Module - 03

Digital Audio

CHAPTER 3

Digital Audio

Syllabus Topic : Basic Sound Concepts - Computer representation of sound

3.1 Basic Sound Concepts

3.1.1 What is Sound ?

- Sound is a physical phenomenon produced by the vibration of matter. The matter can be almost anything: violin string or a block of wood, for example.
- As the matter vibrates, pressure variations are created in the air surrounding it. This alternation of high ar low pressure is propagated through the air in a wave-like motion. When the wave reaches our ears, we hear sound. Fig. 3.1.1 graphs the oscillation of a pressure wave over time.

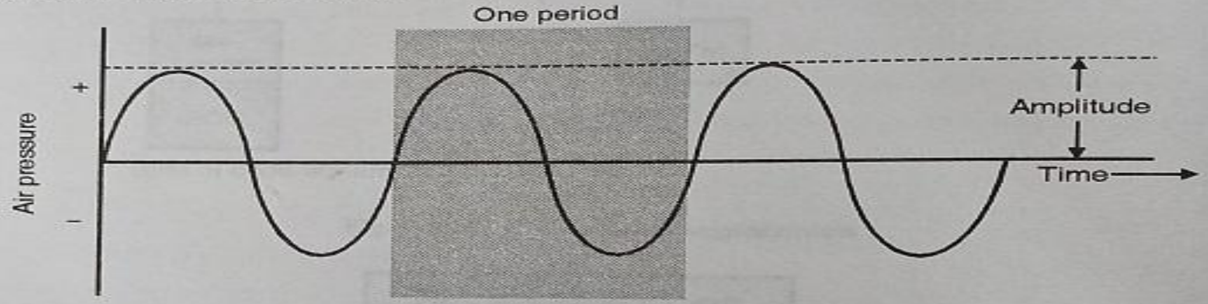

Fig. 3.1.1 : Air Pressure Wave

- The pattern of the pressure oscillation is called a *waveform*. Notice that the waveform in Fig. 3.1.1 repeats th same shape at regular intervals; the shaded area shows one complete shape.
- This portion of the waveform is called a *period*. A waveform with a clearly defined period occurring at regula intervals is called a*periodic waveform*.

- Since they occur naturally, sound waveforms are never as perfectly smooth nor as uniformly periodic as the waveform shown in Fig. 3.1.1.
- However, sounds that display a recognizable periodicity tend to be more musical than those that are nonperiodic.

Here are some sources of periodic and nonperiodic sounds :

Periodic

- Musical instruments other than unpitched percussion
- Vowel sounds
- Bird songs
- Whistling wind

Nonperiodic

- Unpitched percussion instruments
- Consonants, such as "t," "f," and "s"
- Coughs and sneezes
- Rushing water

3.1.2 Characteristics of Sound Waves

Frequency

- **Frequency** is defined as the number of vibrations, oscillations, or cycles in a repeating process occurring per unit time. In the context of sound, it is the number of compressions passing a fixed point of reference in one second. The resulting unit of frequency is called Hertz (Hz).
- Frequency is perceived as pitch. The *frequency* of a sound is the number of times the pressure rises and falls, or oscillates, in a second is measured in *hertz* (Hz). A frequency of 100 Hz means 100 oscillations per second. A convenient abbreviation, kHz for *kilohertz*, is used to indicate thousands of oscillations per second: 1 kHz equals 1000 Hz.
- The frequency range of normal human hearing extends from around 20 Hz up to about 20 kHz.
- The frequency axis is logarithmic, not linear: To traverse the audio range from low to high by equal-sounding steps, each successive frequency increment must be greater than the last.
- For example, the frequency difference between the lowest note on a piano and the note an octave above it is about 27 Hz. Compare this to the piano's top octave, where the frequency difference is over 2000 Hz. Yet, subjectively, the two intervals sound the same.

Amplitude

- **Amplitude** is the maximum change in value of a parameter during the oscillation of a wave. In this unit, that parameter will usually be pressure. In practice, this is the distance between a peak or trough and the x-axis on a graph. A sound also has an *amplitude*, a property subjectively heard as loudness. The amplitude of a sound is the measure of the displacement of air pressure from its mean or quiescent state. The greater the amplitude, the louder the sound.

Phase

The phase of a wave is an expression of how far through its cycle of oscillation it has progressed. Because the mathematical description of wave motion is similar to the mathematical description of motion in a circle, wave phase is expressed in degrees or radians. A wave completes its cycle in 360°, just as a circle is completed in 360°. Half a cycle is 180°, and a quarter cycle is 90°.

The relative phase of two similar waves is a measure of how synchronized they are :

- **In phase** : Similar waves whose peaks (maxima) coincide are said to be in phase. Their phase difference is 0°.
- **Out of phase** : Similar waves whose peaks (maxima) do not coincide are said to be out of phase. If the peaks of one coincide with troughs (minima) of the other, the two waves are said to be 180° out of phase.
- **Period** - The time required for a single wavelength to pass a fixed point of reference. The period of a sound wave is the inverse of its frequency.
- **Wavelength** - The distance between one peak or crest of a sound wave and the next corresponding peak or crest. The product of the wavelength and the frequency of a sound wave yields the velocity of that wave.
- **Waveform** - The detailed way in which a parameter changes during the oscillation of a wave. For sound, that parameter will usually be pressure. In practice, this is the shape of a wave on a graph. Waveform is perceived as timbre

3.1.3 How the Computer Represents Sound ?

The smooth, continuous curve of a sound waveform isn't directly represented in a computer. A computer measures the amplitude of the waveform at regular time intervals to produce a series of numbers. Each of these measurements is called a *sample*. Fig. 3.1.2 illustrates one period of a digitally sampled waveform.

Audio means "of sound" or "of the reproduction of sound". Specifically, it refers to the range of frequencies detectable by the human ear - approximately 20Hz to 20 kHz. It's not a bad idea to memorise those numbers - 20 Hz is the lowest-pitched (bassiest) sound we can hear, 20 kHz is the highest pitch we can hear. Audio work involves the production, recording, manipulation and reproduction of sound waves.

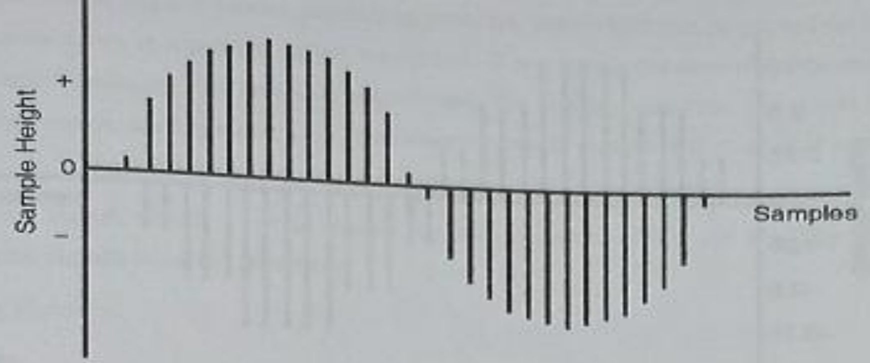

Fig. 3.1.2 : Sampled Waveform

Each vertical bar in Fig. 3.1.2 represents a single sample. The height of a bar indicates the value of that sample.

Sampling Rate

The rate at which a waveform is sampled is called the *sampling rate*. Like frequencies, sampling rates are measured in hertz. The rate can vary typically from 5000-90,000 samples per second. The audio input from a source is sampled several thousand times per second. Each sample is a snapshot of the original signal at a particular time. The CD standard sampling rate of 44100 Hz means that the waveform is sampled 44100 times per second. This may seem a bit excessive, considering that we can't hear frequencies above 20 kHz; however, the highest frequency that a digitally sampled signal can represent is equal to half the sampling rate. So a sampling rate of 44100 Hz can only represent frequencies up to 22050 Hz, a boundary much closer to that of human hearing.

Quantization

A waveform is sampled at discrete times, the value of the sample is also discrete. The *quantization* of a sample value depends on the number of bits used in measuring the height of the waveform. In the case of 8-bit *quantization*, this value is between 0 and 255 (or -128 and 127). In 16-bit digitization, this value is between 0 and 65,535 (or -32,768 and 32,767). Digitized signal can take only certain (discrete) values. The process of digitization introduces noise in a signal. This is related to the number of bits per sample. A higher number of bits used to store the sampled value leads to a more accurate sample, with less noise.

As an extreme example, Fig. 3.1.2 shows the waveform used in the previous example sampled with a 3-bit quantization. This results in only eight possible values : .75, .5, .25, 0, – .25, –.5, –.75, and –1. As you can see, the shape of the waveform becomes less discernible with a coarser quantization. The coarser the quantization, the "buzzier" the sound.

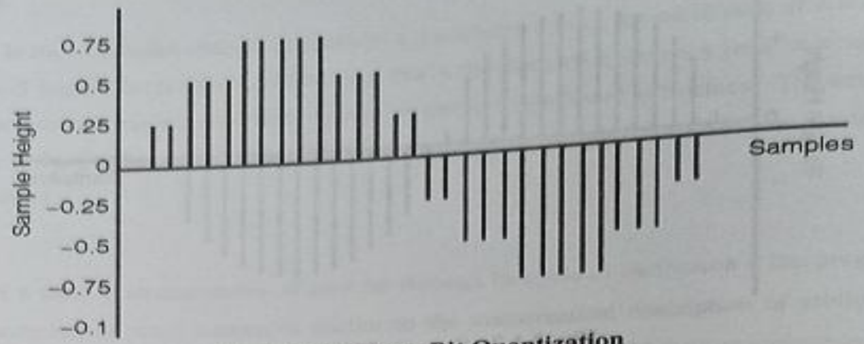

Fig. 3.1.3 : Three-Bit Quantization

Storing Sampled Data

An increased sampling rate and refined quantization improves the fidelity of a digitally sampled waveform; however, the sound will also take up more storage space. Five seconds of sound sampled at 44.1 kHz with a 16-bit quantization uses more than 400,000 bytes of storage a minute will consume more than five megabytes. A number of data compression schemes have been devised to decrease storage while sacrificing some fidelity.

3.1.4 Digitization of Audio

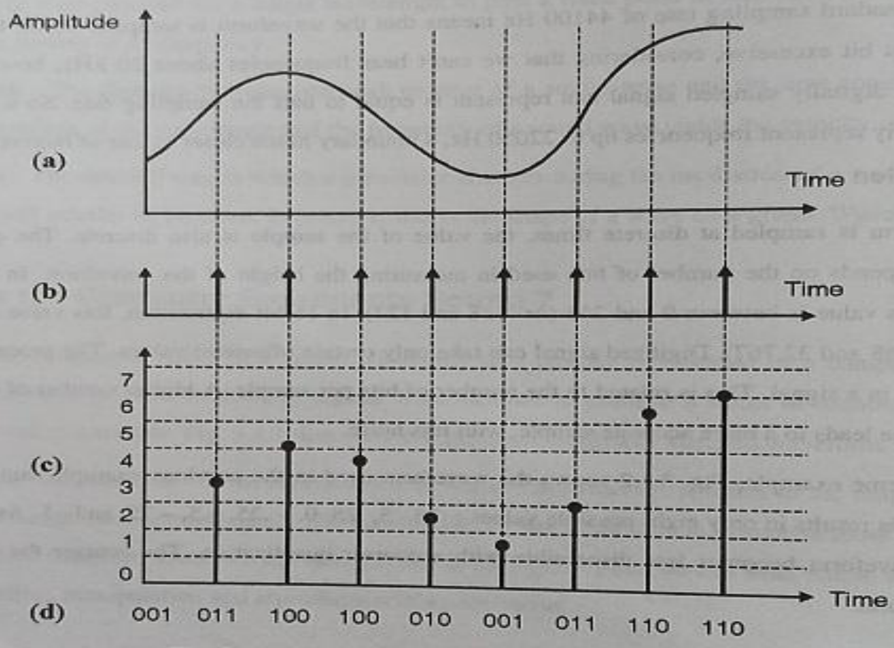

Fig. 3.1.4 : Analog-to-digital conversion process

Sound is produced when objects vibrate producing pressure waves that can be picked up by human's ear. The vibrating pressure waves move in a pattern called waveform. If we graph the intensity or motion of the wave over time, we can get a curve consisting of a series of waveforms. We can say that sound is stored by waves. These are *analog signals*. In other words, analog signals are continuous variable signals that consist of waves.

When the sound is needed to be used in any computer application, we need to converts the air vibrations of sound into an electrical signal, which is called *digital signal* - a stream of 0's and 1's. The process of conversing analog signals to digital signals is called *digitizing*.

(a) Original analog signal;

(b) Sampling pulses;

(c) Sampled values and quantization intervals;

(d) Digitized sequence.

Three stages are involved in digitization of audio : sampling, quantization, and coding

☞ Sampling

The process of converting continuous time into discrete values is called sampling. Fig. 3.1.4(b) and (c) show the sampling process. The time axis is divided into fixed intervals. The reading of the instantaneous value of the analog signal is taken at the beginning of each time interval. This interval is determined by a clock pulse. The frequency of the clock is called the sampling rate or sampling frequency. The sampled value is held constant for the next time interval.

☞ Quantization

The process of converting continuous sample values into discrete values is called quantization. In this process we divide the signal range into a fixed number of intervals. Each interval is of the same size and is assigned a number. In Fig. 3.1.4(c), these intervals are numbered from 0 to 7.

Each sample falls in one of the intervals and is assigned that interval's number. In doing this, each sample has a limited choice of values. In our example, a sample value can only be an integer number between 0 and 7. Before quantization, the last two samples in Fig. 3.1.4(c) have different values. But they have the same value of 6 after quantization. The size of the quantization interval is called the quantization step.

☞ Coding

The process of representing quantized values digitally is called coding Fig. 3.1.4(d). In the above example, eight quantizing levels are used. These levels can be coded using 3 bits if the binary system is used, so each sample is represented by 3 bits. The analog signal in Fig. 3.1.4(a) is represented digitally by the following series of binary numbers: 001, 011, 100, 100, 010, 001, 011, 110, and 110.

3.2 Audio Formats

Sound can be stored in many different formats. Some popular audio formats are listed below and most widel[y] used audio format MIDI, MPEG Audio and wave are described in detail.

The MIDI Format

- The MIDI (Musical Instrument Digital Interface) is a format for sending music information between electroni[c] music devices like synthesizers and PC sound cards.
- The MIDI format was developed in 1982 by the music industry. The MIDI format is very flexible and can be used for everything from very simple to real professional music making.
- MIDI files do not contain sampled sound, but a set of digital musical instructions (musical notes) that can be interpreted by your PC's sound card.
- The downside of MIDI is that it cannot record sounds (only notes). Or, to put it another way : It cannot store songs, only tunes.
- The upside of the MIDI format is that since it contains only instructions (notes), MIDI files can be extremely small. The example above is only 23K in size but it plays for nearly 5 minutes.
- The MIDI format is supported by many different software systems over a large range of platforms. MIDI files are supported by all the most popular Internet browsers.
- Sounds stored in the MIDI format have the extension .mid or .midi.

The RealAudio Format

- The RealAudio format was developed for the Internet by Real Media. The format also supports video.
- The format allows streaming of audio (on-line music, Internet radio) with low bandwidths. Because of the low bandwidth priority, quality is often reduced.
- Sounds stored in the RealAudio format have the extension .rm or .ram.

The AU Format

- The AU format is supported by many different software systems over a large range of platforms.
- Sounds stored in the AU format have the extension .au.

The AIFF Format

- The AIFF (Audio Interchange File Format) was developed by Apple.
- AIFF files are not cross-platform and the format is not supported by all web browsers.
- Sounds stored in the AIFF format have the extension .aif or .aiff.

The SND Format

- The SND (Sound) was developed by Apple.
- SND files are not cross-platform and the format is not supported by all web browsers.
- Sounds stored in the SND format have the extension .snd.

The WAVE Format

- The WAVE (waveform) format is developed by IBM and Microsoft.
- It is supported by all computers running Windows, and by all the most popular web browsers .
- Sounds stored in the WAVE format have the extension .wav.

The MP3 Format (MPEG)

- MP3 files are actually MPEG files. But the MPEG format was originally developed for video by the Moving Pictures Experts Group. We can say that MP3 files are the sound part of the MPEG video format.
- MP3 is one of the most popular sound formats for music recording. The MP3 encoding system combines good compression (small files) with high quality. Expect all your future software systems to support it.
- Sounds stored in the MP3 format have the extension .mp3, or .mpga (for MPG Audio).

3.2.1 WAVE File Format

The WAVE file format is a subset of Microsoft's RIFF specification for the storage of multimedia files. A RIFF file starts out with a file header followed by a sequence of data chunks.

A WAVE file is often just a RIFF file with a single "WAVE" chunk which consists of two sub-chunks - a "fmt " chunk specifying the data format and a "data" chunk containing the actual sample data.