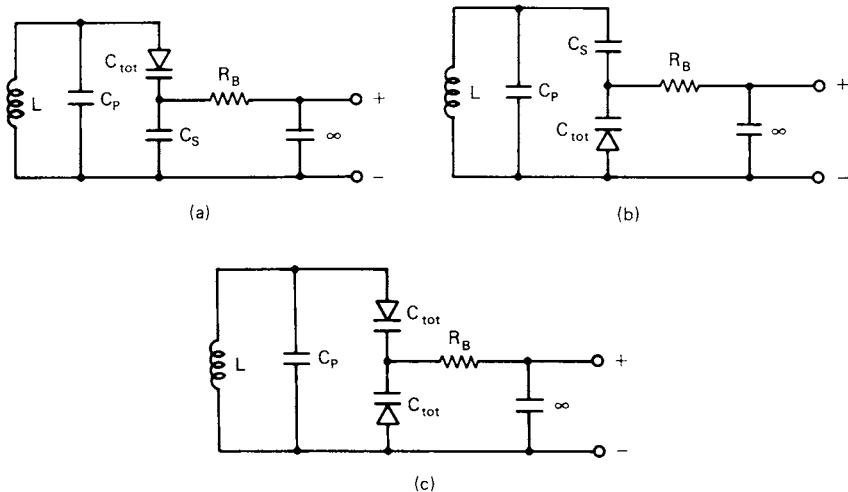


Figure 12.3.17 Typical tuning circuits using varactor diodes as the control element: (a) single diode in the circuit low side, (b) single diode in the circuit high side, (c) two diodes in a series back-to-back arrangement. (From [2]. Used with permission.)

Crystal-Controlled Oscillator

Piezoelectric quartz crystals are the basis for most PLL reference oscillators. Quartz crystals have resonances that are much more stable than the LC circuits discussed so far and also have very high Q . Consequently, quartz crystal resonators are typically used for high-stability fixed-frequency oscillators. A piezoelectric material is one that develops a voltage when it is under a mechanical strain or is placed under strain by an applied voltage. A physical piece of such material, depending upon its shape, can have a number of mechanical resonances. By appropriate shaping and location of the electrodes, one or another resonant mode of vibration can be favored, so that the resonance may be excited by an external voltage.

The crystal exhibits at its frequency of oscillation the equivalent electric circuit shown in Figure 12.3.19. The series resonant circuit represents the effect of the crystal vibrator, and the shunt capacitance is the result of the coupling plates and of capacitance to surrounding metallic objects (such as the metal case). The resonant circuit represents the particular vibrating mode that is excited. If more than one mode can be excited, a more complex circuit would be required to represent the crystal.

The most common type of circuit using a fundamental (AT) crystal is an *aperiodic oscillator*, which has no selective circuits other than the crystal. Such oscillators, often referred to as *parallel resonant oscillators*, use the familiar Pierce and Clapp configurations (see Figure 12.3.20).

12.3.3f AM-FM Demodulation

The function of any receiver is to recover the original information used to modulate the transmitter. This process is referred to as *demodulation*, and the circuits that perform the recovery are

12-60 Radio Receivers

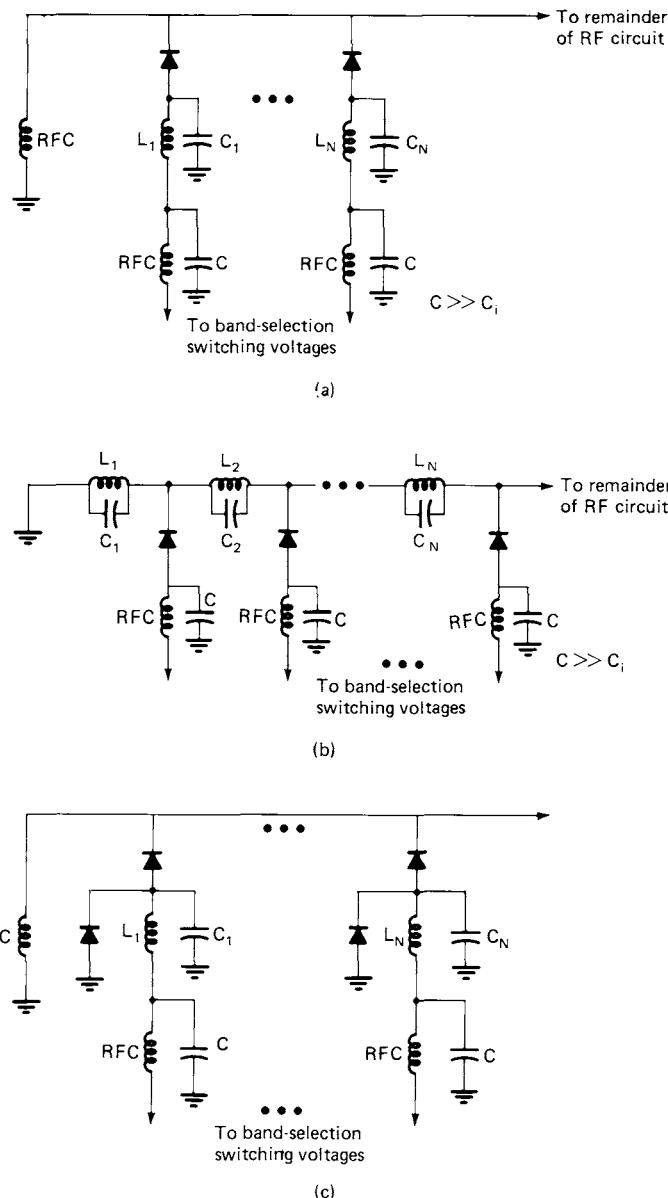


Figure 12.3.18 Typical circuits using diodes for band switching: (a) series diode arrangement, (b) shunt-diode arrangement, (c) use of both series and shunt diodes. (From [1]. Used with permission.)

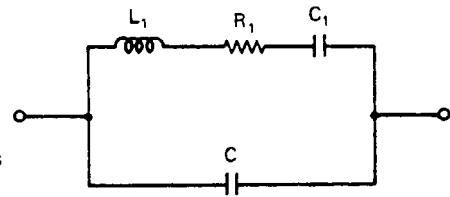


Figure 12.3.19 The equivalent electric circuit of a crystal at resonance (spurious and overtone modes not shown).

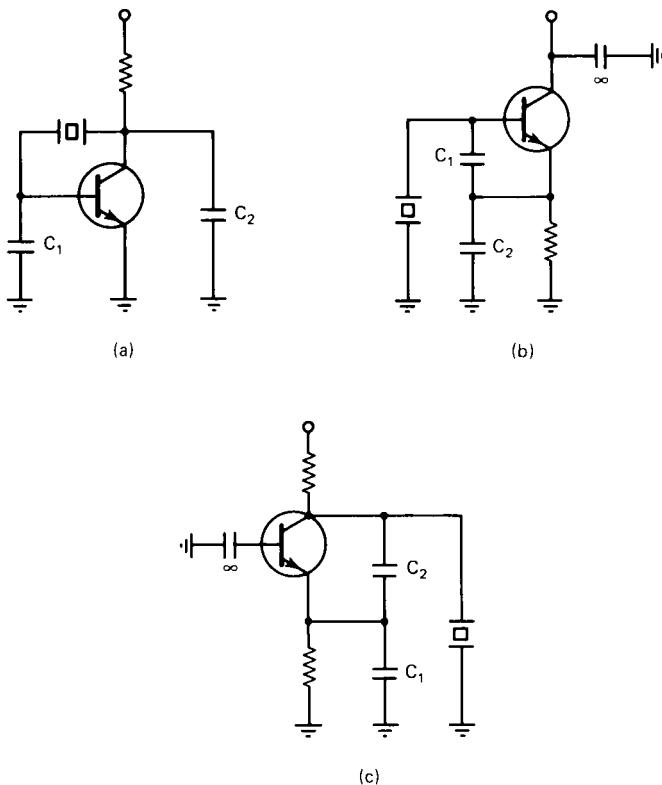


Figure 12.3.20 Common parallel resonant circuits used in fundamental crystal oscillators: (a) Pierce circuit, (b) Clapp circuit, collector grounded, (c) Clapp circuit, base grounded. (From [2]. Used with permission.)

called *demodulators*. The term *detector* is also used, and the demodulator in a single-superheterodyne receiver is sometimes called a *second detector*. Today, however, the term *detector* is seldom used in this fashion.

Because of transmission and reception system distortions and noise caused by thermal, atmospheric, and artificial sources, the demodulated signal is—to some extent—a distorted version of

12-62 Radio Receivers

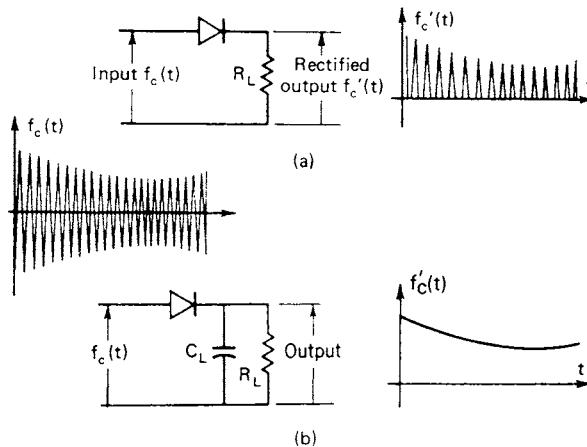


Figure 12.3.21 AM demodulators with idealized waveforms: (a) average demodulator and resulting waveform, (b) envelope detector and resulting waveform. (From [1]. Used with permission.)

the original modulating signal and is corrupted by the addition of noise. It is the obvious goal of the demodulator to minimize these corrupting effects and provide an output signal that is as close to the original modulating waveform as possible.

Analog modulated waves come in a variety of forms, including:

- Conventional AM
- Double-sideband suppressed carrier (DSSC) AM
- Single-sideband (SSB) AM
- Vestigial-sideband (VSB) AM
- Phase modulation (PM)
- Frequency modulation (FM)

For the purposes of this discussion, we will concentrate on the two forms used for AM-FM broadcast transmission.

AM Demodulation

An AM signal is made up of an RF sinusoid whose envelope varies at a relatively slow rate about an average (carrier) level. Any sort of rectifier circuit will produce an output component at the modulation frequency. Figure 12.3.21 illustrates two of the simple diode rectifier circuits that may be used, along with idealized waveforms. The average output of the rectifier of Figure 12.3.21a is proportional to the carrier plus the signal. The circuit exhibits, however, significant output energy at the RF and its harmonics. A low-pass filter is necessary to eliminate these components. If the selected filter incorporates a sufficiently large capacitor at its input, the effect is to produce a peak rectifier, with the idealized waveform shown in Figure 12.3.21b. In this case

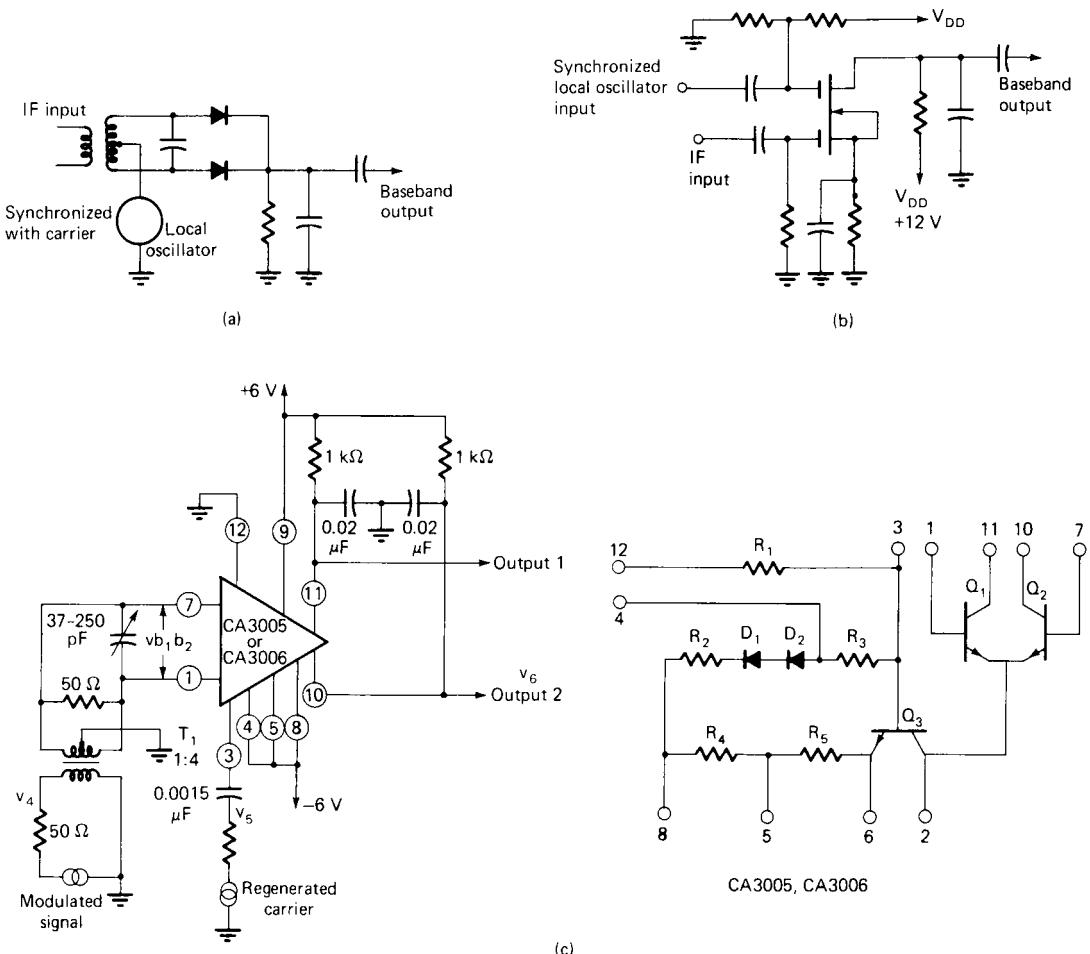


Figure 12.3.22 Three types of synchronous demodulators: (a) diode-based circuit, (b) dual-gate MOSFET-based circuit, (c) bipolar IC (CA3005/CA3006) basic circuit.

the demodulated output is increased from the average of a half a sine wave (0.637 peak) to the full peak, and the RF components are substantially reduced. A peak rectifier used in this way is often referred to as an envelope detector or demodulator. It is the circuit most frequently used for demodulating AM broadcast signals.

AM signals may also be demodulated by using a coherent or synchronous demodulator. This type of demodulator uses a mixer circuit, with an LO signal synchronized in frequency and phase to the carrier of the AM input. Figure 12.3.22 illustrates three approaches to synchronous demodulation. The synchronous component can be generated by an oscillator phase-locked to the carrier, as illustrated in Figure 12.3.23.

12-64 Radio Receivers

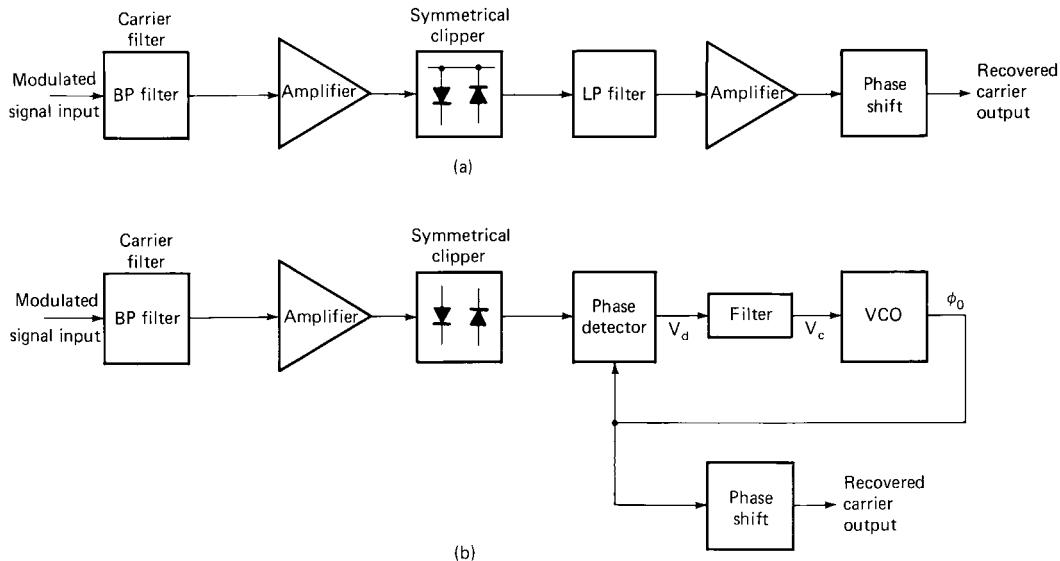


Figure 12.3.23 Two approaches to AM carrier recovery: (a) system based on a filter, clipper, and amplifier; (b) system based on a filter, clipper, and PLL. (From [1]. Used with permission.)

The synchronous demodulator translates the carrier and sidebands to baseband. As long as the LO is locked to the carrier phase, baseband noise results only from the in-phase component of the noise input. Consequently the noise increase and S/N reduction that occur at low levels in the envelope demodulator are absent in the synchronous demodulator. The recovered carrier filtering is narrow band, so that phase lock can be maintained at carrier-to-noise levels below useful modulation output levels. This type of circuit, while better than an envelope demodulator, is not generally used for AM broadcast demodulation because of its complexity. Most stereo AM receivers, however, incorporate synchronous demodulators as part of the decoding circuit.

FM Demodulation

The most common technique for FM demodulation incorporates the use of linear circuits to convert frequency variations to envelope variations, followed by an envelope detector. Another technique used with linear integrated circuits involves the conversion of frequency variations to phase variations that are then applied to a phase demodulator. Still other FM demodulators employ PLLs and frequency-locked loops (FM feedback circuits), or counter circuits whose output is proportional to the rate of zero crossings of the wave. Frequency demodulators are often referred to as discriminators or frequency detectors.

Resonant circuits are used in discriminators to provide adequate sensitivity to small-percent-age frequency changes. To eliminate the dc components, two circuits can be used, one tuned above and one tuned below the carrier frequency. The outputs are demodulated by envelope

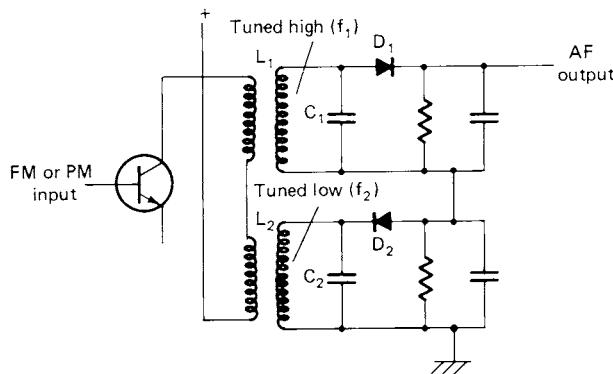


Figure 12.3.24 Schematic diagram of the Travis discriminator. (From [3]. Used with permission.)

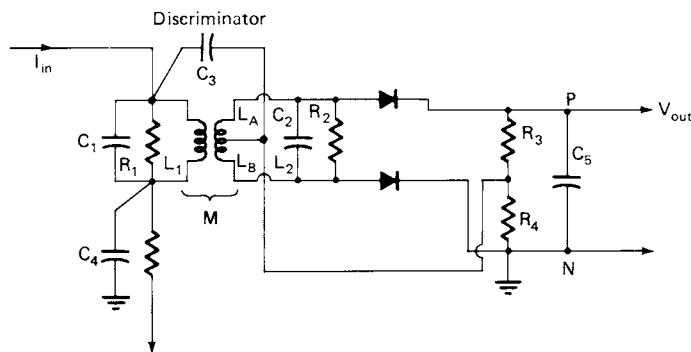


Figure 12.3.25 The Foster-Seeley FM discriminator circuit with tuned primary. (From [1]. Used with permission.)

demodulators and are then subtracted, eliminating the dc component. Voltage sensitivity is doubled compared to the use of a single circuit. This balanced arrangement also eliminates all even-order distortion so that the first remaining distortion term is third-order. Figure 12.3.24 shows one implementation of this scheme, known in the U.S. as the Travis discriminator. Because the circuit depends on the different amplitude responses of two circuits, it has sometimes been called an amplitude discriminator.

The Foster-Seeley discriminator, shown in Figure 12.3.25, is a more common approach to FM demodulation. In this circuit, the voltage across the primary is added to the voltage across each of the two halves of the tuned secondary. At resonance the secondary voltage is in quadrature with the primary voltage, but as the frequency changes, so do the phase shifts. The voltages from the upper and lower halves of the secondary add to the primary voltage in opposition. As the frequency rises, the phase shift increases, and as the frequency falls, it decreases. The opposite phase additions cause the resulting amplitudes of the upper and lower voltages to differ, produc-

12-66 Radio Receivers

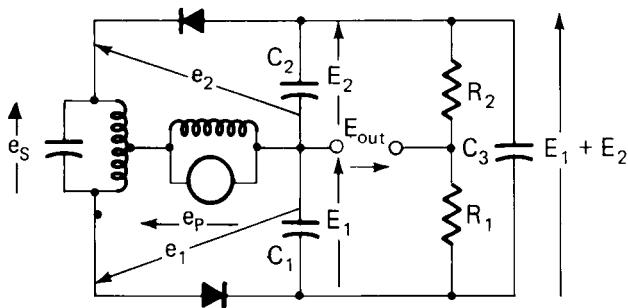


Figure 12.3.26 Schematic diagram of a basic ratio detector circuit. (From [1]. Used with permission.)

ing the discriminator effect. When the primary circuit is also tuned to the center frequency (which produces much higher demodulation sensitivity), the phase of the primary voltage also varies slightly, as does its amplitude. The proper selection of coupling factor is required to produce optimum sensitivity and linearity of the discriminator. Because of the method of arriving at the amplitude difference in this demodulator, it is sometimes referred to as a phase discriminator.

The ratio detector is a variant of the phase discriminator that has an inherent degree of AM suppression. The circuit tolerates less effective limiting in prior circuits and thus can reduce the cost of the receiver. Figure 12.3.26 shows the basic concept of the ratio detector. It resembles the Foster-Seeley circuit, except that the diodes are reversed. The combination of R_1 , R_2 , and C_3 has a time constant that is long compared to the lowest modulating frequency (on the order of 0.1 s for audio modulation). The result is that during modulation, the voltage to the (grounded) center tap across load resistor R_2 is $(E_1 + E_2)/2$, and across R_1 it is $-(E_1 + E_2)/2$. Following the circuit from ground through R_2 and C_2 , we see that the voltage at the center tap of the capacitors is $[(E_1 + E_2)/2] - E_2 = (E_1 - E_2)/2$, or half of the value of the Foster-Seeley discriminator.

The long time constant associated with C_3 reduces the required current from the diodes when $E_1 + E_2$ drops and increases it when $E_1 + E_2$ rises. This changes the load on the RF circuit and causes higher drive when the output falls and lower drive when it rises. This tends to further stabilize the voltage $E_1 + E_2$ against incidental AM. The sum voltage can also be used to generate an AGC signal, so that the prior circuits need not limit. This can be advantageous when a minimum number of circuits is required and the selectivity is distributed.

Amplitude Limiter

Amplitude limiting is essential for FM demodulators using analog circuits. Although solid-state amplifiers tend to limit when the input signal becomes excessive, limiters that make use of this characteristic often limit the envelope asymmetrically. For angle demodulation, symmetrical limiting is desirable. AGC circuits, which can keep the signal output constant over wide ranges of input signals, are unsuitable for limiting because they cannot be designed with sufficiently rapid response to eliminate the envelope variations encountered in angle modulation interference. One or more cascaded limiter stages are required for good FM demodulation.

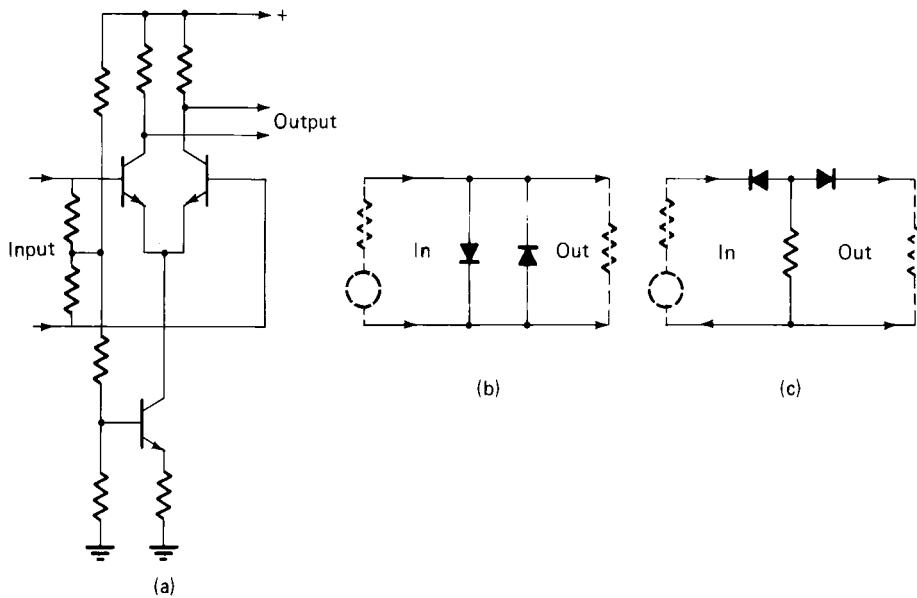


Figure 12.3.27 Typical FM limiter circuits: (a) balanced-transistor amplifier, (b) shunt-diode limiter, (c) series diode limiter. (From [1]. Used with permission.)

Almost any amplifier circuit, when sufficiently driven, provides limiting. However, balanced limiting circuits produce better results than those that are not balanced. In general, current cutoff is more effective than current saturation in producing sharp limiting thresholds. Nonetheless, overdriven amplifiers have been used in many FM systems to provide limiting. If the amplifier is operated with low supply voltage and near cutoff, it becomes a more effective limiter. The transistor differential amplifier shown in Figure 12.3.27a is an excellent limiter when the bias of the emitter load transistor is adjusted to cause cutoff to occur at small base-emitter input levels.

The classic shunt-diode limiter is shown in Figure 12.3.27b. It is important that the off resistance of the diodes be much higher than the driving and load impedances, and the on resistance be much lower. Figure 12.3.27c shows the classic series diode limiter. In this example the diodes are normally biased on, so that they permit current flow between the driver and the load. As the RF input voltage rises, one diode is cut off—and as it falls, the other is cut off. The effectiveness of limiting is determined by the difference in off and on resistances of the diode, compared to the driving and load impedances.

12-68 Radio Receivers

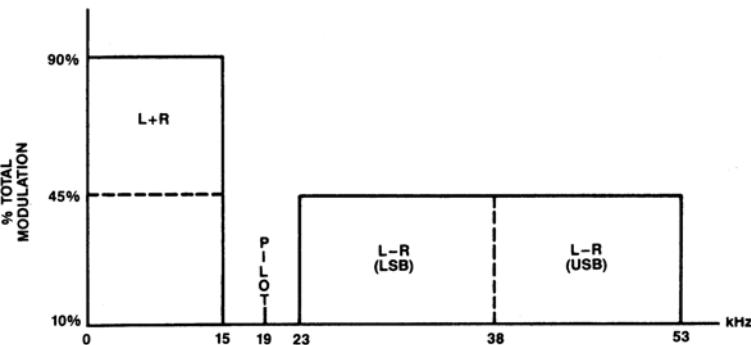


Figure 12.3.28 The composite baseband signal of the FM stereo system.

12.3.4 Stereo Systems

Stereophonic broadcasting has reshaped the radio industry. Stereo FM, introduced during the early 1960s, gave FM broadcasters a powerful new marketing tool. Receiver manufacturers responded with new radios that offered high-quality performance at an affordable price.

Although few people outside the broadcast industry know it, AM broadcasters had been talking about stereo operation as long as FM engineers. It was not until the early 1980s, however, before AM stereo broadcasting became a reality.

12.3.4a FM Stereo

The system devised for broadcasting stereo audio over FM has served the industry well. Key requirements for the scheme were: 1) compatibility with monophonic receivers that existed at the time the standard was developed, and 2) a robust signal that would not be degraded significantly by multipath. Figure 12.3.28 shows the composite baseband that modulates the FM carrier for stereophonic broadcasting. The two-channel baseband has a bandwidth of 53 kHz and is made up of:

- A main channel ($L + R$) signal, which consists of the sum of left plus right audio signals (the same signal broadcast by a monaural station). A fully modulated main channel will modulate the FM transmitter to 45 percent when broadcasting stereo programming.
- A stereophonic subchannel ($L - R$), which consists of a double-sideband AM modulated carrier with a 38-kHz center frequency. The modulating signal is equal to the difference of the left and right audio inputs. The subcarrier is suppressed to conserve modulation capability. As a result, the AM sidebands have the same modulation potential as the main channel. A fully modulated subchannel will modulate the FM transmitter to 45 percent when broadcasting stereo programming.
- A 19-kHz subcarrier pilot, which is one-half the frequency of the stereophonic subcarrier and in phase with it. The pilot supplies the reference signal needed by stereo receivers to reinsert the 38-kHz carrier for demodulation of the double-sideband suppressed carrier transmission. The pilot, in other words, is used to synchronize the decoder circuitry in the receiver to the

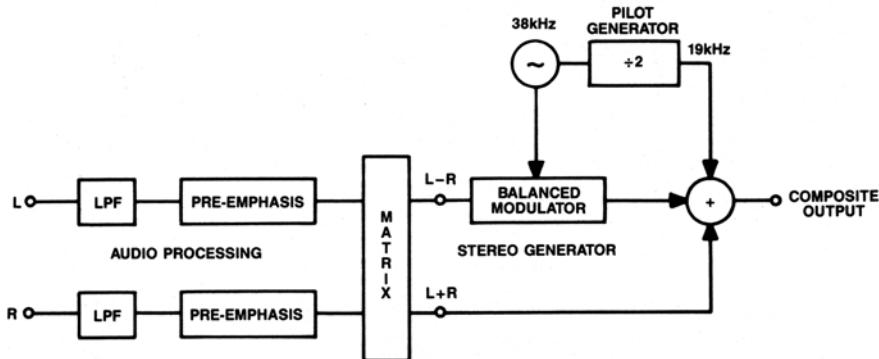


Figure 12.3.29 Functional block diagram of a frequency-division multiplexing (FDM) FM stereo generator.

stereo generator at the transmitter. The frequency tolerance of the pilot is ± 2 Hz. The pilot modulates the transmitter 8 to 10 percent.

Generating the Stereo Signal

Two basic approaches have been used to generate the stereophonic subchannel: *time-division multiplexing* (TDM), the switching method; and *frequency-division multiplexing* (FDM), the matrix method.

A simplified block diagram of the FDM approach is shown in Figure 12.3.29. The left and right audio channels first pass through a preemphasis circuit and low-pass filter. They are then supplied to the matrix, which produces sum and difference components. The audio signals are added to form the L + R main channel signal.

The difference signal is fed to a balanced modulator that generates the L - R subchannel. Because a balanced modulator is used, the 38-kHz carrier is suppressed, leaving only the modulated sidebands. The 19-kHz pilot signal is derived by dividing the 38-kHz oscillator by 2. The main channel, stereophonic sub-channel, and pilot are then combined in the proper (45/45/10 percent) ratio to form the composite baseband.

The TDM method of generating a stereo signal is shown in block diagram form in Figure 12.3.30. The L + R and L - R signals are generated by an electronic switch that is toggled at a 38-kHz rate. The switch samples one audio channel and then the other. Considerable harmonic energy is generated in this process, requiring the use of a low-pass filter. When the harmonics are filtered out, the proper composite waveform results. This approach, while simple and stable, may produce unwanted artifacts, most notably reduced stereo separation, because of the filtering requirements.

An improvement to the basic TDM concept is shown in Figure 12.3.31. By using a soft switch to sample the left and right channels, it is possible to eliminate the low-pass filter and its side-effects. The variable element shown in the figure consists of an electronic attenuator that is capable of swinging between its minimum and maximum attenuation values at a 38-kHz rate. Like the fast-switching TDM system, the L + R and L - R channels are generated in one operation. No

12-70 Radio Receivers

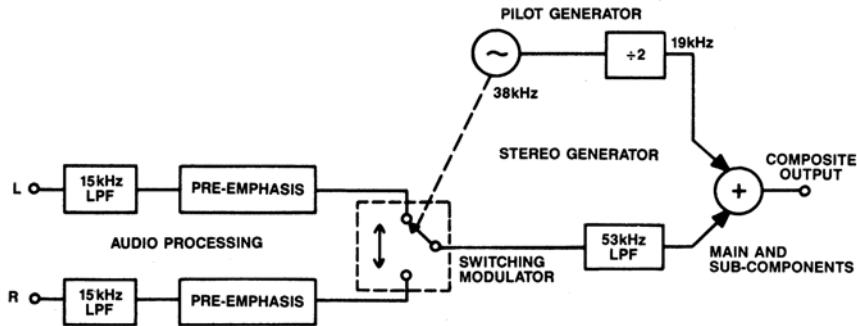


Figure 3.30 Functional block diagram of a time-division multiplexing (TDM) FM stereo generator.

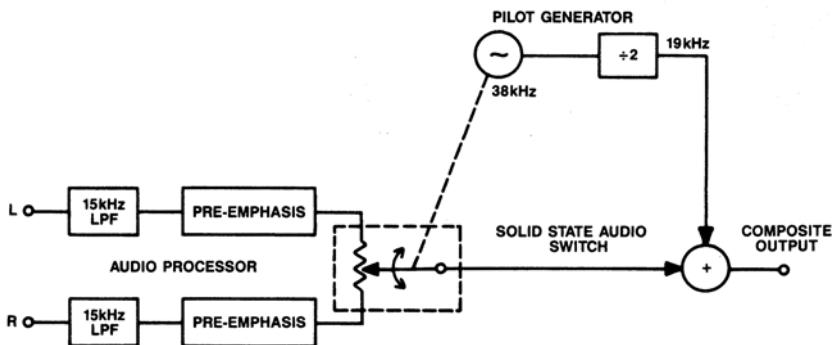


Figure 3.31 Functional block diagram of a time-division multiplexing stereo generator using a variable electronic attenuator.

filter is required at the output of the generator as long as the 38-kHz sine wave is free from harmonics and the variable attenuator has good linearity.

Decoding the Stereo Signal

All stereo FM receivers include a circuit to convert the multiplexed signal at the FM detector to the left and right audio channels originally transmitted by the station. There are a number of ways to accomplish this task. In practice, however, one type of decoder is most commonly found, built around a PLL integrated circuit (see Figure 12.3.32). This system offers both high performance and low cost.

The composite signal from the demodulator is fed to a buffer amplifier and sampled by a PLL within the decoder IC. A voltage controlled oscillator, typically running at 76 kHz (four times the pilot frequency) is locked in phase with the pilot by the error output voltage of the PLL. The oscillator signal is divided by 2, resulting in a square wave at 38 kHz with nearly perfect duty cycle and fast rise and fall times. This signal drives the audio switcher (demultiplexer) to transfer

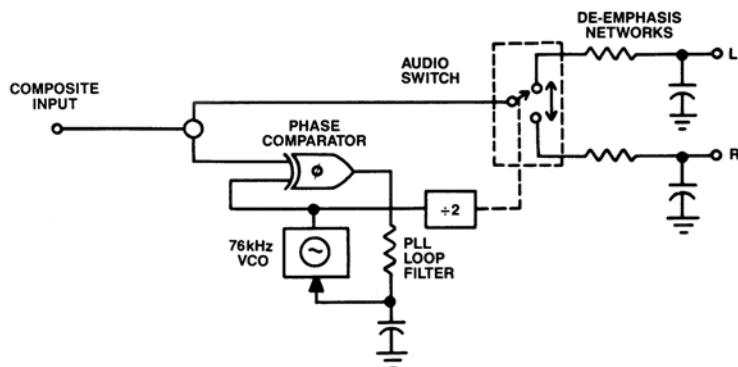


Figure 12.3.32 Block diagram of a stereo decoder using PLL-controlled time-division multiplexing.

the composite baseband to the left and right audio outputs in synchronization with the station's stereo generator. A deemphasis circuit follows the matrix to complement the signal preemphasis at the FM transmitter.

12.3.4b AM Stereo

AM stereo operation was approved by the FCC in 1981 using the C-QUAM (Motorola) system. A modification of simple quadrature modulation, C-QUAM was designed to maintain monophonic compatibility.

The C-QUAM encoder is shown in Figure 12.3.33. As in FM stereo broadcasting, sum and difference signals of the left and right audio inputs are produced. Pure quadrature is generated by taking the $L + R$ and $L - R$ signals and modulating two balanced modulators fed with RF signals that are out of phase by 90° (producing components referred to as I and Q). As shown in the figure, the 90° phase shift is derived by using a Johnson counter, which divides an input frequency (4 times the station carrier frequency) by 4 and provides digital signals precisely 90° out of phase for the balanced modulators. The carrier is inserted directly from the Johnson counter. At the output of the summing network, the result is a pure quadrature AM stereo signal. From there it is passed through a limiter that strips the incompatible AM components from the signal. The output of the limiter is amplified and sent to the transmitter in place of the crystal oscillator.

The left and right audio signals are summed and sent as compatible $L + R$ to the audio input terminals of the transmitter.

Decoding the C-QUAM Signal

C-QUAM AM stereo is decoded by converting the demodulated AM broadcast waveform (which is already close to a quadrature signal) to pure quadrature and then using a quadrature detector to extract the $L - R$ component (see Figure 12.3.34). In order to prepare the received signal for the quadrature demodulator, it has to be converted from the envelope-detector-compatible signal that is broadcast to the original quadrature signal that was not envelope detector compatible. This is accomplished by demodulating the broadcast signal in two ways: with an envelope detector and

12-72 Radio Receivers

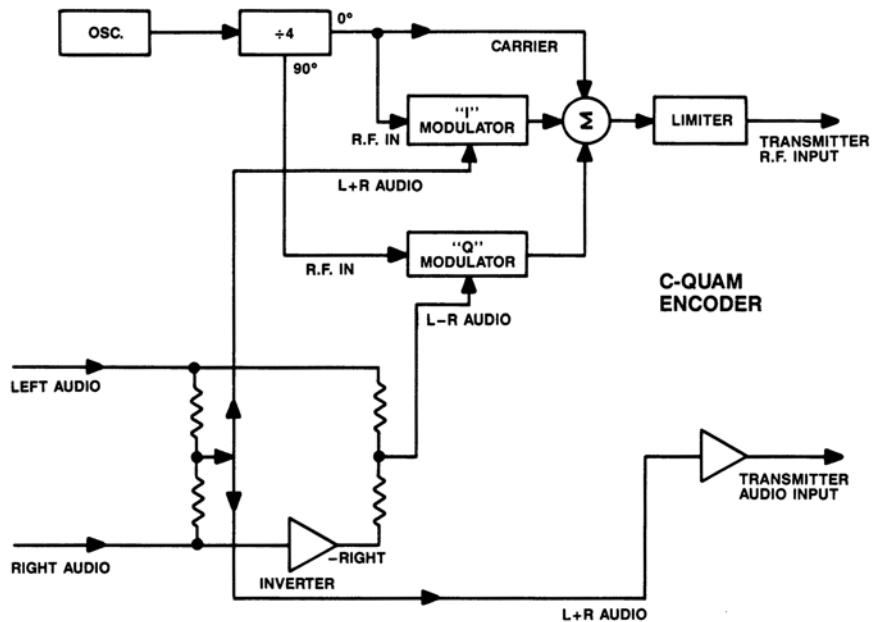


Figure 12.3.33 Block diagram of a C-QUAM AM stereo exciter.

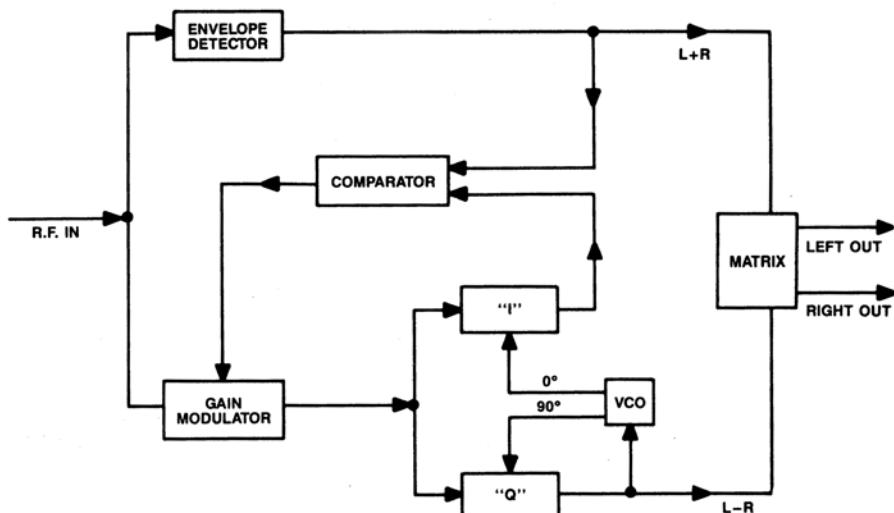


Figure 12.3.34 Block diagram of a C-QUAM AM stereo decoder.

with an I detector. The two signals are compared, and the resultant error is used to gain-modulate the input of the I and Q demodulators.

When the transmitted signal is $L + R$ (monaural, no stereo), it is pure AM (only I sidebands). In this case the envelope detector and the I demodulator see the same thing. There is no error signal, the inverse modulator does nothing, and the signal passes without change. However, when a left- or right-only signal is transmitted, both AM and PM are present, and the input signal is shifted in phase to the I demodulator, which loses some of its I amplitude. The envelope detector sees no difference in the AM because of the phase modulation. When the envelope detector and the I demodulator are compared, there is an error signal. The error signal increases the input level to the detector. This makes the input signal to the I and Q demodulators look like a pure quadrature signal, and the audio output yields the $L - R$ information. The demodulator output is combined with the envelope-detector output in a matrix to reconstruct the left and right audio channels.

12.3.5 References

1. Rohde, Ulrich L., and Jerry C. Whitaker: *Communications Receivers*, 3rd ed., McGraw-Hill, New York, N.Y., 2000.
2. Rohde, Ulrich L.: *Digital PLL Frequency Synthesizers*, Prentice-Hall, Englewood Cliffs, N.J., 1983.
3. Amos, S. W.: "FM Detectors," *Wireless World*, vol. 87, no. 1540, pg. 77, January 1981.

12.3.6 Bibliography

Benson, K. Blair, and Jerry C. Whitaker: *Television and Audio Handbook for Engineers and Technicians*, McGraw-Hill, New York, N.Y., 1990.

Whitaker, Jerry C. (ed.): *NAB Engineering Handbook*, 9th ed., National Association of Broadcasters, Washington, D.C., 1999.

Chapter
12.4
Stereo Television

K. Blair Benson

12.4.1 Introduction

The adoption of the Zenith-dbx system for *multichannel television sound* (MTS) in December 1983 paved the way for the commercial transmission of stereo audio and second-language programming in the U.S. The MTS nomenclature is a mixture of borrowed FM and some new MTS-specific terminology. This chapter discusses the broadcast-equipment specifications that have the most significant impact on MTS-system performance and gives brief explanations of the MTS terminology.

A typical origination-to-end-user block diagram is given in Figure 12.4.1. Alternate paths are shown in the transmitter section, one for discrete encoding and the other for composite encoding in the studio-transmitter link (STL). The transmission components may be grouped into the following categories:

- Audio chain
- Stereo generators
- Composite STL
- Aural exciter

12.4.2 Audio Chain

The audio chain includes distribution amplifiers, mixers, switchers, processors, studio-transmitter link (STL), wiring, and connections between the audio source and the TV MTS generator. The most important performance specifications for equipment in the audio chain are frequency response, signal-to-noise ratio (S/R), distortion, separation, headroom, and channel-to-channel tracking.

12.4.2a Stereo Generator

The TV stereo system is a modification of the standard FM stereo system [1–3]. The main differences are:

12-76 Radio Receivers

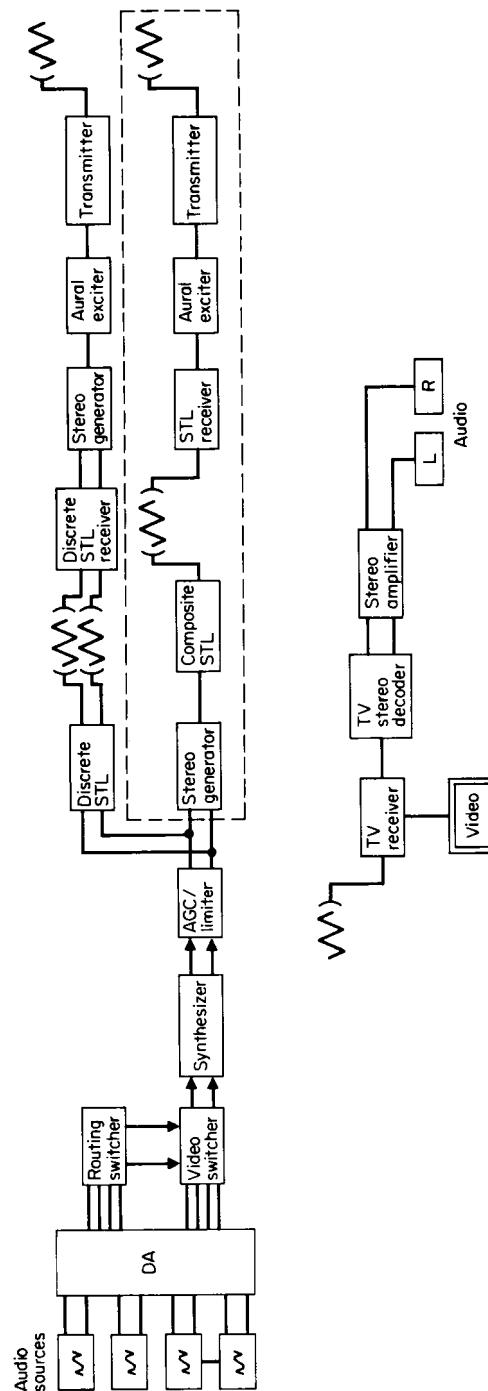


Figure 12.4.1 Typical television MTS system block diagram.

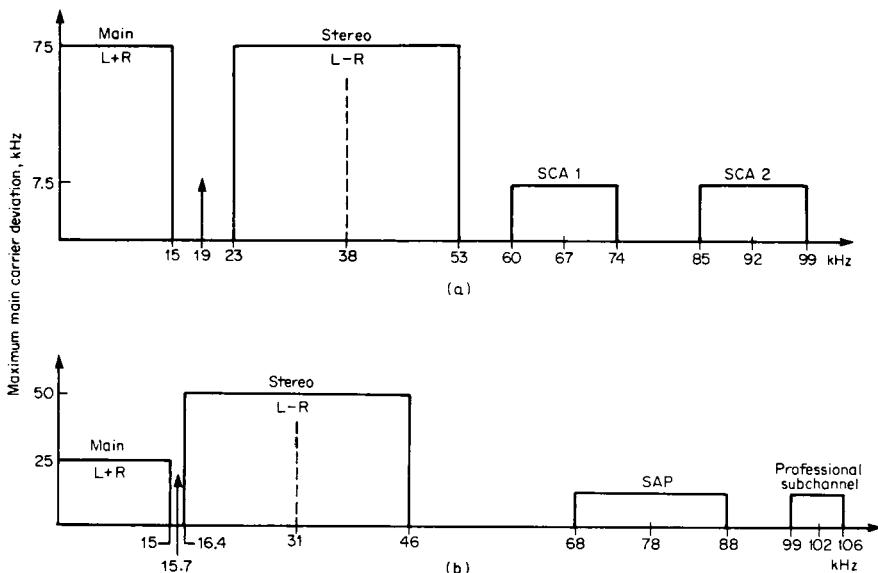


Figure 12.4.2 Baseband frequency allocations: (a) stereophonic FM, (b) BTSC TV stereophonic system.

- 25-kHz deviation for main channel (L + R).
- 50-kHz deviation for subchannel (L – R). This is 2 times, or 6 dB, greater than for the main channel.
- Noise reduction in the subchannel (L – R).
- Pilot frequency equal to f_H (15,734 Hz).

Compatibility with existing monophonic receivers was the reason behind the selection of 25-kHz deviation for the main channel audio. The increased deviation of the subchannel and the use of noise reduction was intended to maintain at least 50 dB S/N in outlying areas. With the noise reduction compandor in the stereo subchannel and the 6-dB increase in level, the L and R S/N should be dependent only on the main-channel S/N, which was found to be approximately 63 dB in system tests. The choice f_H as the pilot was to minimize buzz-beat interference.

Figure 12.4.2 shows the FM-versus-TV stereo baseband spectrum. Table 12.4.1 shows the aural carrier-modulation standards for the TV stereo system. Figure 12.4.3 is a block diagram of a typical TV stereo generator.

Some of the specifications for TV stereo generators are not different from those for other audio equipment except that they cannot be verified without use of a decoder and that the decoder contribution to performance must be recognized.

There are at least two operational modes for TV stereo generators: the normal (noise reduction) operational mode and a test mode called 75- μ s equivalent mode. In the 75- μ s equivalent mode, the noise reduction system in L – R is replaced with a 75- μ s preemphasis network identi-

12-78 Radio Receivers

Table 12.4.1 Modulation Standards for the TV MTS System

Service or Signal	Modulating Signal	Modulating Frequency Range, kHz	Audio Processing or Preemphasis	Subcarrier Frequency ($f_H = 15.734$ kHz)	Subcarrier Modulation Type	Subcarrier Deviation, kHz	Aural Carrier Peak Deviation, kHz
Monophonic	L + R	0.05–15	75 μ s				25*
Pilot				f_H			5
Sterephonic	L – R	0.05–15	BTSC compression	$2f_H$	AM-DSB SC		50*
Second Audio Program		0.05–10	BTSC compression	$5f_H$	FM	10	15
Professional Channel	Voice	0.3–3.4	150 μ s	$6.5f_H$	FM	3	3
	Data	0–1.5	0		FSK		

* Sum does not exceed 50 kHz

cal to that in L + R. This function was included to allow noise, distortion, and separation measurements to be made without the level-dependent degradation caused by noise reduction.

Subchannel Filters

Subchannel filters serve to limit the noise-reduction-control-line bandwidth and to control the out-of-band energy created by the noise reduction circuit. When there is no audio input to the noise reduction circuitry, the spectral compressor gain is at maximum, creating a high-level parabolic noise spectrum which needs to be bandwidth-limited to audio frequencies to prevent spectrum spillover. Because the pilot is separated from the subchannel by only 734 Hz, the filter slope needs to be very sharp to provide pilot protection with flat response to 15,000 Hz.

Any out-of-band information on the noise-reduction-control line (such as f_H) after the subchannel filter will cause the noise reduction circuitry to misencode, causing degradation of received separation and frequency response when decoded. To ensure stereo separation greater than 40 dB, the subchannel filters need to be matched to within 0.08-dB amplitude difference and 1° phase difference. For good overall frequency response, the filter's response should be less than ± 0.2 dB from 50 to 15,000 Hz.

The stereo-generator frequency response should be less than ± 1 dB from 50 to 15,000 Hz (with noise reduction) and ± 0.5 dB without noise reduction to help meet the total-system response goals.

12.4.2b Audio System Specifications and Performance Objectives

Specifications of particular importance for TV stereo audio generator operation include frequency response, separation, crosstalk, spurious suppression, and deviation calibration.

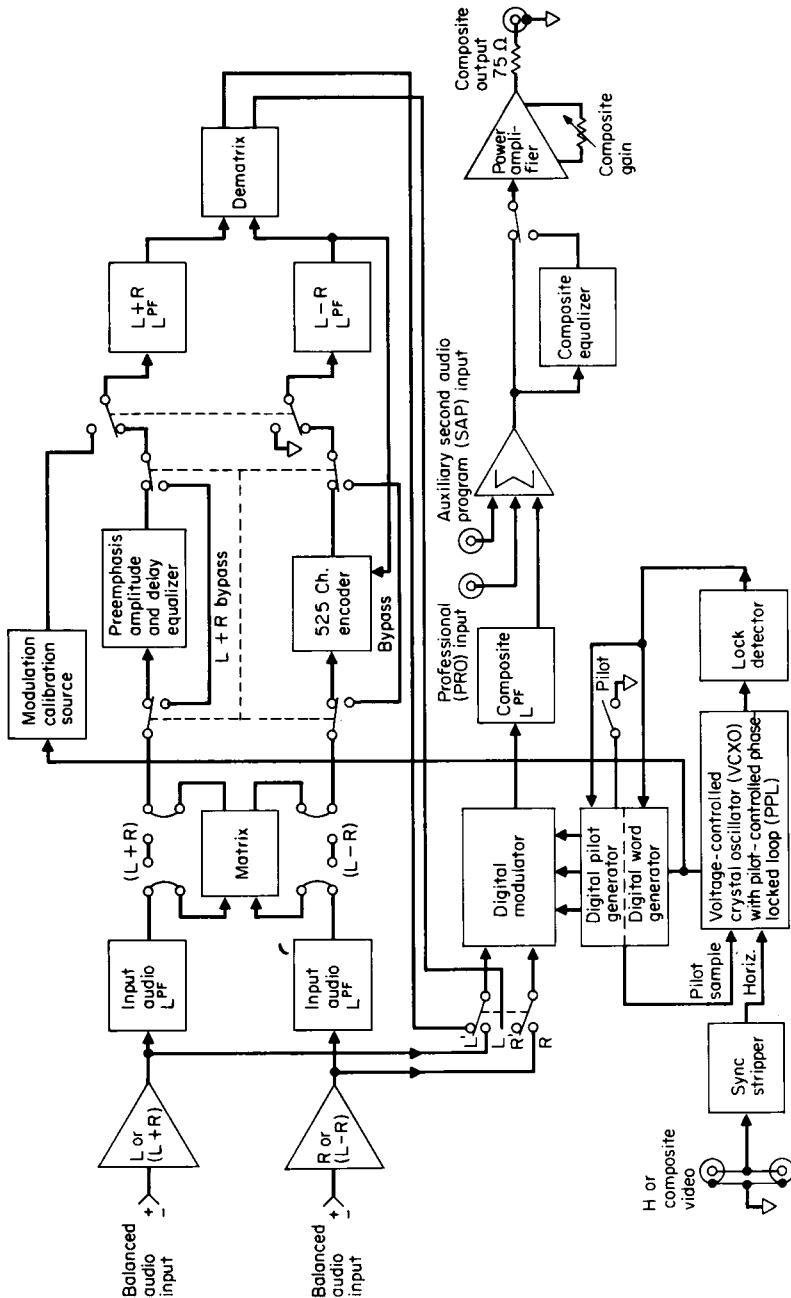


Figure 12.4.3 Stereo-generator block diagram.

12-80 Radio Receivers

Frequency Response

Frequency response performance is the deviation from a constant amplitude across a particular span of frequencies [5–7]. Researchers have found that under some conditions, variations as small as ± 0.2 dB can be discerned, especially in the ear's critical 100– to 10,000-Hz range. Because the ear is less sensitive to response errors at the extremes, an audio-system specification of ± 0.5 dB from 50 to 15,000 Hz and of ± 0.25 dB from 100 to 10,000 Hz is adequate to deliver excellent audio.

A total-system specification of ± 1 dB from 50 Hz to 15,000 Hz, including the stereo encoder and RF chain, is a realizable system performance target. Fortunately, most solid-state mixers and distribution amplifiers have frequency response specifications much better than those given above. Unfortunately, much of the source material, such as videotape and network programming, will not be quite that good unless noise reduction or digital encoding is employed.

Signal-to-Noise Ratio (SNR)

Signal-to-noise ratio is the amplitude difference, usually expressed in decibels, between a standard-level audio signal and the residual noise and hum of the system. The S/N of any equipment in the audio chain should be greater than 65 dB (unweighted). This performance level will ensure that the source S/N or the transmitter FM SNR will be the limiting factor in the system noise contribution. The performance target for the entire system should be greater than 58 dB.

Total Harmonic Distortion (THD)

Total harmonic distortion is the creation by a nonlinear device of spurious signals harmonically related to the audio waveform. Research has shown that although THD levels greater than 1 percent are easily detectable during sine-wave testing, people will tolerate somewhat higher levels of THD on musical material. THD in any part of the audio chain should be less than 0.25 percent from 50 to 15,000 Hz, and the whole audio chain should be less than 0.5 percent THD from 50 to 15,000 Hz. A performance target for the entire system (RF included) should be 1 percent from 50 to 15,000 Hz.

Although harmonics of frequencies greater than 7500 Hz will be attenuated by the 15,000-Hz low-pass filters in the TV MTS generator, the presence of these harmonics in the audio stages before the generator may raise the level of intermodulation-distortion products in the audio signal.

The THD test is sensitive to the S/N of the device under test. If the device has an S/N of 60 dB, the distortion analyzer's best distortion reading will be greater than 0.1 percent ($60 \text{ dB} = 0.001 = 0.1$ percent). An audio spectrum analyzer will allow true THD readings because it will disregard the broadband noise floor of the tested device.

Intermodulation Distortion (IMD)

Intermodulation distortion is the creation by a nonlinear device of spurious signals not harmonically related to the audio waveform. These distortion components are sum-and-difference (beat notes) mixing products that research has shown are more objectionable to listeners than even harmonic-distortion products. The IMD (60 Hz/7000 Hz mixed 4:1 amplitude ratio), measured in accordance with Society of Motion Picture and Television Engineers (SMPTE) guidelines, should be less than 0.25 percent for each part of the audio chain and less than 0.5 percent for the

entire audio chain. The IMD measurement is not greatly affected by the noise floor of the tested device. A performance target for the entire system (including RF) should be 1 percent IMD or less.

Separation

Separation is a specialized definition for signal crosstalk between the left and right channels of a stereo system. The minimum amount of stereo separation needed to define a stable stereo image has been found to be greater than 17 dB at the listener's position. Stereo separation can be degraded at almost any link in the transmission chain, requiring much greater than the nominal 17 dB to be necessary for the audio chain. A separation requirement of greater than 50 dB from 50 to 15,000 Hz should be the minimum for each piece of equipment in the audio chain. The entire audio system should have separation performance of greater than 40 dB from 400 to 15,000 Hz and 35 dB from 50 to 400 Hz. A performance target for the entire system (including RF) should be 30 dB from 100 to 12,000 Hz and 20 dB from 50 to 15,000 Hz.

Headroom

Headroom is the difference (in decibels) between the normal operating level and the maximum output level of the device being tested. In equipment monitored by volume-unit (VU) types of meters, the equipment should have 15 dB of headroom above 0 VU, or normal operating level, to allow for musical peaks that averaging-meter ballistics ignore. When the audio system operates without sufficient headroom, the time average of the program distortion becomes high enough to be audible.

Channel-to-Channel Amplitude and Phase Tracking

Channel-to-channel amplitude and phase tracking is the match between the amplitude-versus-frequency and phase-versus-frequency responses of the stereo channels (L, R). Poor channel-to-channel tracking will result in a wandering or off-center stereo image and monaural summing errors that can be very obvious and distracting to viewers. The most obvious error is a completely out-of-phase condition ($L = -R$) which gives no mono signal ($L + R = 0$). The amplitude and phase match of the L and R channels should be good enough that with both channels driven with the same amplitude signal, but with one channel inverted ($L = -R$) and summed electrically, the residual should be 40 dB below the equivalent $L + R$ level from 50 to 15,000 Hz. The chart in Figure 12.4.4 gives the amplitude and phase requirements to achieve various crosstalk performances.

12.4.2c Stereo Generator Specifications and Performance Objectives

Although various designs of stereo generators require specific performance measurement parameters and techniques, the following general categories apply to all devices and systems.

Separation

The separation specification of a stereo generator can be measured in three ways. Generator performance should be better than the FCC Office of Science and Technology Bulletin 60 numbers given below.

12-82 Radio Receivers

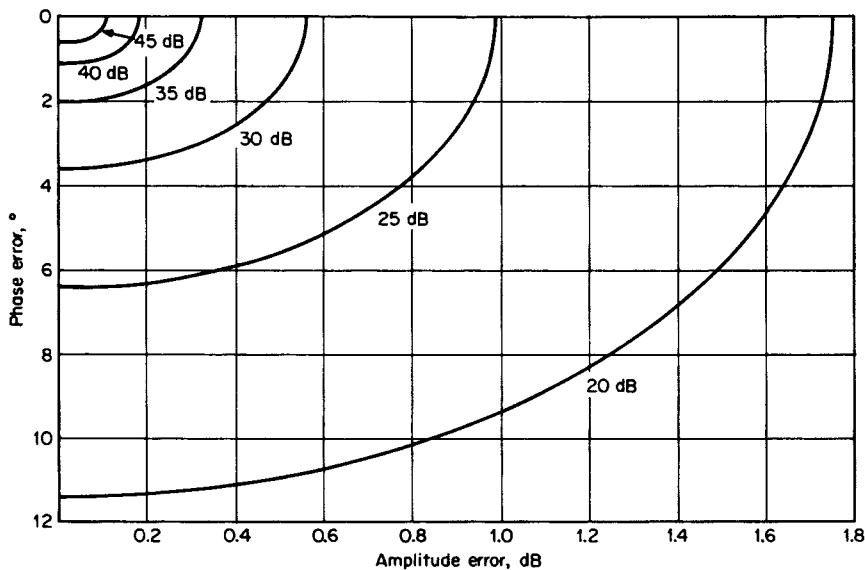


Figure 12.4.4 Crosstalk versus amplitude and phase errors.

- Baseband (FM) generator (no audio filters included): 50-dB minimum from 50 to 15,000 Hz
- Noise reduction bypassed (all audio filters included): 40-dB minimum from 50 to 14,000Hz
- Full-system BTSC mode (all filters and noise reduction included): 30-dB minimum from 100 to 8000 Hz, decreasing to 20 dB at 14,000 Hz, 26-dB minimum at 50 Hz

Linear Crosstalk

Linear crosstalk is leakage from L + R to L – R or L – R to L + R caused by amplitude and phase matching of L and R channels, which will result in a wandering or off-center stereo image and monaural summing errors. The stereo-generator linear crosstalk should be better than 40 dB below 100 percent with all filters in circuit.

Nonlinear Crosstalk

Nonlinear crosstalk is leakage from L + R to L – R or from L – R to L + R caused by distortion products in L + R or L – R. The distortion products generated in the L – R channel cause interference in the SAP (Second Audio Program) or PRO (professional) spectrum. The generator specification should be better than 70 dB below 100 percent.

Spurious Suppression

Spurs are caused by nonlinearities in the AM-DSB (double-sideband) modulation of L – R. These unwanted distortion products can cause noise and whistles in the SAP and PRO channels,

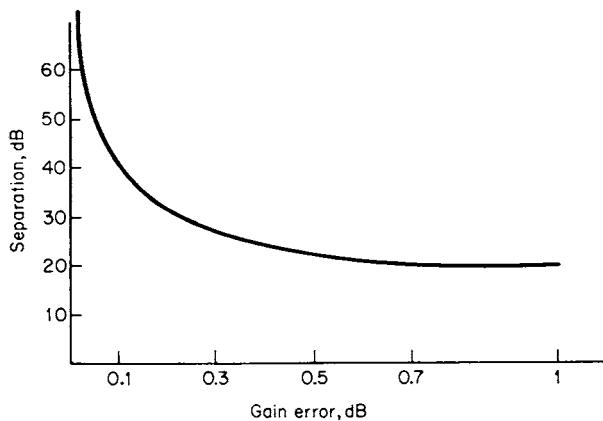


Figure 12.4.5 Stereo separation as a function of composite gain error.

especially since the SAP is 10.5 dB below L – R and the PRO is 20 dB below L – R. Generator spurious suppression should be greater than 75 dB below 100 percent L – R.

Deviation Calibration

Unlike noncompanded FM stereo, the BTSC system requires precise adjustment of aural-deviation calibration for optimum received stereo separation and frequency response. Like some tape noise reduction systems, the encoder-to-decoder levels must be carefully matched for good system performance. For example, an aural-deviation error of 1 dB will degrade separation from perfect to less than 20 dB (Figure 12.4.5). The stereo generator should incorporate a deviation calibration system that allows the aural deviation to be quickly and easily set and checked with an accuracy of better than 1 percent.

12.4.2d Composite STL Specifications and Performance Objectives

To maintain a system separation greater than 40 dB, the composite amplitude response must be within ± 0.08 dB and the composite phase response must be less than $\pm 1^\circ$ from linear phase over the band from 50 to 47,000 Hz.

If a SAP or PRO signal is to be added, the composite amplitude and phase response must be flat (± 1 dB, $\pm 10^\circ$) to 120,000 Hz to prevent crosstalk.

Signal-to-Noise Ratio

The FM signal-to-noise ratio of the STL should be better than 65 dB (unweighted) with 75- μ s deemphasis.

Distortion

The aural composite STL should have distortion of less than 0.1 percent. Any distortion of the baseband signal caused by the STL will have secondary effects of the stereo, SAP, and PRO

12-84 Radio Receivers

crosstalk, which is quite noticeable at the receiver with rather small amounts of distortion added to the baseband. For example, an increase of baseband harmonic distortion from 0.05 to 1.0 percent will increase crosstalk into the SAP by as much as 26 dB.

Amplitude Stability

The encoded TV stereo composite signal is highly sensitive to gain variations in the composite path. As shown in Figure 12.4.5, a gain error of less than 10 percent (1 dB) will reduce the system separation to less than 20 dB. This assumes that the separation was perfect before the gain error was introduced. For good separation, the composite path should have a maximum gain drift of ± 0.2 dB over time and temperature.

12.4.2e Aural Exciter Specifications and Performance Objectives

To maintain a system separation greater than 40 dB, the composite amplitude response must be within ± 0.08 dB and the composite phase response must be less than $\pm 1^\circ$ from linear phase over the band from 50 to 47,000 Hz and over an aural carrier deviation of 50,000 Hz. If a SAP or PRO signal is to be added, the composite amplitude and phase response must be flat (± 1 dB, $\pm 10^\circ$) to 120,000 Hz to prevent crosstalk.

Signal-to-Noise Ratio

The FM signal-to-noise ratio of the aural exciter should be better than 65 dB (unweighted) with 75- μ s deemphasis.

Amplitude Stability

As outlined previously, the encoded TV stereo composite signal is highly sensitive to gain variations in the composite path. For good separation, the aural exciter should have a maximum gain drift of ± 0.2 dB over time and temperature.

12.4.3 References

1. Eilers, C.: "The Zenith Multichannel TV Sound System," *Proc. 38th NAB Eng. Conf.*, National Association of Broadcasters, Washington, D.C., 1984.
2. Federal Communications Commission: *Multichannel Television Sound Transmission and Audio Processing Requirements for the BTSC System*, OST Bulletin 60, FCC, Washington, D.C.
3. Electronic Industries Association: *Multichannel Television Sound—BTSC System Recommended Practices*, EIA, Washington, D.C., ETA Television Systems Bulletin 5, July 1985.
4. Mendenhall, Geoff: "Testing Television Transmission Systems for Multichannel Sound Compatibility," *Proc. 39th NAB Eng. Conf.*, National Association of Broadcasters, Washington, D.C., 1985.
5. Buecklein, R.: "The Audibility of Frequency Response Irregularities," *J. Audio Eng. Soc.*, vol. 29, no. 3, pp. 126–131, March 1981.

6. Plenge, G. H., H. Jakubowski, and P. Schoene: "Which Bandwidth Is Necessary for Optimal Sound Transmission?", presented at the 62d Convention of the Audio Engineering Society, Brussels, March 13–16, 1979.
7. Snow, W. B.: "Audible Frequency Ranges of Music, Speech and Noise," *J. Acoust. Soc. Am.*, vol. 3, pp. 155–166, 1931.

Section**13****Standards and Practices**

A widely used dictionary lists seven meanings for the word “standard,” but only one of these, “anything authorized as a measurement of quantity and quality,” seems to relate to the idea of a standard in the industrial area. The original idea of commercial standards was to be bound legally on units of weights and measure for fairness in trade. In the field of electronics, the standardization of electrical units of measurement would be the counterpart. Today, standards are essentially recommendations for users and/or manufacturers to adhere to basic specifications to allow operational interchangeability in the use of equipment and supplies.

Anyone concerned with interchangeability of equipment or product should be concerned with standards. A prospective user hesitates to purchase equipment that does not conform to recognized interface standards for connectors, input/output levels, control, timing, and test specifications. A manufacturer may find a limited market for a good product if it is not compatible with other equipment in common use.

To most audio professionals, the term “standards” envisions a means of promoting interchange of basic hardware. To others, it evokes thoughts of a slowdown of progress, of maintaining a status quo—perhaps for the benefit of a particular group. Both camps can cite examples to support their viewpoint, but no one can seriously contend that we would better off without standards. Standards promote economies of scale that tend to produce more reliable products at a lower cost.

For most people, the question is: “How do standards affect my life? Do they stifle progress? Do they prevent products from appearing on the market in a timely fashion? Do they discourage alternate technologies that might be beneficial in the long run?” Many would respond affirmatively to one or more of these questions, but consider the upside. Standards ensure that the needs of the user are considered. Interconnection of equipment from different manufacturers is facilitated. The current rollout of digital audio products at a record pace attests to the need for standards. The progress made so far in the digital era would have been wholly impossible without the considerable efforts of organizations such as the AES, ATSC, SMPTE, SCTE, and NAB.

Rapid improvements in technology tend to make many standards technically obsolete by the time they are adopted. But such is the nature of our rapidly expanding technology-based society. There is no need to apologize for this natural phenomena. A standard still provides a stable platform for manufacturers to market their product and assures the user of some degree of compatibility. Technical chaos and anarchy are real possibilities if standards are not adopted in a timely

13-2 Section Thirteen

manner. Only the strongest companies could be expected to survive in an atmosphere where standards are lacking. A successful standard promises a stable period of income to manufacturers while giving users assurance of multiple sources during the active life of the product.

Standardization usually starts within a company as a way to reduce costs associated with parts stocking, design drawings, training, and retraining of personnel. The next level might be a cooperative agreement between firms making similar equipment to use standardized dimensions, parts, and components. Competition, trade secrets, and the *NIH factor* (not invented here) often generate an atmosphere that prevents such an understanding. Enter the professional engineering society, which promises a forum for discussion between users and engineers while down playing the commercial and business aspects.

Of the many standards-setting organizations in the professional video field, the most prominent are:

- Society of Motion Picture and Television Engineers
- Audio Engineering Society
- National Association of Broadcasters
- Electronic Industries Association
- Advanced Television Systems Committee
- Society of Cable and Telecommunications Engineers

Standards, whether for a new radio broadcast system or recorder connection pin assignments, are vital for the continued growth of the audio industry.

In This Section:

Chapter 13.1: Systems Engineering	13-7
Introduction	13-7
Systems Theory	13-7
Systems Engineering	13-8
Functional Analysis	13-9
Synthesis	13-11
Modeling	13-11
Dynamics	13-11
Optimization	13-11
Evaluation and Decision	13-11
Trade Studies	13-12
Description of System Elements	13-14
Phases of a Typical System Design Project	13-15
Design Development	13-15
Electronic System Design	13-16
Detailed Design	13-17
Customer Support	13-19
Budget Requirements Analysis	13-19
Feasibility Study and Technology Assessment	13-20
Planning and Control of Scheduling and Resources	13-20
Project Tracking and Control	13-21

Executive Management	13-22
Project Manager	13-23
Systems Engineer	13-26
References	13-27
Bibliography	13-27
Chapter 13.2: Engineering Documentation	13-29
Introduction	13-29
Basic Concepts	13-29
The Manuals	13-30
Documentation Conventions	13-30
Self Documentation	13-31
Database Documentation	13-32
Graphic Documentation	13-32
Update Procedures	13-33
Equipment Documentation	13-33
Operator/User Documentation	13-34
The True Cost of Documentation	13-35
Bibliography	13-35
Chapter 13.3: Safety Issues	13-37
Introduction	13-37
Electric Shock	13-37
Effects on the Human Body	13-37
Circuit Protection Hardware	13-39
Working with High Voltage	13-39
RF Considerations	13-42
First Aid Procedures	13-42
Operating Hazards	13-42
OSHA Safety Considerations	13-44
Protective Covers	13-44
Identification and Marking	13-44
Grounding	13-45
Beryllium Oxide Ceramics	13-46
Corrosive and Poisonous Compounds	13-48
FC-75 Toxic Vapor	13-48
Nonionizing Radiation	13-49
NEPA Mandate	13-49
Revised Guidelines	13-50
Multiple-user Sites	13-51
Operator Safety Considerations	13-52
X-Ray Radiation Hazard	13-52
Implosion Hazard	13-52
Hot Coolant and Surfaces	13-52
Polychlorinated Biphenyls	13-53
Governmental Action	13-54
PCB Components	13-54
PCB Liability Management	13-56

13-4 Section Thirteen

References	13-56
Bibliography	13-57
Chapter 13.4: Audio/Video Standards	13-59
Standards Relating to Digital Video	13-59
Video	13-59
Audio	13-59
ATSC DTV Standard	13-60
Service Multiplex and Transport Systems	13-60
System Information Standard	13-60
Receiver Systems	13-61
Program Guide	13-61
Program/Episode/Version Identification	13-61
DVB	13-61
General	13-61
Multipoint Distribution Systems	13-62
Interactive Television	13-62
Conditional Access	13-62
Interfaces	13-62
SMPTE Documents Relating to Digital Video and Audio	13-62
General Topics	13-63
Ancillary	13-63
Digital Control Interfaces	13-63
Edit Decision Lists	13-64
Image Areas	13-64
Interfaces and Signals	13-64
Bit-Parallel Interfaces	13-65
Bit-Serial Interfaces	13-65
Scanning Formats	13-66
Monitors	13-66
MPEG-2	13-66
Test Patterns	13-66
Video Recording and Reproduction	13-67
Time and Control Code	13-67
Tape Recording Formats	13-68
SMPTE Documents by Number	13-68
CEA Standards	13-82
Home Networks	13-82

Reference Documents for this Section: _____

Baumgartner, Fred, and Terrence Baun: "Broadcast Engineering Documentation," in *NAB Engineering Handbook*, 9th ed., Jerry C. Whitaker (ed.), National Association of Broadcasters, Washington, D.C., 1999.

Baumgartner, Fred, and Terrence Baun: "Engineering Documentation," in *The Electronics Handbook*, Jerry C. Whitaker (ed.), CRC Press, Boca Raton, Fla., 1996.

- “Current Intelligence Bulletin #45,” National Institute for Occupational Safety and Health, Division of Standards Development and Technology Transfer, February 24, 1986.
- Code of Federal Regulations, 40, Part 761.
- Delatore, J. P., E. M. Prell, and M. K. Vora: “Translating Customer Needs Into Product Specifications”, *Quality Progress*, January 1989.
- DeSantis, Gene: “Systems Engineering Concepts,” in *NAB Engineering Handbook*, 9th ed., Jerry C. Whitaker (ed.), National Association of Broadcasters, Washington, D.C., 1999.
- DeSantis, Gene: “Systems Engineering,” in *The Electronics Handbook*, Jerry C. Whitaker (ed.), CRC Press, Boca Raton, Fla., 1996.
- “Electrical Standards Reference Manual,” U.S. Department of Labor, Washington, D.C.
- Finkelstein, L.: “Systems Theory”, *IEE Proceedings*, vol. 135, Part A, no. 6, July 1988.
- Hammett, William F.: “Meeting IEEE C95.1-1991 Requirements,” *NAB 1993 Broadcast Engineering Conference Proceedings*, National Association of Broadcasters, Washington, D.C., pp. 471–476, April 1993.
- Hammar, Willie: *Occupational Safety Management and Engineering*, Prentice Hall, New York, N.Y.
- Hoban, F. T., and W. M. Lawbaugh: *Readings In Systems Engineering*, NASA, Washington, D.C., 1993.
- Markley, Donald: “Complying with RF Emission Standards,” *Broadcast Engineering*, Intertec Publishing, Overland Park, Kan., May 1986.
- “Occupational Injuries and Illnesses in the United States by Industry,” OSHA Bulletin 2278, U.S. Department of Labor, Washington, D.C, 1985.
- OSHA, “Electrical Hazard Fact Sheets,” U.S. Department of Labor, Washington, D.C, January 1987.
- OSHA, “Handbook for Small Business,” U.S. Department of Labor, Washington, D.C.
- Pfrimmer, Jack, “Identifying and Managing PCBs in Broadcast Facilities,” *1987 NAB Engineering Conference Proceedings*, National Association of Broadcasters, Washington, D.C, 1987.
- “Safety Precautions,” Publication no. 3386A, Varian Associates, Palo Alto, Calif., March 1985.
- Shinners, S. M.: *A Guide to Systems Engineering and Management*, Lexington, 1976.
- Smith, Milford K., Jr., “RF Radiation Compliance,” *Proceedings of the Broadcast Engineering Conference*, Society of Broadcast Engineers, Indianapolis, IN, 1989.
- System Engineering Management Guide*, Defense Systems Management College, Virginia, 1983.
- “Toxics Information Series,” Office of Toxic Substances, July 1983.
- Tuxal, J. G.: *Introductory System Engineering*, McGraw-Hill, New York, N.Y., 1972.
- Whitaker, Jerry C.: *AC Power Systems*, 2nd Ed., CRC Press, Boca Raton, Fla., 1998.

13-6 Section Thirteen

- Whitaker, Jerry C.: G. DeSantis, and C. Paulson: *Interconnecting Electronic Systems*, CRC Press, Boca Raton, Fla., 1993.
- Whitaker, Jerry C.: *Maintaining Electronic Systems*, CRC Press, Boca Raton, Fla. 1991.
- Whitaker, Jerry C.: *Power Vacuum Tubes Handbook*, 2nd Ed., CRC Press, Boca Raton, Fla., 1999.
- Whitaker, Jerry C.: *Radio Frequency Transmission Systems: Design and Operation*, McGraw-Hill, New York, N.Y., 1990.

Chapter

13.1

Systems Engineering

Gene DeSantis

13.1.1 Introduction

Modern systems engineering emerged during World War II as—due to the degree of complexity in design, development, and deployment—weapons evolved into weapon systems. In the sixties, the complexities of the space program made a systems engineering approach to design and problem solving even more critical. Indeed, the Department of Defense and NASA are two of the staunchest practitioners. With the increase in size and complexity of television and nonbroadcast video systems during that same period, the need for a systems approach to planning, designing and building facilities gained increased attention.

Today, large engineering organizations utilize a systems engineering process. Much has been published about system engineering practices in the form of manuals, standards, specifications, and instruction. In 1969, MIL-STD-499 was published to help government and contractor personnel involved in support of defense acquisition programs. In 1974, this standard was updated to MIL-STD-499A, which specifies the application of system engineering principles to military development programs. Likewise, the builders of turnkey television systems and facilities have adopted their own unique systems engineering approaches to projects. The tools and techniques of this processes continue to evolve in order to do each job a little better, save time, and cut costs.

13.1.2 Systems Theory

Although there are other areas of application outside of the broadcast industry, we will be concerned with systems theory as it applies to television systems engineering. We will be concerned with audio, video, RF, control, time code, telecommunications, computer systems, and software. Systems theory can be applied to engineering of all of these elements. Building and vehicle systems—including space planning, power and lighting, environmental control, and safety systems—can all benefit from the systems engineering approach. These systems are made up of component elements that are interconnected and programmed to function together in a facility.

For the purpose of this discussion, a system is defined as a set of related elements that function together as a single entity.

Systems theory consists of a body of concepts and methods that guide the description, analysis, and design of complex entities.

13-8 Standards and Practices

Decomposition is an essential tool of systems theory. The systems approach attempts to apply an organized methodology to completing large complex projects by breaking them down into simpler, more manageable components. These elements are treated separately, analyzed separately, and designed separately. In the end, all of the components are recombined to build the whole.

Holism is an element of systems theory in that the end product is greater than the sum of its component elements. In systems theory, modeling and analytical methods enable all essential effects and interactions within a system and those between a system and its surroundings to be taken into account. Errors resulting from the idealization and approximation involved in treating parts of a system in isolation, or reducing consideration to a single aspect, are thus avoided.

Another holistic aspect of system theory describes *emergent properties*. Properties that result from the interaction of system components, properties that are not those of the components themselves, are referred to as emergent properties.

Although dealing with concrete systems, *abstraction* is an important feature of systems models. Components are described in terms of their function rather than in terms of their form. Graphical models such as block diagrams, flow diagrams, and timing diagrams are commonly used.

Mathematical models may also be employed. Systems theory shows that, when modeled in abstract formal language, apparently diverse kinds of systems show significant and useful *isomorphisms* of structure and function. Similar interconnection structures occur in different types of systems. Equations that describe the behavior of electrical, thermal, fluid, and mechanical systems are essentially identical in form.

Isomorphism of structure and function implies isomorphism of behavior of a system. Different types of systems exhibit similar dynamic behavior such as response to stimulation.

The concept of *hard* and *soft* systems appears in system theory. In hard systems, the components and their interactions can be described by mathematical models. Soft systems can not be described so easily. They are mostly human activity systems that imply unpredictable behavior and non uniformity. They introduce difficulties and uncertainties of conceptualization, description, and measurement. The kinds of system concepts and methodology described previously can not be applied.

13.1.2a Systems Engineering

Systems engineering depends on the use of a process methodology based on systems theory. In order to deal with the complexity of large projects, systems theory breaks down the process into logical steps.

Even though underlying requirements differ from program to program, there is a consistent, logical process that can best be used to accomplish system design tasks. The basic product development process is illustrated in Figure 13.1.1. The systems engineering starts at the beginning of this process to describe the product to be designed. It includes four activities:

- Functional analysis
- Synthesis
- Evaluation and decision
- Description of system elements

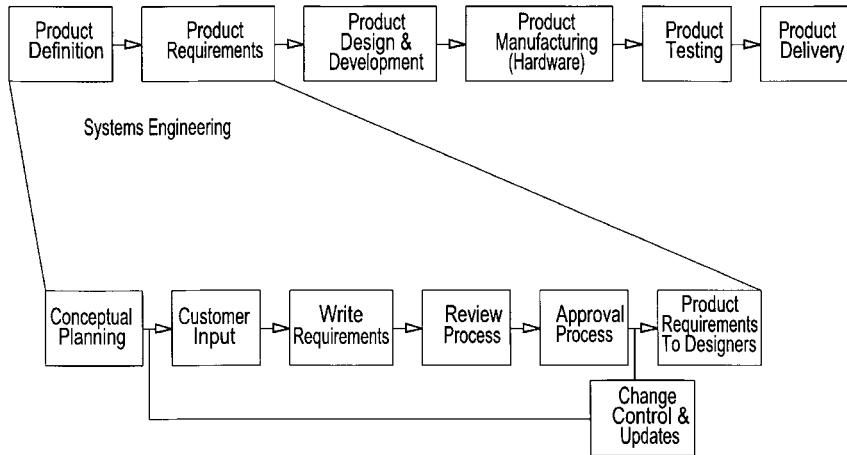


Figure 13.1.1 The product development and documentation process.

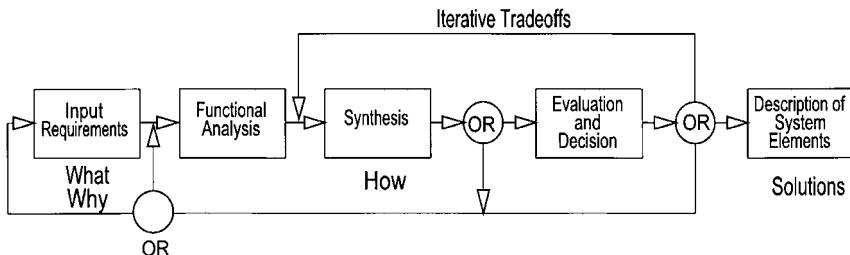


Figure 13.1.2 The systems engineering process. (After [1].)

This process is illustrated in Figure 13.1.2. The process is iterative. That is, with each successive pass, the product element description becomes more detailed. At each stage in the process a decision is made whether to accept, make changes, or return to an earlier stage of the process and produce new documentation. The result of this activity is documentation that fully describes all system elements and which can be used to develop and produce the elements of the system. The systems engineering process does not produce the actual system itself.

13.1.2b Functional Analysis

A *systematic approach* to systems engineering will include elements of systems theory. (See Figure 13.1.3.) To design a product, hardware and software engineers need to develop a vision of the

13-10 Standards and Practices

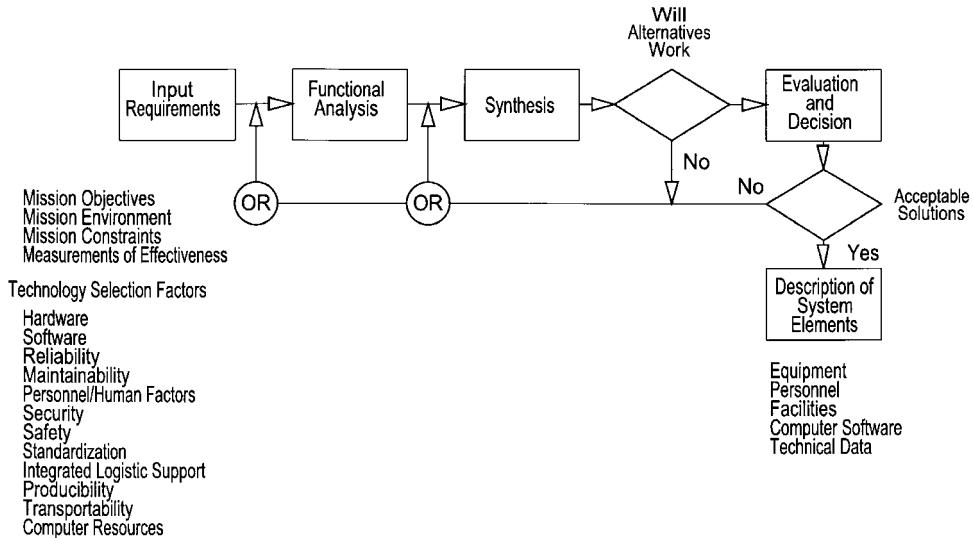


Figure 13.1.3 The systems engineering decision process. (After [1].)

product—the product requirements. These requirements are usually based on customer needs researched by a marketing department. An organized process to identify and validate customer needs will help minimize false starts. System objectives are first defined. This may take the form of a mission statement that outlines the objectives, the constraints, the mission environment, and the means of measuring mission effectiveness.

The purpose of the system is defined, analysis is carried out to identify the requirements and what essential functions the system must perform, and why. The *functional flow block diagram* is a basic tool used to identify *functional needs*. It shows logical sequences and relationships of operational and support functions at the system level. Other functions such as maintenance, testing, logistics support, and productivity may also be required in the functional analysis. The functional requirements will be used during the *synthesis phase* to show the allocation of the functional performance requirements to individual system elements or groups of elements. Following evaluation and decision, the functional requirements provide the functionally oriented data required in the description of the system elements.

Analysis of time critical functions is also a part of this functional analysis process when functions have to take place sequentially, or concurrently, or on a particular schedule. Time line documents are used to support the development of requirements for the operation, testing, and maintenance functions.

13.1.2c Synthesis

Synthesis is the process by which concepts are developed to accomplish the functional requirements of a system. Performance requirements and constraints, as defined by the functional analysis, are applied to each individual element of the system, and a design approach is proposed for meeting the requirements. Conceptual schematic arrangements of system elements are developed to meet system requirements. These documents can be used to develop a description of the system elements and can be used during the *acquisition phase*.

13.1.2d Modeling

The concept of *modeling* is the starting point of synthesis. Because we must be able to weigh the effects of different design decisions in order to make choices between alternative concepts, modeling requires the determination of those quantitative features that describe the operation of the system. We would, of course, like a model with as much detail as possible describing the system. Reality and time constraints, however, dictate that the simplest possible model be selected in order to improve our chances of design success. The model itself is always a compromise. The model is restricted to those aspects that are important in the evaluation of system operation. A model might start off as a simple block diagram with more detail being added as the need becomes apparent.

13.1.2e Dynamics

Most system problems are dynamic in nature. The signals change over time and the components determine the *dynamic response* of the system. The system behavior depends on the signals at a given instant, as well as on the rates of change of the signals and their past values. The term “signals” can be replaced by substituting human factors, such as the number of users on a computer network for example.

13.1.2f Optimization

The last concept of synthesis is *optimization*. Every design project involves making a series of compromises and choices based on relative weighting of the merit of important aspects. The best candidate among several alternatives is selected. Decisions are often subjective when it comes to deciding the importance of various features.

13.1.2g Evaluation and Decision

Program costs are determined by the tradeoffs between operational requirements and engineering design. Throughout the design and development phase, decisions must be made based on evaluation of alternatives and their effects on cost. One approach attempts to correlate the characteristics of alternative solutions to the requirements and constraints that make up the selection criteria for a particular element. The rationale for alternative choices in the decision process are documented for review. Mathematical models or computer simulations can be employed to aid in this evaluation decision making process.

13-12 Standards and Practices

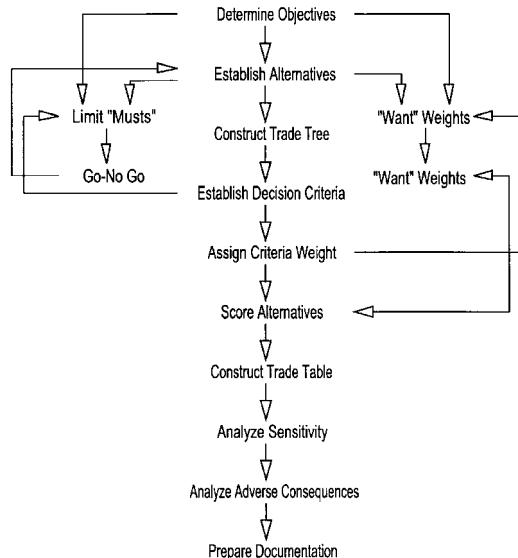


Figure 13.1.4 Trade study process flow.
(After [2].)

13.1.2h Trade Studies

A structured approach is used in the *trade study process* to guide the selection of alternative configurations and insure that a logical and unbiased choice is made. Throughout development, trade studies are carried out to determine the best configuration that will meet the requirements of the program. In the concept exploration and demonstration phases, trade studies help define the system configuration. Trade studies are used as a detailed design analysis tool for individual system elements in the full scale development phase. During production, trade studies are used to select alternatives when it is determined that changes need to be made. Figure 13.1.4 illustrates the relationship of the various of elements that can be employed in a trade study. To provide a basis for the selection criteria, the objectives of the trade study must first be defined. Functional flow diagrams and system block diagrams are used to identify trade study areas that can satisfy certain requirements. Alternative approaches to achieving the defined objectives can then be established.

Complex approaches can be broken down into several simpler areas and a decision tree constructed to show the relationship and dependences at each level of the selection process. This *trade tree*, as it is called, is illustrated in figure 13.1.5. Several trade study areas may be identified as possible candidates for accomplishing a given function. A trade tree is constructed to show relationships and the path through selected candidate trade areas at each level to arrive at a solution.

Several alternatives may be candidates for solutions in a given area. The selected candidates are then submitted to a systematic evaluation process intended to weed out unacceptable candidates. Criteria are determined that are intended to reflect the desirable characteristics. Undesirable characteristics may also be included to aid in the evaluation process. Weights are assigned to each criteria to reflect its value or impact on the selection process. This process is subjective. It

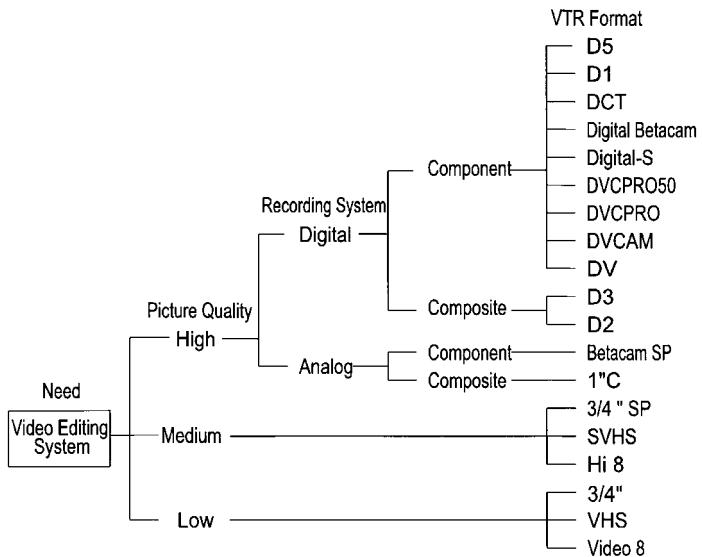


Figure 13.1.5 Trade tree example for a video project.

should also take into account cost, schedule, and hardware availability restraints that may limit the selection.

The criteria data on the candidates is collected and tabulated on a decision analysis work sheet. The attributes and limitations are listed in the first column and the data for each candidate listed in adjacent columns to the right. The performance data is available from vendor specification sheets or can require laboratory testing and analysis to determine. Each attribute is given a relative score from 1 to 10 based on its comparative performance relative to the other candidates. *Utility function graphs* can be used to assign logical scores for each attribute. The utility curve represents the advantage rating for a particular value of an attribute. A graph is made of ratings on the *y* axis versus an attribute value on the *x* axis. Specific scores can then be applied that correspond to particular performance values. The shape of the curve may take into account requirements, limitations, and any other factor that will influence its value regarding the particular criteria being evaluated. The limits to which the curves should be extended should run from the minimum value below which no further benefit will accrue to the maximum value above which no further benefit will accrue.

The scores are filled in on the decision analysis work sheet and multiplied by the weights to calculate the weighted score. The total of the weighted scores for each candidate then determines their ranking. As a general rule, at least a 10 percent difference in score is acceptable as "meaningful."

Further analysis can be applied in terms of evaluating the sensitivity of the decision to changes in the value of attributes, weights, subjective estimates, and cost. Scores should be checked to see if changes in weights or scores would reverse the choice.

13-14 Standards and Practices

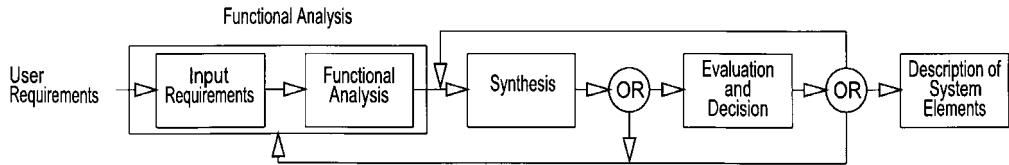


Figure 13.1.6 Basic documentation for systems engineering. (After [1].)

A *trade table* can be prepared to summarize the selection results. Pertinent criteria are listed for each alternative solution. The alternatives may be described in a quantitative manner such as high, medium, or low.

Finally, the results of the trade study are documented in the form of a report that discusses the reasons for the selections and may include the trade tree and the trade table.

There has to be a formal system of change control throughout the systems engineering process to prevent changes from being made without proper review and approval by all concerned parties, and to keep all parties informed. Change control also insures that all documentation is kept up to date and can help to eliminate redundant documents. Finally, change control helps to control project costs.

13.1.2i Description of System Elements

Five categories of interacting system elements can be defined: equipment (hardware), software, facilities, personnel, and procedural data. Performance, design, and test requirements must be specified and documented for equipment, components, and computer software elements of the system. It is necessary to specify environmental and interface design requirements that are necessary for proper functioning of system elements within a facility.

The documentation produced by the systems engineering process controls the evolutionary development of the system. Figure 13.1.6 illustrates the process documentation used by one organization in each step of the systems engineering effort.

The requirements are formalized in written specifications. In any organization, there should be clear standards for producing specifications. This can help reduce the variability of technical content and improve product quality as a result. It is also important to make the distinction here that the product should not be over specified to the point of describing the design or making it too costly. On the other hand, requirements should not be too general or so vague that the product would fail to meet the customer needs. In large departmentalized organizations, commitment to schedules can help assure that other members of the organization can coordinate their time.

The system engineering process does not actually design the system. The system engineering process produces the documentation necessary to define, design, build, and test the system. The technical integrity provided by this documentation ensures that the design requirements for the system elements reflect the functional performance requirements, that all functional perfor-

mance requirements are satisfied by the combined system elements, and that such requirements are optimized with respect to system performance requirements and constraints.

13.1.3 Phases of a Typical System Design Project

The video industry has always been a very dynamic industry as a result of the rapid advancement of electronic technology. The design of a complex modern video facility can be used to illustrate the systems engineering approach.

13.1.3a Design Development

System design is carried out in a series of steps that lead to an operational facility. Appropriate research and preliminary design work is completed in the first phase of the project, the *design and development phase*. It is the intent of this phase to fully delineate all requirements of the project and to identify any constraints. Based on initial concepts and information, the design requirements are modified until all concerned parties are satisfied and approval is given for the final design work to proceed. The first objective of this phase is to answer the following questions.

- What are the functional requirements of the product of this work?
- What are the physical requirements of the product of this work?
- What are the performance requirements of the product of this work?
- Are there any constraints limiting design decisions?
- Will existing equipment be used?
- Is the existing equipment acceptable?
- Will this be a new facility or a renovation?
- Will this be a retrofit or upgrade to an existing system?
- Will this be a stand alone system?

Working closely with the customer's representatives, the equipment and functional requirements of each of the major technical areas of the facility are identified. In the case of facility renovation, the systems engineer's first order of business is to analyze existing equipment. A visit is made to the site to gather detailed information about the existing facility. Usually confronted with a mixture of acceptable and unacceptable equipment, the systems engineer must sort out those pieces of equipment that meet current standards, and determine which items should be replaced. Then, after soliciting input from the facility's technical and operational personnel, the systems engineer develops a list of needed equipment.

One of the systems engineer's most important contributions is the ability to identify and meet the needs of the customer, and do it within the project budget. Based on the customer's initial concepts and any subsequent equipment utilization research conducted by the systems engineer, the desired capabilities are identified as precisely as possible. Design parameters and objectives are defined and reviewed. Functional efficiency is maximized to allow operation by a minimum

13-16 Standards and Practices

number of personnel. Future needs are also investigated at this time. Future technical systems expansion is considered.

After the customer approves the equipment list, preliminary system plans are drawn up for review and further development. If architectural drawings of the facility are available, they can be used as a starting point for laying out an equipment floor plan. The systems engineer uses this floor plan to be certain adequate space is provided for present and future equipment, as well as adequate clearance for maintenance and convenient operation. Equipment identification is then added to the architect's drawings.

Documentation should include, but not be limited to:

- Equipment list with prices
- Technical system functional block diagrams
- Custom item descriptions
- Rack and console elevations
- Equipment floor plans

The preliminary drawings and other supporting documents are prepared to record design decisions and to illustrate the design concepts to the customer. Renderings, scale models, or full size mock ups may also be needed to better illustrate, clarify, or test design ideas.

Ideas and concepts have to be exchanged and understood by all concerned parties. Good communication skills are essential for the team members. The bulk of the creative work is carried out in the design development phase. The physical layout—the look and feel—and the functionality of the facility will all have been decided and agreed upon by the completion of this phase of the project. If the design concepts appear feasible, and the cost is within the anticipated budget, management can authorize work to proceed on the final detailed design.

13.1.3b Electronic System Design

Performance standards and specifications have to be established up-front in a technical facility project. These will determine the performance level of equipment that will be acceptable for use in the system and affect the size of the budget. Signal quality, stability, reliability, and accuracy are examples of the kinds of parameters that have to be specified. Access and processor speeds are important parameters when dealing with computer driven products. The systems engineer has to confirm whether selected equipment conforms to the standards.

At this point it must be determined what functions each component in the system will be required to fulfill, and how each will function together with other components in the system. The management and operation staff usually know what they would like the system to do, and how they can best accomplish it. They have probably selected equipment that they think will do the job. With a familiarity of the capabilities of different equipment, the systems engineer should be able to contribute to this function-definition stage of the process. Questions that need to be answered include:

- What functions must be available to the operators?
- What functions are secondary and therefore not necessary?
- What level of automation should be required to perform a function?

- How accessible should the controls be?

Over-engineering or over-design should be avoided. This serious and costly mistake can be made by engineers and company staff when planning technical system requirements. A staff member may, for example, ask for a seemingly simple feature or capability without fully understanding its complexity or the additional cost burden it may impose on a project. Other portions of the system may have to be compromised in order to implement the additional feature. An experienced systems engineer will be able to spot this type of issue and determine whether the tradeoffs and added engineering time and cost are really justified.

When existing equipment is going to be used, it will be necessary to make an inventory list. This list will be the starting point for developing a final equipment list. Usually, confronted with a mixture of acceptable and unacceptable equipment, the systems engineer must sort out what meets current standards and what should be replaced. Then, after soliciting input from facility technical personnel, the systems engineer develops a summary of equipment needs, including future acquisitions. One of the systems engineer's most important contributions is the ability to identify and meet these needs within the facility budget.

A list of major equipment is prepared. The systems engineer selects the equipment based on experience with the products, and on customer preferences. Often some of the existing equipment may be reused. A number of considerations are discussed with the facility customer to arrive at the best product selection. Some of the major points include:

- Budget restrictions
- Space limitations
- Performance requirements
- Ease of operation
- Flexibility of use
- Functions and features
- Past performance history
- Manufacturer support

The goal here is the specification of equipment to meet the functional requirements of the project efficiently and economically. Simplified block diagrams for the video, audio, control, data, and communication systems are drawn. They are discussed with the customer and presented for approval.

13.1.3c Detailed Design

With the research and preliminary design development completed, the details of the design must now be concluded. The design engineer prepares complete detailed documentation and specifications necessary for the fabrication and installation of the technical systems, including all major and minor components. Drawings must show the final configuration and the relationship of each component to other elements of the system, as well as how they will interface with other building services, such as air conditioning and electrical power. This documentation must communicate the design requirements to the other design professionals, including the construction and installation contractors.

13-18 Standards and Practices

In this phase, the systems engineer develops final, detailed flow diagrams and schematics that show the interconnection of all equipment. Cable interconnection information for each type of signal is taken from the flow diagrams and recorded on the cable schedule. Cable paths are measured and timing calculations are made for signals requiring synchronization, such as video, synchronizing pulses, subcarrier, and digital audio and/or video (when required by the design). These timed cable lengths are entered onto the cable schedule.

The *flow diagram* is a schematic drawing used to show the interconnections between all equipment that will be installed. It is different from a block diagram in that it contains much more detail. Every wire and cable must be included on the drawings.

The starting point for preparing a flow diagram can vary depending upon the information available from the design development phase of the project, and on the similarity of the project to previous projects. If a similar system has been designed in the past, the diagrams from that project can be modified to include the equipment and functionality required for the new system. New models of the equipment can be shown in place of their counterparts on the diagram, and only minor wiring changes made to reflect the new equipment connections and changes in functional requirements. This method is efficient and easy to complete.

If the facility requirements do not fit any previously completed design, the block diagram and equipment list are used as a starting point. Essentially, the block diagram is expanded and details added to show all of the equipment and their interconnections, and to show any details necessary to describe the installation and wiring completely.

An additional design feature that might be desirable for specific applications is the ability to easily disconnect a rack assembly from the system and relocate it. This would be the case if the system were to be prebuilt at a systems integration facility and later moved and installed at the client's site. When this is a requirement, the interconnecting cable harnessing scheme must be well planned in advance and identified on the drawings and cable schedules.

Special custom items need to be defined and designed. Detailed schematics and assembly diagrams are drawn. Parts lists and specifications are finalized, and all necessary details worked out for these items. Mechanical fabrication drawings are prepared for consoles and other custom-built cabinetry.

The design engineer provides layouts of cable runs and connections to the architect. Such detailed documentation simplifies equipment installation and facilitates future changes in the system. During preparation of final construction documents, the architect and the design engineer can firm-up the layout of the technical equipment wire ways, including access to flooring, conduits, trenches, and overhead wire ways.

Dimensioned floor plans and elevation drawings are required to show placement of equipment, lighting, electrical cable ways, duct, conduit, and HVAC ducting. Requirements for special construction, electrical, lighting, HVAC, finishes, and acoustical treatments must be prepared and submitted to the architect for inclusion in the architectural drawings and specifications. This type of information, along with cooling and electrical power requirements, also must be provided to the mechanical and electrical engineering consultants (if used on the project) so they can begin their design calculations.

Equipment heat loads are calculated and submitted to the HVAC consultant. Steps are taken when locating equipment to avoid any excessive heat buildup within the equipment enclosures while maintaining a comfortable environment for the operators.

Electrical power loads are calculated and submitted to the electrical consultant and steps taken to provide for sufficient power, proper phase balance, and backup electricity as required.

13.1.3d Customer Support

The systems engineer can assist in purchasing equipment and help to coordinate the move to a new or renovated facility. This can be critical if a great deal of existing equipment is being relocated. In the case of new equipment, the customer will find the systems engineer's knowledge of prices, features, and delivery times to be an invaluable asset. A good systems engineer will see to it that equipment arrives in ample time to allow for sufficient testing and installation. A good working relationship with equipment manufacturers helps guarantee their support and speedy response to the customer's needs.

The systems engineer can also provide engineering management support during planning, construction, installation, and testing to help qualify and select contractors, resolve problems, explain design requirements, and assure quality workmanship by the contractors and the technical staff.

The procedures described in this section outline an ideal scenario. In reality, management may often try to bypass many of the foregoing steps to save money. This, the reasoning goes, will eliminate unnecessary engineering costs and allow construction to start right away. Utilizing in-house personnel, a small company may attempt to handle the job without professional help. This puts an added burden on the staff who are already working full time taking care of the daily operation of the facility. With inadequate design detail and planning, which can result when using unqualified people, the job of setting technical standards and making the system work then defaults to the construction contractors, in-house technical staff, or the installation contractor. This can result in costly and uncoordinated work-arounds and—of course—delays and added costs during construction, installation, and testing. It makes the project less manageable and less likely to be completed successfully.

The complexity of a project can be as simple as interconnecting a few pieces of equipment together to designing the software for an automated robotic storage system. The size of a technical facility can vary from a small one room operation to a large multimillion dollar facility. Where large amounts of money and other resources are going to be involved, management is well advised to recruit the services of qualified system engineers.

13.1.3e Budget Requirements Analysis

The need for a project may originate with customers, management, operations staff, technicians, or engineers. In any case, some sort of logical reasoning or a specific production requirement will justify the cost. On small projects, like the addition of a single piece of equipment, money only has to be available to make the purchase and cover installation costs. When the need may justify a large project, it is not always immediately apparent how much the project will cost to complete. The project has to be analyzed by dividing it up into its constituent elements. These elements include:

- Equipment
- Materials
- Resources (including money and man hours needed to complete the project)

An executive summary or capital project budget request containing a detailed breakdown of these elements can provide the information needed by management to determine the return on investment, and to make an informed decision on whether or not to authorize the project.

13-20 Standards and Practices

A *capital project budget request* containing the minimum information might consist of the following items:

- Project name. Use a name that describes the result of the project, such as “control room upgrade.”
- Project number (if required). A large organization that does many projects will use a project numbering system of some kind, or may use a budget code assigned by the accounting department.
- Project description. A brief description of what the project will accomplish, such as “design the technical system upgrade for the renovation of production control room 2.”
- Initiation date. The date the request will be submitted.
- Completion date. The date the project will be completed.
- Justification. The reason the project is needed.
- Material cost breakdown. A list of equipment, parts, and materials required for construction, fabrication, and installation of the equipment.
- Total material cost.
- Labor cost breakdown. A list of personnel required to complete the project, their hourly pay rates, the number of hours they will spend on the project, and the total cost for each.
- Total project cost. The sum of material and labor costs.
- Payment schedule. Estimation of individual amounts that will have to be paid out during the course of the project and the approximate dates each will be payable.
- Preparer's name and the date prepared.
- Approval signature(s) and date(s) approved.

More detailed analysis, such as return on investment, can be carried out by an engineer, but financial analysis should be left to the accountants who have access to company financial data.

Feasibility Study and Technology Assessment

In the case where it is required that an attempt be made to implement new technology, and where a determination must be made as to whether certain equipment can perform a desired function, it will be necessary to conduct a *feasibility study*. The systems engineer may be called upon to assess the state-of-the-art in order to develop a new application. An executive summary or a more detailed report of evaluation test results may be required, in addition to a budget estimate, in order to help management make their decision.

Planning and Control of Scheduling and Resources

Several planning tools have been developed for planning and tracking progress toward the completion of projects and scheduling and controlling resources. The most common graphical project management tools are the *Gantt Chart* and the *Critical Path Method* (CPM) utilizing the *Project Evaluation and Review* (PERT) technique. Computerized versions of these tools have greatly enhanced the ability of management to control large projects.

13.1.3f Project Tracking and Control

A project team member may be called upon by the project manager to report the status of work during the course of the project. A standardized project status report form can provide consistent and complete information to the project manager. The purpose is to supply information to the project manager regarding work completed, and money spent on resources and materials.

A project status report containing the minimum information might contain the following items:

- Project number (if required)
- Date prepared
- Project name
- Project description
- Start date
- Completion date (the date this part of the project was completed)
- Total material cost
- Labor cost breakdown
- Preparer's name
- Change Control

After part or all of a project design has been approved and money allocated to build it, any changes may increase or decrease the cost. Factors that effect the cost include:

- Components and materials
- Resources, such as labor and special tools or construction equipment
- Costs incurred because of manufacturing or construction delays

Management will want to know about such changes, and will want to control them. For this reason, a method of reporting changes to management and soliciting their approval to proceed with the change may have to be instituted. The best way to do this is with a *change order request* or change order. A change order includes a brief description of the change, the reason for the change, a summary of the effect it will have on costs, and what effect it will have on the project schedule.

Management will exercise its authority and approve or disapprove each change based upon its understanding of the cost and benefits, and their perceived need for the modification of the original plan. Therefore, it is important that the systems engineer provide as much information and explanation as may be necessary to make the change clear and understandable to management.

A change order form containing the minimum information might include the following items:

- Project number
- Date prepared
- Project name
- Labor cost breakdown

13-22 Standards and Practices

- Preparer's name
- Description of the change
- Reason for the change
- Equipment and materials to be added or deleted
- Material costs or savings
- Labor costs or savings
- Total cost of this change (increase or decrease)
- Impact on the schedule
- Program Management

Systems engineering is both a technical process and a management process. Both processes must be applied throughout a program if it is to be successful. The persons who plan and carry out a project constitute the *project team*. The makeup of a project team will vary depending on the size of the company and the complexity of the project. It is up to management to provide the necessary human resources to complete the project.

13.1.3g Executive Management

The executive manager is the person who can authorize that a project be undertaken. This person can allocate funds and delegate authority to others to accomplish the task. Motivation and commitment is toward the goals of the organization. The ultimate responsibility for a project's success is in the hands of the executive manager. This person's job is to get things done through other people by assigning group responsibilities, coordinating activities between groups, and resolving group conflicts. The executive manager establishes policy, provides broad guidelines, approves the project master plan, resolves conflicts, and assures project compliance with commitments.

Executive management delegates the project management functions and assigns authority to qualified professionals, allocates a capital budget for the project, supports the project team, and establishes and maintains a healthy relationship with project team members.

Management has the responsibility to provide clear information and goals—up front—based upon their needs and initial research. Before initiating a project, the company executive should be familiar with daily operation of the facility, analyze how the company works, how the staff does their jobs, and what tools they need to accomplish the work. Some points that may need to be considered by an executive before initiating a project include:

- What is the current capital budget for equipment?
- Why does the staff currently use specific equipment?
- What function of the equipment is the weakest within the organization?
- What functions are needed but cannot be accomplished with current equipment?
- Is the staff satisfied with current hardware?
- Are there any reliability problems or functional weaknesses?

- What is the maintenance budget, and is it expected to remain steady?
- How soon must the changes be implemented?
- What is expected from the project team?

Only after answering the appropriate questions will the executive manager be ready to bring in expert project management and engineering assistance. Unless the manager has made a systematic effort to evaluate all the obvious points about the facility requirements, the not-so-obvious points may be overlooked. Overall requirements must be broken down into their component parts. Do not try to tackle ideas that have too many branches. Keep the planning as basic as possible. If the company executive does not make a concerted effort to investigate the needs and problems of a facility thoroughly before consulting experts, the expert advice will be shallow and incomplete, no matter how good the engineer.

Engineers work with the information they are given. They put together plans, recommendations, budgets, schedules, purchases, hardware, and installation specifications based upon the information they receive from interviewing management and staff. If the management and staff have failed to go through the planning, reflection, and refinement cycle before those interviews, the company will likely waste time and money.

13.1.3h Project Manager

Project management is an outgrowth of the need to accomplish large complex projects in the shortest possible time, within the anticipated cost, and with the required performance and reliability. Project management is based upon the realization that modern organizations may be so complex as to preclude effective management using traditional organizational structures and relationships. Project management can be applied to any undertaking that has a specific end objective.

The project manager is the person who has the authority to carry out a project. This person has been given the legitimate right to direct the efforts of the project team members. The manager's power comes from the acceptance and respect accorded him or her by superiors and subordinates. The project manager has the power to act, and is committed to group goals.

The project manager is responsible for getting the project completed properly, on schedule and within budget, by utilizing whatever resources are necessary to accomplish the goal in the most efficient manner. The manager provides project schedule, financial, and technical requirement direction and evaluates and reports on project performance. This requires planning, organizing, staffing, directing, and controlling all aspects of the project.

In this leadership role, the project manager is required to perform many tasks including the following:

- Assemble the project organization.
- Develop the project plan.
- Publish the project plan.
- Set measurable and attainable project objectives.
- Set attainable performance standards.
- Determine which scheduling tools (PERT, CPM, and/ or GANTT) are right for the project.

13-24 Standards and Practices

- Using the scheduling tools, develop and coordinate the project plan, which includes the budget, resources, and the project schedule.
- Develop the project schedule.
- Develop the project budget.
- Manage the budget.
- Recruit personnel for the project.
- Select subcontractors.
- Assign work, responsibility, and authority so team members can make maximum use of their abilities.
- Estimate, allocate, coordinate, and control project resources.
- Deal with specifications and resource needs that are unrealistic.
- Decide upon the right level of administrative and computer support.
- Train project members on how to fulfill their duties and responsibilities.
- Supervise project members, giving them day-to-day instructions, guidance, and discipline as required to fulfill their duties and responsibilities.
- Design and implement reporting and briefing information systems or documents that respond to project needs.
- Control the project.

Some basic project management practices can improve the chances for success. Consider the following:

- Secure the necessary commitments from top management to make the project a success.
- Set up an action plan that will be easily adopted by management.
- Use a work breakdown structure that is comprehensive and easy to use.
- Establish accounting practices that help, not hinder, successful completion of the project.
- Prepare project team job descriptions properly up-front to eliminate conflict later on.
- Select project team members appropriately the first time.

After the project is under way, follow these steps:

- Manage the project, but make the oversight reasonable and predictable.
- Get team members to accept and participate in the plans.
- Motivate project team members for best performance.
- Coordinate activities so they are carried out in relation to their importance with a minimum of conflict.
- Monitor and minimize inter-departmental conflicts.
- Get the most out of project meetings without wasting the team's productive time. Develop an agenda for each meeting, and start on time. Conduct one piece of business at a time. Assign

responsibilities where appropriate. Agree on follow-up and accountability dates. Indicate the next step for the group. Set the time and place for the next meeting. End on time.

- Spot problems and take corrective action before it is too late.
- Discover the strengths and weaknesses in project team members and manage them to get the desired results.
- Help team members solve their own problems.
- Exchange information with subordinates, associates, superiors, and others about plans, progress, and problems.
- Make the best of available resources.
- Measure project performance.
- Determine, through formal and informal reports, the degree to which progress is being made.
- Determine causes of and possible ways to act upon significant deviations from planned performance.
- Take action to correct an unfavorable trend, or to take advantage of an unusually favorable trend.
- Look for areas where improvements can be made.
- Develop more effective and economical methods of managing.
- Remain flexible.
- Avoid “activity traps”.
- Practice effective time management.

When dealing with subordinates, each person must:

- Know what they are supposed to do, preferably in terms of an end product.
- Have a clear understanding of what their authority is, and its limits.
- Know what their relationship with other people is.
- Know what constitutes a job well done in terms of specific results.
- Know when and what they are doing exceptionally well.
- Understand that there are just rewards for work well done, and for work exceptionally well done.
- Know where and when they are falling short of expectations.
- Be made aware of what can and should be done to correct unsatisfactory results.
- Feel that their superior has an interest in them as an individual.
- Feel that their superior believes in them and is anxious for them to succeed and progress.

By fostering a good relationship with associates, the manager will have less difficulty communicating with them. The fastest, most effective communication takes place among people with common points of view.

13-26 Standards and Practices

The competent project manager watches what is going on in great detail and can, therefore, perceive problems long before they flow through the paper system. Personal contact is faster than filing out formal forms. A project manager who spends most of his or her time in the management office instead of roaming through the places where the work is being done, is headed for catastrophe.

13.1.3i Systems Engineer

The term *systems engineer* means different things to different people. The systems engineer is distinguished from the engineering specialist, who is concerned with only one aspect of a well-defined engineering discipline in that he must be able to adapt to the requirements of almost any type of system. The systems engineer provides the employer with a wealth of experience gained from many successful approaches to technical problems developed through hands-on exposure to a variety of situations. This person is a professional with knowledge and experience, possessing skills in a specialized and learned field or fields. The systems engineer is an expert in these fields; highly trained in analyzing problems and developing solutions that satisfy management objectives.

A competent systems engineer has a wealth of technical information that can be used to speed up the design process and help in making cost effective decisions. The experienced systems engineer is familiar with proper fabrication, construction, installation, and wiring techniques and can spot and correct improper work.

Training in personnel relations, a part of the engineering curriculum, helps the systems engineer communicate and negotiate professionally with subordinates and management.

Small in-house projects can be completed on an informal basis and, indeed, this is probably the normal routine where the projects are simple and uncomplicated. In a large project, however, the systems engineer's involvement usually begins with preliminary planning and continues through fabrication, implementation, and testing. The degree to which program objectives are achieved is an important measure of the systems engineer's contribution.

During the design process the systems engineer:

- Concentrates on results and focuses work according to the management objectives.
- Receives input from management and staff.
- Researches the project and develops a workable design.
- Assures balanced influence of all required design specialties.
- Conducts design reviews.
- Performs trade-off analyses.
- Assists in verifying system performance.
- Resolves technical problems related to the design, interface between system components, and integration of the system into the facility.

Aside from designing a system, the systems engineer has to answer any questions and resolve problems that may arise during fabrication and installation of the hardware. This person must also monitor the quality and workmanship of the installation. The hardware and software will have to be tested and calibrated upon completion. This too is the concern of the systems engineer. During the production or fabrication phase, systems engineering is concerned with:

- Verifying system capability
- Verifying system performance
- Maintaining the system baseline
- Forming an analytical framework for further analysis

Depending on the complexity of the new installation, the systems engineer may have to provide orientation and operating instruction to the users. During the operational support phase, system engineers:

- Receive input from users
- Evaluate proposed changes to the system
- Establish their effectiveness
- Facilitates the effective incorporation of changes, modifications, and updates

Depending on the size of the project and the management organization, the systems engineer's duties will vary. In some cases the systems engineer may have to assume the responsibilities of planning and managing smaller projects.

13.1.4 References

1. Hoban, F. T., and W. M. Lawbaugh: *Readings In Systems Engineering*, NASA, Washington, D.C., 1993.
2. *System Engineering Management Guide*, Defense Systems Management College, Virginia, 1983.

13.1.5 Bibliography

- Delatore, J. P., E. M. Prell, and M. K. Vora: "Translating Customer Needs Into Product Specifications", *Quality Progress*, January 1989.
- DeSantis, Gene: "Systems Engineering Concepts," in *NAB Engineering Handbook*, 9th ed., Jerry C. Whitaker (ed.), National Association of Broadcasters, Washington, D.C., 1999.
- DeSantis, Gene: "Systems Engineering," in *The Electronics Handbook*, Jerry C. Whitaker (ed.), CRC Press, Boca Raton, Fla., 1996.
- Finkelstein, L.: "Systems Theory", *IEE Proceedings*, vol. 135, Part A, no. 6, July 1988.
- Shinnars, S. M.: *A Guide to Systems Engineering and Management*, Lexington, 1976.
- Tuxal, J. G.: *Introductory System Engineering*, McGraw-Hill, New York, N.Y., 1972.

Chapter

13.2

Engineering Documentation

Fred Baumgartner, Terrence M. Baun

13.2.1 Introduction

Video facilities are designed to have as little down-time as possible. Yet, inadequate documentation is a major contributor to the high cost of systems maintenance and the resulting widespread replacement of poorly documented facilities. The cost of neglecting *hours* of engineering documentation is paid in *weeks* of reconstruction.

Documentation is a management function every bit as important as project design, budgeting, planning, and quality control; it is often the difference between an efficient and reliable facility and a misadventure. If the broadcast engineer does not feel qualified to attempt documentation of a project, the engineer must at the very least oversee and approve the documentation developed by others.

Within the last few years the need for documentation has increased with the complexity of the broadcast systems. Fortunately, the power of documentation tools has kept pace.

13.2.2 Basic Concepts

The first consideration in the documentation process is the complexity of the installation. A basic video editing station may require almost no formal documentation, while a large satellite or network broadcast facility may require computerized databases and a full time staff doing nothing but documentation updates. Most facilities will fall somewhere in the middle of that spectrum.

A second concern is the need for flexibility at the facility. Seldom does a broadcast operation get "completely rewired" because the cabling wears out or fails. More often, it is the supporting documentation that has broken down, frustrating the maintainability of the system. Retroactive documentation is physically difficult and emotionally challenging, and seldom generates the level of commitment required to be entirely successful or accurate; hence, a total rebuild is often the preferred solution to documentation failure. Documentation must be considered a hedge against such unnecessary reconstruction.

Finally, consider efficiency and speed. Documentation is a prepayment of time. Repairs, rerouting, replacements, and reworking all go faster and smoother with proper documentation. If

13-30 Standards and Practices

your installation is one in which any downtime or degradation of service is unacceptable, then budgeting sufficient time for the documentation process is critically important.

Because, in essence, documentation is education, knowing “how much is enough” is a difficult decision. We will begin by looking at the basics, and then expand our view of the documentation process to fit specialized situations.

13.2.2a The Manuals

Even if you never take a pen to paper, you do have one source of documentation to care for, since virtually every piece of commercial equipment comes with a manual.

Place those manuals in a centralized location, and arrange them in an order that seems appropriate for your station. Most engineering shops file manuals alphabetically, but some prefer a filing system based on equipment placement. (For example, production studio equipment manuals would be filed together under a “Production Studio” label and might even be physically located in the referenced studio.) But whatever you do, be consistent. Few things are as frustrating to a technician as being unable to locate a manual when needing something as simple as a part number or the manufacturer’s address.

Equipment manuals are the first line of documentation and deserve our attention and respect.

13.2.2b Documentation Conventions

The second essential item of documentation is the statement of “conventions.” By this we mean a document containing basic information essential to an understanding of the facility, posted in an obvious location and available to all who maintain the plant. Consider the following examples of conventions:

- Where are the equipment manuals and how are they organized?
- What is the architecture of the ground system? Where is the central station ground and is it a star, grid, or other distribution pattern? Are there separate technical and power grounds? Are audio shield grounded at the source, termination, or both locations?
- What is the standard audio input/output architecture? Is this a +8 dBm, +4 dBm, or 0 dBm facility? Is equipment sourced at $600\ \Omega$ terminated at its destination, or left unterminated? Are unbalanced audio sources wired with the shield as ground or is the low side of a balanced pair used for that purpose? How are XLR-type connectors wired—pin 2 high or low?
- How can a technician disconnect utility power to service line voltage wiring within racks? Where are breakers located, and how are they marked? What equipment is on the UPS power, generator, or utility power? Whom do you call for power and building systems maintenance?
- Where are the keys to the transmitter? Is the site alarmed?

For such an essential information source, you will find it takes very little time to generate the conventions document. Keep it short—it is not meant to be a book. A page or two should be sufficient for most installations and, if located in an obvious place, this document will keep the technical staff on track and will save service personnel from stumbling around searching for basic information. This document is the key to preventing many avoidable embarrassments.

The next step beyond the conventions document is a documentation *system*. There are three primary methods:

- Self documentation
- Database documentation
- Graphic documentation

In most cases, a mixture of all three is necessary and appropriate. In addition to documenting the physical plant and its interconnections, each piece of equipment, whether commercially produced or custom made, must be documented in an organized manner.

Self Documentation

In situations where the facility is small and very routine, self documentation is possible. Self documentation relies on a set of standard practices that are repeated. Telephone installations, for example, appear as a mass of perplexing wires. In reality, the same simple circuits are simply repeated in an organized and universal manner. To the initiated, any telephone installation that follows the rules is easy to understand and repair or modify, no matter where or how large. Such a system is truly self documenting. Familiarity with telephone installations is particularly useful, because the telephone system was the first massive electronic installation. It is the telephone system that gave us "relay" racks, grounding plans, demarcation points, and virtually all of the other concepts that are part of today's electronic control or communications facility.

The organization, color codes, terminology, and layout of telephone systems is recorded in minute detail. Once a technician is familiar with the rules of telephone installations, drawings and written documentation are rarely required for routine expansion and repair. The same is true for many parts of other facilities. Certainly, much of the wiring in any given rack of equipment can be self documenting. For example, a video tape recorder will likely be mounted in a rack with a picture monitor, audio monitor, waveform monitor, and vectorscope. The wiring between each of these pieces of equipment is clearly visible, with all wires short and their purpose obvious to any technician familiar with the rules of video. Furthermore, each video cable will conform to the same standards of level or data configuration. Additional documentation, therefore, is largely unnecessary.

By convention, there are rules of grounding, power, and signal flow in all engineering facilities. In general, it can be assumed that in most communications facilities, the ground will be a star system, the power will be individual 20 amp feeds to each rack, and the signal flow within each rack will be from top to bottom. Rules that might vary from facility to facility include color coding, connector pin outs, and rules for shield and return grounding.

To be self documenting, the rules must be determined and all of the technicians working on the facility must know and follow the conventions. The larger the number of technicians, or the higher the rate of staff turn-over, the more important it is to have a readily available document that clearly covers the conventions in use.

One thing must be very clear: a facility that does not have written documentation is not automatically self documenting. Quite the contrary. A written set of conventions and unfaltering adherence to them are the trade marks of a self documenting facility.

While it is good engineering practice to design all facilities to be as self documenting as possible, there are limits to the power of self documentation. In the practical world, self documentation can greatly reduce the amount of written documentation required, but can seldom replace it entirely.

13-32 Standards and Practices

Database Documentation

As facilities expand in size and complexity, a set of conventions will longer answer all of the questions. At some point, a wire leaves an equipment rack and its destination is no longer obvious. Likewise, equipment will often require written documentation as to its configuration and purpose, especially if it is utilized in an uncharacteristic way.

Database documentation records the locations of both ends of a given circuit. For this, each cable must be identified individually. There are two common systems for numbering cables: *ascension numbers*, and *from-to coding*. In ascension numbering schemes, each wire or cable is numbered in increasing order, one, two, three, and so on. In from-to coding, the numbers on each cable represent the source location, the destination location, and normally some identification as to purpose and/or a unique identifier. For example, a cable labeled 31-35-B6 might indicate that cable went from a piece of equipment in rack 31 to another unit in rack 35 and carries black, it is also the sixth cable to follow the same route and carry the same class of signal.

Each method has its benefits. Ascension numbering is easier to assign, and commonly available preprinted wire labels can be used. On the other hand, ascension numbers contain no hints as to wire purpose or path, and for that reason “purpose codes” are often added to the markings.

From-to codes can contain a great deal more information without relying on the printed documentation records, but often space does not permit a full delineation on the tag itself. Here again, supplemental information may still be required in a separate document or database.

Whatever numbering system is used, a complete listing must be kept in a database of some type. In smaller installations, this might simply be a spiral notebook that contains a complete list of all cables, their source, destination, any demarcations, and signal parameters.

Because all cabling can be considered as a transmission line, all cabling involves issues of termination. In some data and analog video applications, it is common for a signal to “loop-through” several pieces of equipment. Breaking or tapping into the signal path often has consequences elsewhere, resulting in unterminated or double terminated lines. While more forgiving, analog audio has similar concerns. Therefore, documentation must include information on such termination.

Analog audio and balanced lines used in instrumentation have special concerns of their own. It is seldom desirable to ground both ends of a shielded cable. Again, the documentation must reflect which end(s) of a given shield is grounded.

In many cases, signal velocity is such that the length of the lines and the resulting propagation delay is critical. In such circumstances, this is significant information that should be retained. In cases where differing signal levels or configurations are used (typical in data and control systems), it is the documentor’s obligation to record those circumstances as well.

However or wherever the database documentation is retained, it represents the basic information that defines the facility interconnections and must be available for updating and duplication as required.

Graphic Documentation

Electronics is largely a graphical language. Schematics and flow charts are more understandable than net lists or cable interconnection lists. Drawings, either by hand—done with the aid of drafting machines and tools—or accomplished on CAD (computer aided design) programs are highly useful in conveying overall facility design quickly and clearly. Normally, the wire numbering

scheme will follow that used in the database documentation, so that the graphic and text documentation can be used together.

CAD drawings are easy to update and reprint. For this reason, documentation via CAD is becoming more popular, even in smaller installations. Because modern CAD programs not only draw but also store information, they can effectively serve as an electronic file cabinet for documentation. While there have been attempts to provide electronic/telecommunications engineering documentation “templates” and corresponding technical graphics packages for CAD programs, most of the work in this area has been done by engineers working independently to develop their own systems. Obviously, the enormous scope of electronic equipment and telecommunications systems make it impractical for a “standard” CAD package to suit every user.

13.2.2c Update Procedures

Because documentation is a dynamic tool, as the facility changes, so too must the documentation. It is common for a technician to “improve” conditions by reworking a circuit or two. Most often this fixes a problem that should be corrected as a maintenance item. But sometimes, it plants a “time bomb” wherein a future change, based on missing or incomplete documentation of the previous work, will cause problems.

It is essential that there be a means of consistently updating the documentation. The most common way of accomplishing this is the mandatory “change sheet.” Here, whenever a technician makes a change it is reported back to those who keep the documentation. If the changes are extensive, the use of the “red-line” drawing and “edit sheet” come into play; the original drawing and database, respectively, are printed and corrected with a “red pen.” This document then is used to update the original documentation.

In some cases, the updating process can be tied to the engineering reports or discrepancy process. Most facilities use some form of “trouble ticket” to track equipment and system performance and to report and track maintenance. This same form may be used to report changes required of the documentation or errors in existing documentation.

13.2.2d Equipment Documentation

Plant documentation does not end when all of the circuit paths in the facility are defined. Each circuit begins and ends at a piece of equipment, which can be modified, reconfigured, or removed from service. Keep in mind that unless the lead technician lives forever, never changes jobs, and never takes a day off, someone unfamiliar with the equipment will eventually be asked to return it to service. For this reason, a documentation file for each piece of equipment must be maintained.

Equipment documentation contains these key elements:

- The equipment manual
- Modification record
- Configuration information
- Maintenance record

The equipment manual is the manufacturer’s original documentation. As mentioned previously, the manuals must be organized in such a manner that they can be easily located. Typically,

13-34 Standards and Practices

the manuals are kept at the site where the equipment is installed (if practical). Remember that equipment with two “ends” such as STLs, RPUs, or remote control systems need manuals at each location!

Of course, if a piece of equipment is of custom construction, there must be particular attention paid to creating a manual. For this reason, a copy of the key schematics and documentation is often attached directly to the equipment. This “built-in manual” may be the only documentation to survive over the years.

Many pieces of equipment, over time, will require modification. Typically, the modification is recorded in three ways. First, internally to the equipment. A simple note glued into the chassis may be suitable, or a marker pen is used to write on a printed circuit board or other component. Second, the changes can be recorded in the manual, either inside the cover, or on the schematic or relevant pages. If the manual serves several machines, this may not be appropriate. If this is the case, a third option is to keep the modification information in a separate *equipment file*.

Equipment files are typically kept in standard file folders, and may be filed with the pertinent equipment manuals. Ideally, the equipment file is started when the equipment is purchased, and should contain purchase date, serial number, all modifications, equipment location(s), and a record of service.

The equipment file is the proper place to keep the configuration information. An increasingly large amount of equipment is microprocessor based or otherwise configurable for a specific mode of operation. Having a record of the machine’s default configuration is extremely helpful when a power glitch (or an operator) reconfigures a machine unexpectedly.

Equipment files should also contain repair records. With most equipment, documenting failures, major part replacements, operating time, and other service-related events serves a valuable purpose. Nothing is more useful in troubleshooting than a record of previous failures, configuration, modifications, and—of course—a copy of the original manual.

13.2.2e Operator/User Documentation

User documentation provides, at its most basic level, instructions on how to use a system. While most equipment manufacturers provide reasonably good instruction and operations manuals for their products, when those products are integrated into a system another level of documentation may be required. Complex equipment may require interface components that need to be adjusted from time to time, or various machines may be incompatible in some data transmission modes—all of which is essential to the proper operation of the system.

Such information often resides only in the heads of certain key users and is passed on by word of mouth. This level of informality can be dangerous, especially when changes take place in the users or maintenance staff, resulting in differing interpretations between operators and maintenance people about how the system normally operates. Maintenance personnel will then spend considerable time tracking hypothetical errors reported by misinformed users.

A good solution is to have the operators write an operating manual, providing a copy to the maintenance department. Such documentation will go a long way toward improving inter-departmental communications and should result in more efficient maintenance as well.

13.2.2f The True Cost of Documentation

There is no question that documentation is expensive—in some cases it can equal the cost of installation. Still, both installation and documentation expenses pale in comparison to the cost of equipment and potential revenue losses resulting from system down-time.

Documentation must be seen as a management and personnel issue of the highest order. Any lapse in the documentation updating process can result in disaster. Procrastination and the resulting lack of follow through will destroy any documentation system and ultimately result in plant failures, extend down-time, and premature rebuilding of the facility.

Making a business case for documentation is similar to making any business case. Gather together all the costs in time, hardware, and software on one side of the equation, and balance this against the savings in time and lost revenue on the other side. Engineering managers are expected to project costs accurately, and the allocation of sufficient resources for documentation and its requisite updating is an essential part of that responsibility.

13.2.3 Bibliography

Baumgartner, Fred, and Terrence Baun: "Broadcast Engineering Documentation," in *NAB Engineering Handbook*, 9th ed., Jerry C. Whitaker (ed.), National Association of Broadcasters, Washington, D.C., 1999.

Baumgartner, Fred, and Terrence Baun: "Engineering Documentation," in *The Electronics Handbook*, Jerry C. Whitaker (ed.), CRC Press, Boca Raton, Fla., 1996.

Chapter

13.3

Safety Issues

Jerry C. Whitaker, Editor-in-Chief

13.3.1 Introduction

Electrical safety is important when working with any type of electronic hardware. Because transmitters and many other systems operate at high voltages and currents, safety is doubly important. The primary areas of concern, from a safety standpoint, include:

- Electric shock
- Nonionizing radiation
- Beryllium oxide (BeO) ceramic dust
- Hot surfaces of vacuum tube devices
- Polychlorinated biphenyls (PCBs)

13.3.2 Electric Shock

Surprisingly little current is required to injure a person. Studies at Underwriters Laboratories (UL) show that the electrical resistance of the human body varies with the amount of moisture on the skin, the muscular structure of the body, and the applied voltage. The typical hand-to-hand resistance ranges between $500\ \Omega$ and $600\ k\Omega$ depending on the conditions. Higher voltages have the capability to break down the outer layers of the skin, which can reduce the overall resistance value. UL uses the lower value, $500\ \Omega$ as the standard resistance between major extremities, such as from the hand to the foot. This value is generally considered the minimum that would be encountered and, in fact, may not be unusual because wet conditions or a cut or other break in the skin significantly reduces human body resistance.

13.3.2a Effects on the Human Body

Table 13.3.1 lists some effects that typically result when a person is connected across a current source with a hand-to-hand resistance of $2.4\ k\Omega$. The table shows that a current of approximately 50 mA will flow between the hands, if one hand is in contact with a 120 V ac source and the

13-38 Standards and Practices

Table 13.3.1 The Effects of Current on the Human Body

Current	Effect
1 mA or less	No sensation, not felt
More than 3 mA	Painful shock
More than 10 mA	Local muscle contractions, sufficient to cause “freezing” to the circuit for 2.5 percent of the population
More than 15 mA	Local muscle contractions, sufficient to cause “freezing” to the circuit for 50 percent of the population
More than 30 mA	Breathing is difficult, can cause unconsciousness
50 mA to 100 mA	Possible ventricular fibrillation
100 mA to 200 mA	Certain ventricular fibrillation
More than 200 mA	Severe burns and muscular contractions; heart more apt to stop than to go into fibrillation
More than a few amperes	Irreparable damage to body tissue

other hand is grounded. The table indicates that even the relatively small current of 50 mA can produce *ventricular fibrillation* of the heart, and perhaps death. Medical literature describes ventricular fibrillation as rapid, uncoordinated contractions of the ventricles of the heart, resulting in loss of synchronization between heartbeat and pulse beat. The electrocardiograms shown in Figure 13.3.1 compare a healthy heart rhythm with one in ventricular fibrillation. Unfortunately, once ventricular fibrillation occurs, it will continue. Barring resuscitation techniques, death will ensue within a few minutes.

The route taken by the current through the body has a significant effect on the degree of injury. Even a small current, passing from one extremity through the heart to another extremity, is dangerous and capable of causing severe injury or electrocution. There are cases where a person has contacted extremely high current levels and lived to tell about it. However, usually when this happens, the current passes only through a single limb and not through the body. In these instances, the limb is often lost, but the person survives.

Current is not the only factor in electrocution. Figure 13.3.2 summarizes the relationship between current and time on the human body. The graph shows that 100 mA flowing through a human adult body for 2 s will cause death by electrocution. An important factor in electrocution, the *let-go range*, also is shown on the graph. This range is described as the amount of current that causes “freezing”, or the inability to let go of the conductor. At 10 mA, 2.5 percent of the population will be unable to let go of a “live” conductor. At 15 mA, 50 percent of the population will be unable to let go of an energized conductor. It is apparent from the graph that even a small amount of current can “freeze” someone to a conductor. The objective for those who must work around electric equipment is how to protect themselves from electric shock. Table 13.3.2 lists required precautions for personnel working around high voltages.

(a)



(b)



Figure 13.3.1 Electrocardiogram of a human heartbeat: (a) healthy rhythm, (b) ventricular fibrillation.

13.3.2b Circuit Protection Hardware

The typical primary panel or equipment circuit breaker or fuse will not protect a person from electrocution. In the time it takes a fuse or circuit breaker to blow, someone could die. However, there are protection devices that, properly used, may help prevent electrocution. The *ground-fault current interrupter* (GFCI), shown in Figure 13.3.3, works by monitoring the current being applied to the load. The GFI uses a differential transformer and looks for an imbalance in load current. If a current (5 mA , $\pm 1\text{ mA}$) begins to flow between the neutral and ground or between the hot and ground leads, the differential transformer detects the leakage and opens up the primary circuit within 2.5 ms .

GFI's will not protect a person from every type of electrocution. If the victim becomes connected to both the neutral and the hot wire, the GFI will not detect an imbalance.

13.3.2c Working with High Voltage

Rubber gloves are commonly used by engineers working on high-voltage equipment. These gloves are designed to provide protection from hazardous voltages or RF when the wearer is working on “hot” ac or RF circuits. Although the gloves may provide some protection from these hazards, placing too much reliance on them can have disastrous consequences. There are several reasons why gloves should be used with a great deal of caution and respect. A common mistake made by engineers is to assume that the gloves always provide complete protection. The gloves found in many facilities may be old or untested. Some may show signs of user repair, perhaps with electrical tape. Few tools could be more hazardous than such a pair of gloves.

Another mistake is not knowing the voltage rating of the gloves. Gloves are rated differently for both ac and dc voltages. For example, a *class 0* glove has a minimum dc breakdown voltage of 35 kV ; the minimum ac breakdown voltage, however, is only 6 kV . Furthermore, high-voltage

13-40 Standards and Practices

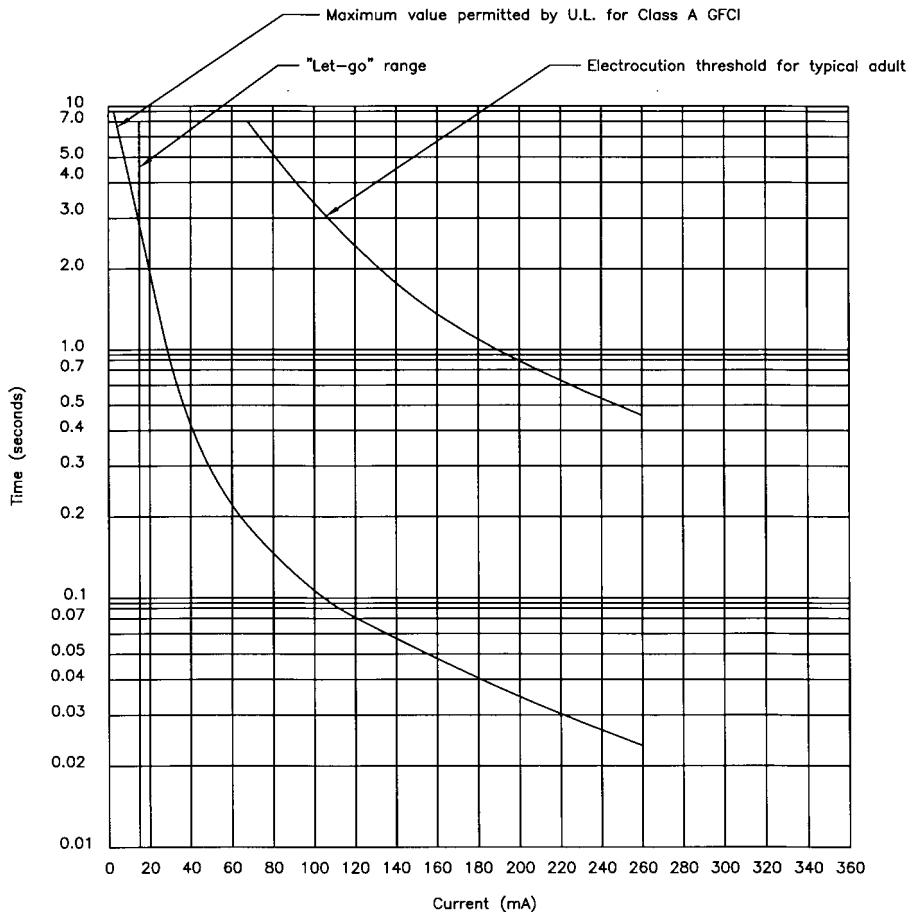


Figure 13.3.2 Effects of electric current and time on the human body. Note the “let-go” range.

rubber gloves are not usually tested at RF frequencies, and RF can burn a hole in the best of them. It is possible to develop dangerous working habits by assuming that gloves will offer the required protection.

Gloves alone may not be enough to protect an individual in certain situations. Recall the axiom of keeping one hand in a pocket while working around a device with current flowing? That advice is actually based on simple electricity. It is not the “hot” connection that causes the problem, but the ground connection that lets the current begin to flow. Studies have shown that more than 90 percent of electric equipment fatalities occurred when the grounded person contacted a live conductor. Line-to-line electrocution accounted for less than 10 percent of the deaths.

Table 13.3.2 Required Safety Practices for Engineers Working Around High-Voltage Equipment

High-Voltage Precautions	
✓	Remove all ac power from the equipment. Do not rely on internal contactors or SCRs to remove dangerous ac.
✓	Trip the appropriate power distribution circuit breakers at the main breaker panel.
✓	Place signs as needed to indicate that the circuit is being serviced.
✓	Switch the equipment being serviced to the local control mode as provided.
✓	Discharge all capacitors using the discharge stick provided by the manufacturer.
✓	Do not remove, short circuit, or tamper with interlock switches on access covers, doors, enclosures, gates, panels, or shields.
✓	Keep away from live circuits.

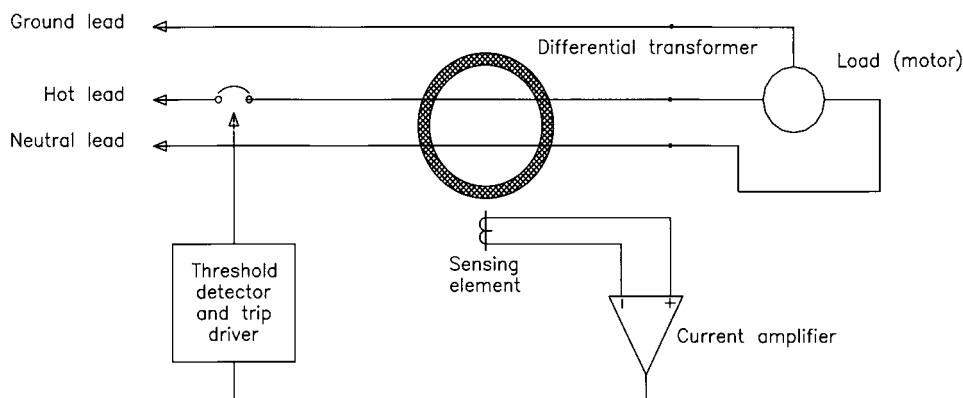


Figure 13.3.3 Basic design of a ground-fault interrupter (GFI).

When working around high voltages, always look for grounded surfaces. Keep hands, feet, and other parts of the body away from any grounded surface. Even concrete can act as a ground if the voltage is sufficiently high. If work must be performed in “live” cabinets, then consider using, in addition to rubber gloves, a rubber floor mat, rubber vest, and rubber sleeves. Although this may seem to be a lot of trouble, consider the consequences of making a mistake. Of course, the best troubleshooting methodology is never to work on any circuit without being certain that no hazardous voltages are present. In addition, any circuits or contactors that normally contain hazardous voltages should be firmly grounded before work begins.

13-42 Standards and Practices

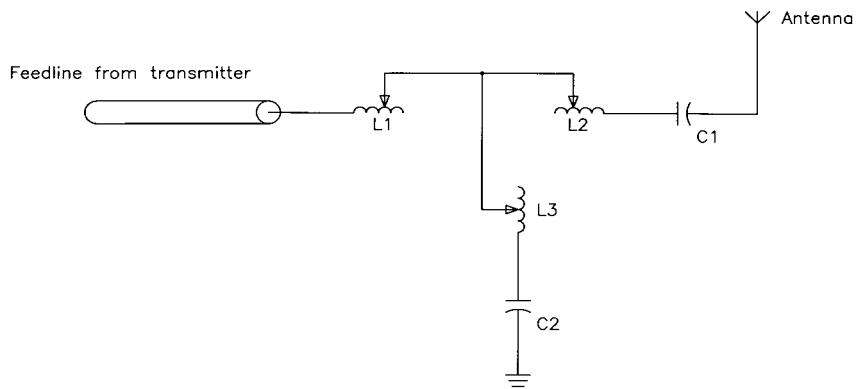


Figure 13.3.4 Example of how high voltages can be generated in an RF load matching network.

RF Considerations

Engineers often rely on electrical gloves when making adjustments to live RF circuits. This practice, however, can be extremely dangerous. Consider the typical load matching unit shown in Figure 13.3.4. In this configuration, disconnecting the coil from either L2 or L3 places the full RF output literally at the engineer's fingertips. Depending on the impedances involved, the voltages can become quite high, even in a circuit that normally is relatively tame.

In the Figure 13.3.4 example, assume that the load impedance is approximately $106 + j202 \Omega$. With 1 kW feeding into the load, the rms voltage at the matching output will be approximately 700 V. The peak voltage (which determines insulating requirements) will be close to 1 kV, and perhaps more than twice that if the carrier is being amplitude-modulated. At the instant the output coil clip is disconnected, the current in the shunt leg will increase rapidly, and the voltage easily could more than double.

13.3.2d First Aid Procedures

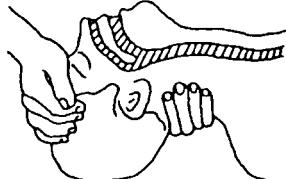
All engineers working around high-voltage equipment should be familiar with first aid treatment for electric shock and burns. Always keep a first aid kit on hand at the facility. Figure 13.3.5 illustrates the basic treatment for victims of electric shock. Copy the information, and post it in a prominent location. Better yet, obtain more detailed information from the local heart association or Red Cross chapter. Personalized instruction on first aid usually is available locally.

13.3.2e Operating Hazards

A number of potential hazards exist in the operation and maintenance of high-power vacuum tube RF equipment. Maintenance personnel must exercise extreme care around such hardware. Consider the following guidelines:

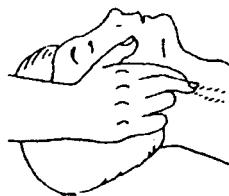
If the victim is not responsive, follow the A-B-Cs of basic life support.

A AIRWAY: If the victim is unconscious, open airway.



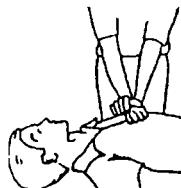
1. Lift up neck
2. Push forehead back
3. Clear out mouth if necessary
4. Observe for breathing

B BREATHING: If the victim is not breathing, begin artificial breathing.



1. Tilt head
 2. Pinch nostrils
 3. Make airtight seal
 4. Provide four quick full breaths
- Check carotid pulse. If pulse is absent, begin artificial circulation. Remember that mouth-to-mouth resuscitation must be commenced as soon as possible.

C CIRCULATION: Depress the sternum 1.2 to 2 inches.



For situations in which there is one rescuer, provide 15 compressions and then 2 quick breaths. The approximate rate of compressions should be 80 per minute.

For situations in which there are two rescuers, provide 5 compressions and then 1 breath. The approximate rate of compressions should be 60 per minute.

Do not interrupt the rhythm of compressions when a second person is giving breaths.

If the victim is responsive, keep warm and quiet, loosen clothing, and place in a reclining position. Call for medical assistance as soon as possible.

Figure 13.3.5 Basic first aid treatment for electric shock.

13-44 Standards and Practices

- Use caution around the high-voltage stages of the equipment. Many power tubes operate at voltages high enough to kill through electrocution. Always break the primary ac circuit of the power supply, and discharge all high-voltage capacitors.
- Minimize exposure to RF radiation. Do not permit personnel to be in the vicinity of open, energized RF generating circuits, RF transmission systems (waveguides, cables, or connectors), or energized antennas. High levels of radiation can result in severe bodily injury, including blindness. Cardiac pacemakers may also be affected.
- Avoid contact with beryllium oxide (BeO) ceramic dust and fumes. BeO ceramic material may be used as a thermal link to carry heat from a tube to the heat sink. Do not perform any operation on any BeO ceramic that might produce dust or fumes, such as grinding, grit blasting, or acid cleaning. Beryllium oxide dust and fumes are highly toxic, and breathing them can result in serious injury or death. BeO ceramics must be disposed of as prescribed by the device manufacturer.
- Avoid contact with hot surfaces within the equipment. The anode portion of many power tubes is air-cooled. The external surface normally operates at a high temperature (up to 250°C). Other portions of the tube also may reach high temperatures, especially the cathode insulator and the cathode/heater surfaces. All hot surfaces may remain hot for an extended time after the tube is shut off. To prevent serious burns, avoid bodily contact with these surfaces during tube operation and for a reasonable cool-down period afterward. Table 13.3.3 lists basic first aid procedures for burns.

13.3.3 OSHA Safety Considerations

The U.S. government has taken a number of steps to help improve safety within the workplace under the auspices of the Occupational Safety and Health Administration (OSHA). The agency helps industries monitor and correct safety practices. OSHA has developed a number of guidelines designed to help prevent accidents. OSHA records show that electrical standards are among the most frequently violated of all safety standards. Table 13.3.4 lists 16 of the most common electrical violations, including exposure of live conductors, improperly labeled equipment, and faulty grounding.

13.3.3a Protective Covers

Exposure of live conductors is a common safety violation. All potentially dangerous electric conductors should be covered with protective panels. The danger is that someone may come into contact with the exposed current-carrying conductors. It is also possible for metallic objects such as ladders, cable, or tools to contact a hazardous voltage, creating a life-threatening condition. Open panels also present a fire hazard.

13.3.3b Identification and Marking

Circuit breakers and switch panels should be properly identified and labeled. Labels on breakers and equipment switches may be many years old and may no longer reflect the equipment actually in use. This is a safety hazard. Casualties or unnecessary damage can be the result of an improp-

Table 13.3.3 Basic First Aid Procedures for Burns (More detailed information can be obtained from any Red Cross office.)

Extensively Burned and Broken Skin	
✓	Cover affected area with a clean sheet or cloth.
✓	Do not break blisters, remove tissue, remove adhered particles of clothing, or apply any salve or ointment.
✓	Treat victim for shock as required.
✓	Arrange for transportation to a hospital as quickly as possible.
✓	If arms or legs are affected, keep them elevated.
✓	If medical help will not be available within an hour and the victim is conscious and not vomiting, prepare a weak solution of salt and soda: 1 level teaspoon of salt and level teaspoon of baking soda to each quart of tepid water. Allow the victim to sip slowly about 4 ounces (half a glass) over a period of 15 minutes. Discontinue fluid intake if vomiting occurs. (Do not offer alcohol.)
Less Severe Burns (First and Second Degree)	
✓	Apply cool (not ice-cold) compresses using the cleanest available cloth article.
✓	Do not break blisters, remove tissue, remove adhered particles of clothing, or apply salve or ointment.
✓	Apply clean, dry dressing if necessary.
✓	Treat victim for shock as required.
✓	Arrange for transportation to a hospital as quickly as possible.
✓	If arms or legs are affected, keep them elevated.

erly labeled circuit panel if no one who understands the system is available in an emergency. If a number of devices are connected to a single disconnect switch or breaker, a diagram should be provided for clarification. Label with brief phrases, and use clear, permanent, and legible markings.

Equipment marking is a closely related area of concern. This is not the same thing as equipment identification. Marking equipment means labeling the equipment breaker panels and ac disconnect switches according to device rating. Breaker boxes should contain a nameplate showing the manufacturer, rating, and other pertinent electrical factors. The intent is to prevent devices from being subjected to excessive loads or voltages.

13.3.3c Grounding

OSHA regulations describe two types of grounding: *system grounding* and *equipment grounding*. System grounding actually connects one of the current-carrying conductors (such as the terminals of a supply transformer) to ground. (See Figure 13.3.6.) Equipment grounding connects all of the noncurrent-carrying metal surfaces together and to ground. From a grounding standpoint, the only difference between a grounded electrical system and an ungrounded electrical system is that the *main bonding jumper* from the service equipment ground to a current-carrying conductor is omitted in the ungrounded system. The system ground performs two tasks:

- It provides the final connection from equipment-grounding conductors to the grounded circuit conductor, thus completing the ground-fault loop.

13-46 Standards and Practices

Table 13.3.4 Sixteen Common OSHA Violations (After [1].)

Fact Sheet	Subject	NEC ¹ Reference
1	Guarding of live parts	110-17
2	Identification	110-22
3	Uses allowed for flexible cord	400-7
4	Prohibited uses of flexible cord	400-8
5	Pull at joints and terminals must be prevented	400-10
6.1	Effective grounding, Part 1	250-51
6.2	Effective grounding, Part 2	250-51
7	Grounding of fixed equipment, general	250-42
8	Grounding of fixed equipment, specific	250-43
9	Grounding of equipment connected by cord and plug	250-45
10	Methods of grounding, cord and plug-connected equipment	250-59
11	AC circuits and systems to be grounded	250-5
12	Location of overcurrent devices	240-24
13	Splices in flexible cords	400-9
14	Electrical connections	110-14
15	Marking equipment	110-21
16	Working clearances about electric equipment	110-16

¹ National Electrical Code

- It solidly ties the electrical system and its enclosures to their surroundings (usually earth, structural steel, and plumbing). This prevents voltages at any source from rising to harmfully high voltage-to-ground levels.

Note that equipment grounding—bonding all electric equipment to ground—is required whether or not the system is grounded. Equipment grounding serves two important tasks:

- It bonds all surfaces together so that there can be no voltage difference among them.
- It provides a ground-fault current path from a fault location back to the electrical source, so that if a fault current develops, it will rise to a level high enough to operate the breaker or fuse.

The National Electrical Code (NEC) is complex and contains numerous requirements concerning electrical safety. The fact sheets listed in Table 13.3.4 are available from OSHA.

13.3.4 Beryllium Oxide Ceramics

Some tubes, both power grid and microwave, contain beryllium oxide (BeO) ceramics, typically at the output waveguide window or around the cathode. Never perform any operations on BeO

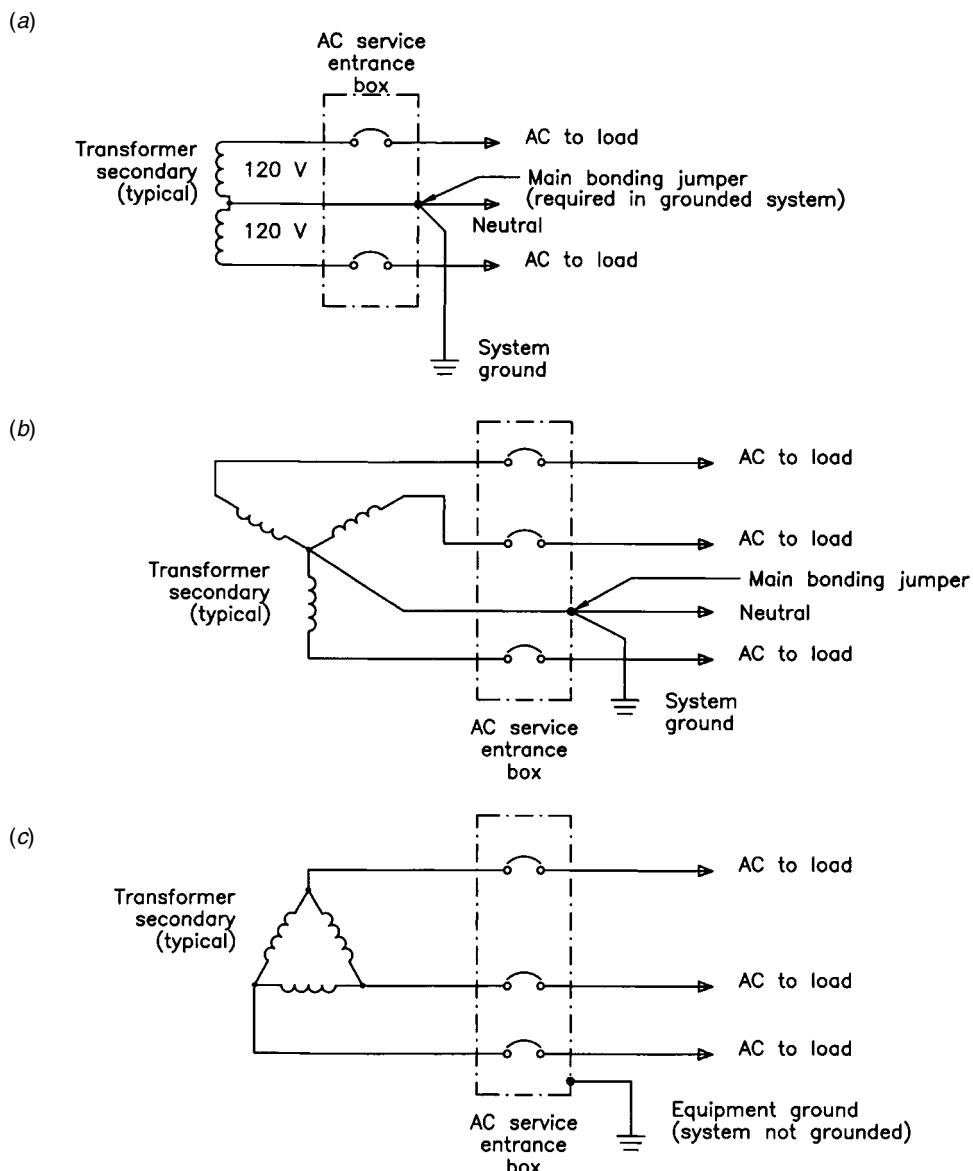


Figure 13.3.6 AC service entrance bonding requirements: (a) 120 V phase-to-neutral (240 V phase-to-phase), (b) 3-phase 208 V wye (120 V phase-to-neutral), (c) 3-phase 240 V (or 480 V) delta. Note that the main bonding jumper is required in only two of the designs.

13-48 Standards and Practices

ceramics that produce dust or fumes, such as grinding, grit blasting, or acid cleaning. Beryllium oxide dust and fumes are highly toxic, and breathing them can result in serious personal injury or death.

If a broken window is suspected on a microwave tube, carefully remove the device from its waveguide, and seal the output flange of the tube with tape. Because BeO warning labels may be obliterated or missing, maintenance personnel should contact the tube manufacturer before performing any work on the device. Some tubes have BeO internal to the vacuum envelope.

Take precautions to protect personnel working in the disposal or salvage of tubes containing BeO. All such personnel should be made aware of the deadly hazards involved and the necessity for great care and attention to safety precautions. Some tube manufacturers will dispose of tubes without charge, provided they are returned to the manufacturer prepaid, with a written request for disposal.

13.3.4a Corrosive and Poisonous Compounds

The external output waveguides and cathode high-voltage bushings of microwave tubes are sometimes operated in systems that use a dielectric gas to impede microwave or high-voltage breakdown. If breakdown does occur, the gas may decompose and combine with impurities, such as air or water vapor, to form highly toxic and corrosive compounds. Examples include Freon gas, which may form lethal *phosgene*, and sulfur hexafluoride (SF_6) gas, which may form highly toxic and corrosive sulfur or fluorine compounds such as *beryllium fluoride*. When breakdown does occur in the presence of these gases, proceed as follows:

- Ventilate the area to outside air
- Avoid breathing any fumes or touching any liquids that develop
- Take precautions appropriate for beryllium compounds and for other highly toxic and corrosive substances
- If a coolant other than pure water is used, follow the precautions supplied by the coolant manufacturer.

FC-75 Toxic Vapor

The decomposition products of FC-75 are highly toxic. Decomposition may occur as a result of any of the following:

- Exposure to temperatures above 200°C
- Exposure to liquid fluorine or alkali metals (lithium, potassium, or sodium)
- Exposure to ionizing radiation

Known thermal decomposition products include *perfluoroisobutylene* (PFIB; $[CF_3]_2 C = CF_2$), which is highly toxic in small concentrations.

If FC-75 has been exposed to temperatures above 200°C through fire, electric heating, or prolonged electric arcs, or has been exposed to alkali metals or strong ionizing radiation, take the following steps:

- Strictly avoid breathing any fumes or vapors.

- Thoroughly ventilate the area.
- Strictly avoid any contact with the FC-75.
- Under such conditions, promptly replace the FC-75 and handle and dispose of the contaminated FC-75 as a toxic waste.

13.3.5 Nonionizing Radiation

Nonionizing radio frequency radiation (RFR) resulting from high-intensity RF fields is a growing concern to engineers who must work around high-power transmission equipment. The principal medical concern regarding nonionizing radiation involves heating of various body tissues, which can have serious effects, particularly if there is no mechanism for heat removal. Recent research has also noted, in some cases, subtle psychological and physiological changes at radiation levels below the threshold for heat-induced biological effects. However, the consensus is that most effects are thermal in nature.

High levels of RFR can affect one or more body systems or organs. Areas identified as potentially sensitive include the ocular (eye) system, reproductive system, and the immune system. Nonionizing radiation also is thought to be responsible for metabolic effects on the central nervous system and cardiac system.

In spite of these studies, many of which are ongoing, there is still no clear evidence in Western literature that exposure to medium-level nonionizing radiation results in detrimental effects. Russian findings, on the other hand, suggest that occupational exposure to RFR at power densities above 1.0 mW/cm^2 does result in symptoms, particularly in the central nervous system.

Clearly, the jury is still out as to the ultimate biological effects of RFR. Until the situation is better defined, however, the assumption must be made that potentially serious effects can result from excessive exposure. Compliance with existing standards should be the minimum goal, to protect members of the public as well as facility employees.

13.3.5a NEPA Mandate

The National Environmental Policy Act of 1969 required the Federal Communications Commission to place controls on nonionizing radiation. The purpose was to prevent possible harm to the public at large and to those who must work near sources of the radiation. Action was delayed because no hard and fast evidence existed that low- and medium-level RF energy is harmful to human life. Also, there was no evidence showing that radio waves from radio and TV stations did not constitute a health hazard.

During the delay, many studies were carried out in an attempt to identify those levels of radiation that might be harmful. From the research, suggested limits were developed by the American National Standards Institute (ANSI) and stated in the document known as ANSI C95.1-1982. The protection criteria outlined in the standard are shown in Figure 13.3.7.

The energy-level criteria were developed by representatives from a number of industries and educational institutions after performing research on the possible effects of nonionizing radiation. The projects focused on absorption of RF energy by the human body, based upon simulated human body models. In preparing the document, ANSI attempted to determine those levels of

13-50 Standards and Practices

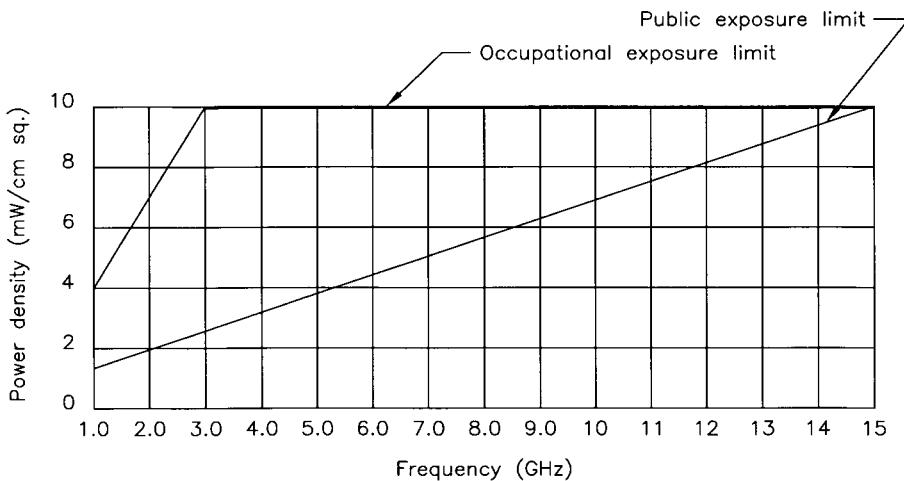


Figure 13.3.7 The power density limits for nonionizing radiation exposure for humans

incident radiation that would cause the body to absorb less than 0.4 W/kg of mass (averaged over the whole body) or peak absorption values of 8 W/kg over any 1 gram of body tissue.

From the data, the researchers found that energy would be absorbed more readily at some frequencies than at others. The absorption rates were found to be functions of the size of a specific individual and the frequency of the signal being evaluated. It was the result of these absorption rates that culminated in the shape of the *safe curve* shown in the figure. ANSI concluded that no harm would come to individuals exposed to radio energy fields, as long as specific values were not exceeded when averaged over a period of 0.1 hour. It was also concluded that higher values for a brief period would not pose difficulties if the levels shown in the standard document were not exceeded when averaged over the 0.1-hour time period.

The FCC adopted ANSI C95.1-1982 as a standard that would ensure adequate protection to the public and to industry personnel who are involved in working around RF equipment and antenna structures.

Revised Guidelines

The ANSI C95.1-1982 standard was intended to be reviewed at 5-year intervals. Accordingly, the 1982 standard was due for reaffirmation or revision in 1987. The process was indeed begun by ANSI, but was handed off to the Institute of Electrical and Electronics Engineers (IEEE) for completion. In 1991, the revised document was completed and submitted to ANSI for acceptance as ANSI/IEEE C95.1-1992.

The IEEE standard incorporated changes from the 1982 ANSI document in four major areas:

- An additional safety factor was provided in certain situations. The most significant change was the introduction of new *uncontrolled* (public) exposure guidelines, generally established

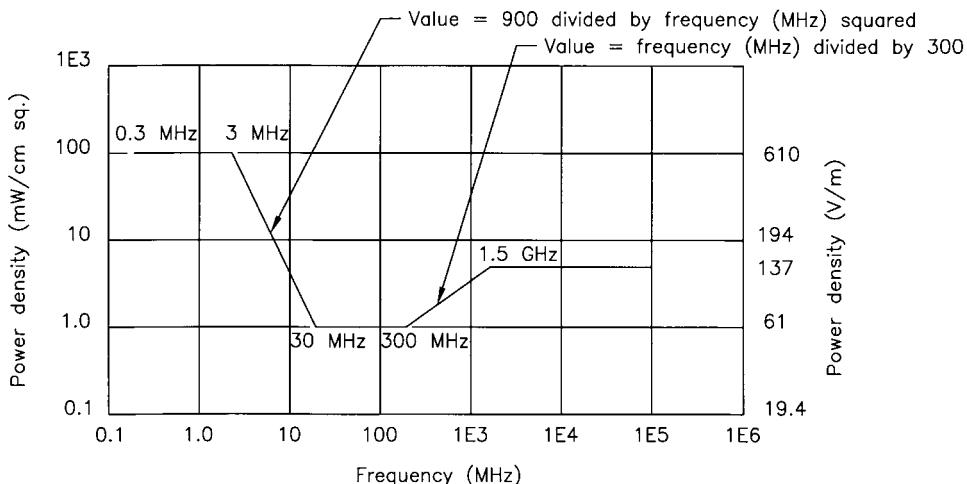


Figure 13.3.8 ANSI/IEEE exposure guidelines for microwave frequencies.

at one-fifth of the *controlled* (occupational) exposure guidelines. Figure 13.3.8 illustrates the concept for the microwave frequency band.

- For the first time, guidelines were included for body currents; examination of the electric and magnetic fields were determined to be insufficient to determine compliance.
- Minor adjustments were made to occupational guidelines, including relaxation of the guidelines at certain frequencies and the introduction of *breakpoints* at new frequencies.
- Measurement procedures were changed in several aspects, most notably with respect to spatial averaging and to minimum separation from reradiating objects and structures at the site.

The revised guidelines are complex and beyond the scope of this handbook. Refer to the ANSI/IEEE document for details.

13.3.5b Multiple-user Sites

At a multiple-user site, the responsibility for assessing the RFR situation—although officially triggered by either a new user or the license renewal of all site tenants—is, in reality, the joint responsibility of all the site tenants. In a multiple-user environment involving various frequencies, and various protection criteria, compliance is indicated when the fraction of the RFR limit within each pertinent frequency band is established and added to the sum of all the other fractional contributions. The sum must not be greater than 1.0. Evaluating the multiple-user environment is not a simple matter, and corrective actions, if indicated, may be quite complex.

13-52 Standards and Practices

Operator Safety Considerations

RF energy must be contained properly by shielding and transmission lines. All input and output RF connections, cables, flanges, and gaskets must be RF leakproof. The following guidelines should be followed at all times:

- Never operate a power tube without a properly matched RF energy absorbing load attached.
- Never look into or expose any part of the body to an antenna or open RF generating tube, circuit, or RF transmission system that is energized.
- Monitor the RF system for radiation leakage at regular intervals and after servicing.

13.3.6 X-Ray Radiation Hazard

The voltages typically used in microwave tubes are capable of producing dangerous X rays. As voltages increase beyond 15 kV, metal-body tubes are capable of producing progressively more dangerous radiation. Adequate X-ray shielding must be provided on all sides of such tubes, particularly at the cathode and collector ends, as well as at the modulator and pulse transformer tanks (as appropriate). High-voltage tubes never should be operated without adequate X-ray shielding in place. The X-ray radiation of the device should be checked at regular intervals and after servicing.

13.3.7 Implosion Hazard

Because of the high internal vacuum in power grid and microwave tubes, the glass or ceramic output window or envelope can shatter inward (implode) if struck with sufficient force or exposed to sufficient mechanical shock. Flying debris could result in bodily injury, including cuts and puncture wounds. If the device is made of beryllium oxide ceramic, implosion may produce highly toxic dust or fumes.

In the event of such an implosion, assume that toxic BeO ceramic is involved unless confirmed otherwise.

13.3.8 Hot Coolant and Surfaces

Extreme heat occurs in the electron collector of a microwave tube and the anode of a power grid tube during operation. Coolant channels used for water or vapor cooling also can reach high temperatures (boiling—100°C—and above), and the coolant is typically under pressure (as high as 100 psi). Some devices are cooled by boiling the coolant to form steam.

Contact with hot portions of the tube or its cooling system can scald or burn. Carefully check that all fittings and connections are secure, and monitor back pressure for changes in cooling system performance. If back pressure is increased above normal operating values, shut the system down and clear the restriction.

For a device whose anode or collector is air-cooled, the external surface normally operates at a temperature of 200 to 300°C. Other parts of the tube also may reach high temperatures, partic-

ularly the cathode insulator and the cathode/heater surfaces. All hot surfaces remain hot for an extended time after the tube is shut off. To prevent serious burns, take care to avoid bodily contact with these surfaces during operation and for a reasonable cool-down period afterward.

13.3.9 Polychlorinated Biphenyls

PCBs belong to a family of organic compounds known as *chlorinated hydrocarbons*. Virtually all PCBs in existence today have been synthetically manufactured. PCBs have a heavy oil-like consistency, high boiling point, a high degree of chemical stability, low flammability, and low electrical conductivity. These characteristics resulted in the widespread use of PCBs in high-voltage capacitors and transformers. Commercial products containing PCBs were widely distributed between 1957 and 1977 under several trade names including:

- Aroclor
- Pyroclor
- Sanotherm
- Pyranol
- Askarel

Askarel is also a generic name used for nonflammable dielectric fluids containing PCBs. Table 13.3.5 lists some common trade names used for Askarel. These trade names typically will be listed on the nameplate of a PCB transformer or capacitor.

PCBs are harmful because once they are released into the environment, they tend not to break apart into other substances. Instead, PCBs persist, taking several decades to slowly decompose. By remaining in the environment, they can be taken up and stored in the fatty tissues of all organisms, from which they are slowly released into the bloodstream. Therefore, because of the storage in fat, the concentration of PCBs in body tissues can increase with time, even though PCB exposure levels may be quite low. This process is called *bioaccumulation*. Furthermore, as PCBs accumulate in the tissues of simple organisms, and as they are consumed by progressively higher organisms, the concentration increases. This process is called *biomagnification*. These two factors are especially significant because PCBs are harmful even at low levels. Specifically, PCBs have been shown to cause chronic (long-term) toxic effects in some species of animals and aquatic life. Well-documented tests on laboratory animals show that various levels of PCBs can cause reproductive effects, gastric disorders, skin lesions, and cancerous tumors.

PCBs may enter the body through the lungs, the gastrointestinal tract, and the skin. After absorption, PCBs are circulated in the blood throughout the body and stored in fatty tissues and a variety of organs, including the liver, kidneys, lungs, adrenal glands, brain, heart, and skin.

The health risk from PCBs lies not only in the PCB itself, but also in the chemicals that develop when PCBs are heated. Laboratory studies have confirmed that PCB by-products, including *polychlorinated dibenzofurans* (PCDFs) and *polychlorinated dibenzo-p-dioxins* (PCDDs), are formed when PCBs or *chlorobenzenes* are heated to temperatures ranging from approximately 900 to 1300°F. Unfortunately, these products are more toxic than PCBs themselves.

13-54 Standards and Practices

Table 13.3.5 Commonly Used Trade Names for PCB Insulating Material

Apirolio	Abestol	Askarel	Aroclor B	Chlorextol	Chlophen
Chlorinol	Clorphon	Diaclor	DK	Dykanol	EEC-18
Elemex	Eucarel	Fenclor	Hyvol	Inclor	Inerteen
Kanechlor	No-Flamol	Phenodlor	Pydraul	Pyralene	Pyranol
Pyroclor	Sal-T-Kuhl	Santothern FR	Santovac	Solvil	Thermin

13.3.9a Governmental Action

The U.S. Congress took action to control PCBs in October 1975 by passing the Toxic Substances Control Act (TSCA). A section of this law specifically directed the EPA to regulate PCBs. Three years later the Environmental Protection Agency (EPA) issued regulations to implement the congressional ban on the manufacture, processing, distribution, and disposal of PCBs. Since that time, several revisions and updates have been issued by the EPA. One of these revisions, issued in 1982, specifically addressed the type of equipment used in industrial plants and transmitting stations. Failure to properly follow the rules regarding the use and disposal of PCBs has resulted in high fines and even jail sentences.

Although PCBs are no longer being produced for electrical products in the United States, there are thousands of PCB transformers and millions of small PCB capacitors still in use or in storage. The threat of widespread contamination from PCB fire-related incidents is one reason behind the EPA's efforts to reduce the number of PCB products in the environment. The users of high-power equipment are affected by the regulations primarily because of the widespread use of PCB transformers and capacitors. These components usually are located in older (pre-1979) systems, so this is the first place to look for them. However, some facilities also maintain their own primary power transformers. Unless these transformers are of recent vintage, it is quite likely that they too contain a PCB dielectric. Table 13.3.6 lists the primary classifications of PCB devices.

13.3.9b PCB Components

The two most common PCB components are transformers and capacitors. A PCB transformer is one containing at least 500 ppm (parts per million) PCBs in the dielectric fluid. An Askarel transformer generally has 600,000 ppm or more. A PCB transformer may be converted to a *PCB-contaminated device* (50 to 500 ppm) or a *non-PCB device* (less than 50 ppm) by having it drained, refilled, and tested. The testing must not take place until the transformer has been in service for a minimum of 90 days. Note that this is *not* something a maintenance technician can do. It is the exclusive domain of specialized remanufacturing companies.

PCB transformers must be inspected quarterly for leaks. If an impervious dike is built around the transformer sufficient to contain all of the liquid material, the inspections can be conducted yearly. Similarly, if the transformer is tested and found to contain less than 60,000 ppm, a yearly inspection is sufficient. Failed PCB transformers cannot be repaired; they must be properly disposed of.

If a leak develops, it must be contained and daily inspections begun. A cleanup must be initiated as soon as possible, but no later than 48 hours after the leak is discovered. Adequate records

Table 13.3.6 Definition of PCB Terms as Identified by the EPA

Term	Definition	Examples
PCB	Any chemical substance that is limited to the biphenyl molecule that has been chlorinated to varying degrees, or any combination of substances that contain such substances.	PCB dielectric fluids, PCB heat-transfer fluids, PCB hydraulic fluids, 2,2',4-trichlorobiphenyl
PCB article	Any manufactured article, other than a PCB container, that contains PCBs and whose surface has been in direct contact with PCBs.	Capacitors, transformers, electric motors, pumps, pipes
PCB container	A device used to contain PCBs or PCB articles, and whose surface has been in direct contact with PCBs.	Packages, cans, bottles, bags, barrels, drums, tanks
PCB article container	A device used to contain PCB articles or equipment, and whose surface has not been in direct contact with PCBs.	Packages, cans, bottles, bags, barrels, drums, tanks
PCB equipment	Any manufactured item, other than a PCB container or PCB article container, that contains a PCB article or other PCB equipment.	Microwave systems, fluorescent light ballasts, electronic equipment
PCB item	Any PCB article, PCB article container, PCB container, or PCB equipment that deliberately or unintentionally contains, or has as a part of it, any PCBs.	
PCB transformer	Any transformer that contains PCBs in concentrations of 500 ppm or greater.	
PCB contaminated	Any electric equipment that contains more than 50, but less than 500 ppm of PCBs. (Oil-filled electric equipment other than circuit breakers, reclosers, and cable whose PCB concentration is unknown must be assumed to be PCB-contaminated electric equipment.)	Transformers, capacitors, contaminated circuit breakers, reclosers, voltage regulators, switches, cable, electromagnets

must be kept of all inspections, leaks, and actions taken for 3 years after disposal of the component. Combustible materials must be kept a minimum of 5 m from a PCB transformer and its enclosure.

As of October 1, 1990, the use of PCB transformers (500 ppm or greater) was prohibited in or near commercial buildings when the secondary voltages are 480 V ac or higher.

The EPA regulations also require that the operator notify others of the possible dangers. All PCB transformers (including PCB transformers in storage for reuse) must be registered with the local fire department. The following information must be supplied:

- The location of the PCB transformer(s).
- Address(es) of the building(s) and, for outdoor PCB transformers, the location.
- Principal constituent of the dielectric fluid in the transformer(s).

13-56 Standards and Practices

- Name and telephone number of the contact person in the event of a fire involving the equipment.

Any PCB transformers used in a commercial building must be registered with the building owner. All building owners within 30 m of such PCB transformers also must be notified. In the event of a fire-related incident involving the release of PCBs, the Coast Guard National Spill Response Center (800-424-8802) must be notified immediately. Appropriate measures also must be taken to contain and control any possible PCB release into water.

Capacitors are divided into two size classes, *large* and *small*. A PCB small capacitor contains less than 1.36 kg (3 lbs) of dielectric fluid. A capacitor having less than 100 in³ also is considered to contain less than 3 lb of dielectric fluid. A PCB large capacitor has a volume of more than 200 in³ and is considered to contain more than 3 lb of dielectric fluid. Any capacitor having a volume between 100 and 200 in³ is considered to contain 3 lb of dielectric, provided the total weight is less than 9 lb. A PCB *large high-voltage capacitor* contains 3 lb or more of dielectric fluid and operates at voltages of 2 kV or greater. A *large low-voltage capacitor* also contains 3 lb or more of dielectric fluid but operates below 2 kV.

The use and servicing of PCB small capacitors is not restricted by the EPA unless there is a leak. In that event, the leak must be repaired or the capacitor disposed of. Disposal may be handled by an approved incineration facility, or the component may be placed in a specified container and buried in an approved chemical waste landfill. Items such as capacitors that are leaking oil greater than 500 ppm PCBs should be taken to an EPA-approved PCB disposal facility.

13.3.9c PCB Liability Management

Properly managing the PCB risk is not particularly difficult; the keys are understanding the regulations and following them carefully. Any program should include the following steps:

- Locate and identify all PCB devices. Check all stored or spare devices.
- Properly label PCB transformers and capacitors according to EPA requirements.
- Perform the required inspections and maintain an accurate log of PCB items, their location, inspection results, and actions taken. These records must be maintained for 3 years after disposal of the PCB component.
- Complete the annual report of PCBs and PCB items by July 1 of each year. This report must be retained for 5 years.
- Arrange for necessary disposal through a company licensed to handle PCBs. If there are any doubts about the company's license, contact the EPA.
- Report the location of all PCB transformers to the local fire department and to the owners of any nearby buildings.

The importance of following the EPA regulations cannot be overstated.

13.3.10 References

1. National Electrical Code, NFPA #70.

13.3.11 Bibliography

- “Current Intelligence Bulletin #45,” National Institute for Occupational Safety and Health, Division of Standards Development and Technology Transfer, February 24, 1986.
- Code of Federal Regulations, 40, Part 761.
- “Electrical Standards Reference Manual,” U.S. Department of Labor, Washington, D.C.
- Hammett, William F.: “Meeting IEEE C95.1-1991 Requirements,” *NAB 1993 Broadcast Engineering Conference Proceedings*, National Association of Broadcasters, Washington, D.C., pp. 471–476, April 1993.
- Hammar, Willie: *Occupational Safety Management and Engineering*, Prentice Hall, New York, N.Y.
- Markley, Donald: “Complying with RF Emission Standards,” *Broadcast Engineering*, Intertec Publishing, Overland Park, Kan., May 1986.
- “Occupational Injuries and Illnesses in the United States by Industry,” OSHA Bulletin 2278, U.S. Department of Labor, Washington, D.C., 1985.
- OSHA, “Electrical Hazard Fact Sheets,” U.S. Department of Labor, Washington, D.C., January 1987.
- OSHA, “Handbook for Small Business,” U.S. Department of Labor, Washington, D.C.
- Pfrimmer, Jack, “Identifying and Managing PCBs in Broadcast Facilities,” *1987 NAB Engineering Conference Proceedings*, National Association of Broadcasters, Washington, D.C., 1987.
- “Safety Precautions,” Publication no. 3386A, Varian Associates, Palo Alto, Calif., March 1985.
- Smith, Milford K., Jr., “RF Radiation Compliance,” *Proceedings of the Broadcast Engineering Conference*, Society of Broadcast Engineers, Indianapolis, IN, 1989.
- “Toxics Information Series,” Office of Toxic Substances, July 1983.
- Whitaker, Jerry C.: *AC Power Systems*, 2nd Ed., CRC Press, Boca Raton, Fla., 1998.
- Whitaker, Jerry C.: G. DeSantis, and C. Paulson: *Interconnecting Electronic Systems*, CRC Press, Boca Raton, Fla., 1993.
- Whitaker, Jerry C.: *Maintaining Electronic Systems*, CRC Press, Boca Raton, Fla. 1991.
- Whitaker, Jerry C.: *Power Vacuum Tubes Handbook*, 2nd Ed., CRC Press, Boca Raton, Fla., 1999.
- Whitaker, Jerry C.: *Radio Frequency Transmission Systems: Design and Operation*, McGraw-Hill, New York, N.Y., 1990.

Chapter
13.4
Audio/Video Standards

Jerry C. Whitaker, Editor-in-Chief

13.4.1 Standards Relating to Digital Video

The following references provide additional information on digital television in general, and the ATSC and DVB standards in particular.

13.4.1a Video

- ISO/IEC IS 13818-1, International Standard (1994), MPEG-2 Systems
- ISO/IEC IS 13818-2, International Standard (1994), MPEG-2 Video
- ITU-R BT.601-4 (1994), Encoding Parameters of Digital Television for Studios
- SMPTE 274M-1995, Standard for Television, 1920×1080 Scanning and Interface
- SMPTE 293M-1996, Standard for Television, 720×483 Active Line at 59.94 Hz Progressive Scan Production, Digital Representation
- SMPTE 294M-1997, Standard for Television, 720×483 Active Line at 59.94 Hz Progressive Scan Production, Bit-Serial Interfaces
- SMPTE 295M-1997, Standard for Television, 1920×1080 50 Hz, Scanning and Interface
- SMPTE 296M-1997, Standard for Television, 1280×720 Scanning, Analog and Digital Representation, and Analog Interface

13.4.1b Audio

- ATSC Standard A/52 (1995), Digital Audio Compression (AC-3)
- AES 3-1992 (ANSI S4.40-1992), AES Recommended Practice for digital audio engineering—Serial transmission format for two-channel linearly represented digital audio data
- ANSI S1.4-1983, Specification for Sound Level Meters
- IEC 651 (1979), Sound Level Meters

13-60 Standards and Practices

IEC 804 (1985), Amendment 1 (1989), Integrating/Averaging Sound Level Meters

13.4.2 ATSC DTV Standard

The following documents form the basis for the ATSC digital television standard.

13.4.2a Service Multiplex and Transport Systems

- ATSC Standard A/52 (1995), Digital Audio Compression (AC-3)
- ISO/IEC IS 13818-1, International Standard (1994), MPEG-2 Systems
- ISO/IEC IS 13818-2, International Standard (1994), MPEG-2 Video
- ISO/IEC CD 13818-4, MPEG Committee Draft (1994), MPEG-2 Compliance

13.4.2b System Information Standard

- ATSC Standard A/52 (1995), Digital Audio Compression (AC-3)
- ATSC Standard A/53 (1995), ATSC Digital Television Standard
- ATSC Standard A/80 (1999), Modulation And Coding Requirements For Digital TV (DTV) Applications Over Satellite
- ISO 639, Code for the Representation of Names of Languages, 1988
- ISO CD 639.2, Code for the Representation of Names of Languages: alpha-3 code, Committee Draft, dated December 1994
- ISO/IEC 10646-1:1993, Information technology—Universal Multiple-Octet Coded Character Set (UCS) — Part 1: Architecture and Basic Multilingual Plane
- ISO/IEC 11172-1, Information Technology—Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s—Part 1: Systems
- ISO/IEC 11172-2, Information Technology—Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s—Part 2: Video
- ISO/IEC 11172-3, Information Technology—Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s—Part 3: Audio
- ISO/IEC 13818-3:1994, Information Technology—Coding of moving pictures and associated audio—Part 3: Audio
- ISO/CD 13522-2:1993, Information Technology—Coded representation of multimedia and hypermedia information objects—Part 1: Base notation
- ISO/IEC 8859, Information Processing—8-bit Single-Octet Coded Character Sets, Parts 1 through 10
- ITU-T Rec. H. 222.0 / ISO/IEC 13818-1:1994, Information Technology—Coding of moving pictures and associated audio—Part 1: Systems

ITU-T Rec. H. 262 / ISO/IEC 13818-2:1994, Information Technology—Coding of moving pictures and associated audio—Part 2: Video

ITU-T Rec. J.83:1995, Digital multi-programme systems for television, sound, and data services for cable distribution

ITU-R Rec. BO.1211:1995, Digital multi-programme emission systems for television, sound, and data services for satellites operating in the 11/12 GHz frequency range

13.4.2c Receiver Systems

47 CFR Part 15, FCC Rules

EIA IS-132, EIA Interim Standard for Channelization of Cable Television

EIA IS-23, EIA Interim Standard for RF Interface Specification for Television Receiving Devices and Cable Television Systems

EIA IS-105, EIA Interim Standard for a Decoder Interface Specification for Television Receiving Devices and Cable Television Decoders

13.4.2d Program Guide

ATSC Standard A/53 (1995), ATSC Digital Television Standard

ANSI/EIA-608-94 (1994), Recommended Practice for Line 21 Data Service

ISO/IEC IS 13818-1, International Standard (1994), MPEG-2 Systems

13.4.2e Program/Episode/Version Identification

ATSC Standard A/53 (1995), Digital Television Standard

ATSC Standard A/65 (1998), Program and System Information Protocol for Terrestrial Broadcast and Cable

ATSC Standard A/70 (1999), Conditional Access System for Terrestrial Broadcast

ISO/IEC IS 13818-1, International Standard (1994), MPEG-2 systems

13.4.3 DVB

The following documents form the basis of the DVB digital television standard. For additional information contact DVB at www.dvb.org.

13.4.3a General

Digital Satellite Transmission Systems, ETS 300 421

Digital Cable Delivery Systems, ETS 300 429

13-62 Standards and Practices

- Digital Terrestrial Broadcasting Systems, ETS 300 744
- Digital Satellite Master Antenna Television (SMATV) Distribution Systems, ETS 300 473
- Specification for the Transmission of Data in DVB Bitstreams, TS/EN 301 192
- Digital Broadcasting Systems for Television, Sound and Data Services; Subtitling Systems, ETS 300 743
- Digital Broadcasting Systems for Television, Sound and Data Services; Allocation of Service Information (SI) Codes for Digital Video Broadcasting (DVB) Systems, ETR 162

13.4.3b Multipoint Distribution Systems

- Digital Multipoint Distribution Systems at and Above 10 GHz, ETS 300 748
- Digital Multipoint Distribution Systems at or Below 10 GHz, ETS 300 749

13.4.3c Interactive Television

- Return Channels in CATV Systems (DVB-RCC), ETS 300 800
- Network-Independent Interactive Protocols (DVB-NIP), ETS 300 801
- Interaction Channel for Satellite Master Antenna TV (SMATV), ETS 300 803
- Return Channels in PSTN/ISDN Systems (DVB-RCT), ETS 300 802
- Interfacing to PDH Networks, ETS 300 813
- Interfacing to SDH Networks, ETS 300 814

13.4.3d Conditional Access

- Common Interface for Conditional Access and Other Applications, EN50221
- Technical Specification of SimulCrypt in DVB Systems, TS101 197

13.4.3e Interfaces

- DVB Interfaces to PDH Networks, prETS 300 813
- DVB Interfaces to SDH Networks, prETS 300 814

13.4.4 SMPTE Documents Relating to Digital Video and Audio

The following documents relating to digital television have been approved (or are pending at this writing) by the Society of Motion Picture and Television Engineers. For additional information, contact the SMPTE at www.smpte.org.

13.4.4a General Topics

- AES/EBU Emphasis and Preferred Sampling Rate, EG 32
- Alignment Color Bar Signal, EG 1
- Audio: Linear PCM in MPEG-2 Transport Stream, SMPTE 302M
- Camera Color Reference Signals, Derivation of, RP 176
- Color, Equations, Derivation of, RP 177
- Color, Reference Pattern, SMPTE 303M
- Wide-Screen Scanning Structure, SMPTE RP 199

13.4.4b Ancillary

- AES/EBU Audio and Auxiliary Data, SMPTE 272M
- Camera Positioning by Data Packets, SMPTE 315M
- Data Packet and Space Formatting, SMPTE 291M
- DTV Closed-Caption Server to Encoder Interface, SMPTE 333M
- Error Detection and Status Flags, RP 165
- Format for Non-PCM Audio and Data in an AES3 Serial Digital Audio Interface, SMPTE 337M
- Format for Non-PCM Audio and Data in an AES3 Serial Digital Audio Interface—ATSC A/52 (AC-3) Data Type, SMPTE 340M
- Format for Non-PCM Audio and Data in an AES3 Serial Digital Audio Interface—Captioning Data Type, SMPTE 341M
- Format for Non-PCM Audio and Data in an AES3 Serial Digital Audio Interface—Data Types, SMPTE 338M
- Format for Non-PCM Audio and Data in an AES3 Serial Digital Audio Interface—Generic Data Types, SMPTE 339M
- HDTV 24-bit Digital Audio, SMPTE 299M
- LTC and VITC Data as HANC Packets, RP 196
- Time and Control Code, RP 188
- Transmission of Signals Over Coaxial Cable, SMPTE 276M

13.4.4c Digital Control Interfaces

- Common Messages, RP 172
- Control Message Architecture, RP 138
- Electrical and Mechanical Characteristics, SMPTE 207M
- ESlan Implementation Standards, EG 30

13-64 Standards and Practices

ESlan Remote Control System, SMPTE 275M
 ESlan Virtual Machine Numbers, RP 182
 Glossary, Electronic Production, EG 28
 Remote Control of TV Equipment, EG 29
 Status Monitoring and Diagnostics, Fault Reporting, SMPTE 269M
 Status Monitoring and Diagnostics, Processors, RP 183
 Status Monitoring and Diagnostics, Protocol, SMPTE 273M
 Supervisory Protocol, RP 113
 System Service Messages, RP 163
 Tributary Interconnection, RP 139
 Type-Specific Messages, ATR, RP 171
 Type-Specific Messages, Routing Switcher, RP 191
 Type-Specific Messages, VTR, RP 170
 Universal Labels for Unique ID of Digital Data, SMPTE 298M
 Video Images: Center, Aspect Ratio and Blanking, RP 187
 Video Index: Information Coding, 525- and 625-Line, RP 186

13.4.4d Edit Decision Lists

Device Control Elements, SMPTE 258M
 Storage, 3-1/2-in Disk, RP 162
 Storage, 8-in Diskette, RP 132
 Transfer, Film to Video, RP 197

13.4.4e Image Areas

8 mm Release Prints, TV Safe Areas, RP 56
 16 mm and 35 mm Film and 2 × 2 slides, SMPTE 96
 Review Rooms, SMPTE 148
 Safe Areas, RP 27.3

13.4.4f Interfaces and Signals

12-Channel for Digital Audio and Auxiliary Data, SMPTE 324M
 Checkfield, RP 178
 Development of NTSC, EG 27

Key Signals, RP 157

NTSC Analog Component 4:2:2, SMPTE 253M

NTSC Analog Composite for Studios, SMPTE 170M

Pathological Conditions, EG 34

Bit-Parallel Interfaces

1125/60 Analog Component, RP 160

1125/60 Analog Composite, SMPTE 240M

1125/60 High-Definition Digital Component, SMPTE 260M

NTSC Digital Component, SMPTE 125M

NTSC Digital Component, 16 × 9 Aspect Ratio, SMPTE 267M

NTSC Digital Component 4:4:4:4 Dual Link, RP 175

NTSC Digital Component 4:4:4:4 Single Link, RP 174

NTSC Digital Composite, SMPTE 244M

Bit-Serial Interfaces

4:2:2p and 4:2:0p Bit Serial, SMPTE 294M

540 Mbits/s Serial Digital Interface, SMPTE 344M

Digital Component 4:2:2 AMI, SMPTE 261M

Digital Component S-NRZ, SMPTE 259M

Digital Composite AMI, SMPTE 261M

Digital Composite, Error Detection Checkwords/Status Flag, RP 165

Digital Composite, Fiber Transmission System, SMPTE 297M

Digital Composite, S-NRZ, SMPTE 259M

Element and Metadata Definitions for the SDTI-CP, SMPTE 331M

Encapsulation of Data Packet Streams over SDTI (SDTI-PF), SMPTE 332M

HDTV, SMPTE 292M

High Data-Rate Serial Data Transport Interface (HD-SDTI), SMPTE 348M

HDTV, Checkfield, RP 198

Jitter in Bit Serial Systems, RP 184

Jitter Specification, Characteristics and Measurements, EG 33

Jitter Specification, Measurement, RP 192

SDTI Content Package Format (SDTI-CP), SMPTE 326M

Serial Data Transport Interface, SMPTE 305.2M

13-66 Standards and Practices

Time Division Multiplexing Video Signals and Generic Data over High-Definition Interfaces,
SMPTE 346M

Vertical Ancillary Data Mapping for Bit-Serial Interface, SMPTE 334M

Scanning Formats

1280×720 Scanning, SMPTE 296M

1920×1080 Scanning, 60 Hz, SMPTE 274M

1920×1080 Scanning, 50 Hz, SMPTE 295M

720×483 Digital Representation, SMPTE 293M

13.4.4g Monitors

Alignment, RP 167

Colorimetry, RP 145

Critical Viewing Conditions, RP 166

Receiver Monitor Setup Tapes, RP 96

13.4.4h MPEG-2

4:2:2 Profile at High Level, SMPTE 308M

4:2:2 P@HL Synchronous Serial Interface, SMPTE 310M

Alignment for Coding, RP 202

MPEG-2 Video Elementary Stream Editing Information, SMPTE 328M

MPEG-2 Video Recoding Data Set, SMPTE 327M

MPEG-2 Video Recoding Data Set—Compressed Stream Format, SMPTE 329M

Opportunistic Data Broadcast Flow Control, SMPTE 325M

Splice Points for the Transport Stream, SMPTE 312M

Transport of MPEG-2 Recoding Information as Ancillary Data Packets, SMPTE 353M

Transporting MPEG-2 Recoding Information Through 4:2:2 Component Digital Interfaces,
SMPTE 319M

Transporting MPEG-2 Recoding Information Through High-Definition Digital Interfaces,
SMPTE 351M

Unique Material Identifier (UMID), SMPTE 330M

13.4.4i Test Patterns

Alignment Color Bars, EG 1

Camera Registration, RP 27.2
Deflection Linearity, RP 38.1
Mid-Frequency Response, RP 27.5
Operational Alignment, RP 27.1
Safe Areas, RP 27.3
Telecine Jitter, Weave, Ghost, RP 27.4

13.4.4j Video Recording and Reproduction

Audio Monitor System Response, SMPTE 222M
Channel Assignments, AES/EBU Inputs, EG 26
Channel Assignments and Magnetic Masters to Stereo Video, RP 150
Cassette Bar Code Readers, EG 31-1995
Data Structure for DV-Based Audio, Data, and Compressed Video, SMPTE 314M
Loudspeaker Placement, HDEP, RP 173
Relative Polarity of Stereo Audio Signals, RP 148
Tape Care, Handling, Storage, RP 103
Telecine Scanning Capabilities, EG 25

Time and Control Code

Binary Groups, Date and Time Zone Transmissions, SMPTE 309M
Binary Groups, Storage and Transmission, SMPTE 262M
Directory Index, Auxiliary Time Address Data, RP 169
Directory Index, Dialect Specification of Page-Line, RP 179
Specifications, TV, Audio, Film, SMPTE 12M
Time Address Clock Precision, EG 35
Vertical Interval, 4:2:2 Digital Component, SMPTE 266M
Vertical Interval, Encoding Film Transfer Information, 4:2:2 Digital, RP 201
Vertical Interval, Location, RP 164
Vertical Interval, Longitudinal Relationship, RP 159
Vertical Interval, Switching Point, RP 168

13-68 Standards and Practices

Tape Recording Formats

SMPTE Documents Relating to Tape Recording Formats (Courtesy of SMPTE.)

	B	C	D-1	D-2	D-3	D-5	D-6	D-7 (1)	D-9 (2)	E (3)	G (4)	H (5)	L (6)	M-2
Basic system parameters														
525/60	15M	18M	EG10	EG20	264M	279M	277M	306M	316M	21M			RP144	RP158
625/50					265M	279M	277M	306M	316M	21M			229M	249M
Record dimensions	16M	19M	224M	245M	264/5M	279M	277M				32M			
Characteristics											32M			
Video signals	RP84	RP86									32M			
Audio and control signals	17M	20M	RP155	RP155	264/5M	279M	278M				32M			
Data and control record			227M	247M	264/5M	279M	278M				32M			
Tracking control record						279M	277M				32M			
Pulse code modulation audio														
Time and control recording	RP93			228M	248M	264/5M	279M	278M					230M	251M
Audio sector time code, equipment type information				RP181										
Nomenclature		18M	EG21		EG21					21M		32M		
Index of documents					EG22						RP142	RP142	RP142	
Stereo channels	RP142	RP142		RP148	RP148						RP148	RP148	RP148	
Relative polarity	RP148	RP148		RP148	RP148						35M	32M	238M	250M
Tape	25M	25M		225M	246M	264/5M	279M	277M						
Reels	24M	24M												
Cassettes			226M	226M	263M	263M	226M	307M	317M	22M	35M	32M	238M	250M
Small										31M				
Bar code labeling														
Dropout specifications	RP121	RP121		RP156	RP156									
Reference tape and recorder														
System parameters	29M													
Tape	26M	26M												
High-definition														
Encoding process/data format														
Transmission over 360 Mb/s														
Notes:														
1DVCPRO, 2 Digital S, 3 U-matic, 4 Beta, 5 VHS, 6 Betacam														

13.4.4k SMPTE Documents by Number

Documents without year designation were at the proposal stage at printing.

SMPTE Standards, Recommended Practices, and Engineering Guidelines by Number

SMPTE Standards	
SMPTE 1-1996	Video Recording—2-in Magnetic Recording Tape
SMPTE 3-1998	Television Analog Recording—Frequency Response and Operating Level of Recorders and Reproducers—Audio 1 Record on 2-in Tape Operating at 15 and 7.5 in/s
SMPTE 4-1995	Television Analog Recording—2-in Magnetic Tape for Quadruplex Recording—Speed
SMPTE 5-1995	Television Analog Recording—2-in Reels
SMPTE 6-1998	Video Recording—2-in Quadruplex Tape—Video, Audio and Tracking-Control Records
SMPTE 7-1999	Motion-Picture Film (16-mm)—Camera Aperture Image and Usage
SMPTE 8-1995	Video Recording—Quadruplex Recorders Operating at 15 in/s—Audio Level and Multifrequency Test Tape

SMPTE Standards, Recommended Practices, and Engineering Guidelines by Number

SMPTE 11-1995	Video Recording—Quadruplex Recorders Operating at 7.5 in/s—Audio Level and Multifrequency Test Tape
SMPTE 12M-1999	Television, Audio, and Film—Time and Control Code
SMPTE 15M-1998	Television Analog Recording—1-in Type B Helical Scan—Basic System Parameters
SMPTE 16M-1998	Television Analog Recording—1-in Type B Helical Scan—Records
SMPTE 17M-1998	Television Analog Recording—1-in Type B Helical Scan—Frequency Response and Operating Level
SMPTE 18M-1996	Television Analog Recording—1-in Type C- Basic System and Transport Geometry Parameters
SMPTE 19M-1996	Television Analog Recording—1-in Type C--Records
SMPTE 20M-1996	Television Analog Recording—1-in Type C Recorders and Reproducers—Longitudinal Audio Characteristics
SMPTE 21M-1997	Video Recording—3/4-in Type E Helical Scan—Records
SMPTE 22M-1997	Video Recording—3/4-in Type E Helical Scan—Cassette
SMPTE 24M-1996	Television Recording—1-in Reels
SMPTE 25M-1995	Video Recording—1-in Magnetic Recording Tape
SMPTE 26M-1995	Video Recording—1-in Helical-Scan Recorders—Raw Stock for Reference Tapes
SMPTE 29M-1995	Television Analog Recording—1-in Type B Reference Recorders—Basic System and Transport Geometry
SMPTE 30M-1995	Television Analog Recording—1-in Type B Reference Recorders—Records on Reference Tapes
SMPTE 31M-1995	Television Analog Recording—3/4-in Type E—Small Video Cassette
SMPTE 32M-1998	Video Recording—1/2-in Type H—Cassette, Tape, and Records
SMPTE 35M-1997	Television Analog Recording—1/2-in Type G—Cassette and Tape
SMPTE 37M-1994	Motion-Picture Equipment—Raw Stock Cores
SMPTE 40-1997	Motion-Picture Film (35-mm)—Release Prints—Photographic Audio Records
SMPTE 41-1999	Motion-Picture Film (16-mm)—Prints—Photographic Audio Records
SMPTE 48-1995	Motion-Picture Film (16-mm)—Picture and Sound Contact Printing—Printed Areas
SMPTE 55-2000	Motion-Picture Film—35- and 16-mm Audio Release Prints—Leaders and Cue Marks
SMPTE 56-1996	Motion-Picture Film—Nomenclature for Studios and Processing Laboratories
SMPTE 59-1998	Motion-Picture Film (35-mm)—Camera Aperture Images and Usage
SMPTE 73-1998	Motion-Picture Film (32-mm)—35-mm Film Perforated 32-mm, 2R
SMPTE 74-1993	Motion-Picture Cameras (16- and 8-mm)—Zero Point for Focusing Scales (R1998)
SMPTE 75M-1994	Motion-Picture Film—Raw Stock—Designation of A and B Windings
SMPTE 76-1996	Motion-Picture Cameras—16-mm and 8-mm Threaded Lens Mounts
SMPTE 83-1996	Motion-Picture Film (16-mm)—Edge Numbers—Location and Spacing
SMPTE 86-1996	Motion-Picture Film—Magnetic Audio Records—Two, Three, Four, and Six Records on 35-mm and One Record on 17.5-mm Magnetic Film
SMPTE 87M-1996	Motion-Picture Film (16-mm)—100-Mil Magnetic Striping
SMPTE 93-1998	Motion-Picture Film (35-mm)—Perforated BH

13-70 Standards and Practices

SMPTE Standards, Recommended Practices, and Engineering Guidelines by Number

SMPTE 96-1999	Television—35- and 16-mm Motion-Picture Film and 2 × 2-in Slides—Scanned Area and Photographic Image Area for 4:3 Aspect Ratio
SMPTE 97-1999	Motion-Picture Film (16-mm)—200-Mil Position-Edge Magnetic Audio Record
SMPTE 101-1998	Motion-Picture Film (16-mm)—Perforated 2R-3000—Magnetic Striping
SMPTE 102-1997	Motion-Picture Film (35-mm)—Perforated CS-1870
SMPTE 109-1998	Motion-Picture Film (16-mm)—Perforated 1R and 2R
SMPTE 111-1996	Motion-Picture Film (35-mm)—Prints Made on Continuous Contact Printers—Exposed Areas for Picture and Audio
SMPTE 112-1999	Motion-Picture Film (16-mm)—100-Mil Magnetic Audio Record
SMPTE 117M-1996	Motion-Picture Film—Photographic Audio Record—Spectral Diffuse Density
SMPTE 119-1999	Motion-Picture Film (70-mm)—Perforated 65-mm, KS-1870
SMPTE 125M-1995	Television—Component Video Signal 4:2:2—Bit-Parallel Digital Interface
SMPTE 127-1994	Motion-Picture Film (16-mm)—Magnetic Photographic Audio Records—Magnetic Striping of Prints
SMPTE 137-2000	Motion-Picture Film (35-mm)—Release Prints—Four Magnetic Audio Records
SMPTE 139-1996	Motion-Picture Film (35-mm)—Perforated KS
SMPTE 143-1994	Motion-Picture Film (8-mm Type R)—Length of Film on Camera Spools—25-Ft Capacity
SMPTE 145-1999	Motion-Picture Film (65-mm)—Perforated KS
SMPTE 146M-1996	Motion-Picture Film—16- and 8-mm Reversal Color Camera Films—Determination of Speed
SMPTE 148-1991	Motion-Picture Film—35- and 16-mm Prints for Television Transmission—Film Image Area for Review Room Viewing
SMPTE 149-1999	Motion-Picture Film (8-mm Type S)—Perforated 1R
SMPTE 151-1998	Motion-Picture Film (8-mm Type S)—16-mm Film Perforated 8-mm Type S, (1-3)
SMPTE 152-1994	Motion-Picture Film (70-mm)—Projectable Image Area
SMPTE 153-1996	Motion-Picture Film (8-mm Type S)—16-mm Film Perforated 8-mm Type S (1-4)—Printed Areas
SMPTE 154-1998	Motion-Picture Film (8-mm Type S)—Projectable Image Area and Projector Usage
SMPTE 157-1999	Motion-Picture Film (8-mm Type S)—Camera Aperture Image and Usage
SMPTE 159.1-1996	Motion-Picture Film (8-mm Type S)—Model 1 Camera Cartridge, Cartridge-Camera Interface, and Take-Up Core Drive
SMPTE 159.2-1996	Motion-Picture Film (8-mm Type S)—Model 1 Camera Cartridge Aperture, Camera Aperture Profile, Film Position, Pressure Pad, and Flatness
SMPTE 160M-1995	Motion-Picture Equipment (8-mm Type S)—Projection Reels—100- to 312-mm Diameter
SMPTE 161-1998	Motion-Picture Film (8-mm Type S)—Magnetic Striping
SMPTE 162-1998	Motion-Picture Film (8-mm Type S)—16-mm Film Perforated 8-mm Type S, (1-4)—Magnetic Striping
SMPTE 163-1998	Motion-Picture Film (8-mm Type S)—35-mm Film Perforated 8-mm Type S, 5R—Magnetic Striping
SMPTE 164-1993	Motion-Picture Film (8-mm Type S)—Magnetic Audio Record—Position, Dimensions, and Reproducing Speed
SMPTE 165-1999	Motion-Picture Film (35-mm)—Perforated 8-mm Type S, 5R (1-3-5-7-0)

SMPTE Standards, Recommended Practices, and Engineering Guidelines by Number

SMPTE 166-1999	Motion-Picture Film (8-mm Type S)—Exposure Control and Stock Identification—Sound and Silent Camera Cartridge Notches
SMPTE 168-1996	Motion-Picture Film (16-mm)—Perforated 8-mm Type S, (1-4)
SMPTE 169-1997	Motion-Picture Film (35-mm)—Perforated 8-mm Type S, 2R-1664 (1-0)
SMPTE 170M-1999	Television—Composite Analog Video Signal—NTSC for Studio Applications
SMPTE 171-1996	Motion-Picture Film (35-mm)—Perforated 16-mm, 3R (1-3-0)
SMPTE 173-1999	Motion-Picture Equipment (8-mm Type R)—Double 8-mm Camera Spools—100-Ft Capacity
SMPTE 174-1994	Motion-Picture Equipment (16-mm)—Camera Spools—50- to 400-ft Capacity
SMPTE 176-1999	Motion-Picture Film (8-mm Type S)—16-mm Film Perforated 2R-1667 (1-3)—Magnetic Striping
SMPTE 177-1995	Motion-Picture Film (35-mm)—Four-Track Magnetic Audio Release Prints—Magnetic Striping
SMPTE 179-1996	Motion-Picture Film (8-mm Type S)—35-mm Film Perforated 2R (1-0) and 5R (1-3-5-7-0)—Printed Areas
SMPTE 181-1996	Motion-Picture Film (8-mm Type S)—16-mm Film Perforated 8-mm Type S (1-3)—Printed Areas
SMPTE 183M-1996	Motion-Picture Film—Photographic Audio Level Test Films—Measurement of Photoelectric Output Factor
SMPTE 184M-1998	Motion-Picture Film—Raw Stock Identification and Labeling
SMPTE 185-1993	Motion-Picture Film (70-mm)—Six Magnetic Records on Release Prints—Position, Dimensions, Reproducing Speed, and Identity
SMPTE 188M-1994	Motion-Picture Equipment (8-mm Type S)—Model II Camera Cartridges (15-m Capacity)—Camera-Run Film Length
SMPTE 189M-1994	Motion-Picture Equipment (8-mm Type S)—Model II Camera Cartridges—Loaded Film Location
SMPTE 190M-1994	Motion-Picture Equipment (8-mm Type S)—Model II Camera Cartridges—Cartridge-Camera Fit and Core
SMPTE 191M-1994	Motion-Picture Equipment (8-mm Type S)—Model II Camera Cartridges—Slots, Projections, and Cartridge Hole
SMPTE 192-1997	Motion-Picture Equipment (35-mm)—Shipping Reels for Prints
SMPTE 194-1997	Motion-Picture Film (35-mm)—Projector Usage—Release Prints Having Four Perforations per Frame
SMPTE 195-2000	Motion-Picture Film (35-mm)—Motion-Picture Prints—Projectable Image Area
SMPTE 196M-1995	Motion -Picture Film—Indoor Theater and Review Room Projection—Screen Luminance and Viewing Conditions
SMPTE 197-1998	Motion-Picture Film (8-mm Type S)—50-Ft Model 1 Sound Camera Cartridge—Cartridge, Cartridge-Camera Interface, and Take-Up Core
SMPTE 198-1998	Motion-Picture Film (8-mm Type S)—50-Ft Model 1 Sound Camera Cartridge—Aperture, Pressure Pad, and Film Position
SMPTE 199-1998	Motion-Picture Film (8-mm Type S)—50-Ft Model 1 Sound Camera Cartridge—Pressure Pad Flatness and Camera Aperture Profile
SMPTE 200M-1998	Motion-Picture Equipment (8-mm Type S)—Model 1 Camera Cartridge—Camera Run Length, Perforation Cutout, and End-of-Run Notch
SMPTE 201M-1996	Motion-Picture Film (16-mm)—Type W Camera Aperture Image

13-72 Standards and Practices

SMPTE Standards, Recommended Practices, and Engineering Guidelines by Number

SMPTE 202M-1998	Motion-Pictures—B-Chain Electroacoustic Response—Dubbing Theaters, Review Rooms, and Indoor Theaters
SMPTE 203-1998	Motion-Picture Film (35-mm)—Prints—Two-Track Photographic Audio Records
SMPTE 205-1993	Motion-Picture Equipment (8-mm Type S)—Model 1 Camera Cartridge—Interface and Take-Up Core Drive (200-ft Capacity) (R1998)
SMPTE 206-1998	Motion-Picture Equipment (8-mm Type S)—Model 1 Sound Camera Cartridge—Aperture, Profile, Film Position, Pressure Pad, and Flatness (200-ft Capacity)
SMPTE 207M-1997	Television—Digital Control Interface—Electrical and Mechanical Characteristics
SMPTE 208M-1992	Motion-Picture Film—35- and 16-mm Magnetic Audio Records—Recorded Characteristics (R1998)
SMPTE 209M-1996	Motion-Picture Film (8-mm Type S)—Magnetic Audio Records—Recorded Characteristic
SMPTE 210M-1999	Motion-Picture Film (16-mm)—Magnetic Audio Records—Two Records on 16-mm Magnetic Film
SMPTE 211M-1996	Motion-Picture Film—16- and 35-mm Variable-Area Photographic Audio Records—Signal-to-Noise Ratio
SMPTE 212M-1995	Motion-Picture Equipment (8-mm Type S)—Projection Reels—75-mm Diameter
SMPTE 214M-1999	Motion-Picture Film (35-mm)—Photographic Audio Reproduction Characteristic
SMPTE 215-1995	Motion-Picture Film (65-mm)—Camera Aperture Image
SMPTE 216-1998	Motion-Picture Film (35-mm)—Recorded Characteristic of Magnetic Audio Records—Four Track Striped Release Prints
SMPTE 217-1998	Motion-Picture Film (70-mm)—Recorded Characteristic of Magnetic Audio Records—Striped Release Prints
SMPTE 218M-1996	Motion-Picture Film (16-mm)—200-mil Center-Position Magnetic Audio Record
SMPTE 220-1996	Motion-Picture and Television Equipment—Camera Mounting Connections—1/4-inch-20 Thread and 3/8-inch-16 Thread Tripod Screws
SMPTE 221-1998	Motion-Picture Film (70-mm)—Six-Track Audio Release Prints—Magnetic Striping
SMPTE 222M-1994	Television—Control and Review Rooms—Monitor System Electroacoustic Response
SMPTE 223M-1996	Motion-Picture Film—Safety Film
SMPTE 224M-1996	Television Digital Component Recording—19-mm Type D-1—Tape Record
SMPTE 225M-1996	Television Digital Component Recording—19-mm Type D-1—Magnetic Tape
SMPTE 226M-1996	Television Digital Recording—19-mm Tape Cassettes
SMPTE 227M-1996	Television Digital Component Recording—19-mm Type D-1—Helical Data and Control Records
SMPTE 228M-1996	Television Digital Component Recording—19-mm Type D-1—Time and Control Code and Cue Records
SMPTE 229M-1996	Television Analog Recording—1/2-in Type L—Records
SMPTE 230M-1996	Television Analog Recording—1/2-in Type L—Electrical Parameters, Control Code, and Tracking Control
SMPTE 231-1999	Motion-Picture Film (8-mm Type R)—Camera Aperture Image and Usage
SMPTE 233-1998	Motion-Picture Film (16-mm)—Projectable Image Area and Projector Usage
SMPTE 234-1998	Motion-Picture Film (8-mm Type R)—Projectable Image Area and Projector Usage
SMPTE 235-1998	Motion-Picture Equipment (16-mm)—Projection Reels—200- to 2300-ft Capacity

SMPTE Standards, Recommended Practices, and Engineering Guidelines by Number

SMPTE 236-1998	Motion-Picture Equipment (8-mm Type R)—Projection Reels
SMPTE 237-1998	Motion-Picture Film (35-mm)—Perforated DH-1870
SMPTE 238M-1998	Television Analog Recording—1/2-in Type L—Tapes and Cassettes
SMPTE 239-1999	Motion-Picture Film (16-mm)—Perforated 8-mm Type R, 2R
SMPTE 240M-1999	Television—Signal Parameters—1125-Line High-Definition Production Systems
SMPTE 241-1995	Motion-Picture Equipment—35- and 70-mm Projection Reels
SMPTE 242-1993	Motion-Picture Equipment (35-mm)—Universal Intermittent Sprockets
SMPTE 243M-1993	Motion-Picture Equipment—35- and 70-mm Projection Lenses and Mounts (R1998)
SMPTE 244M-1995	Television—System M/NTSC Composite Video Signals—Bit-Parallel Digital Interface
SMPTE 245M-1993	Television Digital Recording—19-mm Type D-2 Composite Format—Tape Record
SMPTE 246M-1993	Television Digital Recording—19-mm Type D-2 Composite Format—Magnetic Tape
SMPTE 247M-1993	Television Digital Recording—19-mm Type D-2 Composite Format—Helical Data and Control Records
SMPTE 248M-1993	Television Digital Recording—19-mm Type D-2 Composite Format—Cue Record, and Time and Control Code Record
SMPTE 249M-1996	Television Analog Recording—1/2-in Type M-2—Records
SMPTE 250M-1996	Television Analog Recording—1/2-in Type M-2—Tapes and Cassettes
SMPTE 251M-1996	Television Analog Recording—1/2-in Type M-2—Electrical Parameters of Video, Audio, Time and Control Code, and Tracking Control
SMPTE 252M-1996	Television Analog Recording—1/2-in Type M-2—Pulse Code Modulation Audio
SMPTE 253M-1998	Television—Three-Channel RGB Analog Video Interface
SMPTE 254-1998	Motion-Picture Film (35-mm)—Manufacturer-Printed, Latent Image Identification Information
SMPTE 256M-1996	Television—Specifications for Video Tape Leader
SMPTE 257-1998	Motion-Picture Film (35-mm)—Stereoscopic Prints with Vertically Positioned Subframes—Projectable Image Areas
SMPTE 258M-1993	Television—Transfer of Edit Decision Lists
SMPTE 259M-1997	Television—10-Bit 4:2:2 Component and $4f_{sc}$ Composite Digital Signals—Serial Digital Interface
SMPTE 260M-1999	Television—1125/60 High-Definition Production System—Digital Representation and Bit-Parallel Interface
SMPTE 261M-1993	Television—10-Bit Serial Digital Television Signals: 4:2:2 Component and $4f_{sc}$ NTSC Composite—AMI Transmission Interface
SMPTE 262M-1995	Television, Audio, and Film—Binary Groups of Time and Control Codes—Storage and Transmission of Data
SMPTE 263M-1996	Television Digital Recording—1/2-in Type D-3 Composite and 1/2-in Type D-5 Component Formats—Tape Cassette
SMPTE 264M-1998	Television Digital Recording—1/2-in Type D-3 Composite Format—525/60
SMPTE 265M-1993	Television Digital Recording—1/2-in Type D-3 Composite Format—625/50
SMPTE 266M-1994	Television—4:2:2 Digital Component Systems—Digital Vertical Interval Time Code
SMPTE 267M-1995	Television—Bit-Parallel Digital Interface—Component Video Signal 4:2:2 16 × 9 Aspect Ratio

13-74 Standards and Practices

SMPTE Standards, Recommended Practices, and Engineering Guidelines by Number

SMPTE 268M-1994	File Format for Digital Moving-Picture Exchange (DPX)
SMPTE 269M-1999	Television—Fault Reporting in Television Systems
SMPTE 270-1994	Motion-Picture Film (65-mm)—Manufacturer-Printed Latent Image Identification Information
SMPTE 271-1994	Motion-Picture Film (16-mm)—Manufacturer-Printed Latent Image Identification Information
SMPTE 272M-1994	Television—Formatting AES/EBU Audio and Auxiliary Data into Digital Video Ancillary Data Space
SMPTE 273M-1995	Television—Status Monitoring and Diagnostics Protocol
SMPTE 274M-1998	Television—1920 × 1080 Scanning and Analog and Parallel Digital Interfaces for Multiple Picture Rates
SMPTE 275M-1995	Television and Audio Equipment—Eslan-1 Remote Control System
SMPTE 276M-1995	Television—Transmission of AES-EBU Digital Audio Signals Over Coaxial Cable
SMPTE 277M-1996	Television Digital Recording—19-mm Type D-6—Helical Data, Longitudinal Index, Cue, and Control Records
SMPTE 278M-1996	Television Digital Recording—19-mm Type D-6—Content of Helical Data, and Time and Control Code Records
SMPTE 279M-1996	Digital Video Recording—1/2-in Type D-5 Component Format—525/60 and 625/50
SMPTE 291M-1998	Television—Ancillary Data Packet and Space Formatting
SMPTE 292M-1998	Television—Bit-Serial Digital Interface for High-Definition Television Systems
SMPTE 293M-1996	Television—720 × 483 Active Line at 59.94-Hz Progressive Scan Production—Digital Representation
SMPTE 294M-1997	Television—720 × 483 Active Line at 59.94-Hz Progressive Scan Production—Bit-Serial Interfaces
SMPTE 295M-1997	Television—1920 × 1080 50-Hz—Scanning and Interface
SMPTE 296M-2001	Television—1280 × 720 Scanning, Analog and Digital Representation, and Analog Interface
SMPTE 297M-2000	Television—Serial Digital Fiber Transmission System for ANSI/SMPTE 259M Signals
SMPTE 298M-1997	Television—Universal Labels for Unique Identification of Digital Data
SMPTE 299M-1997	Television—24-Bit Digital Audio Format for HDTV Bit-Serial Interface
SMPTE 300-1997	Motion-Picture Color Print Film (35-mm)—Manufacturer-Printed Latent Image Identification Information
SMPTE 301M-1999	Motion-Picture Film—Theater Projection Leader, Trailer, and Cue Marks
SMPTE 302-1998	Television—Mapping of AES3 Data into MPEG-2 Transport Stream
SMPTE 303M	Television—Color Reference Pattern
SMPTE 304M-1998	Television—Broadcast Cameras—Hybrid Electrical and Fiber-Optic Connector
SMPTE 305M-2-2000	Television—Serial Data Transport Interface
SMPTE 306M-1998	Television Digital Recording—6.35-mm Type D-7 Component Format—Video Compression at 25 Mb/s—525/60 and 625/50
SMPTE 307M-1998	Television Digital Recording—6.35-mm Type D-7 Component Format—Tape Cassette
SMPTE 308M-1998	Television—MPEG-2 4:2:2 Profile at High Level

SMPTE Standards, Recommended Practices, and Engineering Guidelines by Number

SMPTE 309M-1999	Television—Transmission of Date and Time Zone Information in Binary Groups of Time and Control Code
SMPTE 310M-1998	Television—Synchronous Serial Interface for MPEG-2 Digital Transport Streams
SMPTE 311M-1998	Television—Hybrid Electrical and Fiber Optical Camera Cable
SMPTE 312M-1999	Television—Splice Points for MPEG-2 Transport Streams
SMPTE 313-1999	Motion-Picture Film (65-mm)—Manufacturer-Printed Latent Image Identification Information—120 Perforation Repeat
SMPTE 314M-1999	Television—Data Structure for DV-Based Audio, Data, and Compressed Video—25 and 50 Mbit/s
SMPTE 315M-1999	Television—Camera Positioning Information Conveyed by Ancillary Data Packets
SMPTE 316M-1999	Television Digital Recording—12.65-mm Type D-9 Component Format—Video Compression—525/60 and 625/50
SMPTE 317M-1999	Television Digital Recording—12 .65-mm Type D-9 Component Format—Tape Cassette
SMPTE 318M-1999	Television and Audio—Reference Signals for the Synchronization of 59.94- or 50-Hz Related Video and Audio Systems in Analog and Digital Areas
SMPTE 319M-2000	Television—Transporting MPEG-2 Recoding Information through 4:2:2 Component Digital Interfaces
SMPTE 320M-1999	Television—Channel Assignments and Levels on Multichannel Audio Media
SMPTE 321M-1999	Television—Data Stream Format for the Exchange of DV-Based Audio, Data, and Compressed Video Over a Serial Data Transport Interface
SMPTE 322M-1999	Television—Format for Transmission of DV Compressed Video, Audio, and Data Over a Serial Data Transport Interface
SMPTE 323M-1999	Motion-Picture Film—Channel Assignments and Levels on Multichannel Audio Media
SMPTE 324M-2000	Television—12-Channel Serial Interface for Digital Audio and Auxiliary Data
SMPTE 325M-1999	Digital Television—Opportunistic Data Broadcast Flow Control
SMPTE 326M-2000	Television—SDTI Content Package Format (SDTI-CP)
SMPTE 327M-2000	Television—MPEG-2 Video Recoding Data Set
SMPTE 328M-2000	Television—MPEG-2 Video Elementary Stream Editing Information
SMPTE 329M-2000	Television—MPEG-2 Video Recoding Data Set—Compressed Stream Format
SMPTE 330M-2000	Television—Unique Material Identifier (UMID)
SMPTE 331M-2000	Television—Element and Metadata Definitions for the SDTI-CP
SMPTE 332M-2000	Television—Encapsulation of Data Packet Streams over SDTI (SDTI-PF)
SMPTE 333M-1999	Television—DTV Closed-Caption Server to Encoder Interface
SMPTE 334M-2000	Television—Vertical Ancillary Data Mapping for Bit-Serial Interface
SMPTE 337M-2000	Television—Format for Non-PCM Audio and Data in an AES3 Serial Digital Audio Interface
SMPTE 338M-2000	Television—Format for Non-PCM Audio and Data in an AES3 Serial Digital Audio Interface—Data Types
SMPTE 339M-2000	Television—Format for Non-PCM Audio and Data in an AES3 Serial Digital Audio Interface—Generic Data Types
SMPTE 340M-2000	Television—Format for Non-PCM Audio and Data in an AES3 Serial Digital Audio Interface—ATSC A/52 (AC-3) Data Type

13-76 Standards and Practices

SMPTE Standards, Recommended Practices, and Engineering Guidelines by Number

SMPTE 341M-2000	Television—Format for Non-PCM Audio and Data in an AES3 Serial Digital Audio Interface—Captioning Data Type
SMPTE 342M-2000	Television—HD-D5 Compressed Video 1080i and 720p Systems—Encoding Process and Data Format
SMPTE 344M-2000	Television—540 Mbits/s Serial Digital Interface
SMPTE 346M-2000	Television—Time Division Multiplexing Video Signals and Generic Data over High-Definition Interfaces
SMPTE 348M-2000	Television—High Data-Rate Serial Data Transport Interface (HD-SDTI)
SMPTE 351M-2000	Television—Transporting MPEG-2 Recoding Information through High-Definition Digital Interface
SMPTE 353M-2000	Transport of MPEG-2 Recoding Information as Ancillary Data Packets
SMPTE Recommended Practices	
RP 6-1994	Recorded Carrier Frequencies and Preemphasis Characteristics for 2-in Quadruplex Video Magnetic Tape Recording for 525-Line/60-Field Television Systems (R1999)
RP 9-1995	Dimensions of Double-Frame 35-mm 2 × 2 Slides for Precise Applications in Television
RP 11-1994	Tape Vacuum Guide Configuration and Position for Quadruplex Video Magnetic Tape Recording (R1999)
RP 12-1997	Screen Luminance for Drive-In Theaters
RP 14-1997	Plotting Data from Sensitometric Strips Exposed on Type Ib (Intensity Scale) Sensitometers
RP 15-1997	Calibration of Densitometers Used for Black-and-White Photographic Density Measurement
RP 16-1993	Specifications of Tracking Control Record for 2-in Quadruplex Video Magnetic Tape Recordings (R1997)
RP 17-1964	Photographic Recording Technique for Measuring High-Speed Camera Image Unsteadiness (R1997)
RP 18-1995	Specifications for Test Film for Subjective Checking of 16-mm Motion-Picture Audio Projectors
RP 19-1995	Specifications for 8-mm Type R Registration Test Film
RP 20-1995	Specifications for 16-mm Registration Test Film
RP 21-1997	Dimensions of 35- and 70-mm Motion-Picture Rewind Spindles
RP 24-1997	Dimensions for 16-mm Motion-Picture Camera Spindles
RP 25-1995	Audio and Picture Synchronization on Motion-Picture Film Relative to the Universal Leader for Magnetic and Photographic Records
RP 27.1-1989	Specifications for Operational Alignment Test Pattern for Television
RP 27.2-1989	Specifications for Operational Registration Test Pattern for Multiple-Channel Television Cameras
RP 27.3-1989	Specifications for Safe Action and Safe Title Areas Test Pattern for Television Systems
RP 27.4-1994	Specifications for an Operational Test Pattern for Checking Jitter, Weave and Travel Ghost in Television Projectors
RP 27.5-1989	Specifications for Mid-Frequency Response Test Pattern for Television
RP 32-1995	Specifications for 8-mm Type S Test Film for Projectors and Printers
RP 34-1998	Dimensions for 16-mm Motion-Picture Projector Reel Spindles
RP 36-1999	Positioning the Headwheel and Adjacent Tape Guides for 2-in Quadruplex Video Magnetic Tape Recorders

SMPTÉ Standards, Recommended Practices, and Engineering Guidelines by Number

RP 38.1-1989	Specifications for Deflection Linearity Test Pattern for Television
RP 39-1993	Specifications for Maintaining an Emulsion-In Orientation on Theatrical Release Prints (R1997)
RP 40-1995	Specifications for 35-mm Projector Alignment and Screen Image Quality Test Film
RP 45-1972	Use and Care of Sound Test Films (R1987)
RP 47-1999	Electronic Method of Dropout Detection and Counting
RP 48-1999	Lubrication of 16- and 8-mm Motion-Picture Prints
RP 49-1995	Leaders for 8-mm Type R and S Motion-Picture Release Prints Used in Continuous-Loop Cartridges
RP 50-1995	Dimensions for 8-mm Type S Motion-Picture Projector Reel Spindles
RP 51-1995	Screen Luminance and Viewing Conditions for 8-mm Review Rooms
RP 53-1993	Scene-Change Methods for Printing 35-mm, 16-mm and 8-mm Type S Motion-Picture Film (R1997)
RP 54-1999	Edge Numbering on 16-mm Release Prints
RP 55-1997	8-mm Type S Sprocket Design
RP 56-1995	Safe Action and Safe Title Areas for 8-mm Release Prints
RP 58-1995	Nomenclature for Devices Enclosing 8-mm Motion-Picture Film for Projection
RP 59-1999	Color and Luminance of Review Room Screens for Viewing Motion-Picture Materials Intended for Slides or Film Strips
RP 63-1997	Specifications for Sound-Focusing Test Film for 16-mm Audio Reproducers, Photographic Type
RP 64-1999	Specifications for Audio-Focusing Test Film for 35-mm Audio Reproducers, Photographic Type
RP 65-2000	Motion-Picture Enlargement/Reduction Ratios
RP 67-1997	Specifications for Buzz-Track Test Film for 16-mm Motion-Picture Audio Reproducers, Photographic Type
RP 68-1997	Specifications for Buzz-Track Test Film for 35-mm Motion-Picture Photographic Audio Reproducers
RP 69-1997	Specifications for Scanning-Beam Uniformity Test Film for 35-mm Motion-Picture Audio Reproducers
RP 70-1997	Specifications for Flutter Test Film for 16-mm Audio Reproducers, Photographic Type
RP 73-1992	8-mm Type R (Regular 8) Sprocket Design (R1997)
RP 74-1992	16-mm Sprocket Design (R1997)
RP 75-1997	Specifications for Flutter Test Film for 35-mm Studio Audio Reproducers, Magnetic Type
RP 76-1997	Specifications for Flutter Test Film for 16-mm Audio Reproducers, Magnetic Type
RP 77-1994	Specifications for Azimuth Test Film for 35-mm Studio Audio Reproducers, Magnetic Type
RP 78-1997	Specifications for Azimuth Test Film for 16-mm Audio Projectors, Magnetic Type
RP 79-1999	Specifications for Flutter Test Film for 35-mm Four-Track Striped Release Print Audio Reproducers, Magnetic Type
RP 81-1999	Specifications for Scanning-Beam Uniformity Test Film for 16-mm Motion-Picture Photographic Audio Reproducers
RP 82-1995	Specifications for 16-mm Projector Alignment and Screen Image Quality Test Film

13-78 Standards and Practices

SMPTÉ Standards, Recommended Practices, and Engineering Guidelines by Number

RP 83-1996	Specifications of Tracking-Control Record for 1-in Type B Helical-Scan Television Analog Recording
RP 84-1996	Reference Carrier Frequencies and Preemphasis Characteristics for 1-in Type B Helical-Scan Television Analog Recording
RP 85-1999	Tracking-Control Record for 1-in Type C Helical-Scan Television Tape Recording
RP 86-1991	Video Record Parameters for 1-in Type C Helical-Scan Television Tape Recording (R1995)
RP 87-1999	Reference Carrier Frequencies, Preemphasis Characteristic and Audio and Control Signals for 3/4-in Type E Helical-Scan Video Tape Cassette Recording
RP 90-1999	Specifications for Type U Audio Level and Multifrequency Test Film for 16-mm Audio Reproducers, Magnetic Type
RP 91-1997	Specifications for 70-mm Projector Alignment and Screen Image Quality Test Film
RP 92-1995	Specifications for Audio Level and Multifrequency Test Films for 8-mm Type S Audio Reproducers, Magnetic Type
RP 93-1999	Requirements for Recording American National Standard Time and Control Code for 1-in Type B Helical-Scan Video Tape Recorders
RP 94-2000	Gain Determination of Front Projection Screens
RP 95-1994	Installation of Gain Screens
RP 96-1993	Specifications for Subjective Reference Tapes for Helical-Scan Video Tape Reproducers for Checking Receiver/Monitor Setup
RP 97-1997	Specifications for Flutter Test Film for 35-mm Audio Reproducers, Photographic Type
RP 98-1995	Measurement of Screen Luminance in Theaters
RP 103-1995	Care, Storage, Operation, Handling and Shipping of Video Tape for Television
RP 104-1994	Cross-Modulation Tests for Variable-Area Photographic Audio Tracks
RP 105-1995	Method for Determining the Degree of Jump and Weave in 70-, 35,- and 16-mm Motion-Picture Projected Images
RP 106-1994	Film Tension in 35-mm Motion-Picture Systems Operating Under 0.9 m/s (180 ft/min)
RP 107-1995	Video and Audio Reference Tape for 1-in Type B Helical-Scan Format
RP 109-1994	Spectral Response of Photographic Audio Reproducers for 8-mm Type S Motion-Picture Film (R1999)
RP 110-1992	Specifications for an Alignment Test Film for Anamorphic Attachments to 35-mm Motion-Picture Projectors
RP 111-1999	Dimensions for 70-, 65- and 35-mm Motion-Picture Film Splices
RP 113-1996	Supervisory Protocol for Digital Control Interface
RP 114-1994	Dimensions of Photographic Control and Data Record on 16-mm Motion-Picture Film (R1999)
RP 115-1997	Dimensions of Photographic Control and Data Record on 35-mm Motion-Picture Release Prints
RP 116-2000	Dimensions of Photographic Control and Data Record on 35-mm Motion-Picture Camera Negatives
RP 117-1994	Dimensions of Magnetic Control and Data Record on 8-mm Type S Motion-Picture Film
RP 120-1994	Measurement of Intermodulation Distortion in Motion-Picture Audio Systems
RP 121-1997	Tape Dropout Specifications for 1-in Types B and C Video Tape Recorders/Reproducers

SMPTÉ Standards, Recommended Practices, and Engineering Guidelines by Number

RP 122-1993	Dimensions of Cemented Splices on 8-mm Type S Motion-Picture Film, Projection Type (R1997)
RP 123-1997	Dimensions of Tape Splices on 8-mm Type S Motion-Picture Film, Projection Type
RP 124-1998	Insertion Pivot for Studio Lighting Units and Mating Holders for Use with Standing and Hanging Support Systems
RP 127-1999	Specifications for Type U Audio Level and Multifrequency Test Film for 35-mm Studio Audio Reproducers, Magnetic Full-Coat Type
RP 128-1997	Specifications for Audio Level and Multifrequency Test Film for 70-mm Striped Six-Track Release Print Audio Reproducers, Magnetic Type
RP 129-1995	Requirements for 35-, 16- and 8-mm Type S Tape Splices on Magnetic Audio Recording Motion-Picture Film
RP 130-1995	Dimensions of Tape Splices on 16-mm and 8-mm Type R Motion-Picture Film, Projection Type
RP 131-1994	Storage of Motion-Picture Films
RP 132-1994	Storage of Edit Decision Lists on 8-in Flexible Diskette Media
RP 133-1999	Specifications for Medical Diagnostic Imaging Test Pattern for Television Monitors and Hard-Copy Recording Cameras (R1995)
RP 134-1994	Polarity for Analog Audio Magnetic Recording and Reproduction (R1999)
RP 135-1999	Use of Binary User Groups in Motion-Picture Time and Control Codes
RP 136-1999	Time and Control Codes for 24, 25, or 30 Frame-Per-Second Motion-Picture Systems
RP 138-1996	Control Message Architecture for Digital Control Interface
RP 139-1997	Tributary Interconnection for Digital Control Interface
RP 140-1995	Position of Photographic Audio Record for Routine Test Signals
RP 141-1995	Background Acoustic Noise Levels in Theaters and Review Rooms
RP 142-1997	Stereo Audio Track Allocations and Identification of Noise Reduction for Video Tape Recording
RP 143-1999	Specifications for Type U Audio Level and Multifrequency Test Film for 35-mm Striped Four-Track Release Print Audio Reproducers, Magnetic Type
RP 144-1999	Basic System and Transport Geometry Parameters for 1/2-in Type L Format
RP 145-1999	SMPTÉ C Color Monitor Colorimetry
RP 148-1987	Relative Polarity of Stereo Audio Signals (R1997)
RP 149-1992	Dimensions of Transverse Cemented Splices on 16-mm and 8-mm Type R Motion-Picture Film (R1997)
RP 150-2000	Channel Assignments and Test Leader for Magnetic Film Masters Intended for Transfer to Video Media Having Stereo Audio
RP 151-1999	Lubrication of 35-mm Motion-Picture Prints for Projection
RP 152-1994	Edge Identification of Leader and Picture for 35-mm Release Prints (R1999)
RP 153-1999	Method for Measuring 35- and 70-mm Shutter Efficiency
RP 155-1997	Audio Levels for Digital Audio Records on Digital Television Tape Recorders
RP 156-1999	Bar Code Labeling for Type D-1 Component and Type D-2 Composite Cassette Identification
RP 157-1995	Key Signals
RP 158-1999	Basic System and Transport Geometry Parameters for 1/2-in Type M-2 Format

13-80 Standards and Practices**SMPTÉ Standards, Recommended Practices, and Engineering Guidelines by Number**

RP 159-1995	Vertical Interval Time Code and Longitudinal Time Code Relationship
RP 160-1997	Three-Channel Parallel Analog Component High-Definition Video Interface
RP 161-1999	Logic Design for Decoding Digital Audio Control Words in D-1 Helical Data and Control Records
RP 162-1993	Storage of Edit Decision Lists on 3-1/2 in Disks
RP 163-1992	Television—System Service Messages
RP 164-1996	Location of Vertical Interval Time Code
RP 165-1994	Error Detection Checkwords and Status Flags for Use in Bit-Serial Digital Interfaces for Television
RP 166-1995	Critical Viewing Conditions for Evaluation of Color Television Pictures
RP 167-1995	Alignment of NTSC Color Picture Monitors
RP 168-1993	Definition of Vertical Interval Switching Point for Synchronous Video Switching
RP 169-1995	Television, Audio and Film Time and Control Code—Auxiliary Time Address Data in Binary Groups—Dialect Specification of Directory Index Locations
RP 170-1993	Video Tape Recorder Type-Specific Messages for Digital Control Interface
RP 171-1993	Type-Specific Messages for Digital Control Interface of Analog Audio Tape Recorders
RP 172-1993	Common Messages for Digital Control Interface
RP 174-1993	Bit-Parallel Digital Interface for 4:4:4:4 Component Video Signal (Single Link)
RP 175-1997	Digital Interface for 4:4:4:4 Component Video Signals (Dual Link)
RP 176-1997	Derivation of Reference Signals for Television Camera Color Evaluation
RP 177-1993	Derivation of Basic Television Color Equations (R1997)
RP 178-1996	Serial Digital Interface Checkfield for 10-Bit 4:2:2 Component and $4f_{sc}$ Composite Digital Signals
RP 179-1994	Dialect Specification of Page-Line Directory Index for Television, Audio, and Film Time and Control Code for Video-Assisted Film Editing
RP 180-1999	Spectral Conditions Defining Printing Density in Motion-Picture Negative and Intermediate Films
RP 181-1999	Audio Sector Time Code and Equipment-Type Information for 19-mm Type D-1 Digital Component Recording
RP 182-1995	List of Virtual Machine Numbers for ESbus and ESLan Systems
RP 183-1995	Monitoring and Diagnostics Processors
RP 184-1996	Specification of Jitter in Bit-Serial Digital Systems
RP 185-1995	Classification of Projection Depth of Focus
RP 186-1995	Video Index Information Coding for 525- and 625-Line Television Systems
RP 187-1995	Center, Aspect Ratio and Blanking of Video Images
RP 188-1999	Transmission of Time Code and Control Code in the Ancillary Data Space of a Digital Television Data Stream
RP 189-1996	Organization of DPX Files on TAR Tapes
RP 190-1996	Care and Preservation of Audio Magnetic Recordings
RP 191-1996	Routing Switcher Type-Specific Messages for Remote Control of Broadcast Equipment
RP 192-1996	Jitter Measurement Procedures in Bit-Serial Digital Interfaces
RP 193	Test Patterns and Test Images for DPX Leader

SMPTE Standards, Recommended Practices, and Engineering Guidelines by Number

RP 194-1998	Film Negative Cutter's Conform List
RP 195-1998	Use of the Reference Mark in Manufacturer-Printed Latent Image Key Numbers for Unambiguous Film Frame Identification
RP 196-1997	Transmission of LTC and VITC Data as HANC Packets in Serial Digital Television Interfaces
RP 197-1998	Film to Video Transfer List
RP 198-1998	Bit-Serial Digital Checkfield for Use in High-Definition Interfaces
RP 199-1999	Mapping of Pictures in Wide-Screen (16:9) Scanning Structure to Retain Original Aspect Ratio of the Work
RP 200-1999	Relative and Absolute Sound Pressure Levels for Motion-Picture Multichannel Sound Systems
RP 201-1999	Encoding Film Transfer Information Using Vertical Interval Time Code
RP 202	Video Alignment for MPEG-2 Coding
RP 203-2000	Real Time Opportunistic Data Flow Control in an MPEG-2 Transport Emission Multiplex
RP 204-2000	SDTI-CP MPEG Decoder Templates
RP 205-2000	Application of Unique Material Identifiers in Production and Broadcast Environments
RP 206-1999	Opportunistic Data Flow Control Using Ethernet as a Control Channel in an MPEG-2 Transport Emissions Multiplex
RP 209-2000	Format for Transmission of HD-D5 Compressed Video and Audio Data over 360 Mb/s Serial Digital Interface
RP 211-2000	Implementation of 24P, 25P and 30P Segmented Frames for 1920 x 1080 Production Format

SMPTE Engineering Guidelines

EG 1-1990	Alignment Color Bar Test Signal for Television Picture Monitors
EG 2-1999	Edge Identification of Motion-Picture Raw Stock Containers
EG 3-1994	Projection for Technical Conferences
EG 5-1994	Projected Image Quality of 70-, 35-, and 16-mm Motion-Picture Projection Systems
EG 7-1994	Audio Sync Pulse for 8-mm Type S Cameras, Magnetic Audio Recorders, and Rerecording Projectors (R1999)
EG 8-1993	Specifications for Motion-Picture Camera Equipment Used in Space Environment (R1997)
EG 9-1995	Audio Recording Reference Level for Post-Production of Motion-Picture Related Materials
EG 10-1996	Tape Transport Geometry Parameters for 19-mm Type D-1 Television Digital Component Recording
EG 12-1994	Control of Basic Parameters in the Manufacture of SMPTE Photographic and Magnetic Audio Test Films (R1999)
EG 13-1986	Use of Audio Magnetic Test Films (R1997)
EG 14-1999	Acoustical Background Noise Levels in Dubbing Stages
EG 15-1987	Recording Level for Dialog in Motion-Picture Production
EG 16-1997	Measurement Methods for Motion-Picture Camera Acoustical Noise—Field Method
EG 17-1997	B-Chain Electroacoustical Response for Preparing Magnetic Masters for Transfer to 16-mm or 35/32-mm Monaural Photographic Film
EG 18-1994	Design of Effective Cine Theaters

13-82 Standards and Practices

SMPTE Standards, Recommended Practices, and Engineering Guidelines by Number

EG 20-1997	Tape Transport Geometry Parameters for 19-mm Type D-2 Composite Format for Television Digital Recording
EG 21-1997	Nomenclature for Television Digital Recording of 19-mm Type D-1 Component and Type D-2 Composite Formats
EG 22-1997	Description and Index of Documents for 19-mm Type D-2 Composite Television Digital Recording
EG 23-1996	Transfer of Two-Channel Stereo Audio from Audio Magnetic Film or Tape to Video Tape
EG 24-1995	Video and Audio Alignment Tapes and Procedures for 1-in Type C Helical-Scan Television Analog Recorders
EG 25-1996	Telecine Scanning for Film Transfer to Television
EG 26-1995	Audio Channel Assignments for Digital Television Tape Recorders with AES/EBU Digital Audio Input
EG 27-1994	Supplemental Information for ANSI/SMPTE 170M and Background on the Development of NTSC Color Standards (R1999)
EG 28-1993	Annotated Glossary of Essential Terms for Electronic Production
EG 29-1993	Remote Control of Television Equipment
EG 30-1995	Implementation of ESlan Standards
EG 31-1995	Considerations for Cassette Bar Code Readers
EG 32-1996	Emphasis of AES/EBU Audio in Television Systems and Preferred Audio Sampling Rate
EG 33-1998	Jitter Characteristics and Measurements
EG 34-1999	Pathological Conditions in Serial Digital Video Systems
EG 35-1999	Time and Control Code Time Address Clock Precision for Television, Audio and Film
EG 36	Transformations Between Television Component Color Signals

13.4.5 CEA Standards

The following documents relating to radio and television broadcasting have been adopted by the Consumer Electronics Association. For additional information contact the CEA at www.cea.org.

13.4.5a Home Networks

ANSI/EIA-600.10: Introduction to the CEBus Standard

The EIA-600 Specification covers the overall topology of the EIA-600 network and the detailed topology for each individual medium used; the electrical and physical specifications for the media usable by EIA-600; the physical interface from a device to the medium and the signaling method specifications to be used on the medium; the protocol to be used for network access and the description of the control message format; and a command language that allows all devices to communicate a common set of functions to be performed. Aspects of the overall EIA-600 network that are not addressed in this specification are operation and maintenance of the network. This standard establishes a minimal set of rules for compliance. It does not rule out extended services that may be provided, as long as the rules of this standard are adhered to within the system.

It is, in fact, the intention of the standards to permit extended services (defined by users) to exist.

ANSI/EIA-600.31: Power Line Physical Layer and Medium Specification (ANSI/EIA-600.31-97)

This document is the preliminary specification for the CEBus Power Line (PL) Physical Layer and Media portion of the Physical Layer and Media Specifications of EIA-600. Its purpose is to present the information necessary for the development of a PL physical network and devices to communicate and share information over the network. This is one of a series of documents covering the various media that comprise the CEBus standard.

ANSI/EIA-600.32: Twisted Pair Physical Layer and Medium Specification (ANSI/EIA-600.32-97)

This document is the specification for the CEBus Twisted Pair (TP) Physical Layer and Medium. Its purpose is to present all the information necessary for the development of a TP physical network and devices to communicate and share information over that network in an orderly manner. This is one of a series of documents covering the various media that comprise the CEBus standard.

ANSI/EIA-600.33: Coax Cable Physical Layer and Medium Specification (ANSI/EIA-600.33-97)

This document is the preliminary specification for the CEBus Coax (CX) Physical Layer and Medium. Its purpose is to present all the information necessary for the development of a CX physical network and devices to communicate and share information over that network in an orderly manner. This is one of a series of documents covering the various media that comprise the CEBus standard.

ANSI/EIA-600.34: IR Physical Layer and Medium Specification (ANSI/EIA-600.34-97)

This document is a preliminary specification for the CEBus Infrared (IR) Physical Layer and Medium portion of the Physical Layer and Medium specifications of EIA-600. Its purpose is to present all the information necessary for the development of a IR physical network and devices to communicate and share information over that network to and from IR and other CEBus media in an orderly manner. This is one of a series of documents covering the various media that comprise the CEBus standard.

ANSI/EIA-600.35: RF Physical Layer and Medium Specification (ANSI/EIA-600.35-97)

This document is the preliminary specification for the CEBus Radio Frequency (RF) Physical Layer and Medium portion of the Physical Layer and Medium specifications of EIA-600. Its purpose is to present all of the information necessary for the development of a RF physical layer for the CEBus device. This is one of a series of documents covering various media that comprise the CEBus standard.

ANSI/EIA-600.37: Symbol-Encoding Sublayer

This document describes the portion of the Node Physical Layer that interfaces to the Medium Access Control (MAC) Sublayer and to Layer System Management (LSM). This sublayer is called the Symbol Encoding (SE) Sublayer.

13-84 Standards and Practices

ANSI/EIA-600.38: Power Line/Radio Frequency Symbol Encoding Sublayer

This document describes the portion of the Power Line or RF Physical Layer that interfaces to the Medium Access Control (MAC) Sublayer and to Layer System Management (LSM). This sublayer is called the Power Line/RF Symbol Encoding (PL/RF SE) Sublayer.

ANSI/EIA-600.41: Description of the Data Link Layer (ANSI/EIA-600.41-97)

This document provides a prose description of the Data Link Layer Design for the CEBus Network. The intent of this document is to be descriptive, rather than provide a formal specification, and contains a discussion of the Data Link Layer interfaces to the Network Layer and Physical Layer, as well as a functional description of the Data Link Layer.

ANSI/EIA-600.42: Node Medium Access Control Sublayer (ANSI/EIA-600.42-97)

This part of the CEBus standard is a technical specification of the services and protocol for the Node Medium Access Control Sublayer.

ANSI/EIA-600.43: Node Logical Link Control Sublayer (ANSI/EIA-600.43-97)

This part of the CEBus standard is a technical specification of the services and protocol for the Node Logical Link Control Sublayer.

ANSI/EIA-600.45: Node Network Layer Specification

This document is the CEBus Node Network Layer part of EIA-600.

ANSI/EIA-600.46: Node Application Layer Specification

This document is the CEBus Node Application Layer part of EIA-600.

ANSI/EIA-600.81: Common Application Language (CAL) Specification (ANSI/EIA-600.81-97)

This document describes the basic framework of CAL. It is intended as an introduction to CAL operation and syntax that stresses the object-oriented aspects of CAL. It is believed that the object-oriented methodology offers the best means of understanding the complex interaction between devices, controls, and controllers present in the CEBus environment.

ANSI/EIA-600.82: CAL Context Description (ANSI/EIA-600.82-97)

This document describes the contexts, or main subsystems within a device, supported by the Common Application Language (CAL).

ANSI/EIA-633.10: Introduction to EIA-600 Conformance Specification

This standard is concerned with conformance of an implementation to the associated protocol specifications contained in EIA-600. A dual audience is expected for this standard. The first is laboratories that may be interested in acting as conformance testing agencies. Such agencies are tasked with converting these requirements into a hardware/software test system. The second audience is the set of designers of EIA-600 compatible products. Any designer of compatible products should understand the importance of the tests described in this standard, as they relate to implementing EIA-600. The second group should read this standard while considering whether its hardware and/or software implementation is likely to pass the stated tests.

ANSI/EIA-633.31: Power Line Physical Layer Conformance Specification

This portion of the conformance standard specifies tests to determine conformance of a Node's Power Line (PL) PL Physical Layer to IS-60. Part one of this standard provides an overview of the conformance philosophy. The reader is urged to review that material before attempting to use the details provided in this part.

ANSI/EIA-633.32: Twisted Pair Physical Layer Conformance (ANSI/EIA-633.32-97)

This standard specifies tests to determine conformance of a device's Twisted Pair Physical Layer to EIA-600.

ANSI/EIA-633.34: Infrared Physical Layer Conformance (ANSI/EIA-633.34-97)

This standard specifies tests to determine conformance of a Node's IR Physical Layer to EIA-600.

ANSI/EIA-633.37: Symbol Encoding Sublayer Physical Layer Conformance (ANSI/EIA-633.37-97)

This standard specifies tests to determine conformance of a Node's Symbol Encoding Sublayer to EIA-600.

ANSI/EIA-633.38: PL and RF Symbol Encoding Physical Layer Conformance (ANSI/EIA-633.38-97)

This standard specifies tests to determine conformance of a Node's Power Line or RF Symbol Encoding Sublayer to EIA-600.

ANSI/EIA-633.46: Node Application Layer Conformance Specification

This portion of the conformance standard specifies tests to determine conformance of a Node's Application Layer to EIA-600.

ANSI/EIA-633.81: CAL Conformance Specification

This portion of the conformance standard specifies tests to determine conformance of a Node's CAL to EIA-600.81. Part one of this standard provides an overview of the conformance philosophy. The reader is urged to review that material before attempting to use the details provided in this part.

ANSI/EIA-709.1-A: Control Network Protocol Specification (ANSI/EIA-709.1-A-99)

This specification applies to a communication protocol for networked control systems. The protocol provides peer-to-peer communication for networked control and is suitable for implementing both peer-to-peer and master-slave control strategies. This specification describes services in layers 2-7. In the layer 2 specification, it also describes the data link layer and the MAC sub-layer interface to the physical layer. The physical layer provides a choice of transmission media. The interface described in this specification supports multiple transmission media at the physical layer.

13-86 Standards and Practices

ANSI/EIA-709.2-A: Control Network Power Line (PL) Channel Specification (ANSI/EIA-709.2-A-99)

This document specifies the Control Network Power Line (PL) Channel and serves as a companion document to the EIA-709.1 Control Network Protocol Specification. Its purpose is to present the information necessary for the development of a PL physical network and nodes to communicate the share information over the network. This is one of a series of documents covering the various media that comprise the EIA-709 Standard.

ANSI/EIA-709.3: Free-Topology Twisted-Pair Channel Specification

This document specifies the EIA-709.3 free-topology twisted-pair channel and serves as a companion document to the EIA-709.1 Control Network Protocol Specification. The channel supports communication at 78.125 kbits/s between multiple nodes, each of which consists of a transceiver, a protocol processor, and application processor, a power supply, and application electronics.

ANSI/EIA-709.4: Fiber-Optic Channel Specification

In conjunction with ANSI/EIA-709.1-A Control Network Protocol Specification, EIA-709.4 defines a complete 7-layer protocol stack for communications on an EIA-709.4 single-fiber (half-duplex) fiber-optic channel. EIA-709.4 specifies the physical layer (OSI Layer 1) requirements for the EIA-709.4 fiber-optic channel which encompasses the interface to the Media Access Control (MAC) layer and the interface to the medium. The single-fiber channel implemented as specified in EIA-709.4 allows two nodes to communicate bidirectionally across a single piece of fiber cable.

ANSI/EIA-721.1: Generic Common Application Language (Generic CAL) Specification

This document describes the basic framework of Generic CAL. It is intended as an introduction to Generic CAL operation and syntax that stresses the object-oriented aspects of Generic CAL. It is believed that the object-oriented methodology offers the best means of understanding the complex interaction between devices, controls, and controllers present in a Generic Network environment.

ANSI/EIA-721.2: Generic CAL Context Description

This document describes the contexts, or main subsystems within a device, supported by the Generic Common Application Language (Generic CAL).

ANSI/EIA-721.3: Node Application Layer Specification

This Application Layer consists of four main elements. The application Process is the interface to the Application Layer. Services are provided by the Generic Common Application Language (Generic CAL) Element to the User Element of the Application Process. Generic CAL is the language framework through which Resource Allocation and Control functions are executed. Services are provided by the Message Transfer Element to the Generic CAL Element. The Message Transfer Element interfaces to the lower layers of the Generic Network either directly or through the Association Control Element. The lower layers are representative of some home automation networks. Additional OSI layers may be included. An adaptation layer may be required between the Generic CAL Application Layer and the Generic Network lower layers.

ANSI/EIA-721.4: Generic Common Application Language Quality of Service

This specification for Generic CAL consists of an Application Layer containing a command language and a Message Transfer Service Element. The specifications of the lower OSI layers are not within the scope of this standard. However, the services provided by the lower layers affect the performance and composition of messages issued from the Application Layer. These lower layer service options are collectively called the Quality-of-Service (QOS) available from the communications protocol. This portion of EIA-721 standard describes the lower layer QOS options that may impact the Application Layer. Recommended capabilities are specified. Also, a mechanism to convey these options to the Generic CAL Application Layer using Layer System Management functions is presented.

ANSI/EIA-776.1: CEBus-EIB Router Communications Protocol—Description of the CEBus-EIB Router

This document describes the operation of a CEBus-EIB Router. This document is not intended to define how a router should operate, but to provide an overview of the operation and the coordination of various router elements.

ANSI/EIA-776.2: CEBus-EIB Router Communications Protocol—CEBus-EIB Router Medium Access Control Sublayer

The CEBus-EIB Router Medium Control (MAC) Sublayer is almost identical to the CEBus or EIB Node MAC Sublayer corresponding to the “CEBus Side” or the Router. The differences are in the way the Router does address matching on a received packet and on the information exchanged in some of the service primitives. Rather than copy the Node MAC specification here and make minor changes, the Router MAC is specified by exception to the Node MAC for both the CEBus and EIB Specifications.

ANSI/EIA-776.3: CEBus-EIB Router Communications Protocol—CEBus-EIB Router Logical Link Control Sublayer

This section specifies the CEBus-EIB Router Logical Link Control Sublayer interfaces to the Router Network Layer and to the Layer System Management. The interfaces are described in terms of service primitives which are abstract interfaces across a layer boundary. A service primitive represents an exchange of information into or out of a layer. Although service primitives are defined using a format similar to that of programming language procedure calls, no implementation technique is implied.

ANSI/EIA-776.4: “CEBus-EIB Router Communications Protocol—CEBus-EIB Router Network Layer

The CEBus-EIB Router Network Layer is conceptually divided into several elements, each performing distinct well-defined services. Each element may be thought of as an independent process that communicates with the other elements and protocol layers through specified interfaces.

ANSI/EIA-776.5: CEBus-EIB Router Communications Protocol—The EIB Communications Protocol

EIB is a control system for related applications in homes and buildings. The EIB system offers standardized basic and system components, e.g., Bus Coupling Units (BCU), Power Supply Units (PSU), Bus Interface Modules (BIM), Routers, and RS-232 data interfaces. EIB offers the

13-88 Standards and Practices

capability of constructing devices in a modular form using system devices like BCU or BIM that support communications-specific functions. A standardized interface called Physical External Interface (PEI) reduces the expense of developing EIB devices and allows them to be exchanged.

EIA/CEA-844: XML Encoding of Generic Common Application Language

This standard specifies the encoding of Generic Common Application Language (CAL) into XML. It is based on ANSI/EIA-721 and EIA-851.

EIA/CEA-851: VHN Home Network

The R-7.4 VHN Home Network Standard defines a flexible and open network architecture and communications protocol specification for digital devices in the home.

EIA-600 CEBus SET: EIA Home Automation System (CEBus)

Provides the necessary specifications for the Consumer Electronic Bus (CEBus), a local communications and control network designed specifically for the home. The CEBus network provides a standardized communication facility for exchange of control information as data among devices and services in the home. The major motivations for its development were: a) to develop a universal low cost method for devices in the home to communicate regardless of manufacturer; b) to allow the introduction of new products and services to the home at minimal cost and confusion to consumers; c) to meet the majority of anticipated home control requirements with a single multi-media network standard; and d) to minimize the redundancy of control and operation methods among devices and equipment in the home. CEBus is intended to handle existing and anticipated control communication requirements at minimum practical costs consistent with a broad spectrum of residential applications. It is intended for such functions as remote control, status indication, remote instrumentation, energy management, security systems, and entertainment device coordination. These situations require economical connection to a shared local communication network carrying relatively short digital messages. A major objective of this specification is compatibility. It is intended that every implementation of CEBus be able to co-exist with every other implementation; that every device that meets this specification can communicate with all other CEBus devices; and the language used for control functions will be understood by all devices. This version includes portions of IS-60 that have not been revised as well as new ANSI approved updates.