Chapter 1
Multimedia

1.1 Introduction

Multi - many; much; multiple

Medium - a substance regarded as the means of transmission of a force or effect; a channel or system of communication, information, or entertainment

Medium

- *Means for distribution and presentation of information*
- *Classification based on perception (text, audio, video) is appropriate for defining multimedia*

Multimedia is a combination of text, graphics, sound, animation, and video that is delivered interactively to the user by electronic or digitally manipulated means.

Text:

- A broad term for something that contains words to express something.
- Text is the most basic element of multimedia.
- A good choice of words could help convey the intended message to the users (keywords).

Graphics:

- Two-dimensional figure or illustration.
- Could be produced manually (by drawing, painting, carving, etc.) or by computer graphics technology.
- Used in multimedia to show more clearly what a particular information is all about (diagrams, picture).

Audio:

- Produced by vibration, as perceived by the sense of hearing.
- In multimedia, audio could come in the form of speech, sound effects and also music score.

Animation:

- The illusion of motion created by the consecutive display of images of static elements.
- In multimedia, animation is used to further enhance/enriched the experience of the user to further understand the information conveyed to them.

Video:

- Is the technology of capturing, recording, processing, transmitting, and reconstructing moving pictures.
- Video is more towards photo realistic image sequence/live recording as in comparison to animation.
- Video also takes a lot of storage space. So plan carefully before you are going to use it.

Multimedia is the media that uses multiple forms of information content and information processing (e.g. text, audio, graphics, animation, and video interactivity) to inform or entertain the user. *Multimedia* also refers to the use of electronic media to store and experience multimedia content. Multimedia is similar to traditional mixed media in fine art, but with a broader scope. The term "rich media" is synonymous for interactive multimedia.

Medium: An intervening substance through which some thing is transmitted or carried on.

Computer System Medium:

1. Text
2. Image
3. Sound
4. Video

Representation Dimension of media:

Media are divided into two types in respect to time in their representation space:

1. *Time independent (discrete):* Information is expressed only in its individual value. E.g. : text, image etc.

2. *Time dependent (continuous):* Information is expressed not only it's individual value, but also by the time of its occurrences. E.g.: sound and video.

Multimedia system is defined by computer controlled, integrated production, manipulation, presentation, storage and communication of independent information, which is encoded at least through a continuous and discrete media.

Multimedia Building Blocks:

Any multimedia application consists any or all of the following components:

1. Text:

Text and symbols are very important for communication in any medium. With the recent explosion of the Internet and World Wide Web, text has become more the important than ever. Web is HTML (Hyper text Markup language) originally designed to display simple text documents on computer screens, with occasional graphic images thrown in as illustrations.

2. Audio:

Sound is perhaps the most element of multimedia. It can provide the listening pleasure of music, the startling accent of special effects or the ambience of a mood-setting background.

3. Images:

Images whether represented analog or digital plays a vital role in a multimedia. It is expressed in the form of still picture, painting or a photograph taken through a digital camera.

4. Animation:

Animation is the rapid display of a sequence of images of 2-D artwork or model positions in order to create an illusion of movement. It is an optical illusion of motion due to the phenomenon of persistence of vision, and can be created and demonstrated in a number of ways.

5. Video:

Digital video has supplanted analog video as the method of choice for making video for multimedia use. Video in multimedia are used to portray real time moving pictures in a multimedia project

Digital Representation:

Multimedia regards content and technologies dealing with a combination of different content forms/modalities, e.g. speech, audio, text, video, images, 3D models, etc.

Analog Signal:

An analog signal is any variable signal, continuous in both time and amplitude.

Digitization:

Digitization is the process of expressing analog data in digital form.

Analog data implies 'continuity' while digital data is concerned with discrete states, e.g. symbols, digits.

Interactive Multimedia:

When the user is given the option of controlling the elements.

Hyper Media:

A combination of hypertext, graphics, audio, video, (linked elements) and interactivity culminating in a complete non-linear computer-based experience.

Linear Vs Non-Linear

Linear:

A multimedia project is defined as linear when:

- It is not interactive
- Users have no control over the content that is being showed to them.
- Example: A movie, A non-interactive lecture/demo show

Non-Linear:

A multimedia project is defined as Non-linear when:

- It is interactive
- Users have control over the content that is being showed to them.
- Users are given navigational control.

- Example: Games

1.2 The Media Aspect

Perception Media:

Perception media help human to sense their environment. The central question is: *How human perceive information in a computer environment?* The answer is through seeing and hearing.

Seeing:

For the perception of information through seeing the usual such as *text, image and video* are used.

Hearing:

For the perception of information through hearing media such as *music, noise and speech* are used.

Representation Media:

Representation media are defined by internal computer representation of information. The central question is: *How the computer information is coded?* The answer is that various formats are used to represent media information in a computer.

i. Text, character is coded in ASCII code.
ii. Graphics are coded according to CEPT or CAPTAIN video text standard.
iii. Image can be coded as JPEG format.
iv. Audio video sequence can be coded in different TV standard format (PAL, NTSC, SECAM and stored in the computer in MPEG format).

Presentation Media:

Presentation media refer to the tools and devices for the input and output of the information. The central question is: *Through which the information is delivered by the computer and is introduced to the computer?* The answer is:

Output media:

Paper, screen and speaker are the output media.

Input media:

Keyboard, mouse, camera, microphone are the input media.

Storage Media:

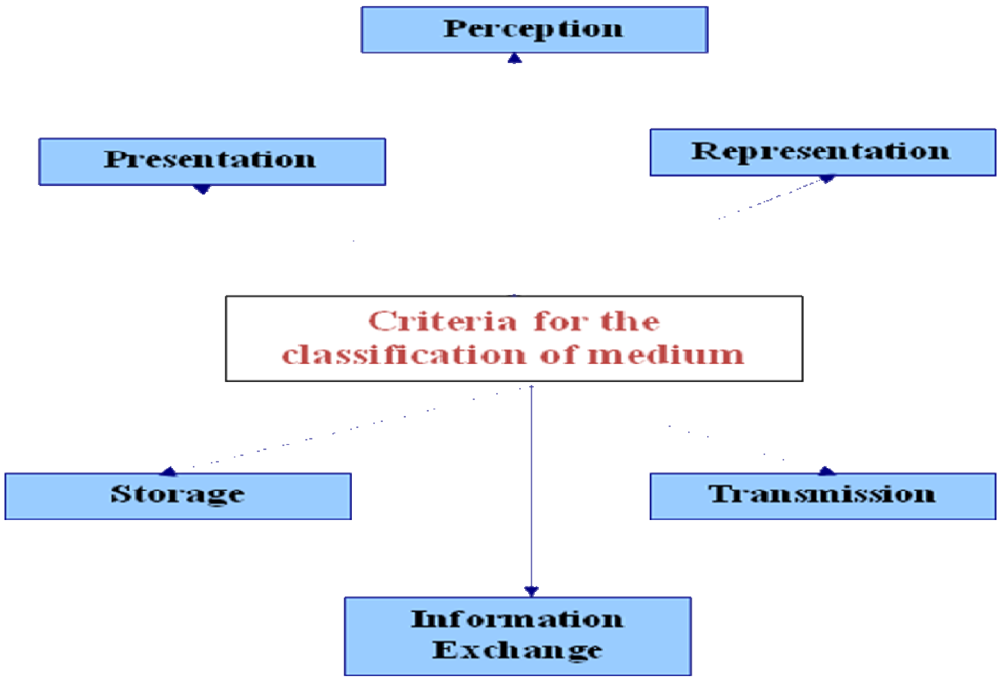
Storage media refer to the data carrier which enables storage of information. The central question is: *How will information be stored?* The answer is hard disk, CD-ROM, etc.

Transmission Media:

Transmission media are the different information carrier that enables continuous data transmission. The central question is: *Over what will the information be transmitted?* The answer is co-axial cable and fiber optics, as well as free air space transmission, which is used for wireless traffic.

Information Exchange Media:

Information exchange media includes all information carrier for transmission, i.e. all storage and transmission media. The central question is: *Which information carrier will be used for information exchange between different places?* The answer is combine uses of storage and transmission media. For example: Electronic Mailing System.



1.3 Main Properties of Multimedia

- **Combination of media:**
  - *Continuous and discrete.*

    A simple text processing program with incorporated images is often called a multimedia application because two media are processed through one program. But one should talk about multimedia only when both continuous and discrete media are utilized.
- **Levels of media-independence:**
  - *Some media types (audio/video) may be tightly coupled, others may not.*

    In general, there is a request for independence of different media, but multimedia may require several levels of independence. On the one hand, a computer controlled video recorder stores audio and video information, but there is an inherently tight connection the two types of media.
- **Computer supported integration:**
  - *Timing, spatial and semantic synchronization*

A text processing program that supports text, table calculations and video clips does not satisfy the demand for integration if program supporting the connection between the data cannot be established. A high integration level is accomplished if changing the content of a table row causes corresponding video scene and text changes.

Everything can be presented with video and sound that is presented with text and graphics today.

- **Communication capability:**

  Communication-capable multimedia systems must be approached. Multimedia information cannot only be created, processed, presented and stored, but also distributed above the single computer's boundary.

1.4 Media Concepts

Each medium defines:

- ❖ *Representation values*
- ❖ *Representation space*
- ❖ *Representation dimensions*

Representation values **-** determine the information representation of different media

- Continuous representation values (e.g. electro-magnetic waves)
- Discrete representation values(e.g. text characters in digital form)

Representation space determines the surrounding where the media are presented.

- Visual representation space (e.g. paper, screen)
- Acoustic representation space (e.g. stereo)

Representation dimensions of a representation space are:

- ■ Spatial dimensions
- ■ Temporal dimensions

Spatial dimensions:

- two dimensional (2D graphics)
- three dimensional (holography)

Temporal dimensions:

Time independent (document) - Discrete media

Information consists of a sequence of individual elements without a time component.

Time dependent (movie) - Continuous media

Information is expressed not only by its individual value but also by its time of occurrence.

1.5 Traditional Data Stream Characteristics

Distributed multimedia communication systems data of discrete and continuous media are broken into individual units (packets) and transmitted.

Data Stream

- Sequence of individual packets that are transmitted in a time-dependant fashion.
- Transmission of information carrying different media leads to data streams with varying features:
  - Asynchronous:-
  - Synchronous:-
  - Isochronous:-

Data Stream Characteristics

Asynchronous transmission mode: -

- Provides for communication with no time restriction
- Packets reach receiver as quickly as possible, e.g. protocols for email transmission

Synchronous transmission mode: -

- Defines a maximum end-to-end delay for each packet of a data stream.
- May require intermediate storage
- E.g. audio connection established over a network.

Isochronous transmission mode: -

- ❖ Defines a maximum and a minimum end-to-end delay for each packet of a data stream. Delay jitter of individual packets is bounded.
- ❖ E.g. transmission of video over a network.
- ❖ Intermediate storage requirements reduced.

Data Stream characteristics for continuous media can be based on:

The Time Intervals Between a Complete Transmission of Consecutive Packets:

- ➢ *Strongly periodic data streams* - constant time interval
- ➢ *Weakly periodic data streams* - periodic function with finite period.
- ➢ *Aperiodic data streams* - all other possibilities of transmission with respect to time interval.

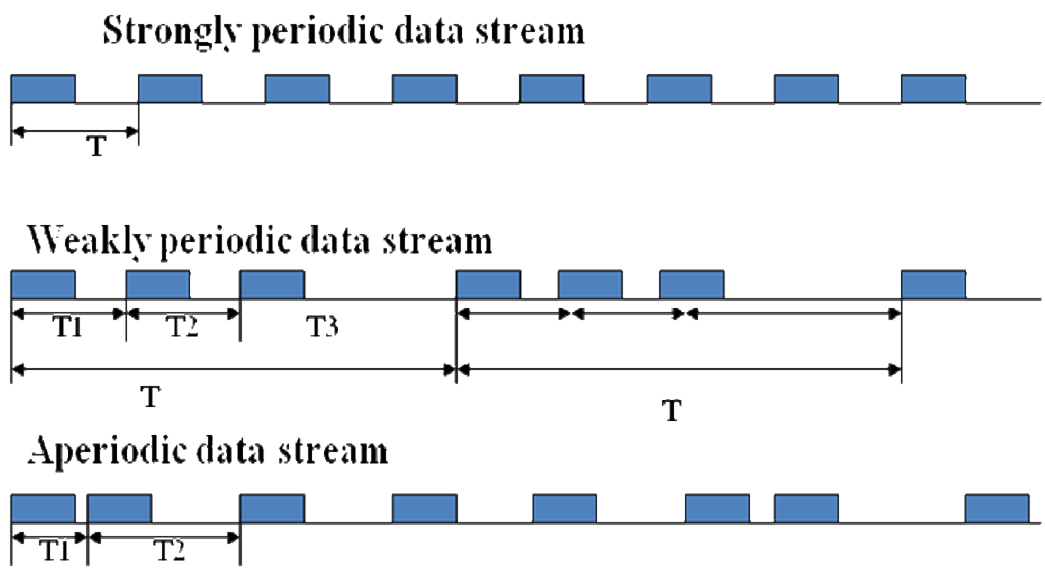


Figure: Classification based on time intervals

Data size - amount of consecutive packets:

- ➢ *Strongly regular data streams* - constant amount of data
- ➢ *Weakly regular data streams* - varies periodically with time
- ➢ *Irregular data streams* – the amount of data is neither constant nor changes according to a periodic function.

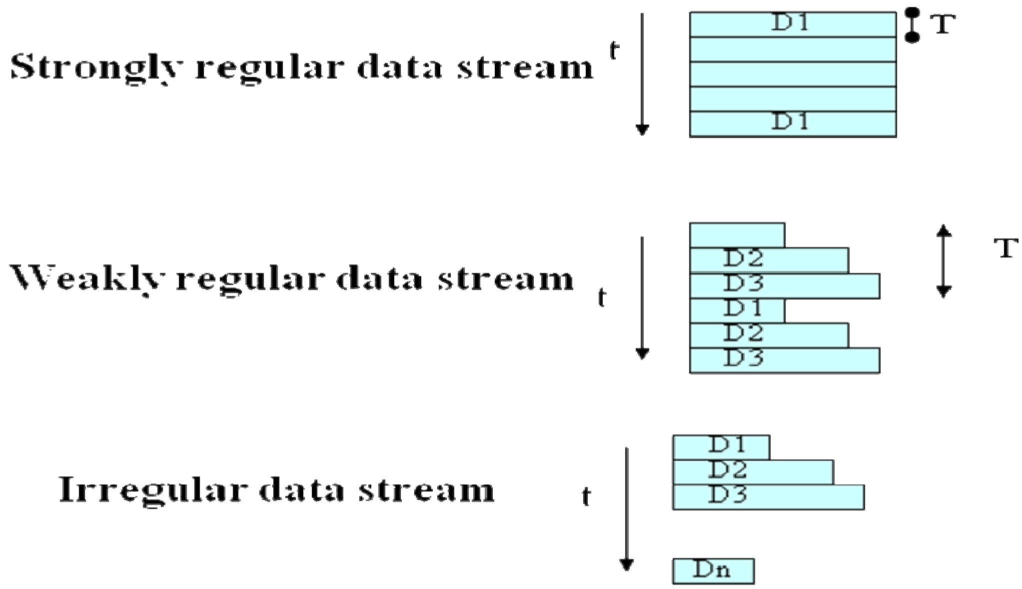


Figure: Classification based on packet size

Continuity (Contiguous Packets):

- *Continuous data streams* – the packets are transmitted without intermediate gaps.
- *Discrete data streams* – gaps exist among the packets.

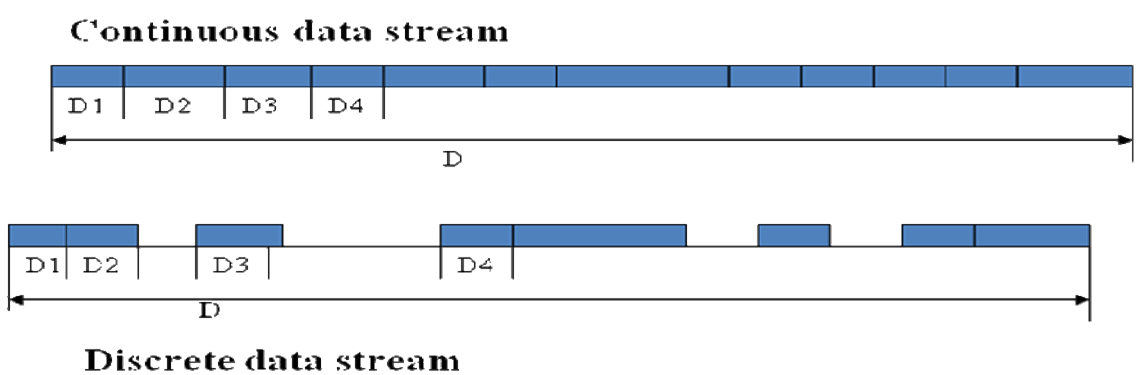


Figure: Classification based on continuity

1.6 Information Units

Continuous media consist of a time-dependent sequence of individual information units called Logical Data Units (LDU).

- a symphony consists of independent sentences
- a sentence consists of notes
- notes are sequences of samples

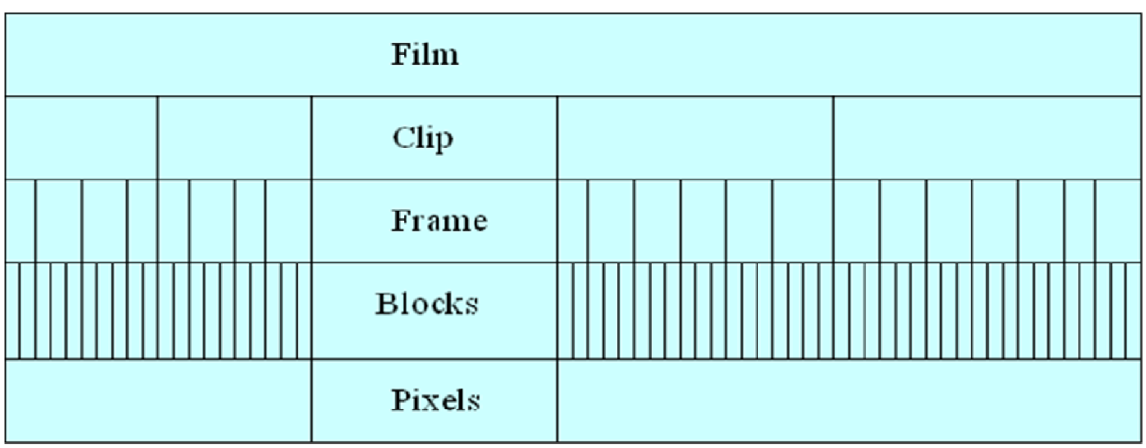
Granularity of LDUs

- symphony, sentence, individual notes, grouped samples, individual samples
- film, clip, frame, raster, pixel

Duration of LDU:

- Open LDU - duration not known in advance
- Closed LDU - predefined duration

Granularity of Logical Data Units



Applications of Multimedia

Multimedia finds its application in various areas including, but not limited to, advertisements, art, education, entertainment, engineering, medicine, mathematics, business, scientific research and spatial, temporal applications.
A few application areas of multimedia are listed below:

Creative industries

Creative industries use multimedia for a variety of purposes ranging from fine arts, to entertainment, to commercial art, to journalism, to media and software services provided for any of the industries listed below. An individual multimedia designer may cover the spectrum throughout their career. Request for their skills range from technical, to analytical and to creative.

Commercial

Much of the electronic old and new media utilized by commercial artists is multimedia. Exciting presentations are used to grab and keep attention in advertising. Industrial, business to business, and interoffice communications are often developed by creative services firms for advanced multimedia presentations beyond simple slide shows to sell ideas or liven-up training. Commercial multimedia developers may be hired to design for governmental services and nonprofit services applications as well.

Entertainment and Fine Arts

In addition, multimedia is heavily used in the entertainment industry, especially to develop special effects in movies and animations. Multimedia games are a popular pastime and are software programs available either as CD-ROMs or online. Some video games also use multimedia features.
Multimedia applications that allow users to actively participate instead of just sitting by as passive recipients of information are called *Interactive Multimedia*.

Education

In Education, multimedia is used to produce computer-based training courses (popularly called CBTs) and reference books like encyclopedia and almanacs (directory). A CBT lets the user go through a series of presentations, text about a particular topic, and associated illustrations in various information formats.

Edutainment is an informal term used to describe combining education with entertainment, especially multimedia entertainment.

Engineering

Software engineers may use multimedia in Computer Simulations for anything from entertainment to training such as military or industrial training.

Multimedia for software interfaces are often done as collaboration between creative professionals and software engineers.

Industry

In the Industrial sector, multimedia is used as a way to help present information to shareholders, superiors and coworkers. Multimedia is also helpful for providing employee training, advertising and selling products all over the world via virtually unlimited web-based technologies.

Mathematical and Scientific Research

In Mathematical and Scientific Research, multimedia is mainly used for modeling and simulation. For example, a scientist can look at a molecular model of a particular substance and manipulate it to arrive at a new substance.

Representative research can be found in journals such as the Journal of Multimedia.

Medicine

In Medicine, doctors can get trained by looking at a virtual surgery or they can simulate how the human body is affected by diseases spread by viruses and bacteria and then develop techniques to prevent it.

Multimedia in Public Places

In hotels, railway stations, shopping malls, museums, and grocery stores, multimedia will become available at stand-alone terminals or kiosks to provide information and help. Such installation reduce demand on traditional information booths and personnel, add value, and they can work around the clock, even in the middle of the night, when live help is off duty.

A menu screen from a supermarket kiosk that provide services ranging from meal planning to coupons. Hotel kiosk list nearby restaurant, maps of the city, airline schedules, and provide guest services such as automated checkout.

Printers are often attached so users can walk away with a printed copy of the information. Museum kiosk are not only used to guide patrons through the exhibits, but when installed at each exhibit, provide great added depth, allowing visitors to browser though richly detailed information specific to that display.

Chapter 2
Sound and Audio

2.1 Basic Sound Concept

Sound is perhaps the most important element of multimedia. It is meaningful "speech" in any language, from a whisper to a scream. It can provide the listening pleasure of music, the startling accent of special effects or the ambience of a mood setting background. Sound is the terminology used in the analog form, and the digitized form of sound is called as audio. *It is basically a pattern formed in the vibration or movement of molecules of air.* When a sound is made, air molecules move out from the source in waves.

Sound is a physical phenomenon produced by the vibration of matter. The matter can be almost anything: a 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 and low pressure is propagated through the air in a wave-like motion. When the wave reaches our ears, we hear a sound.

- *Sound is a continuous wave that travels through the air.*
- *The wave is made up of pressure differences.*
- *Sound is detected by measuring the pressure level at a location.*
- *Sound waves have normal wave properties (reflection, refraction, diffraction etc.)*

Sound Transmission

- Sound is transmitted by molecules bumping into each other.
- Sound is a continuous wave that travels through air.
  Sound is detected by measuring the pressure level at a point.

Receiving

- Microphone in sound field moves according to the varying pressure exerted on it.
- Transducer converts energy into a voltage level (i.e. energy of another form - electrical energy)

Sending

- Speaker transforms electrical energy into sound waves.

Frequency of a sound wave

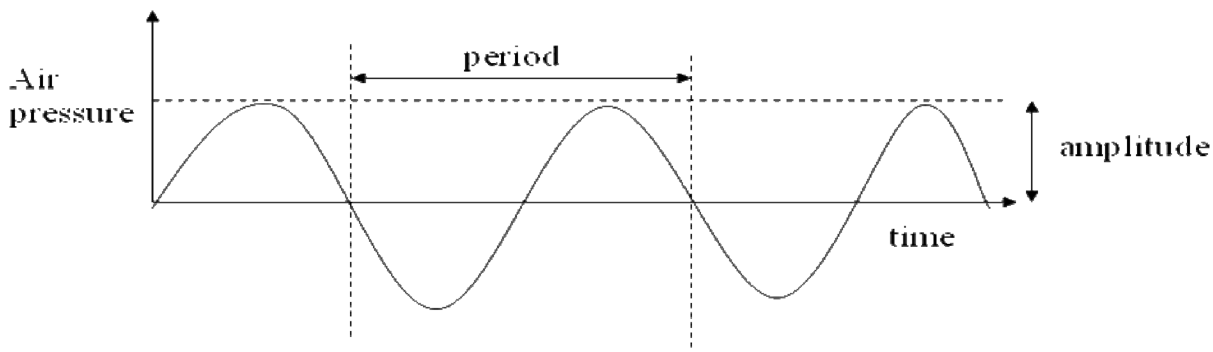


Figure 2.1: Oscillation of an air pressure wave.

The pattern of the pressure oscillation is called a *waveform*. The waveform in repeats the same shape at regular intervals and this portion of the waveform is called a *period*. A waveform with a clearly defined period occurring at regular 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 figure 2.1. 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

Frequency

The *frequency* of a sound is the reciprocal value of the period; it represents the number of times the pressure rises and falls, or oscillates, in a second and is measured in *hertz* (Hz) or *cycles per second (cps)*. 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. Represents the number of periods in a second and is measured in hertz (Hz) or cycles per second.

Wavelength is the distance travelled in one cycle

- 20Hz is 56 feet, 20KHz is 0.7 in.

Frequency represents the number of periods in a second (measured in hertz, cycles/second).

Human hearing frequency range: 20Hz - 20Khz, voice is about 500Hz to 2Khz.

The frequency range is divided into:

Infrasound	from 0 - 20 Hz
Human range	from 20Hz - 20KHz
Ultrasound	from 20kHz - 1GHz
Hypersound	from 1GHz - 10THz

Amplitude

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.

- Amplitude of a sound is the measure of the displacement of the air pressure wave from its mean or quiescent state.
- Subjectively heard as loudness. Measured in decibels.

0 db -	essentially no sound heard
35 db -	quiet home
70 db -	noisy street
120db -	discomfort

Computer Representation of Audio

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*. Figure illustrates one period of a digitally sampled waveform.

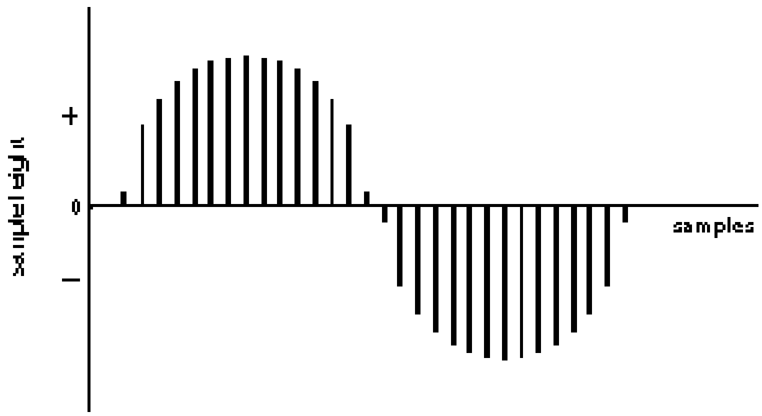


Figure 2.2: Sampled Waveform

Each vertical bar in Figure 2-2 represents a single sample. The height of a bar indicates the value of that sample.

The mechanism that converts an audio signal into digital samples is called an *analog-to-digital converter*, or *ADC*. To convert a digital signal back to analog, you need a *digital-to-analog converter*, or *DAC*.

- ❖ A transducer converts pressure to voltage levels.
- ❖ Convert analog signal into a digital stream by discrete sampling.
    - Discretization both in time and amplitude (quantization).
- ❖ In a computer, we sample these values at intervals to get a vector of values.
- ❖ A computer measures the amplitude of the waveform at regular time intervals to produce a series of numbers (samples).

Sampling Rate

The rate at which a waveform is sampled is called the *sampling rate*. Like frequencies, sampling rates are measured in hertz. 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.

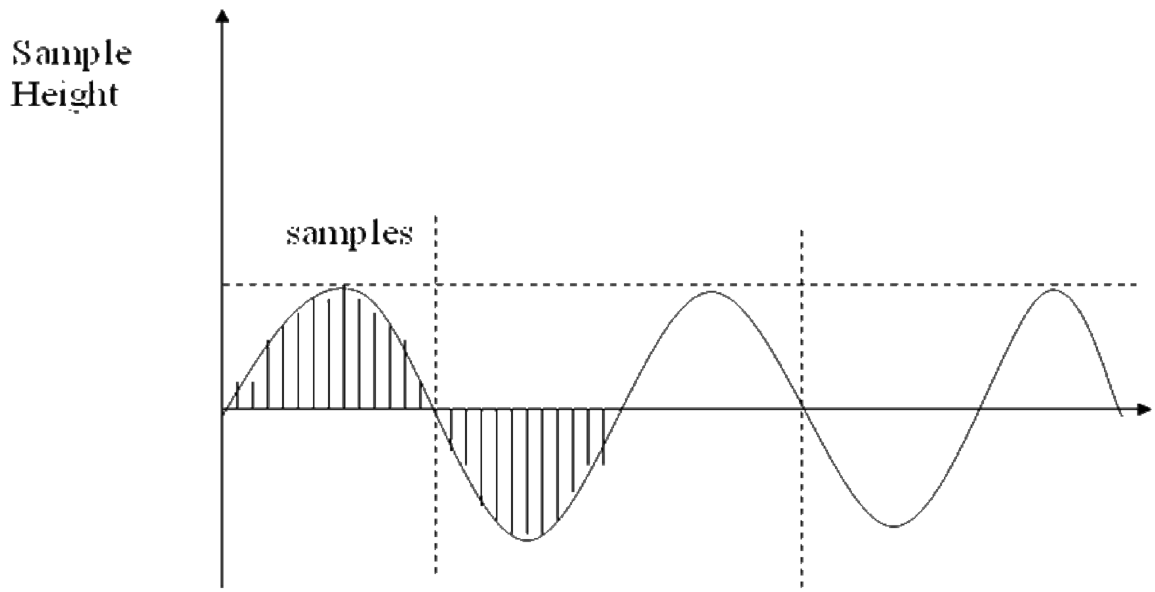


Figure 2.3: Sampling

- Rate at which a continuous wave is sampled (measured in Hertz)
    - CD standard - 44100 Hz, Telephone quality - 8000 Hz.
- Direct relationship between sampling rate, sound quality (fidelity) and storage space.

Quantization

Just as 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. An 8-bit quantization yields 256 possible values; 16-bit CD quality quantization results in over 65000 values. As an extreme example, Figure 2-3 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.

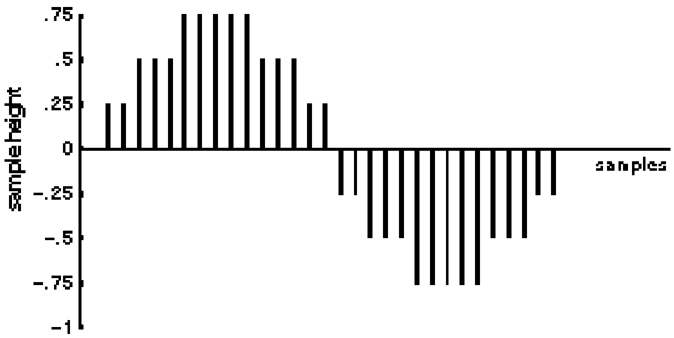


Figure 2.4: Three - bit quantization

Nyquist Sampling Theorem

If a signal f(t) is sampled at regular intervals of time and at a rate higher than twice the highest significant signal frequency, then the samples contain all the information of the original signal.

Example**:**

- ❖ Actual playback frequency for CD quality audio is 22050 Hz
- ❖ Because of Nyquist Theorem - we need to sample the signal twice, therefore sampling frequency is 44100 Hz.

Data Rate of a Channel

Noiseless Channel

Nyquist proved that if any arbitrary signal has been run through a low pass filter of bandwidth H, the filtered signal can be completely reconstructed by making only 2H (exact) samples per second. If the signal consists of V discrete levels, Nyquist's theorem states:

*Max datarate = 2 *H log_2 V bits/sec*

- Noiseless 3kHz channel with quantization level 1 bit cannot transmit binary signal at a rate exceeding 6000 bits per second.

Noisy Channel

Thermal noise present is measured by the ratio of the signal power S to the noise power N (signal-to-noise ratio S/N).

Max datarate - H log_2 (1+S/N)

Audio Formats

Audio formats are characterized by four parameters:

- *Sample rate:* Sampling frequency
- *Encoding:* audio data representation
  - ✓ µ-law encoding corresponds to CCITT G.711 - standard for voice data in telephone companies in USA, Canada, Japan
  - ✓ A-law encoding - used for telephony elsewhere.
  - ✓ A-law and µ-law are sampled at 8000 samples/second with precision of 12bits, compressed to 8-bit samples.
  - ✓ Linear Pulse Code Modulation(PCM) - uncompressed audio where samples are proportional to audio signal voltage.
- *Precision:* number of bits used to store audio sample
  - ✓ µ-law and A-law - 8 bit precision, PCM can be stored at various precisions, 16 bit PCM is common.
- *Channel:* Multiple channels of audio may be interleaved at sample boundaries.

- *Available on UNIX*
  - au (SUN file format), wav (Microsoft RIFF/waveform format), al (raw a-law), u (raw u-law)…
- *Available on Windows-based systems (RIFF formats)*
  - wav, midi (file format for standard MIDI files), avi
- *RIFF (Resource Interchange File Format)*
  - tagged file format (similar to TIFF).. Allows multiple applications to read files in RIFF format
- *RealAudio, MP3 (MPEG Audio Layer 3)*

Digital Audio

Digital audio is created when a sound wave is converted into numbers – a process referred to as digitizing. It is possible to digitize sound from a microphone, a synthesizer,

existing tape recordings, live radio and television broadcasts, and popular CDs. You can digitize sounds from a natural source or prerecorded.
Digitized sound is sampled sound.

2.2 Basic Music (MIDI) Concepts

The relationship between music and computer has become more and more important, especially considering the development of MIDI (Musical Instrument Digital Interface) and its important contributions in the music industry today.
MIDI (Musical Instrument Digital Interface) is a communication standard developed for electronic musical instruments and computers. MIDI files allow music and sound synthesizers from different manufacturers to communicate with each other by sending messages along cables connected to the devices.

- *MIDI doesn't directly describe musical sound*
- *MIDI is not a language*
- *It is a data communications protocol*

Creating your own original score can be one of the most creative and rewarding aspects of building a multimedia project, and MIDI (Musical Instrument Digital Interface) is the quickest, easiest and most flexible tool for this task.
The process of creating MIDI music is quite different from digitizing existing audio. To make MIDI scores, however you will need sequencer software and a sound synthesizer.
The MIDI keyboard is also useful to simply the creation of musical scores. An advantage of structured data such as MIDI is the ease with which the music director can edit the data.
A MIDI file format is used in the following circumstances:

- Digital audio will not work due to memory constraints and more processing power requirements
- When there is high quality of MIDI source
- When there is no requirement for dialogue.

Computer Representation of Music

MIDI (Music Instrument Digital Interface) is a standard that manufacturers of musical instruments use so that instruments can communicate musical information via computers.

The MIDI interface consists of:

- *Hardware* - physical connection b/w instruments, specifies a MIDI port (plugs into computers serial port) and a MIDI cable.
- *Data format* - has instrument specification, notion of beginning and end of note, frequency and sound volume. Data grouped into MIDI messages that specify a musical event.
- *An instrument that satisfies both is a MIDI device (*e.g. synthesizer)
- *MIDI software applications* include a music recording and performance applications, musical notations and printing applications, music education etc.

MIDI Reception Modes:

- *Mode 1: Omni On/Poly - usually for testing devices*
- *Mode 2: Omni On/Mono - has little purpose*
- *Mode 3: Omni Off/Poly - for general purpose*
- *Mode 4: Omni Off/Mono - for general purpose*

Omni On/Off: respond to all messages regardless of their channel

Poly/Mono: respond to multiple/single notes per channel

The first half of the mode name specifies how the MIDI device monitors the incoming MIDI channels.

If Omni is turned on, the MIDI device monitors all the MIDI channels and responds to all channel messages, no matter which channel they are transmitted on.

If Omni is turned off, the MIDI device responds only to channel messages sent on the channel(s) the device is set to receive.

The second half of the mode name tells the MIDI device how to play notes coming in over the MIDI cable.

If the option Poly is set, the device can play several notes at a time.

If the mode is set to Mono, the device plays notes like a monophonic synthesizer-one note at a time.

MIDI Devices

The heart of any MIDI system is the MIDI synthesizer device. Most synthesizer have the following common components:

- *Sound Generators:*

  Sound generators do the actual work of synthesizing sound; the purpose of the rest of the synthesizer is to control the sound generators.
- *Microprocessors:*

  The microprocessor communicates with the keyboard to know what notes the musician is playing, and with the control panel to know what commands the musician wants to send to microprocessor. The microprocessor send and receives MIDI message.
- *Keyboard:*

  The keyboard affords the musician's direct control of the synthesizer.
- *Control Panel:*

  The control panel controls those functions that are not directly concerned with notes and durations (controlled by the keyboard).
- *Auxiliary Controllers:*

  Auxiliary controllers are available to give more control over the notes played on the keyboard. Two very common variables on a synthesizer are pitch bend and modulation.
- *Memory:*

  Synthesizer memory is used to store patches for the sound generators and settings on the control panel.

MIDI Messages

MIDI messages transmit information between MIDI devices and determine what kinds of musical events can be passed from device to device.

MIDI messages are divided into two different types:

(1) Channel Messages:

Channel messages go only to specified devices. There are two types of channel messages:

(i) *Channel Voice Messages:*

Send actual performance data between MIDI devices. Example: Note On, Note Off, Channel Pressure, Control Change etc.

(ii) Channel Mode Messages:

Determine the way that a receiving MIDI device responds to channel voice messages. Example: Local Control, All Notes Off, Omni Mode Off etc.

(2) System Messages:

System messages go to all devices in a MIDI system because no channel numbers are specified. There are three types of system messages:

(i) System Real-time Messages:

System real time messages are very short and simple, consisting of only one byte. They carry extra data with them. Example: System Reset, Timing Clock etc.

(ii) System Common Messages:

System common messages are commands that prepare sequencers and synthesizers to play a song. Example: Song Select, Tune Request etc.

(iii) System Exclusive Messages:

System exclusive messages allow MIDI manufacturers to create customized MIDI messages to send between their MIDI devices.

MIDI Software

The software applications generally fall into four major categories:

(i) Music recording and performance applications:

- Provides functions such as recording of MIDI messages
- Editing and playing back the messages in performance.

(ii) Musical notation and printing applications:

- Allows writing music using traditional musical notation.

- Print the music on paper for live performance or publication.

(iii) *Synthesizer patch editors and librarians:*

- Allow information storage of different synthesizer patches in the computer's memory and disk drives.
- Editing of patches in the computer.

(iv) *Music education applications:*

- Teach different aspects of music using the computer monitor, keyboard and other controllers of attached MIDI instruments.

The main issue in current MIDI-based computer music systems is interactivity.

The processing chain of interactive computer music systems can be conceptualized in three stages:

- ❖ *The sensing stage,* when data are collected from controller reading gesture information from human performers on stages.
- ❖ *The processing stage,* when the computer reads and interprets information coming from the sensors and prepares data for the response stage.
- ❖ *The response stage,* when the computer and some collection of sound-producing devices share in realizing a musical output.

2.3 Speech

Speech can be "perceived", "understood" and "generated" by humans and also by machines. A human adjusts himself/herself very efficiently to different speakers and their speech habits.

The brain can recognize the very fine line between speech and noise.

The human speech signal comprises a subjective lowest spectral component known as the pitch, which is not proportional to frequency.

The human ear is most sensitive in the range from 600 Hz to 6000 Hz.

Speech signal have two properties which can be used in speech processing:

- ➢ *Voice speech signals* show during certain time intervals almost periodic behavior.

- *The spectrum of audio signals* shows characteristic maxima, which are mostly 3-5 frequency bands.

Speech Generation

Generated speech must be understandable and must sound natural. The requirement of understandable speech is a fundamental assumption, and the natural sound of speech increases user acceptance.

Basic Notions

- The lowest periodic spectral component of the speech signal is called the *fundamental frequency*. It is present in a voiced sound.
- A *phone* is the smallest speech unit, such as the *m* of *mat* and *b* of *bat* in English, that distinguish one utterance or word from another in a given language.
- *Allophones* mark the variants of a phone. For example, the aspirated *p* of *pit* and the unaspirated *p* of *spit* are allophones of the English phoneme *p*.
- The *morph* marks the smallest speech unit which carries a meaning itself. Therefore, *consider* is a morph, but *reconsideration* is not.
- A *voiced sound* is generated through the vocal cords. *m,v* and *l* are examples of voiced sounds. The pronunciation of a voiced sound depends strongly on each speaker.
- During the generation of an *unvoiced sound*, the vocal cords are opened. *f* and *s* are unvoiced sounds. Unvoiced sounds are relatively independent from the speaker.

Exactly, there are:

- *Vowels* - a speech sound created by the relatively free passage of breath through the larynx and oral cavity, usually forming the most prominent and central sound of a syllable (e.g., *u* from hunt);
- *Consonants* - a speech sound produced by a partial or complete obstruction of the air stream by any of the various constrictions of the speech organs (e.g.,

voiced consonants, such as *m* from *mother*, fricative voiced consonants, such as *v* from *voice*, fricative voiceless consonants, such as *s* from *nurse*, plosive consonants, such as *d* from daily and affricate consonants, such as *dg* from knowledge, or *ch* from *chew*).

Reproduced Speech Output

There are two way of speech generation/output performed by *time-dependent sound concatenation* and a *frequency-dependent sound concatenation*.

Time-dependent Sound Concatenation

Individual speech units are composed like building blocks, where the composition can occur at different levels. The individual phones are understood as speech units. The individual phones of the word *curmb*. It is possible with just a few phones to create an unlimited vocabulary.

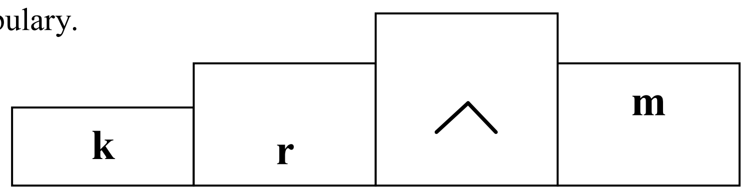


Figure 2.5: Phone sound concatenation.

Two phones can constitute a diphone (from di-phone). Figure 2.6 shows the word *crumb*, which consist of an ordered se of diphones.

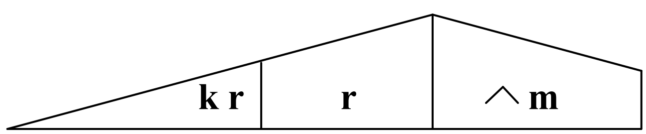


Figure 2.6: Diphone sound concatenation.

To make the transition problem easier, syllables can be created. The speech is generated through the set of syllables. Figure 2.7 shows the syllable sound of the word *crumb.*

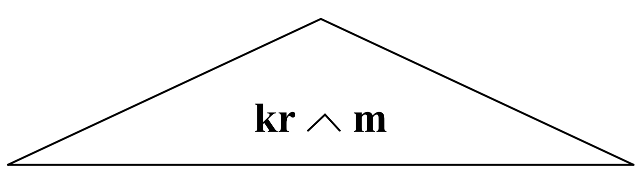


Figure 2.7: Syllable sound.

The best pronunciation of a word is achieved through storage of the whole word. This leads towards synthesis of the speech sequence (Figure 2.8).

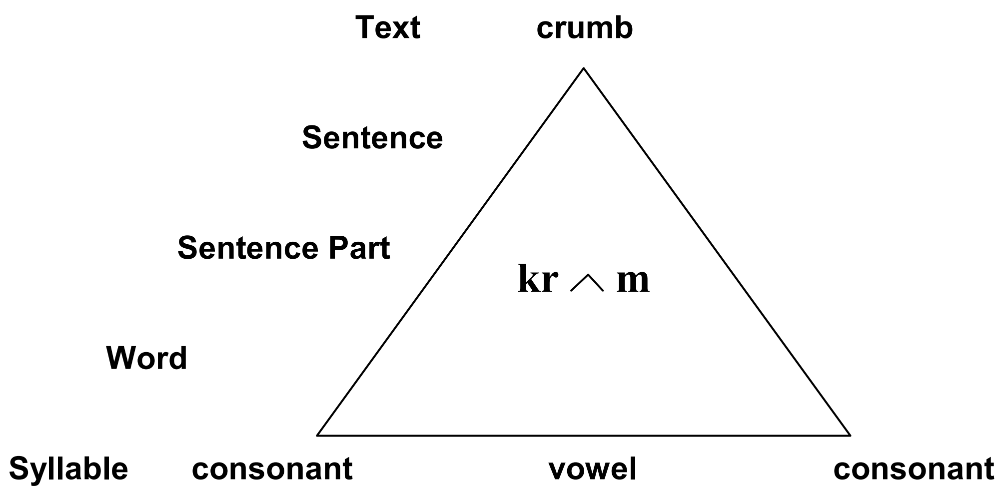


Figure 2.8 Word sound concatenation.

Frequency-dependent Sound Concatenation

Speech generation/output can also be based on a frequency-dependent sound concatenation, e.g. through a formant-synthesis. Formants are frequency maxima in the spectrum of the speech signal. Formant synthesis simulates the vocal tract through a filter.

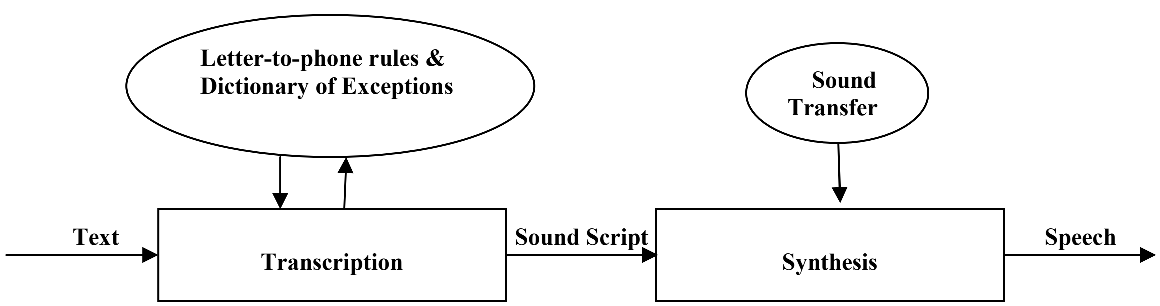


Figure 2.9: Components of a speech synthesis system with time-dependent sound concatenation.

In the first step, transcription is performed, in which text is translated into sound script. Most transcription methods work here with letter-to-phone rules and a Dictionary of Exceptions stored in a library. The generation of such a library in work-extensive, but using the interactive control the user it can be improved continuously. The user recognizes the formula deficiency in the transcription and improves the pronunciation manual.

In the second step, the sound script is translated into a speech signal. Time or frequency-dependent concatenation can follow. While the first step is always a software solution, the second step is most often implemented with signal processors or even dedicated processors.

Speech Analysis

Speech analysis/input deals with the research areas shown in Figure 2.10:

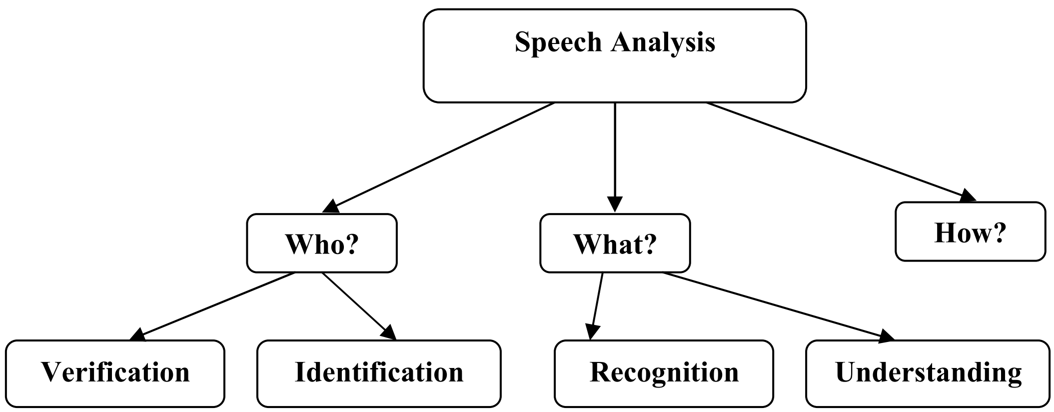


Figure 2.10: Research areas of speech analysis.

❖ Human speech has certain characteristics determined by a speaker. Hence, speech analysis can serve to analyze *who* is speaking, i.e. *to recognize a speaker* for his/her *identification* and *verification*.
  - The computer identifies and verifies the speaker using an acoustic fingerprint.
  - An acoustic fingerprint is a digitally stored speech *probe* (e.g., certain statement) of a person.

❖ Another task of speech analysis is to analyze what has been said, i.e., to recognize and understand the speech signal itself. Based on speech sequence, the corresponding text is generated. This can lead to a speech-controlled typewriter, a translation system or part of a workplace for the handicapped.

- Another area of speech analysis tries to research speech patterns with respect to *how* a certain statement was said. For example, a spoken sentence sounds differently if a person is *angry* or *calm*.

Speech Recognition

Speech recognition is the process of converting an acoustic signal, captured by a microphone or a telephone, to set of words.

Speech recognition is the process of converting spoken language to written text or some similar form.

Speech recognition is the foundation of human computer interaction using speech.

Speech recognition in different contexts:

- Dependent or independent on the speaker.
- Discrete words or continuous speech.
- Small vocabulary or large vocabulary.
- In quiet environment or noisy environment.

Natural Language Understanding (NLU) is a process of analysis of recognized words and transforming them into data meaningful to computer. Other words, NLU is a computer based system that "understands" human language. NLU is used in combination with speech recognition.

Basic Terms and Concepts

- *Utterance* is any stream of speech between two periods of silence.
- *Pronunciation* is what the speech engine thinks a word should sound like.
- *Grammars* define a domain (of words) within which recognition engine works.
- *Vocabulary (dictionary)* a list of words (utterances) that can be recognized by the speech recognition engine.
- *Training* is the process of adapting the recognition system to a speaker.
- *Accuracy* is the measure of recognizer's ability to correctly recognize utterances.
- *Speaker Dependence*
    - *Speaker dependent system* is designed for only one user (at the time).
    - *Speaker independent system* is designed for variety of speakers.

Types of Speech Recognition

Speech recognizers are divided into several different classes according to the type of utterance that they can to recognize:

- Isolated words,
- Connected words,
- Continuous speech (computer notation)
- Spontaneous (natural) speech
- Voice Verification
- Voice Identification

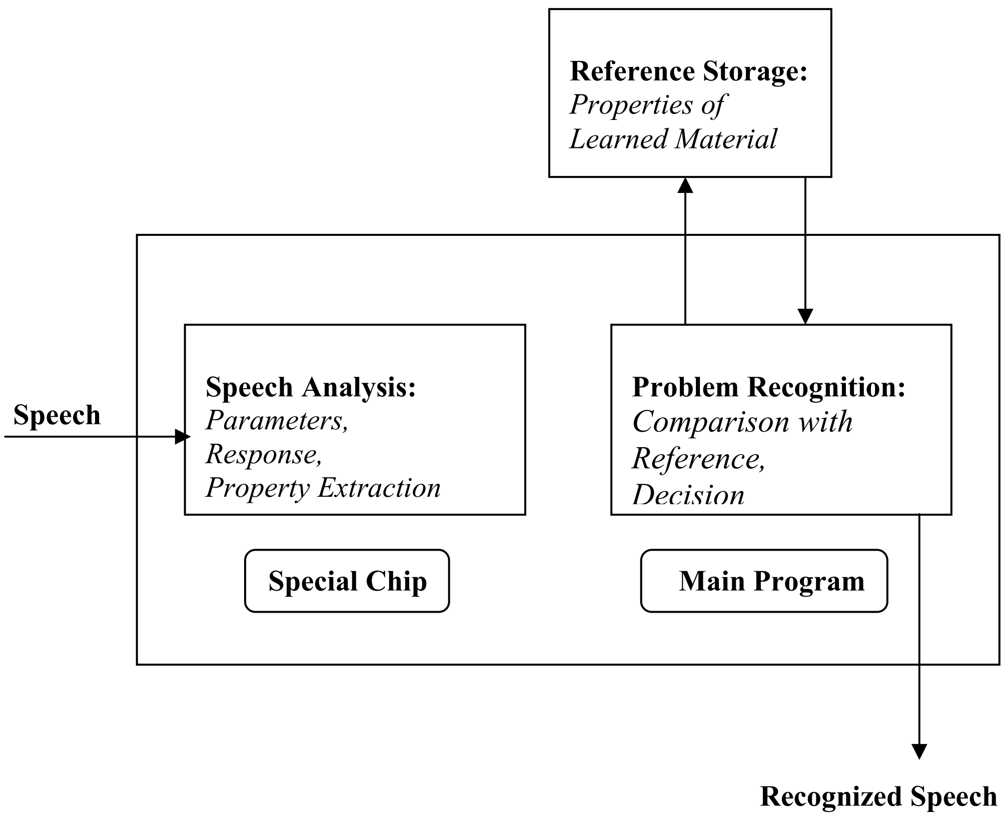


Figure 2.11: Speech recognition system: task division into system components, using the basic principle "Data Reduction Through Property Extraction."

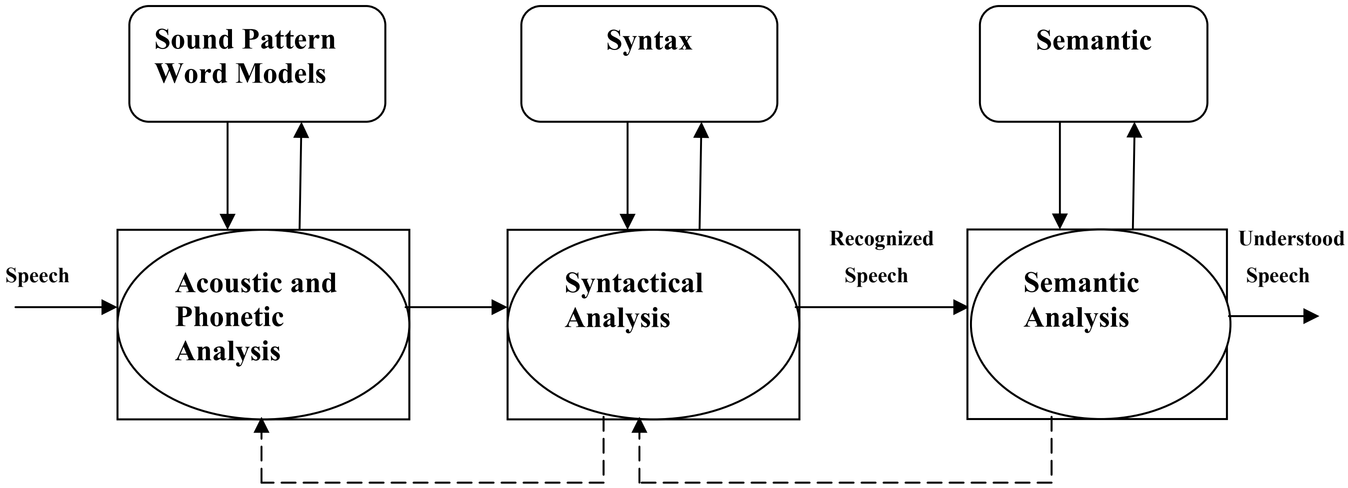


Figure 2.12: Components of speech recognition and understanding.

The first step, the principle is applied to a sound pattern and /or word model. An acoustical and phonetical analysis is performed.

The second step, certain speech units go through syntactical analysis; thereby, the errors of the previous step can be recognized. Very often during the first step, no unambiguous decisions can be made. In this case, syntactical analysis provides additional decision help and the result is a *recognized speech*.

The third step deals with the semantics of the previously recognized language. Here the decision errors of the previous step can be recognized and corrected with other analysis methods. The result of this step is an understood speech.

Chapter 3
Images and Graphics

3.1 Basic Images Concept

An image is a spatial representation of an object, a two dimensional or three dimensional scenes or another image.

Abstractly, an image is a continuous function defining a rectangular region of a plane.

- *Intensity image:* proportional to radiant energy received by a sensor/detector.
- *Range image:* line of sight distance from sensor position.

An image can be thought of as a function with resulting values of the light intensity at each point over a planar region.

Digital Image Representation

For computer representation, function (e.g. intensity) must be sampled at discrete intervals.

Sampling quantizes the intensity values into discrete intervals.

- Point at which an image is sampled are called picture elements or pixels.
- Resolution specifies the distance between points accuracy.

A digital image is represented by a matrix of numeric values each representing a quantized intensity value.

A digital image is a numeric representation (normal binary) of two dimensional images.

When I is a two-dimensional matrix, then $I\ (r,\ c)$ is the intensity value at the position corresponding to row r and column c of the matrix.

Intensity value can be represented by bits for black and white images (binary valued images), 8 bits for monochrome imagery to encode color or grayscale levels, 24 bit (RGB).

Image Formats

There are different kinds of image formats in the literature. We shall consider the image format that comes out of an image frame grabber, i.e., the captured image format, and the format when images are stored, i.e., the stored image format.

(i) Captured Image Format

(ii) Stored Image Format

Captured Image Format: -

The image format is specified by two main parameters: spatial resolution, which is specified as pixels x pixels (e.g. 640x480) and color encoding, which is specified by bits per pixel. Both parameter values depend on hardware and software for input/output of images.

Stored Image Format: -

When we store an image, we are storing a two-dimensional array of values, in which each value represents the data associated with a pixel in the image. For a bitmap, this value is a binary digit.

A *bitmap* is a simple information matrix describing the individual dots that are the smallest elements of resolution on a computer screen or other display or printing device.

Image file format include:

- GIF (Graphic Interchange Format)
- X11 bitmap
- Postscript
- JPEG (Joint Picture Expert Group)
- TIFF (Tagged Image File Format) etc.

There are many file formats used to store bitmaps and vectored drawing. Following is a list of few image file formats.

Format	Extension
Microsoft Windows DIB	.bmp .dib .rle
Microsoft Palette	.pal
Autocad format 2D	.dxf
JPEG	.jpg
Windows Meta file	.wmf
Portable network graphic	.png
Compuserve gif	.gif
Apple Macintosh	.pict .pic .pct

Graphics Format

Graphic image formats are specified through graphics primitives and their attributes.

- *Graphic primitive* – line, rectangle, circle, ellipses, specification 2D and 3D objects.
- *Graphic attribute* – line style, line width, color.

Graphics formats represent a higher level of image representation, i.e., they are not represented by a pixel matrix initially.

- *PHIGS (Programmer's Hierarchical Interactive Graphics)*
- *GKS (Graphical Kernel System)*

3.2 Computer Image Processing

Image processing is any form of signal processing for which the input is an image, such as a photograph or video frame; the output of image processing may be either an image or a set of characteristics or parameters related to the image.

Image processing usually refers to digital image processing, but optional and analog image processing also are possible.

Computer graphics concern the pictorial synthesis of real or imaginary objects from their computer-based models.

The related field of image processing treats the converse process: the analysis of scenes, or the reconstruction of model from pictures of 2D or 3D objects.

Image Synthesis

Image synthesis is an integral part of all computer user interfaces is indispensable for visualizing 2D, 3D and higher dimensional objects. Areas as diverse as education, science, engineering, medicine, advertising and entertainment all rely on graphics.

Dynamic in Graphics

Graphics are not confined to static pictures. Picture can be dynamically varied; for example, a user can control animation by adjusting the speed, portion of the total scene inn view, amount of detail shown, etc.

Motion Dynamic:

With motion dynamic, objects can be moved and enabled with respect to a stationary observer.

Update Dynamic:

Update dynamic is the actual change of the shape, color, or other properties of the objects being viewed.

The Framework of Interactive Graphics System

Image can be generated by video digitizer cards that capture NTSC (PAL) analog signals and create a digital image.

Graphical images are generated using interactive graphics systems.

The high-level conceptual framework of almost any interactive graphics system consists of three software components: *an application model, an application program and a graphics system*, and a hardware component: graphics hardware.

Application Model:

The application model represents the data or objects to be picture on the screen; it is stored in an application database. The model is an application-specific and is created independency of any particular display system.

Application Program:

The application program handles user input. It produces views by sending to the third component, the graphics system, a series of graphics output commands that contain both a detailed geometric description of what is to be viewed and the attributes describing how the objects should appear.

Graphics System:

The graphics system is responsible for actually producing the picture from the detailed descriptions and for passing the user's input to the application program for processing.

The graphics system is an intermediary component between the application program and the display hardware.

Graphics Hardware:

At the hardware level, a computer receives input from interaction devices and output images to display devices.

Input:

Current input technology provide us with the ubiquitous mouse, the data tablet and transparent, touch sensitive panel mounted on the screen.

The other graphics input are track-balls, space-balls or the data glove.

Track-ball can be made to sense rotation about the vertical axis in addition to the about two horizontal axes.

A space-ball is a rigid sphere containing strain gauges. The user pushes or pulls the sphere in any direction, providing 3D translation and orientation.

The data glove records hand position and orientation as well as finger movements. It is a glove covered with small, lightweight sensors.

Output: Raster Display

- ❖ Most common type of graphic monitors using raster scan display type CRT
- ❖ Point plotting device
- ❖ Based on TV technology
- ❖ Electron beam is swept across the screen, one row at a time from top to bottom, starting at the upper left corner of the display
- ❖ Process is repeated until the entire screen is covered, and the beam is then returned to the upper left corner to start a new scan
- ❖ Beam intensity is turned ***on*** and ***off*** to create a pattern of illuminated spots
- ❖ Pictures are dynamically stored in a piece of memory known as ***frame buffer*** or ***refresh buffer***
- ❖ This buffer holds the set of intensity values all the screen points (pixels)
- ❖ Requirement to control the intensity of the screen positions:
    - ➢ Simple black and white system:

- 1 bit per pixel (bitmap)

➢ Color system:

- 24 bits/pixel (maximum no. of color representation, pixmap)

❖ ***Frame buffer*** or ***refresh buffer*** (storage) requirements:

➢ Large storage

e.g. 24 bits/pel, screen resolution of 1024x1024 requires 3mb of RAM

➢ Refresh rate: 60 to 80 frames per second

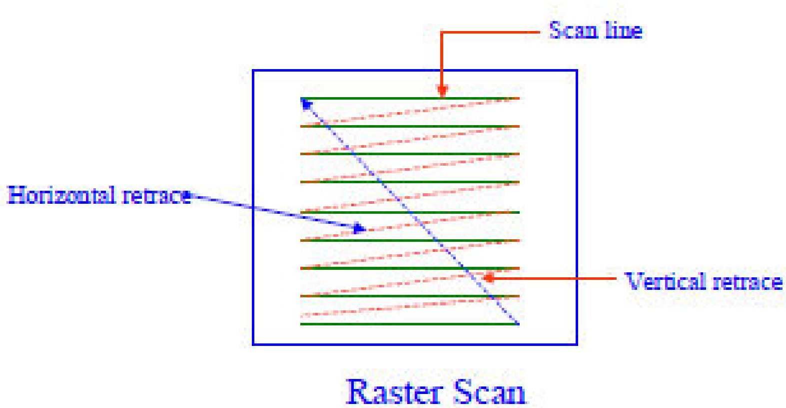


Figure 3.1 Raster Scan

Advantages of Raster Scan Display:

❖ Capable of presenting bright pictures

❖ Unaffected by picture complexity

❖ Suitable for showing dynamic motion

❖ Lower cost

❖ Ability to display areas filled with solid colors or patterns

Disadvantages of Raster Scan Display:

❖ Requires large amount of memory (RAM)

❖ Produced "stair stepped" appearance of diagonal lines on the image (known as ***aliasing*** effect)

❖ True line cannot be represented exactly due to the discretization of the display surface (discrete nature of pixel representation)

Image Analysis

Image analysis is concerned with techniques for extracting descriptions from images that are necessary for high-level scene analysis methods.

Image analysis techniques include computation of perceived brightness and color, partial or complete recovery of three-dimensional data in the scene, location of discontinuities corresponding to objects in the scene and characterization of the properties of uniform regions in the image.

Image processing includes image enhancement, pattern detection and recognition and scene analysis and computer vision.

Image enhancement deals with improving image quality by eliminating noise or by enhancing contrast.

Pattern detection and recognition deal with detecting and clarifying standard patterns and finding distortions from these patterns.

Scene analysis and computer vision deal with recognizing and reconstructing 3D models of a scene from several 2D images.

Image Recognition Steps

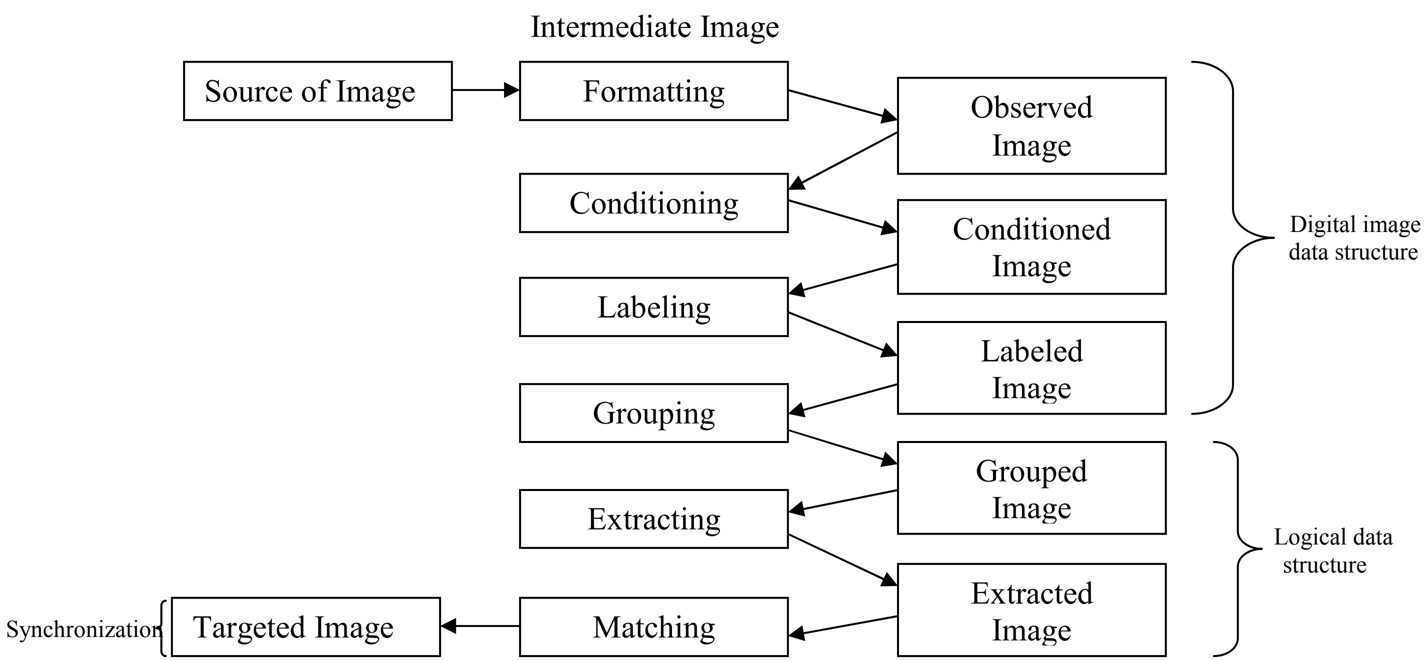


Figure 3.2 Image Recognition Steps

Formatting

Capturing an image from a camera and bringing it into a digital form. It means that we will have a digital representation of an image in the form of pixels.

Conditioning

In image, there is usually uninteresting object introduced during digitize as noise. In conditioning, interesting objects are highlighted by suppressing or analyzing uninteresting in systematic or patterned variations. Conditioning is typically applied uniformly and is context-independent.

Labeling

The informative pattern has structure as a spatial arrangement of events, each spatial event being a set of connected pixels. Labeling determines in what kinds of spatial events each pixel participates.

E.g. edge detection technique

Edge detection technique determines continuous adjacent pairs which differ in intensity or color. Another labeling operation must occur after edge detection, namely thresholding.

Thresholding specifies which edges should be accepted and which should not; the thresholding operation filters only the significant edges from the image and labels them.

Grouping

It can turn edges into line by determining edges belongs to same spatial event. A grouping operation, where edges are grouped into lines, is called line filtering. The grouping operation involves a change of logical data structure.

Extracting

Generating list of properties from set of pixel in spatial event. Extraction can also measure topological or spatial relationship between two or more grouping.

Matching

After the completion of the extracting operation, the events occurring on the image have been identified and measured but the events in and of themselves have no meaning.

It is the matching operation that determines the interpretation of some related set of image events, associating these events with some given three dimensional object or two-dimensional shape.

The classic example is template matching, which compares the examined pattern with stored models (templates) of known patterns and chooses the best match.

Image Transmission

Image transmission takes into account transmission of digital images through computer networks. There are several requirements on the networks when images are transmitted:

- ❖ The network must accommodate bursty data transport because image transmission is bursty (The burst is caused by the large size of the image).
- ❖ Image transmission requires reliable transport.
- ❖ Time-dependence is not a dominant characteristic of the image in contrast to audio/video transmission.

Image size depends on the image representation format used for transmission. There are several possibilities:

Raw Image Data Transmission

The image is generated through a video digitizer and transmitted in its digital format.

$$Size = Spatial\_resolution \times Pixel\_quantization$$

For example, the transmission of an image with a resolution of 640 x 480 pixels and pixel quantization of 8 bit per pixel requires transmission of 307,200 bytes through the network.

Compressed Image Data Transmission

The image is generated through a video digitizer and compressed before transmission. The reduction of image size depends on the compression method and compression rate.

JPEG (Joint Picture Expert Group) & MPEG (Motion Picture Expert Group)

Symbolic Image Data Transmission

The image is represented through symbolic data representation as image primitives (e.g. 2D or 3D geometric representation), attributes and other control information.

3.3 Image Enhancement

Enhancement is the process an image so that the result is more suitable than the original image for a specific application.

Enhancement approaches:

- *Spatial domain:*

  Spatial domain techniques are techniques that operate directly on pixels.

- *Frequency domain:*

  Frequency domain techniques are based on modifying the Fourier transform of an image.

g(x,y)=T[f(x,y)]

f(x,y): input image, g(x,y): processed image

T: an operator

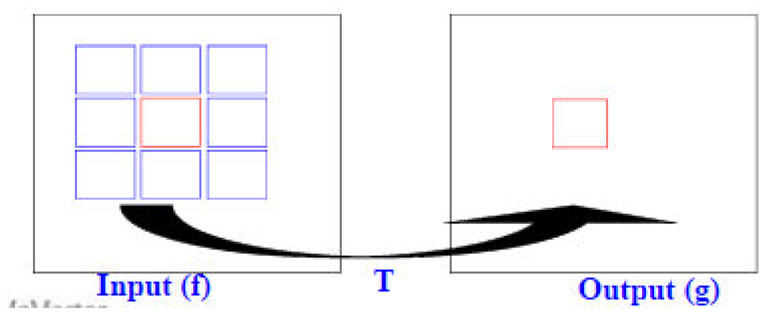


Figure 3.3: Background of Spatial Domain

Spatial Domain: Point Processing

s=T(r)

r: gray-level at (x,y) in original image f(x,y)

s: gray-level at (x,y) in processed image g(x,y)

T is called gray-level transformation or mapping

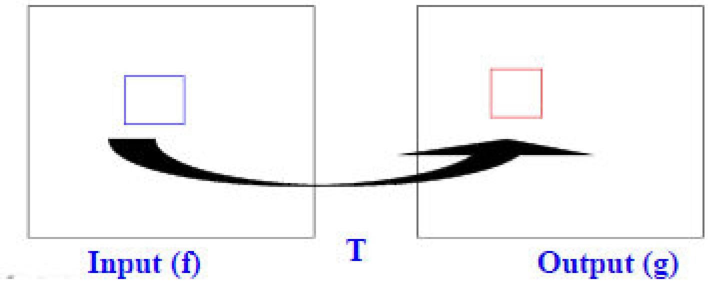
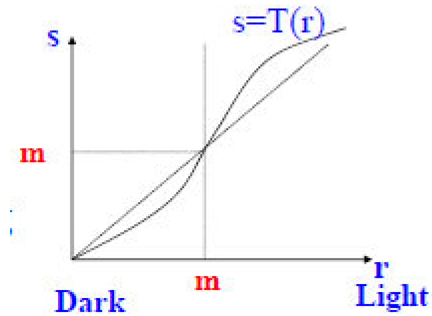


Figure 3.4: Spatial Domain: Point Processing

Contrast Stretching: to get an image with higher contrast than the original image. The gray levels below *m* are darkened and the levels above *m* are brightened.



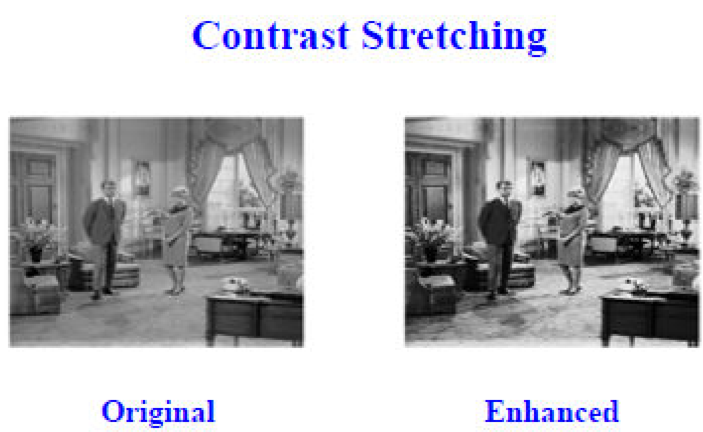


Figure 3.5: Contrast Stretching

Chapter 4
Video and Animation

4.1 Basic Concept

Video is the technology of electrically capturing, recording, processing, storing, transmitting, and reconstructing is a sequence of still images representing series in motion.

Visual representation is shown an idea or image that is presented in a particular way to have its meaning or symbolism.

Digital video has supplanted analog video as the method of choice for making video for multimedia use. While broadcast stations and professional production and postproduction houses remain greatly invested in analog video hardware, digital video gear produces excellent finished products at a fraction of the cost of analog. A digital camcorder directly connected to a computer workstation eliminates the image-degrading analog-to-digital conversion step typically performed by expensive video capture cards, and brings the power of nonlinear video editing and production to everyday users.

Video Signal Representation

Video signal representation consists of three aspects:

Visual Representation:

The main objective of visual representation is to offer the viewer a sense of presence in the scene and of participation in the events portrayed.

Transmission:

Video signals are transmitted to the receiver through a single television channel.

Digitization:

Digitization is the process of analog to digital conversion, sampling of gray (color) level, quantization.

Visual Representation

The main objective of visual representation is to offer the viewer a sense of presence in the scene and of participation in the events portrayed. To meet the main objective, the televised image should convey the spatial and temporal content of the scene. Importance measures are:

1. *Vertical detail and viewing distance:*

   The geometry of the field occupied by the television image is based on the ratio of the picture width W to height H. It is called aspect ratio.

   Aspect ratio: ratio of picture width and height (4/3 = 1.33 is the conventional aspect ratio)

   Viewing angle = Viewing distance/Picture height (D/H)

2. *Horizontal detail and picture width:*

   Picture width (Conventional TV service) = 4/3 x picture height

3. *Total detail content of the image:*
   - ❖ Number of pixels presented separately in the picture height = vertical resolution
   - ❖ Number of pixels in the picture width = vertical resolution x aspect ratio
   - ❖ The product of the number of elements vertically and horizontally equals the total number of picture elements in the image.

4. *Perception of depth:*
   - ❖ In natural vision, this is determined by angular separation of images received by the two eyes of the viewer.
   - ❖ In the flat image of TV, focal length of lenses and changes in depth of focus in a camera influence depth perception.

5. *Luminance and chrominance:*
   - ❖ Color-vision - achieved through 3 signals, proportional to the relative intensities of RED, GREEN and BLUE.

- Color encoding during transmission uses one LUMINANCE and two CHROMINANCE signals

6. *Temporal aspect of resolution:*

   Motion resolution is a rapid succession of slightly different frames. For visual reality, repetition rate must be high enough (a) to guarantee smooth motion and (b) persistance of vision extends over interval between flashes(light cutoff b/w frames).

7. *Continuity of motion:*
   - Motion continuity is achieved at a minimal 15 frames per second; is good at 30 frames/sec; some technologies allow 60 frames/sec.
   - NTSC standard provides 30 frames/sec - 29.97 Hz repetition rate.
   - PAL standard provides 25 frames/sec with 25Hz repetition rate.

8. *Flicker effect:*

   Flicker effect is a periodic fluctuation of brightness perception. To avoid this effect, we need 50 refresh cycles/sec. Display devices have a display refresh buffer for this.

9. *Temporal aspect of video bandwidth:*

   Temporal aspect of video bandwidth depends on rate of the visual system to scan pixels and on human eye scanning capabilities.

Transmission

Video signals are transmitted to receivers through a single television channel. To encode color, a video signal is a composite of three signals. For transmission purposes, a video signal consists of one luminance and two chrominance signals.

- Video bandwidth is computed as follows
  - 700/2 pixels per line X 525 lines per picture X 30 pictures per second
  - Visible number of lines is 480.

- Intermediate delay between frames is
  - 1000ms/30fps = 33.3ms
- Display time per line is
  - 33.3ms/525 lines = 63.4 microseconds
- The transmitted signal is a composite signal
  - Consists of 4.2Mhz for the basic signal and 5Mhz for the color, intensity and synchronization information.

Color Encoding:

- A camera creates three signals
  - RGB (red, green and blue)
- For transmission of the visual signal, we use three signals: 1 luminance (brightness-basic signal) and 2 chrominance (color signals).
  - In NTSC, luminance and chrominance are interleaved
  - Goal at receiver
    - separate luminance from chrominance components
    - avoid interference between them prior to recovery of primary color signals for display.

RGB signal

- for separate signal coding
- consists of 3 separate signals for red, green and blue colors. Other colors are coded as a combination of primary color. (R+G+B = 1) --> neutral white color.

YUV signal

- separate brightness (luminance) component Y and
- color information (2 chrominance signals U and V)
  - $Y = 0.3R + 0.59G + 0.11B$
  - $U = (B-Y) * 0.493$
  - $V = (R-Y) * 0.877$
- Resolution of the luminance component is more important than U,V
- Coding ratio of Y, U, V is 4:2:2

YIQ signal

- similar to YUV - used by NTSC format
  - $Y = 0.3R + 0.59G + 0.11B$
  - $U = 0.60R - 0.28G + 0.32\ B$
  - $V = 0.21R - 0.52g + 0.31B$

Composite signal

- All information is composed into one signal
- To decode, need modulation methods for eliminating interference b/w luminance and chrominance components.

Digitization

Before a picture or motion video can be processed by a computer or transmitted over a computer network, it need to be converted from analog to digital representation.

Digitization is the representation of an object, image, sound, document or a signal (usually analog signal) by a discrete set of its points or samples.

Digitization = Sampling + Quantization

Sampling is the reduction of a continuous signal to a discrete signal.

- Refers to sampling the gray/color level in the picture at MXN array of points.
- Once points are sampled, they are quantized into pixels
  - sampled value is mapped into an integer
  - quantization level is dependent on number of bits used to represent resulting integer, e.g. 8 bits per pixel or 24 bits per pixel.
- Need to create motion when digitizing video
  - digitize pictures in time
  - obtain sequence of digital images per second to approximate analog motion video.

Computer Video Format

The computer video format depends on the input and output devices for the motion video medium.

Current video digitizers differ in digital image (frame) resolution, quantization and frame rate (frames/s).

The output of the digitalized motion video depends on the display device. The most often used displays are raster displays, which store display primitives in a refresh buffer in terms of their component pixels.

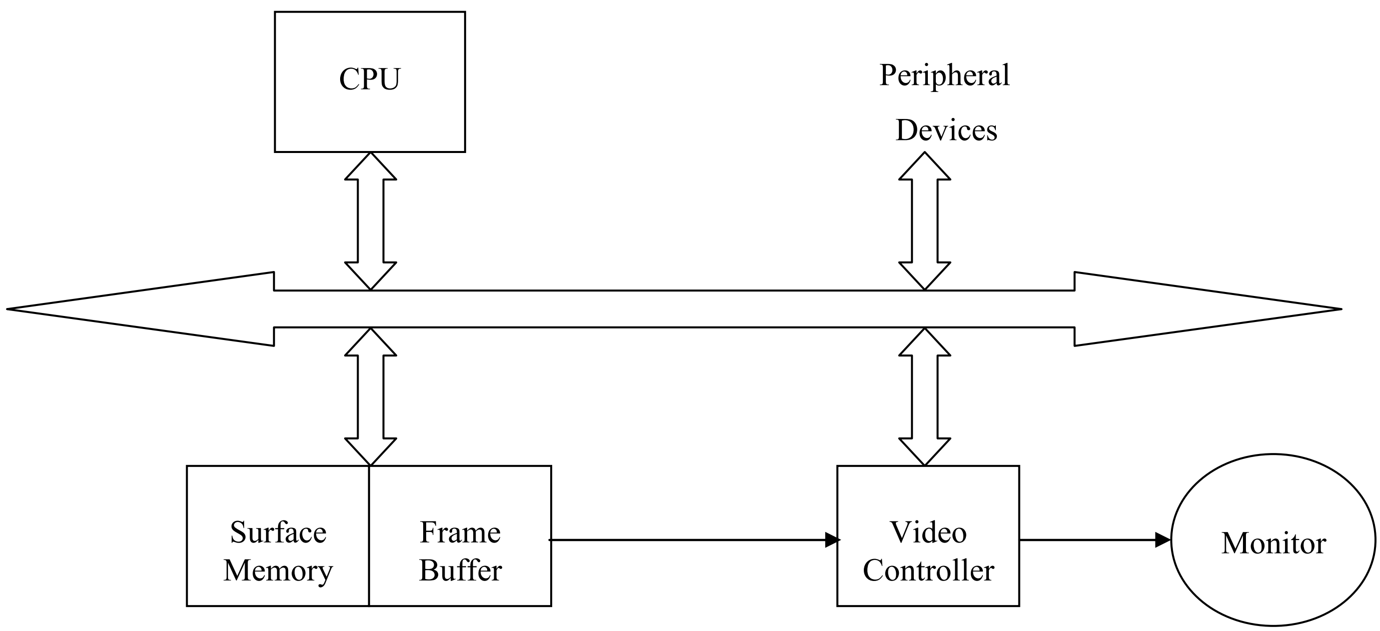


Figure 4.1: A common raster display system architecture

The video controller displays the image stored in the frame buffer, accessing the memory through a separate access port as often as the raster scan rate dictates. The constant refresh of the display is its most important task. Because of the disturbing flicker effect, the video controller cycles through the frame buffer, one scan line at a time, typically 60 times/second.

Some computer video controller standards are given here an example. Each of these systems supports different resolution and color presentation.

The Color Graphics Adapter (CGA):

The CGA has a resolution of 320x200 pixels with simultaneous presentation of four colors. Therefore, the storage capacity per image is:

$$300 \times 200\ pixels \times \frac{2\,\frac{bits}{pixels}}{8\,\frac{bits}{byte}}$$

$$= 16{,}000\ bytes$$

The Enhanced Graphics Adapter (EGA):

The EGA has a resolution of 640×350 pixels with 16-color presentation. Therefore, the storage capacity per image is:

$$640 \times 350\ pixels \times \frac{4\,\frac{bits}{pixels}}{8\,\frac{bits}{byte}}$$

$$= 112{,}000\ bytes$$

The Video Graphics Adapter (VGA):

The EGA has a resolution of 640×480 pixels with 16-color presentation. In this case, 256 colors can be displayed simultaneously. The monitor is controlled through an RGB output. The storage capacity per image is:

$$640 \times 480\ pixels \times \frac{8\,\frac{bits}{pixels}}{8\,\frac{bits}{byte}}$$

$$= 307{,}200\ bytes$$

The 8514/ A Display Adapter Mode:

A Display Adapter Mode can present 256 colors with a resolution of 1024×768 pixels. The storage capacity per image is:

$$1024 \times 768\ pixels \times \frac{8\,\frac{bits}{pixels}}{8\,\frac{bits}{byte}}$$
$$= 786{,}432\ bytes$$

The Extended Graphics Array (XGA):

The XGA supports a resolution of 640×480 pixels and 65,000 different colors. With the resolution of 1024×768 pixels, 256 colors can be presented. In this case, we have the same storage capacity per image as the 8514/A adapter.

The Super VGA (SVGA):

The SVGA offers resolutions up to 1024×768 pixels and color formats up to 24 bits per pixel. The storage capacity per image is:

$$1024 \times 768\ pixels \times \frac{24\,\frac{bits}{pixels}}{8\,\frac{bits}{byte}}$$
$$= 2{,}359{,}296\ bytes$$

Television

Television is the most important application that has driven the development of motion video. Television is a telecommunication medium for transmitting and receiving moving images that can be monochrome (black and white) or colored, with or without accompanying sound. Television may also refer specifically to a television set, television programming or television transmission.

Conventional Systems

Conventional system used in black and white and color television. Conventional television systems employ the following standards:

NTSC (National Television Systems Committee)

- NTSC developed in U.S., is the oldest and most widely used television standard.

- The color carrier is used with approximately 4.429 MHZ or with approximately 3.57 MHZ.
- NTSC uses a quadrature amplitude modulation with a suppressed color carrier and work with a motion frequency of approximately 30 Hz.
- 4×3 Aspect ratio.
- 525 lines
- 30 frames per second.
- Scanned in fields.

PAL (Phase Alternating Line)

- Invented by W.Bruch (Telefunken) in 1963.
- It is used in parts of Western Europe.
- The basic principle of PAL is a quadrature amplitude modulation similar to NTSC, but the color carrier is not suppressed.
- PAL is an analogue television color encoding system used in broadcast television systems in many countries.
- 4×3 Aspect ratio.
- 625 lines
- 25 frames per second.
- Scanned in fields.
- There are slight variations: PAL-B, PAL-G, PAL-H and PAL-N.
- Used in continental Europe and parts of Africa, Middle East and South America.
- More Lines = Better Resolution
- Fewer Frame/fields = More Flicker

SECAM (Sequential Color and Memory)

- SECAM is a standard used in France and Eastern Europe.
- In contrast to NTSC and PAL, it is based on frequency modulation.
- It uses a motion frequency of 25 Hz and each picture has 625 lines.
- SECAM is an analog color television system first used in france.

Enhanced Systems

Enhanced Definition Television Systems (EDTV) are conventional systems modified to offer improved vertical and/or horizontal resolution. EDTV are an intermediate solution, to digital interactive television system and their coming standards.

HDTV (High Definition Television)

- The next generation of TV is known as HDTV.
- HDTV is a digital system.
- 16:9 Aspect ratio.
- Permits several level of picture resolution similar to that of High Quality Computer Monitors, with 720 or 1080 line (1280×720 pixels or 1920×1080 pixels).
- Range from 24 to 60 frame per second, progressive or interlaced scan.
- Uses MPEG-2 compression to squeeze a 19 Megabit per second data flow so that it can be accommodated by a standard broadcast TV channel of 6 MHz bandwidth.

Digital coding are essential in the design and implementation of HDTV. There are two kinds of possible digital codings: composite coding and component coding.

Composite Coding:

The simplest possibility for digitizing video signal is to sample the composite analog video signal. Here, all signal components are converted together into a digital representation. The composite coding of the color television signal depends on the television standard. Using component coding, the sampling frequency is not coupled with the color carrier frequency.

Component Coding:

The principle of component coding consists of separate digitization of various image components or planes; for example, coding of luminance and color difference (chrominance) signals. These digital signals can be transmitted together using multiplexing. The luminance signal is sampled with 13.5 MHz as it is more important than the chrominance signal.

4.3 Basic Concepts of Animation

Animation is the rapid display of a sequence of images of 2-D or 3-D artwork or model positions in order to create an illusion of movement.

It is an optical illusion of motion due to the phenomenon of persistence of vision, and can be created and demonstrated in a number of ways.

The most common method of presenting animation is as a motion picture or video program, although several other forms of presenting animation also exist.

Animation is anything that moves on your screen like a cartoon character. It is the visual art of creating the illusion of motion through the successive display of still images with slightly perceptible changes in positioning of images. Animation is the illusion of movement.

Animating = making something appear to move that doesn't move itself

Animation = a motion picture made from a series of drawings simulating motion by means of slight progressive changes in the drawings

The result of animation is a series of still images assembled together in time to give the appearance of motion

Animation is the art of movement expressed with images that are not taken directly from reality. In animation, the illusion of movement is achieved by rapidly displaying many still images or frames in sequence.

4.4 Types & Techniques of Animation

When you create an animation, organize its execution into a series of logical steps. First, gather up in your mind all the activities you wish to provide in the animation; if it is complicated, you may wish to create a written script with a list of activities and required objects. Choose the animation tool best suited for the job. Then build and tweak your sequences; experiment with lighting effects. Allow plenty of time for this phase when you are experimenting and testing. Finally, post-process your animation, doing any special rendering and adding sound effects.

Cel Animation

The term cel derives from the clear celluloid sheets that were used for drawing each frame, which have been replaced today by acetate or plastic. Cels of famous animated cartoons have become sought-after, suitable-for-framing collector's items. Cel animation artwork begins with key frames (the first and last frame of an action). For example, when an animated figure of a man walks across the screen, balances the weight of his entire body on one foot and then the other in a series of falls and recoveries, with the opposite foot and leg catching up to support the body.

The animation techniques made famous by Disney use a series of progressively different on each frame of movie film which plays at 24 frames per second. A minute of animation may thus require as many as 1,440 separate frames. The term cel derives from the clear celluloid sheets that were used for drawing each frame, which is been replaced today by acetate or plastic. Cel animation artwork begins with key frames.

Computer Animation

Computer animation programs typically employ the same logic and procedural concepts as cel animation, using layer, key frame, and tweening techniques, and even borrowing from the vocabulary of classic animators. On the computer, paint is most often filled or drawn with tools using features such as gradients and antialiasing. The word links, in computer animation terminology, usually means special methods for computing RGB pixel values, providing edge detection, and layering so that images can blend or otherwise mix their colors to produce special transparencies, inversions, and effects. Computer Animation is same as that of the logic and procedural concepts as cel animation and use the vocabulary of classic cel animation– terms such as layer, Key frame, and tweening. The primary difference between the animation software program is in how much must be drawn by the animator and how much is automatically generated by the software In 2D animation the animator creates an object and describes a path for the object to follow. The software takes over, actually creating the animation on the fly as the program is being viewed by your user. In 3D animation the animator puts his effort in creating the models of individual and designing the characteristic of their shapes and surfaces. Paint is most often filled or drawn with tools using features such as gradients and anti- aliasing.

Kinematics

It is the study of the movement and motion of structures that have joints, such as a walking man. Inverse Kinematics is in high-end 3D programs, it is the process by which you link objects such as hands to arms and define their relationships and limits. Once those relationships are set you can drag these parts around and let the computer calculate the result.

Morphing

Morphing is popular effect in which one image transforms into another. Morphing application and other modeling tools that offer this effect can perform transition not only between still images but often between moving images as well the morphed images were built at a rate of 8 frames per second, with each transition taking a total of 4 seconds.

4.5 Principles of Animation

Animation is possible because of a a biological phenomenon known as *persistence of vision a*nd the psychological phenomenon called *phi* .

An object seen by the human eye remains chemically mapped on the eye's retina for a brief time after viewing. This makes it possible for a series of images that are changed very slightly and very rapidly, one after the other, seem like continuous motion.

1. *Squash and Stretch*: - Defining the rigidity & mass of an object by distorting its shape during an action.
2. *Timing:* - Spacing actions to define the weight & size of objects & the personality of characters.
3. *Anticipation:* - The preparation for an action.

4. *Staging:* - Presenting an idea so that it is unmistakably clear.
5. *Follow Through & Overlapping Action:* - The termination of an action & establishing its relationship to the next action.
6. *Straight Ahead Action & Pose-To-Pose Action:* - The two contrasting approaches to the creation of movement.
7. *Slow In and Out:* - The spacing of in-between frames to achieve subtlety of timing & movements.
8. *Arcs:* - The visual path of action for natural movement.
9. *Exaggeration:* - Accentuating the essence of an idea via the design & the action.
10. *Secondary Action:* - The Action of an object resulting from another action
11. *Appeal:* - Creating a design or an action that the audience enjoys watching.
12. *Solid Drawing:* - Knowing them can dramatically improve one's ability to create good, strong poses and compose them with well crafted environments.

4.6 Animation Languages:

There are many different languages for describing animation, and new ones are constantly being developed. They fall into three categories:

(i) Linear-list Notations:

In linear-list notations for animation each event in the animation is described by a starting and ending frame number and an action that is to take place (event). The actions typically take parameters, so a statement such as

42, 53, B, ROTATE "PALM", 1, 30

means "between frames 42 and 53, rotate the object called PALM about axis 1 by 30 degrees, determining the amount of rotation at each frame from tabled B".

(ii) General-purpose Language:

The values of a variables in the language can be used as parameters to the routines, which perform the animation. ASAS is an example of such a language. It is built on top of LISP, and its primitive entities include vectors, colors, polygons, solids, groups, points of view, subworlds and lights. ASAS also includes a wide range of geometric transformations that operate on objects. The ASAS program fragment below describes an animated sequence in which an object called my-cube is spun while the camera pans.

(**grasp** my-cube); The cube becomes the current object
(**cw** 0.05); Spin it clockwise by a small amount
(**grasp** camera); Make the camera the current object
(right panning-speed); Move it to the right

(iii) Graphical Languages:

Graphical animation languages describe animation in a more visual way. These languages are used for expressing, editing and comprehending the simultaneous changes taking place in an animation. The principal notion in such languages is substitution of a visual paradigm for a textual one. Rather than explicitly writing out descriptions of actions, the animator provides a picture of the action. Example of such systems and languages are *GENESYS, DIAL* and S-Dynamics System.

4.7 Method of Controlling Animation

Controlling animation is independent of the language used for describing it. Animation control mechanisms can employ different techniques.

Full Explicit Control

Explicit control is the simplest type of animation control. The animator provides a description of everything that occurs in the animation, either by specifying simple changes, such as scaling, translation, and rotation, or by providing key frame information and interpolation methods to use between key frames.

Procedural Control

Procedural control is based on communication between various objects to determine their properties.

Physical-based systems, the position of one object may influence the motion of another.

Action-based systems, the individual actors may pass their position to other actors to affect the other actors' behaviors.

Constraint-based Systems

Support a hierarchy of constraints and to provide motion where constraints are specified by the dynamics of physical bodies and structural characteristics of materials is a subject of active research.

Tracking Live Action

Trajectories of objects in the course of an animation can also be generated by tracking live action. Traditional animation uses rotoscoping. Tracking live-action technique is to attach some sort of indicator to key points on a person's body.

4.8 Transmission of Animation

Transmission over computer networks may be performed using one of two approaches:

The symbolic representation

The symbolic representation of animation objects is transmitted together with the operation commands performed on the object, and at the receiver side the animation is displayed. The transmission time is short because animated object is smaller in byte size than its pixmap representation, but the display time at the receiver takes longer because the scan-converting operation has to be performed at the receiver side.

The pixmap representation

The pixmap representation of the animated objects is transmitted and displayed on the receiver side. The transmission time is longer than the symbolic representation but the display time is shorter because the scan-conversion of the animated objects is avoided at the receiver side.

Chapter 5
Data Compression

5.1 Data Compression and Coding Fundamentals

Data compression is the process of converting an input data stream or the source stream or the original raw data into another data stream that has a smaller size. For example text compression, image compression, audio compression and video compression.

There are two types of data compression:

1. *Lossy Compression*
2. *Lossless Compression*

Lossy Compression

Lossy compression algorithms is normally not to reproduce an exact copy of the source information after decompression but rather a version of it which is perceived by the recipient as a true copy.

In lossy compression some information is lost during the processing, where the image data is stored into important and unimportant data. The system then discards the unimportant data.

It provides much higher compression rates but there will be some loss of information compared to the original source file. The main advantage is that the loss cannot be visible to eye or it is visually lossless. Visually lossless compression is based on knowledge about color images and human perception.

Lossless Compression

In this type of compression no information is lost during the compression and the decompression process. Here the reconstructed image is mathematically and visually identical to the original one. It achieves only about a 2:1 compression ratio.

This type of compression technique looks for patterns in strings of bits and then expresses them more concisely.

Lossless compression algorithm the aim is to be transmitted in such a way that, when the compressed information is decompressed, therefore is no loss information.

Techniques of Data Compression

There are three important techniques of data compression:

1. *Basic Technique*
2. *Statistical Technique*
3. *Dictionary Method*

Basic Technique

These are the techniques, which have been used only in the past. The important basic techniques are *run length encoding* and *move to front encoding*.

Run Length Encoding:

The basic idea behind this approach to data compression is this: if a data item occurs n consecutive times in the input stream replace the n occurrences with a single pair. The n consecutive occurrences of a data item are called run length of n and this approach is called run length encoding or RLE.

Move to Front Coding:

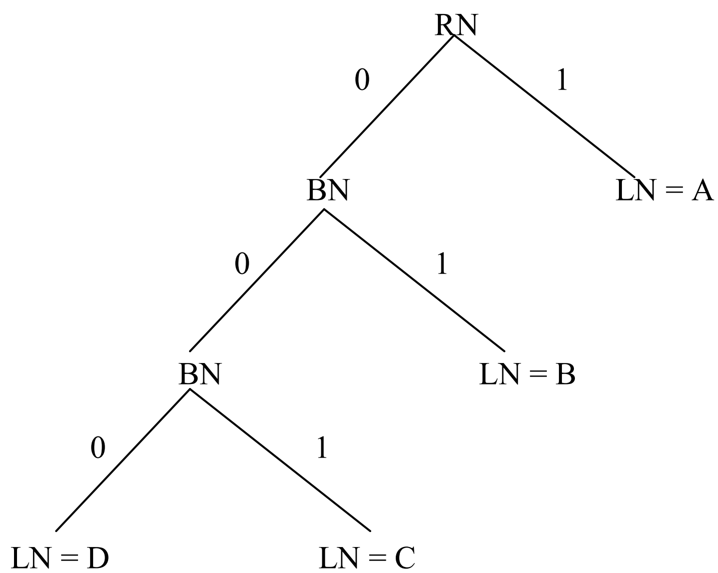
The basic idea of this method is to maintain the alphabet A of symbols as a list where frequently occurring symbols are located near the front. A symbol 'a' is encoded as the number of symbols that precede it in this list. Thus if A = ('t', 'h', 'e', 's') and the next symbol in the input stream to be encoded is 'e', it will be encoded as '2' since it is preceded by two symbols. The next step is that after encoding 'e' the alphabet is modified to A = ('e', 't', 'h', 's'). This move to front step reflects the hope that once 'e' has been read from the input stream it will read many more times and will at least for a while be a common symbol.

Huffman Coding

A commonly used method for data compression is Huffman coding. The method starts by building a list of all the alphabet symbols in descending order of their probabilities. It then construct a tree with a symbol at every leaf from the bottom up.

The Character string to be transmitted is first analyzed and the character types and their relative frequency determined.

A Huffman (code) tree is a binary tree with branches assigned the value 0 or 1. As each branch divides, a binary value of 0 or 1 is assigned to each new branch: a binary 0 for the left branch and a binary 1 for the right branch.



Where RN = Root Node, BN = Branch Node, LN = Leaf Node

A = 1, B = 01, C = 001, D = 000

Figure 5.1: Huffman Tree

Arithmetic Coding

In this method the input stream is read symbol and appends more to the code each same time a symbol is input and processed. To understand this method it is useful to imagine resulting code as a number in the range [0,1] that is the range of real numbers from 0 to 1 not including one. The first step is to calculate or at least to estimate the frequency of occurrence of each symbol.

LZ77 (Sliding Window)

LZ77 and LZ78 are the names for the two lossless data compression algorithms published in papers by Abraham Lempel and Jacob Ziv in 1977 and 1978. They are also known as LZ1 and LZ2 respectively. They are both dictionary coders. LZ77 is the Sliding Window compression algorithm.

The main idea of this method is to use part previously seen input stream as the dictionary. The encoder maintains a window to the input stream and shifts the input in that window from right to left as strings of symbols are being encoded. The method is thus based on "Sliding Window". The window is divided into two parts that is search buffer, which is the current dictionary and lookahed buffer, containing text yet to be encoded.

Source, Entropy and Hybrid Coding

Entropy Coding

- lossless encoding
- used regardless of media's specific characteristics
- data taken as a simple digital sequence
- decompression process regenerates data completely
- e.g. run-length coding, Huffman coding, Arithmetic coding

Source Coding

- lossy encoding
- takes into account the semantics of the data
- degree of compression depends on data content.
- E.g. content prediction technique - DPCM, delta modulation

Hybrid Coding (used by most multimedia systems)

- combine entropy with source encoding
- E.g. JPEG, H.263, DVI (RTV & PLV), MPEG-1, MPEG-2, MPEG-4

Major Steps of Data Compression

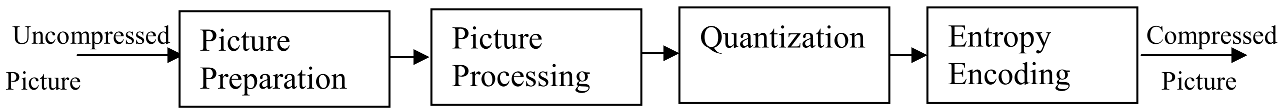


Figure 5.2: Major Steps of Data Compression

Picture Preparation:

Preparation includes analog to digital conversion and generating an appropriate digital representation of the information. An image is divided into blocks of 8x8 pixels, and represented by a fixed number of bits per pixel.

Picture Processing:

Processing is actually the first step of the compression process which makes use of sophisticated algorithms. A transformation from the time to the frequency domain can be performed using DCT. In the case of motion video compression, interframe coding uses a motion vector for each 8x8 blocks. Motion video computation for digital video.

Quantization:

Quantization process the results of the previous step. It specifies the granularity of the mapping of real numbers into integers. This process results in a reduction of precision. For example they could be quantized using a different number of bits per coefficient. For example 12 bits for real values, 8 bits for integer value.

Entropy Encoding:

Entropy encoding is usually the last step. It compresses a sequential digit data stream without loss. For example, a sequence of zeroes in a data stream can be compressed by specifying the number of occurrences followed by the zero itself.

JPEG (Joint Photographic Expert Group)

JPEG is a commonly used method of lossy compression for digital photography (image). The JPEG lossy compression scheme is one of the most popular and versatile compression schemes in widespread use. It’s ability to attain considerable size reductions with minimal visual impact with relative light computational requirements and the ability to fine tune the compression level to suit the image at hand has made it the standard for continuous tone still images.

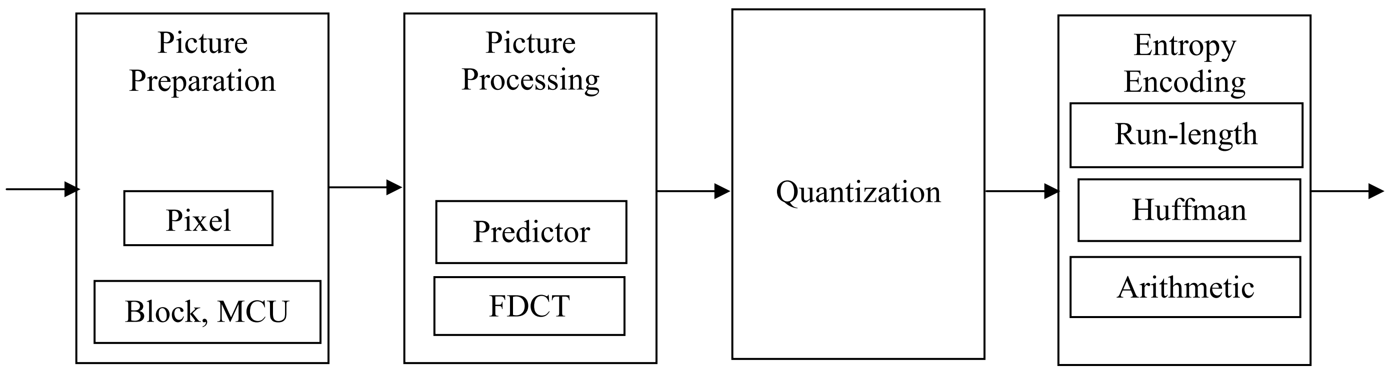


Figure 5.3: Steps of the JPEG compression process.

Requirements on JPEG implementations

JPEG Image Preparation

- Blocks, Minimum Coded Units (MCU)

JPEG Image Processing

- Discrete Cosine Transformation (DCT)

JPEG Quantization

- Quantization Tables

JPEG Entropy Encoding

- Run-length Coding/Huffman Encoding

Additional Requirements - JPEG

- JPEG implementation is independent of image size and applicable to any image and pixel aspect ratio.
- Image content may be of any complexity (with any statistical characteristics).
- JPEG should achieve very good compression ratio and good quality image.
- From the processing complexity of a software solution point of view: JPEG should run on as many available platforms as possible.

- Sequential decoding (line-by-line) and progressive decoding (refinement of the whole image) should be possible.

Variants of Image Compression

Four different modes

Lossy Sequential DCT based mode

- Baseline process that must be supported by every JPEG implementation.

Expanded Lossy DCT based mode

- enhancements to baseline process

Lossless mode

- low compression ratio
- allows perfect reconstruction of original image

Hierarchical mode

- accomodates images of different resolutions

Image Preparation

The image preparation model in JPEG is very general.

It is not based on

- 9-bit YUV encoding
- fixed number of lines, columns
- mapping of encoded chrominance

It is independent from image Parameters such as *image size*, *image* and *pixel ratio*.

Source image consists of 1 to 255 components (planes).

- For example, each component Ci (1≤ i ≤ 255) may be assigned to YUV, RGB or YIQ signals.

All pixels of all components within the same image are coded with the same number of bits.

- Lossy modes of JPEG use precision of 8-12 bits per pixel.
- Lossless mode uses precision of 2 up to 12 bits per pixel.

- If JPEG application makes use of any other number of bits, then application must perform a suitable image transformation to the well-defined number of bits/pixel (JPEG standard).

Images are divided into data units, called blocks

- Lossy modes operate on blocks of 8X8 pixels.
- Lossless modes operate on data units equal to 1 pixel.
- DCT transformation operates on blocks.

Data units are processed component by component and passed to image processing.

Processing of data units per component can be

- Non-interleaved data ordering
- Left to right, top to bottom
- Interleaved data ordering

Interleaved Data units of different components are combined to minimum coded units (MCUs).

If image has the same resolution, then MCU consists of exactly one data unit for each component.

- Decoder displays the image MCU by MCU.

If image has different resolution for single components, then reconstruction of MCUs is more complex.

- For each component, determine regions of the data units. Each component has same number of regions, MCU corresponds to one region.
- Data units in a region are ordered left-right, top-bottom
- Build MCU

JPEG standard - only 4 components can be encoded in interleaved mode

Bound on length of MCU

- MCU consists of at most 10 data units.

After image preparation, uncompressed image samples are grouped into data units of 8x8 pixels and passed to the JPEG encoder

- order of the data units is defined by the MCUs
- precision 8 bits/pixel represents the baseline mode
- values are in the range of [0,255];

Image Processing

First step:

- Pixel values are shifted (ZERO-SHIFT) into the range [-128,127] with 0 in the center.

  Values in the 8x8 pixel are defined by Syx with y,x in the range [0,7] and there are 64 sampled values Syx in each block.
- DCT maps values from time to frequency domain.

  1D Forward Discrete Cosine Transformation

$$S(u) = \frac{C(u)}{2}\sum_{x=0}^{7} S(x)\cos\left(\frac{(2x+1)u\pi}{16}\right)$$

$S(x) - 1D\ sampled\ value,$
$C(u) - scaling\ coefficient,$
$S(u) - 1D\ DCT\ coefficient\ (transforrms\ S(x) into\ frequency\ domain)$

  2D Forward Discrete Cosine Transformation

$$S(u,v) = \frac{1}{4}C(u)C(v)\sum_{x=0}^{7}\sum_{y=0}^{7} S(y,x)\cos\left(\frac{(2x+1)u\pi}{16}\right)\cos\left(\frac{(2x+1)v\pi}{16}\right)$$

$C(u), C(v) = {}^{1}/_{\sqrt{2}}\ for\ u,v = 0, C(u), C(v) = 1\ otherwise$
$S(y,x) - 2D\ sampled\ values$
$C(u), C(v) - scaling\ coefficients$
$S(v,u) - 2D\ DCT\ coefficients$

S(v,u) coefficients:

- S(0,0) includes the lowest frequency in both directions and it is called the DC coefficient. S(0,0) determines the fundamental color of the BLOCK(64 pixels). For this coefficient the frequency is equal 0 in both directions.
- S(0,1)…S(7,7) are called AC coefficients. Their frequency is non-zero in one or both directions. There exist many AC coefficients with a value around 0.

Factoring

- By computing the DCT coefficients, we can use factoring; the problem will be reduced to a series of 1D FDCTs.

$$S(v,u) = \frac{1}{4}\sum_{x=0}^{7} C(u)\cos\left(\frac{(2x+1)u\pi}{16}\right)\left(\sum_{y=0}^{7} C(v)\cos\left(\frac{(2y+1)v\pi}{16}\right)S(y,x)\right)$$

Quantization

GOAL: To throw out bits.

Example:

- 101101 = 45 (6 bits).
- We can truncate this to 4 bits: 1011 - 11
- or 3 bits 101 = 5 (original value - 40) or 110 = 6 (value = 48)

Uniform quantization is achieved by dividing the DCT coefficient value S(v,u) by N and rounding the result.

In S(v,u) how many bits do we throw away?

ANSWER: *Use quantization tables*

Entropy Encoding

- After image processing we have quantized DC and AC coefficients.
- Initial step of entropy encoding is to map 8x8 plane into 64 element vector

H.261

- H.261 is a video coding standard published by the ITU-T in 1990.
- It is the most widely used international video compression standard for video coding.
- H.261 is usually used in conjunction with other control and framing standards.
- The H.261 standard describes the video coding/decoding methods for the video portion of an audiovisual service
- Designed for data rates of p*64 kbps, where p is in the range 1-30

- Targeted for circuit-switched networks (ISDN was the communication channel considered within the framework of the standard)
- Defines two picture formats:
  - CIF (352x288) and QCIF(176x144)

Picture Formats Supported

Picture format	Luminance pixels	Luminance lines	H.261 support	Uncompressed bitrate (Mbit/s)			
				10 frames/s		30 frames/s	
				Grey	Colour	Grey	Colour
QCIF	176	144	Yes	2.0	3.0	6.1	9.1
CIF	352	288	Optional	8.1	12.2	24.3	36.5

- The H.621 encoding algorithm is a combination of:
  - inter-picture prediction (to remove temporal redundancy)
  - transform coding (to remove spatial redundancy)
  - motion vectors (for motion compensation)
- The three main elements are:
- Prediction: blocks are intra- or inter-coded
  - Intra-coded blocks stand alone
  - Inter-coded blocks are based on differences between the previous frame and the current one
- Block Transformation:
  - each block (inter and intra) is composed into 8x8 blocks and processed by a 2-D DCT function
- Quantization & Entropy Coding:
  - achieves further compression by representing DCT coefficients with only the necessary precision
  - entropy encoding (non-lossy) using Huffman encoding
- Coding algorithm is a hybrid of:
  - Inter-picture prediction - removes temporal redundancy
  - Transform coding - removes the spatial redundancy

- <u>Motion compensation</u> – uses motion vectors to help the codec compensate for motion
- Data rate can be set between 40 Kbit/s and 2 Mbit/s
- Input signal format
  - CIF (Common Intermediate Format) and QCIF (Quarter CIF) available
- Bit rate
  - The target bit rate is ~ 64Kbps to 1920Kbps

MPEG (Motion/Moving Pictures Expert Group)

Motion Pictures Expert Group (MPEG) standards are digital video encoding processes that coordinate the transmission of multiple forms of media (multimedia). MPEG is a working committee that defines and develops industry standards for digital video systems. These standards specify the data compression and decompression process and how they are delivered on digital broadcast systems. MPEG is a part of International Standards Organization (ISO).

Started in 1988, the MPEG project was developed by a group of hundreds of experts under the auspices of the ISO (International Standardization Organization) and the IEC (International Electrotechnical Committee). The name MPEG is an acronym for Moving Pictures Experts Group. MPEG is a method for video compression, which involves the compression of digital images and sound, as well as synchronization of the two. There currently are several MPEG standards. MPEG-1 is intended for intermediate data rates, on the order of 1.5 Mbit/s. MPEG-2 is intended for high data rates of at least 10 Mbit/s. MPEG-3 was intended for HDTV compression but was found to be redundant and was merged with MPEG-2. MPEG-4 is intended for very low data rates of less than 64 Kbit/s. A third international body, the ITU-T, has been involved in the design of both MPEG-2 and MPEG-4. This section concentrates on MPEG-1 and discusses only its image compression features.

MPEG standard consists of both video and audio compression. MPEG standard includes also many technical specifications such as image resolution, video and audio synchronization, multiplexing of the data packets, network protocol, and so on. Here we consider only the video compression in the algorithmic level. The MPEG algorithm relies on two basic techniques

- Block based *motion compensation*
- DCT based compression

MPEG itself does not specify the encoder at all, but only the structure of the decoder, and what kind of bit stream the encoder should produce. Temporal prediction techniques with motion compensation are used to exploit the strong temporal correlation of video signals. The motion is estimated by predicting the current frame on the basis of certain previous and/or forward frame. The information sent to the decoder consists of the compressed DCT coefficients of the residual block together with the *motion vector*. There are three types of pictures in MPEG:

- Intra-pictures (*I*)
- Predicted pictures (*P*)
- Bidirectionally predicted pictures (*B*)

Figure 1 demonstrates the position of the different types of pictures. Every *N*th frame in the video sequence is an *I*-picture, and every *M*th frame a *P*-picture. Here *N*=12 and *M*=4. The rest of the frames are *B*-pictures.

Compression of the picture types:

Intra pictures are coded as still images by DCT algorithm similarly than in JPEG. They provide access points for random access, but only with moderate compression. *Predicted pictures* are coded with reference to a past picture. The current frame is predicted on the basis of the previous *I*- or *P*-picture. The residual (difference between the prediction and the original picture) is then compressed by DCT. Bidirectional pictures are similarly coded than the *P*-pictures, but the prediction can be made both to a past and a future frame which can be *I*- or *P*-pictures. Bidirectional pictures are never used as reference.

The pictures are divided into 16×16 macroblocks, each consisting of four 8□8 elementary blocks. The *B*-pictures are not always coded by bidirectional prediction, but four different prediction techniques can be used:

- Bidirectional prediction
- Forward prediction
- Backward prediction
- Intra coding.

The choice of the prediction method is chosen for each macroblock separately. The bidirectional prediction is used whenever possible. However, in the case of sudden camera movements, or a breaking point of the video sequence, the best predictor can sometimes be given by the forward predictor (if the current frame is before the breaking point), or backward prediction (if the current frame is after the breaking point). The one that gives the best match is chosen. If none of the predictors is good enough, the macroblock is coded by intra-coding. Thus, the *B*-pictures can consist of macroblock coded like the *I*-, and *P*-pictures.

The intra-coded blocks are quantized differently from the predicted blocks. This is because intra-coded blocks contain information in all frequencies and are very likely to produce 'blocking effect' if quantized too coarsely. The predicted blocks, on the other hand, contain mostly high frequencies and can be quantized with more coarse quantization tables.

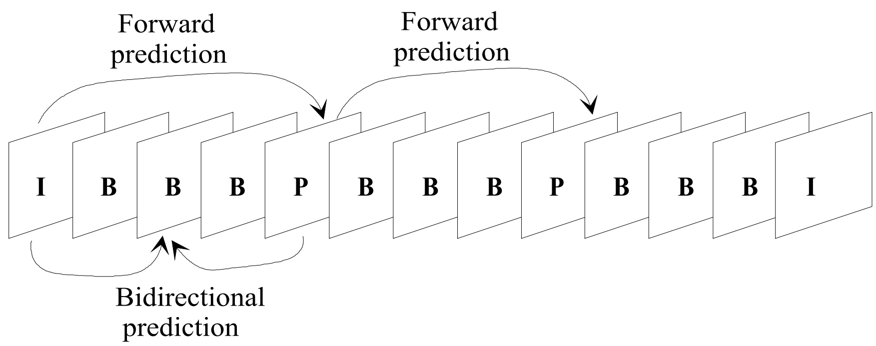


Figure 5.4: Interframe coding in MPEG.

Motion estimation:

The prediction block in the reference frame is not necessarily in the same coordinates than the block in the current frame. Because of motion in the image sequence, the most suitable predictor for the current block may exist anywhere in the reference frame. The *motion estimation* specifies where the best prediction (best match) is found, whereas *motion compensation* merely consists of calculating the difference between the reference and the current block.

The motion information consists of one vector for forward predicted and backward predicted macroblocks, and of two vectors for bidirectionally predicted macroblocks. The MPEG standard does not specify how the motion vectors are to be computed, however, *block matching techniques* are widely used. The idea is to find in the reference frame a similar macroblock to the macroblock in the current frame (within a predefined search range). The candidate blocks in the reference frame are compared to the current one. The one minimizing a cost function measuring the mismatch between the blocks, is the one which is chosen as reference block.

Exhaustive search where all the possible motion vectors are considered are known to give good results. Because the full searches with a large search range have such a high computational cost, alternatives such as *telescopic* and *hierarchical searches* have been investigated. In the former one, the result of motion estimation at a previous time is used as a starting point for refinement at the current time, thus allowing relatively narrow searches even for large motion vectors. In hierarchical searches, a lower resolution representation of the image sequence is formed by filtering and subsampling. At a reduced resolution the computational complexity is greatly reduced, and the result of the lower resolution search can be used as a starting point for reduced search range conclusion at full resolution.

MPEG in the RealWorld

MPEG has found a number of applications in the real world, including:

- *Direct Broadcast Satellite.* MPEG video streams are received by a dish/decoder, which unpacks the data and synthesizes a standard NTSC television signal.
- *Cable Television.* Trial systems are sending MPEG-II programming over cable television lines.
- *Media Vaults.* Silicon Graphics, Storage Tech, and other vendors are producing on-demand video systems, with twenty file thousand MPEG-encoded films on a single installation.
- *Real-Time Encoding.* This is still the exclusive province of professionals. Incorporating special-purpose parallel hardware, real-time encoders can cost twenty to fifty thousand dollars.

Compare and contrast between H.261 & MPEG

H.261	MPEG
■ based on JPEG ■ encodes video only ■ lossy algorithm with compression in space and time ■ uses I and P-frames ■ uses DCT on 8×8 blocks ■ best for video with little motion (eg. video conferencing) ■ optimized for bandwidth efficiency and low delay	■ based on H.261 and JPEG ■ encodes audio & video ■ lossy algorithm with compression in space and time ■ uses I, P, and B-frames ■ uses DCT on 8×8 blocks ■ designed to handle moving picture components ■ less bandwidth efficient

H.261	MPEG
■ Only 1, 2 or 3 skipped pictures allowed. ■ Pixel accurate motion vectors. ■ Typical motion vector range is +/-7 pixels. ■ Used mostly in interactive applications. End-to-end delay is very critical.	■ No restrictions on skipped pictures. ■ Sub-pixel accurate motion vectors. ■ Typical motion vector range is +/-15 pixels ■ The end-to-end coding delay is not critical.

DVI (Digital Video Interactive)

Digital Video Interactive (DVI) is a technology that includes coding algorithms. The fundamental components are a VLSI (Very Large Scale Integration) chip set for the video subsystem, a well-specified data format for audio and video files, an application user interface to the audio-visual kernel (Audio Video Kernel, the kernel software interface to the DVI hardware) and compression, as well as decompression, algorithms.

DVI can process data, text, graphics, still images, video and audio. The original essential characteristic was the asymmetric technique of video compression and decompression known as Presentation-Level Video (PLV).

Concerning audio, the demand for a hardware solution that can be implemented at a reasonable price is met by using a standard signal processor. Processing of still images and video is performed by a video processor. The video hardware of a DVI board is shown in Figure. It consists of two VLSI chips containing more than 265,000 transistors each. This Video Display Processor (VDP) consists of the pixel processor VDP1 and the display processor VDP2. VDP1 processes bitmaps and in programmed in microcode; VDP2 generates analog RGB signals out of the different bitmap formats and its configuration is also programmable. The coupling of the processors is carried out by the Video-RAM (VRAM). An important characteristic is the capability for microprogramming. It allows one to change and adapt the compression and decompression algorithms to new developments without investing in new hardware.

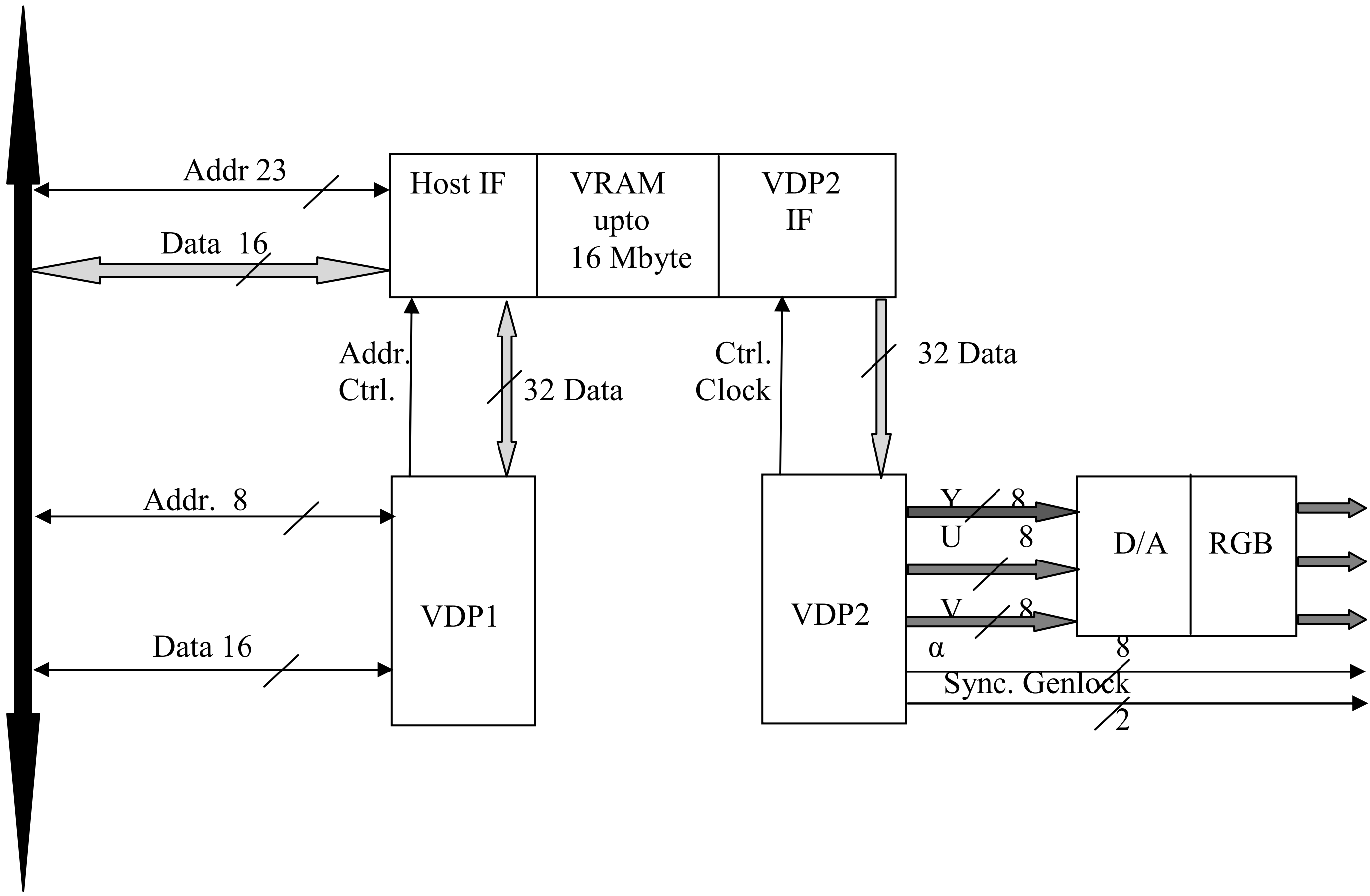


Figure 5.5: DVI Video Processing

Chapter 6

Optical Storage Media

In principle, optical storage media use the intensity of reflected laser light as an information source. A laser beam of approximately 780 nm wave length can be focused at approximately 1µm. In a polycarbonate substrate layer we encounter holes, corresponding to the coded data, which are called pits. The areas between these pits are called lands. Figure 1 shows a cut through an optical disk along a track. In the middle of the figure, the lands and pits are schematically presented.

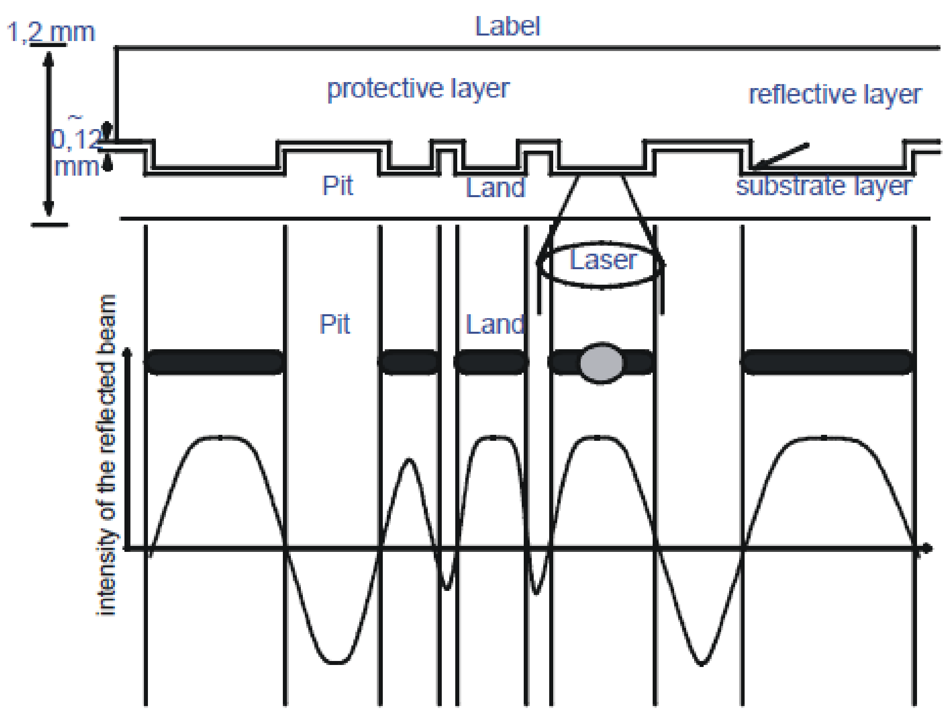


Figure 6.1: Cut through an optical disk along the data trace. A schematic presentation with the layers (above), the "lands" and the "pits" (in the middle), and the signal (below).

The substrate layer is covered with a thin reflective layer. The laser beam is focused on the reflective layer from the substrate layer. Therefore, the reflected beam has a strong intensity at the lands. The pits have a depth of 0.12 µm (from the substrate surface). The laser beam is lightly scattered at the pits, meaning it is reflected with a weak intensity. The signal, shown in figure, denotes schematically the intensity of the reflected beam-a horizontal line is drawn as the threshold value. Hence, according to figure, a compact disk consists of:

The label

The protective layer

The reflective layer

The substrate layer

An optical disk consists of a sequential order of these pits and lands allocated in one track. Figure 2 shows an enlarged cut of such a structure.

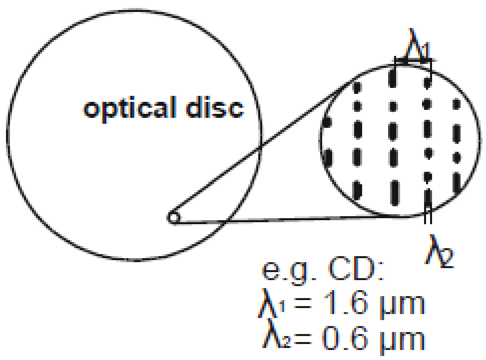


Figure 6.2: Data on a CD as an example of an optical disk (track with "lands" and "pits").

Information is stored in a spiral-shaped track:

- Series of **pits** and **lands** in substrate layer
- Transition from pit to land and from land to pit: ‘1’
- Between transitions: sequence of ‘0’ s
- 16000 turns/inch (tpi)

Reading: Laser focused onto reflective layer

- **Lands** - almost totally reflect the light
- **Pits** – scatter the light

The track is a spiral. In the case of a CD, the distance between the track is 1.6 µm. The track width of each pit is 0.6 µm. The pits themselves have different lengths. Using these measurements, the main advantage of the optical disk in comparison to magnetic disks is that on the former 1.66 data bits per μm can be stored. This results in a data density of 1,000,000 bits per mm^2, which implies 16,000 tracks per inch. In comparison, a floppy disk has 96 tracks per inch.

Advantages of Optical Storage Media

High data density

- 1.66 data bits/μm of track
- Inter-track density: 16000 tpi; compare diskette at 96 tpi

Long term storage

- Insensitive to magnetic/electric interference
- Insensitive to dust, scratches

Low probability of head crashes

- Distance between head and substrate surface > 1 mm

Adequate error correction

- allows handling of many defects

Perception quality

- e.g., each digital music disc is exactly equivalent to the master

Video Disks and Other WORMs

The video disk, in the form of Laser Vision, serves as the output of motion pictures and audio. The data are stored in an analog-coded format on the disk; the reproduced data meet the highest quality requirements. The Laser Vision disk has a diameter of approximately 30 cm and stores approximately 2.6 Gigabytes.

Video disk was designed as Read Only Memory (ROM), many different write-once optical storage systems have come out, known as the Write Once Read Many (WORM) disk. An example is the Interactive Video Disk. This disk is played at a Constant Angular Velocity (CAV). On each side, 36 minutes of audio and video at a rate of 30 frames per second can be stored and retrieved. One can also store around 54,000 studio quality images per side.

A write once read many or WORM drive is a data storage device where information, once written, cannot be modified. On ordinary data storage devices, the number of times data can be modified is not limited, except by the rated lifespan of the device, as modification involves physical changes that may cause wear to the device. The "read many" aspect is unremarkable, as modern storage devices permit unlimited reading of data once written.

WORM devices are useful in archiving information when users want the security of knowing it has not been modified since the initial write, which might imply tampering. The CD-R and DVD-R optical disks for computers are common WORM devices.

Write-once storage media have a capacity between 600 Mega Bytes (MB) and approximately 8 Gigabytes. The diameter of the disks is between 3.5 and 14 inches. The main advantage of a WORM disk, compared to other mass storage media, is the ability to store large amounts of data which may not be changed later, i.e., an archive which is secure. To increase capacity, juke-boxes are available, which allow the stocking of several disks and lead to capacities of over 20 Gigabytes.

Compact Disk Digital Audio

The first CD format was of course that which defined the audio CD used in all regular CD players, called *CD Digital Audio* or *CD-DA* for short. The specifications for this format were codified in the first CD standard, the so-called "red book" that was developed by Philips and Sony, the creators of the original compact disk technology. The "red book" was published in 1980, and actually specifies not just the data format for digital audio but also the physical specifications for compact disks: the size of the media, the spacing of the tracks, etc. In a sense, then, all of the subsequent standards that came after CD-DA build on the "red book" specification, since they use the same specifications for the media and how it is read. They also base their structure on the original structure created for CD audio.

Data in the CD digital audio format is encoded by starting with a source sound file, and sampling it to convert it to digital format. CD-DA audio uses a sample rate of 44.1 kHz, which is roughly double the highest frequency audible by humans (around 22 kHz.) Each sample is 16 bits in size, and the sampling is done in stereo. Therefore, each second of sound takes (44,100 * 2 * 2) bytes of data, which is 176,400 bytes.

Audio data is stored on the disk in blocks, which are also sometimes called sectors. Each block holds 2,352 bytes of data, with an additional number of bytes used for error detection and correction, as well as control structures. Therefore, 75 blocks are required for each second of sound. On a standard 74-minute CD then, the total amount of storage is (2,352 * 75 * 74 * 60), which is 783,216,000 bytes or about 747 MB. From this derives the handy rule of thumb that a minute of CD audio takes about 10 MB, uncompressed.

CD-DA: Characteristics

Audio data rate

The audio data rate can be easily derived from the given samples frequency of 44.1 kHz and the 16-bit linear quantization. The stereo-audio signal obeys the pulse-code modulation rules and the following audio data rate is derived:

- Sampling frequency: 44,100 Hz
- Quantization: 16 bits
- Pulse code modulation (PCM), uniform quantization
- Audio data rate = 1,411,200 bit/s = (~ 1.4 Mbit/s) (stereo)

$$Audiodatarate_{CD-DA} = 16\,\frac{bits}{samples} \times 2\,channels \times 44100\,\frac{samples}{s \times channel}$$

$$= 1{,}411{,}200\,\frac{bits}{s} = 1{,}411{,}200\,\frac{bits/s}{8\,bitst/byte}$$

$$= 176.4\,\frac{kbytes}{s} \cong 172.3\,\frac{Kbytes}{s}$$

Quality

Analog LPs and cassette tapes have a signal-to-noise ratio between 50 dB and 60 dB. The quality of the CD-DA is substantially higher. As a first approximation, we can assume 6 dB per bit during the sampling process. Hence, with 16-bit linear sampling, we obtain the following:

- Signal to noise ratio (S/N): ~ 6 dB/bit, 16 bit quantization => S/N exactly 98 dB
- Compare LP, tape: S/N 50-60 dB

$$S/N_{CD\text{-}DA} \cong 6\frac{dB}{bit} \times 16\,bits = 96\,dB\,(decibel)$$

The signal-to-noise ratio is exactly 98 dB.

Capacity (without error correction data)

A CD-DA play time is at least 74 minutes. With this value, the capacity of a CD-DA can be easily determined. The following example shows the computation of a capacity for pure audio data without taking into consideration additional information such as error correction:

- Playback time: maximal 74 min
- Raw capacity =74 min x 1,411,200 bit/s = 6265728000 bit ~ 747 Mbyte

$$Capacity_{CD\text{-}DA} = 74\text{ min} \times 1{,}411{,}200\frac{bits}{s} = 6{,}265{,}728{,}000\,bits$$

$$= 6{,}265{,}728{,}000\,bits \times \frac{1}{8\frac{bits}{byte}} \times \frac{1}{1024\frac{bytes}{Kbyte}} \times \frac{1}{\frac{Kbytes}{Mbyte}} \cong 747\ Mbytes$$

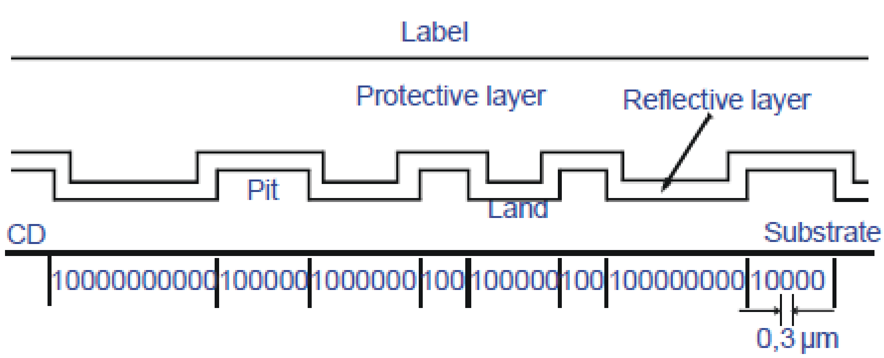


Figure 6.3: Lands and pits with their related digital data stream.

Length of pits and lands: multiples of 0.3 µm.

CD-DA: Eight-to-Fourteen Modulation

Restricted laser resolution

Requires a minimal distance between transitions (pit to land, land to pit): at least two "0"s between two "1"s

Generation (adaptation) of the clock signal is driven by transitions

Requires a maximal distance between transitions (pit to land, land to pit): not more than 10 consecutive "0"

For these reasons, the bits written on a CD-DA, in the form of pits and lands, do not directly corresponds to the actual information; before recording, Eight-to-Fourteen Modulation is applied. Using this transformation, the regularity of the minimal and maximal distances is met.

- An 8 bit data value is encoded using 14 bits
- 267 combinations fulfill the criteria above, 256 are chosen. Criterion: efficient implementation with a small number of gates.

Example for the code conversion table

Audio Bits	**Modulation Bits**
00000000	01001000100000
00000001	10000100000000

But: a concatenation of two independent 14-bit values could lead to a violation of:

- minimum distance of 2 bits between Ones
- maximum distance of 10 bits between Ones

Therefore, three additional ***merging (filling) bits*** are inserted between two consecutive modulation symbols so that the required regularity can be met. The filling bits are chosen depending on the neighboring bits.

CD-DA: Eight-to-Fourteen Modulation Example

Audio Bits	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
Modulation Bits	0 1 0 0 1 0 0 0 1 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0
Filling Bits	0 1 0 1 0 0
Channel Bits	0 1 0 0 1 0 0 1 0 0 0 1 0 0 0 0 0 1 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0
On the CD-DA	l p p p l l l p p p p l l l l l l p p p l l l l l p p p p p p p p p

CD-DA: Error Handling

The goal of error handling on a CD-DA is the detection and correction of typical error patterns. A typical error, a consequence of a scratch and /or pollution, can be characterized as a burst error.

A two stage error correction is implemented according to the *Reed Solomon algorithm*:

First level: byte level, EDC and ECC. Two groups, each with four correction bytes for 24 data bytes:

- 1st group: correction of single byte errors
- 2nd group: correction of double byte errors, detection of additional errors

Second level: frame interleaving

- frame: 588 channel bits for 24 audio data bytes
- distribution of consecutive data bytes and corresponding ECC bytes over adjacent frames

Compact Disk Read Only Memory

The Compact Disk Read Only Memory (CD-ROM) was designed as the storage format for general computer data - in addition to uncompressed audio data.

CD-DA provides a suitable means for the handling of typical errors caused by damage or dust. The CD-DA specification became the basis of a **family** of optical storage media.

But not conceived for:

- video (different ECC, EDC scheme required)
- discrete data (error rate too high)
- simultaneous play back of various media

For computers there is a need for storage of:

- Data, audio, compressed audio and video

The Yellow Book CD-ROM Standard

- CD-ROM mode 1: for any data
- CD-ROM mode 2: for compressed audio and video data
- But cannot be combined on a single track

Within a single track:

- Only CD-DA audio or only CD-ROM specific data

Mixed Mode Disc:

- Data tracks at the beginning
- Subsequent tracks for audio data

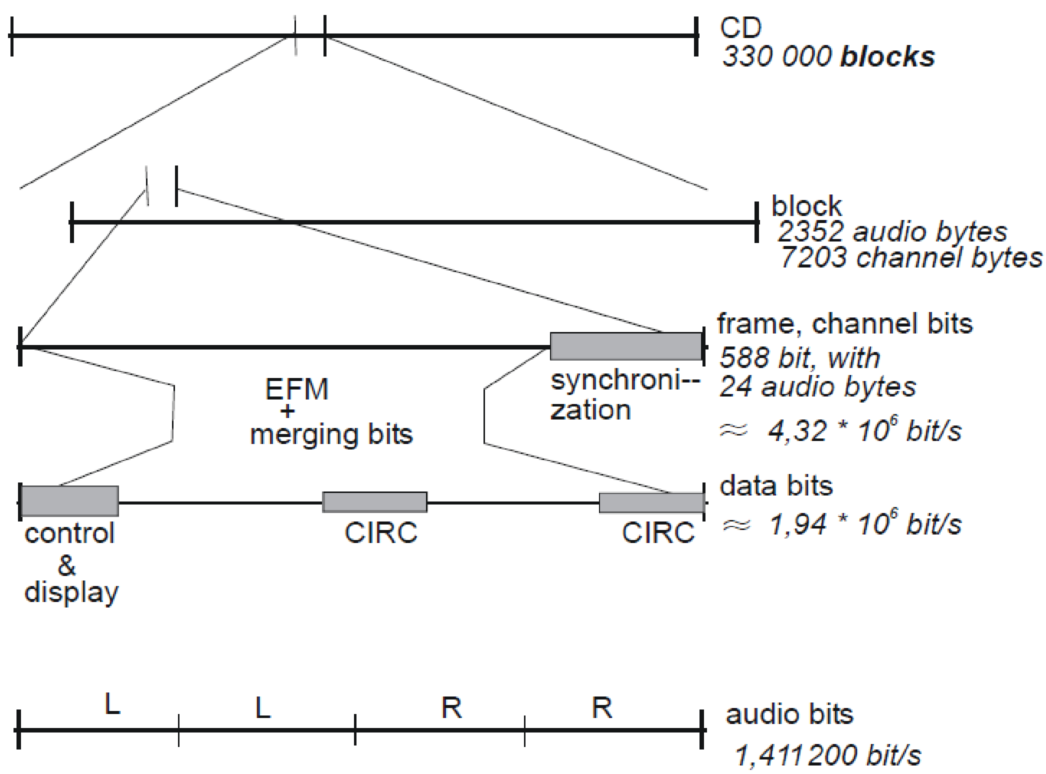


Figure 6.4: CD-ROM data hierarchy with audio data.

CD-ROM: Structure

Fine granularity for random access

- Tracks and Index Points not sufficient
- Structure with a higher resolution: the **block**
- Blocks contain a fixed number of frames

Disk structure

- 1 block = 32 frames
- 75 blocks/s (for a single-speed CD-ROM)
- 1411200 bit/s / 75 blocks/s / 8 bits/byte = 2352 bytes/block

Allows for

- Random access
- Better EDC, ECC

CD-ROM Mode

The CD-ROM was specified with the following goal: it should serve to hold uncompressed CD-DA data and computer data. This goal is achieved by introducing two modes: mode 1 and mode 2.

CD-ROM Mode 1

The block contains 2,048 bytes for information storage out of the available 2,352 bytes.

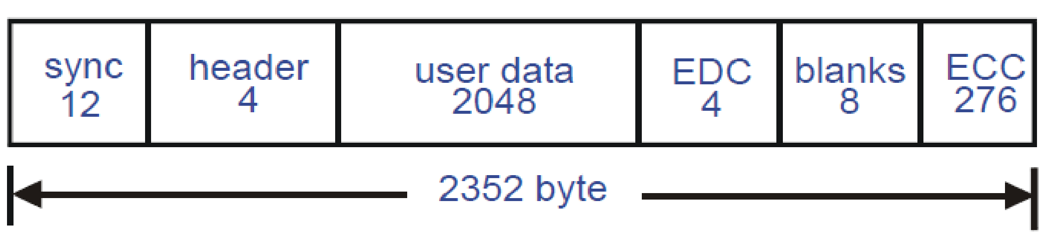


Figure 6.5: CD-ROM mode 1 block layout

The 2,352 bytes are split into the following groups:

- *12 bytes for synchronization; i.e.*, for the detection of the block beginning.

- *4 bytes for the header,* which carries an unambiguous specification of the block. The first byte stores minutes, the second byte stores seconds and the third byte contains the block number. The fourth byte includes the mode specification.
- *2,048 bytes* for the user data.
- *4 bytes* for error detection.
- *8 unused bytes.*
- *276 bytes* for error correction.

A CD-ROM contains 333,000 blocks to be played in 74 minutes. The capacity of a CD-ROM with all blocks in mode 1 can be computed as follows:

$$Capacity_{CD-ROM \bmod e1} = 333,000\ blocks \times 2048\ \frac{bytes}{block} = 681,984,000\ bytes$$

$$= 681,984,000 \times \frac{1}{1024\ \frac{bytes}{Kbytes}} \times \frac{1}{1024\ \frac{Kbytes}{Mbyte}} \approx 660\ Mbytes$$

The data rate in mode 1 is:

$$Rate_{CD-ROM \bmod e1} = 2,048\ \frac{bytes}{Blocks} \times 75\ \frac{Blocks}{s} = 153.6\ \frac{Kbytes}{s} \equiv 150\ \frac{Kbytes}{s}$$

CD-ROM Mode 2

CD-ROM mode 2 holds data of any media. The data layout of a CD-ROM block in mode 2 is shown in Figure 6.

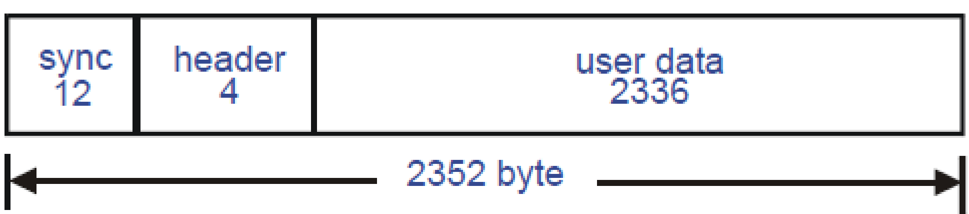


Figure 6.6: CD-ROM mode 2 block layout

The capacity and data rate of a CD-ROM with all blocks in mode 2 be computed as follows:

$$Capacity_{CD-ROM\ \mathrm{mod}\ e2} = 333{,}000\ blocks \times 2336\ \frac{bytes}{block} \approx 777{,}888{,}000\ bytes$$

$$Rate_{CD-ROM\ \mathrm{mod}\ e2} = 2336\ \frac{bytes}{Blocks} \times 75\, blocks\ /\ s \approx 175.2\, Kbyets\ /\ s$$

CD-ROM Extended Architecture

The Compact Disk Read Only Memory/Extended Architecture (CD-ROM/XA) standard was established by N.V. Phillips and the Sony and Microsoft Corporations and is based on the CD-ROM specification.

CD-ROM/XA differentiates blocks with form 1 and form 2 formats. This is similar to the CD-ROM modes:

CD-ROM/XA Form 1

This CD-ROM mode 2 XA format provides improved error detection and correction. Analogous to the CD-ROM mode 1, four bytes are needed for detection and 276 bytes for correction. Contrary to CD-ROM mode 1, the unused eight bytes of mode 1 are used for subheaders. Figure 7 shows a block where 2,048 bytes are used as data.

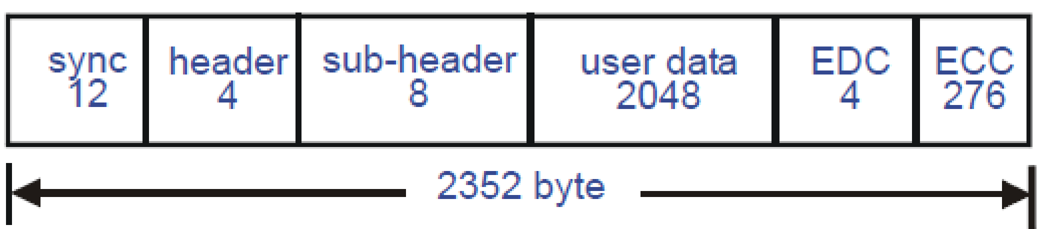


Figure 6.7: CD-ROM/XA block layout according to "Green Book" - layout of a CD-ROM block in mode 2, form 1

CD-ROM/XA Form 2

This CD-ROM mode 2 XA format allows 13% more storage capacity out of the entire block size (2,352 bytes) for user data, which means 2,324 bytes for user data. This is gained at the expense of worse error handling. In these form 2 blocks, compressed data of different media, including audio and video, can be stored.

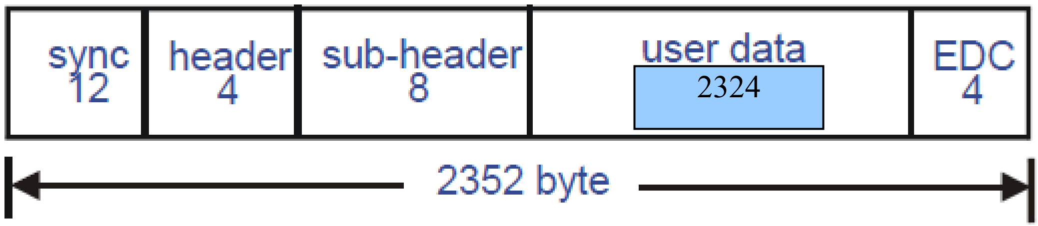


Figure 6.8: CD-ROM/XA block layout according to the "Green Book" - layout of a CD-ROM block in mode 2, form 2.

Principle of the CD-WO

The Compact Disk Write Once (CD-WO), like WORM (Write Once Read Many), allows the user to write once to a CD and afterwards to read it many times.

In the case of a CD-WO, an absorption layer exists between the substrate and the reflective layer. This layer can be irreversible change of the reflection characteristics by heating up the absorption layer ("burning"). The absorption layer in the pre-grooved track is heated to above 250^0C with a laser three to four times the intensity of a reading player.

CD-MO: Compact Disc Magneto Optical

The Compact Disk Magneto Optical (CD-MO) has a high storage capacity and allows on to write multiple times to the CD.

The magnetic-optical method is based on the polarization of the magnetic field where the polarization is caused by a heat.

To be written, the block (sector) is heated to above 150^0C. Simultaneously, a magnetic field approximately 10 times the strength of the earth's magnetic field is created. The individual dipoles in the material are then polarized according to this magnetic field. Hereby, a pit corresponds to a low value of the magnetic field. A land is coded through a high value of the magnetic field.

After the CD is irradiated with a laser beam, the polarization of the light changes corresponding to the existing magnetization. Using this process, the read operation is executed.

For a delete activity a constant magnetic field is created in the area of a block and the sector is simultaneously heated.

Feature:

- Write data
- Read data
- Erase data
- Rewrite data

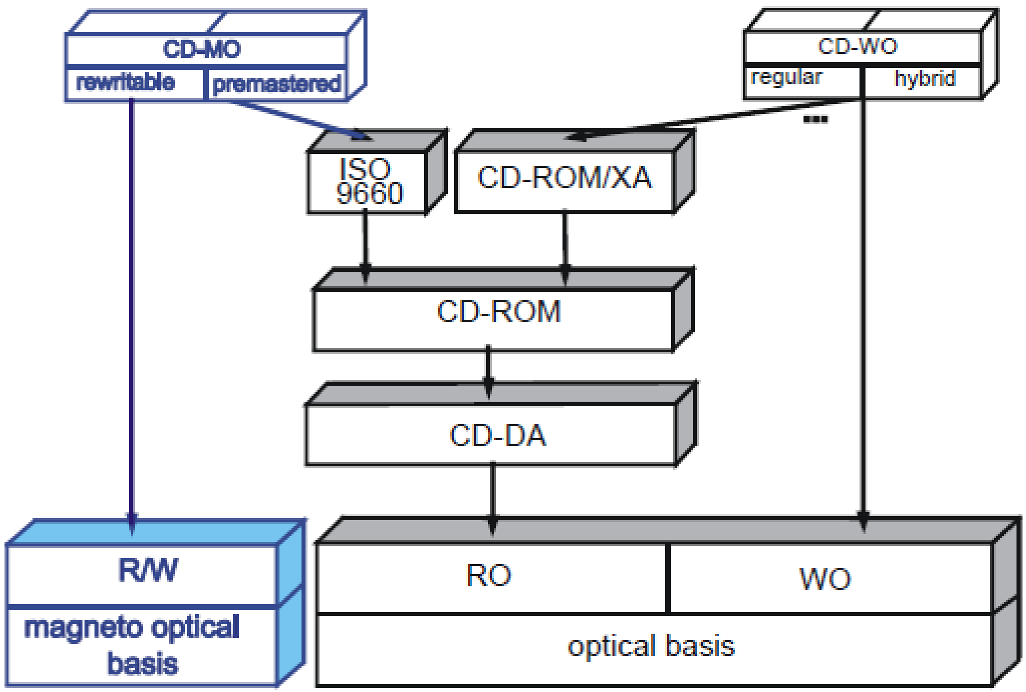


Figure 6.9: CD-WO and CD-M

Chapter 7

Computer Technology and Multimedia Operating Systems (MOS)

Multimedia Operating System

The operating system is the shield of the computer hardware against all software components. It provides a comfortable environment for the execution of programs, and it ensures effective utilization of the computer hardware. The operating system offers various services related to the essential resources of a computer: CPU, main memory, file system, storage, networking device and all input and output devices. *Operating system is the layer between hardware and the applications.*

Multimedia application demands that human perceive these media in a natural, error free-way. *An audio application will involve:*

- I/O devices: microphone, speakerphone
- CPU: processing the data
- Memory: temporarily store the data
- Network: real-time transmission of the data

The key to the Multimedia OS is the real-time processing of the continuous data:

- *CPU management:* appropriate scheduling is necessary.
- *Memory management:* guaranteed timing delay and buffer management.
- *File system:* allows transparent and guaranteed continuous retrieval of audio and video data to any application using the file system.
- To guarantee real-time processing of the media data, one concept used is resource reservation. The resource is reserved prior to the execution of the application. This has to be performed for all components along the data path of a multimedia application.

Real-Time Multimedia OS

A real-time process is a process which delivers the results of the processing in a given time-span.

Real-time system:

"A system in which the correctness of a computation depends not only on obtaining the right result, but also upon providing the result on time."

Deadline:-

A deadline represents the latest acceptable time for the presentation of a processing result. It marks the border between normal (correct) and anomalous (failing) behavior. A real-time system has both hard and soft deadlines.

Soft deadline:

The term soft deadline is often used for a deadline which cannot be exactly determined and which failing to meet does not produce and unacceptable result. If violated does not result in unacceptable results.

Hard deadline:

Hard deadlines should never be violated. A hard deadline violation is a system failure. Hard deadlines are determined by the physical characteristics of real-time processes.

Characteristics of Real-time Systems

- Predictably fast response to time-critical events and accurate timing information. For example in the control system of a nuclear power plant, the response to a malfunction must occur within a well-defined period to avoid a potential disaster.
- High degree of schedulability. Schedulability refers to the degree of resource utilization at which, the deadline of each time-critical task can be taken into account.
- Stability under transient overload. Under system overload, the processing of critical tasks must be ensured. These critical tasks are vital to the basic functionality provided by the system.

Some traditional real-time applications include manufacturing and monitoring systems. New applications include multimedia system, surveillance, etc.

Characteristics of Multimedia OS

- More fault-tolerant – compare to the real-time system for nuclear power plant control, the system in a video playback product will cause less damage if some errors occur.
- Deadlines tend to be soft – for example, small errors in video playback timing is not noticeable.
- Schedulability consideration is much easier because the media streams tend to be periodic (results of sampling) and consistent.
- Bandwidth requirement is not always stringent – more compression or lower resolution can always be used to achieve lower bit rates.

Resource Management

Multimedia is a real-time application. This implies a certain amount of data must be handled within a specified time frame. Multimedia requires a tremendous amount of resources to accommodate. Audio and video files consume large space on secondary storage. Computing and processing multimedia files requires many CPU cycles and efficient I/O. Some argue today's workstations and networks do not meet the requirements of current multimedia applications.

Essentially, they cannot ensure consistent, on-time data delivery. There are three reasons for this:

- Capacity of system resources is too low (performance is too low)
- The existing resources are not assigned to tasks efficiently (resource scheduling is poor).
- Access to resources is not controlled properly to avoid conflict (resource reservation is insufficient).

Multimedia systems with integrated audio and video processing are at the limit of their capacity, even with data compression and utilization of new technologies. Current computers do not allow processing of data according to their deadlines without any resource reservation and real-time process management. Processing is this context refers to any kind of manipulation and communication of data. This stage of development is known as the window of insufficient resources.

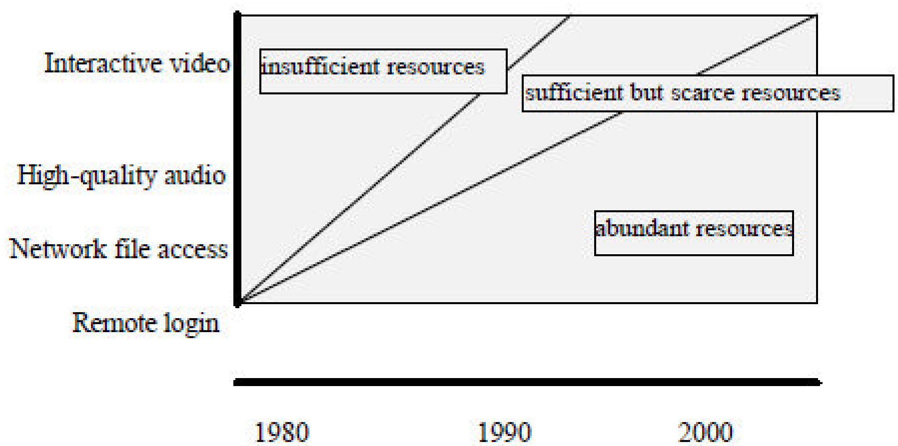


Figure: Window of insufficient resources.

Resources:

A resource is a system entity required by tasks for manipulating data. Each resource has a set of distinguishing characteristics classified using the following scheme:

- Active resources and passive resource:

  Active resources: An active resource is the CPU or a network adapter for protocol processing; it provide service.

  Passive resources: A passive resource is the main memory, communication bandwidth or a file systems; it denotes some system capability required by active resources.
- Resources can be used exclusively by one process at a time or shared between various processes.

Active resources: Active resources are often ***exclusive***.

Passive resources: Passive resources can usually be ***shared*** among processes.

- A resource that exists only once in the system is known as a single, otherwise it is a multiple resource. In a transputer-based multiprocessor system, the individual CPU is a multiple resource.

Process Management

Process management deals with the resource main processor. The capacity of this resource is specified as processor capacity. The process manager maps single processes onto resources according to a specified scheduling policy such that all processes meet their requirements. In most systems, a process under control of the process manager can adopt one of the following states:

- In the initial state, no process is assigned to the program. The process in the idle state.
- If a process is waiting for an event, i.e., the process lacks one of the necessary resources for processing, it is the blocked state.
- If all necessary resources are assigned to the process, it is ready to run. The process only needs the processor for the execution of the program.
- A process is running as long as the system processor is assigned to it.

The process manager is the scheduler. This component transfers a process into the ready to run state by assigning it a position in the respective queue of the dispatcher, which is the essential part of the operating system kernel. The dispatcher manages the transition from ready to run to run. In most operating systems, the next process to run is chosen according to a priority policy. Between processes with the same priority, the one with the longest ready time is chosen.

File Systems

The file system is said to be the most visible part of an operating system. Most programs write or read files. Their program code, as well as user data, are stored in files. The organization of the file system is an important factor for the usability and convenience of the operating system. A file is a sequence of information held as a unit for storage and use in a computer system.

Files are stored in secondary storage, so they can be used by different applications.

The file system provides access and control functions for the storage and retrieval of files. From the user's viewpoint, it is important how the file system allows file organization and structure. The internals, which are more important in our context i.e, the organization of the file system, deal with the representation of information in files, their structure and organization in secondary storage.

Chapter 8
Documentation, Hypertext and MHEG

Documents

A document consists of a set of structural information that can be in different forms of media, and during presentation can be generated or recorded. A document is aimed at the perception of a human, and is accessible for computer processing.

A multimedia document is a document which is comprised of information coded in at least one continuous (time-dependent) medium and in one discrete (time-independent) medium. Integration of the different media is given through a close relation between information units. This is also called synchronization. A multimedia document is closely related to its environment of tools, data abstractions, basic concepts and document architecture.

Document Architecture

Exchanging documents entails exchanging the document content as well as the document structure. This requires that both documents have the same document architectures are the Standard Generalized Markup Language (SGML) and the Open Document Architecture (ODA). There are also proprietary document architectures, such as DEC's Document Content Architecture (DCA) and IBM's Mixed Object Document Content Architecture (MO:DCA).

Information architectures use their data abstractions and concepts. A document architecture describes the connections among the individual elements represented as models (e.g., presentation model, manipulation model). The elements in the document architecture and their relations are shown in Figure.

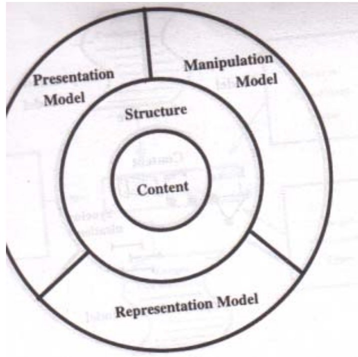


Figure: Document architecture and its elements.

The manipulation model describes all the operations allowed for creation, change and detection of multimedia information. The representation model defines: (1) the protocols for exchanging this information among different computers; and, (2) the formats for storing the data. It includes the relations between the individual information elements which need to be considered during presentation. It is important to mention that an architecture may not include all described properties, respectively models.

Manipulation of Multimedia Data

The user becomes most aware of multimedia documents through tools for manipulation of multimedia data, such as editors, desktop publishing programs and other text processing programs.

A document undergoes the process shown in Figure. The information included in a document belongs to a certain document type, e.g., a business letter or an internal memorandum. The same document can belong to other types which mainly influence the final representation. The transformation from the actual information to its final representation behaves according to rules specific to the document architecture.

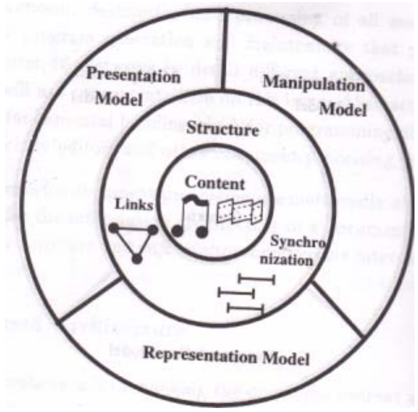


Figure: A multimedia document architecture and its constituent elements.

The processing cycles of a traditional document and an interactive multimedia presentation are analogous, as shown in figure.

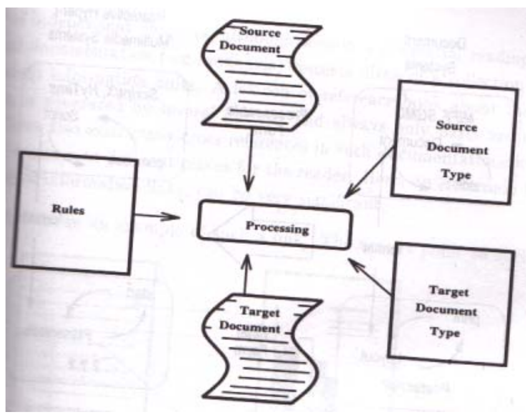


Figure: Processing of a document: from the information to the presentation

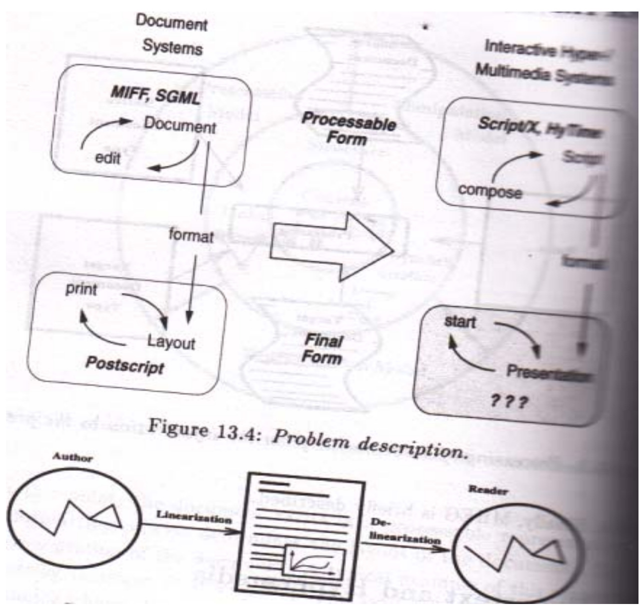


Figure: Information transmission

Currently, an author edits a document with a text editor. Thus, he or she uses the system's character set (e.g., ASCII) as the actual content of a document, as well as a hidden language available in most interactive editors for structural description (e.g., SGML).

Hypertext, Hypermedia and Multimedia

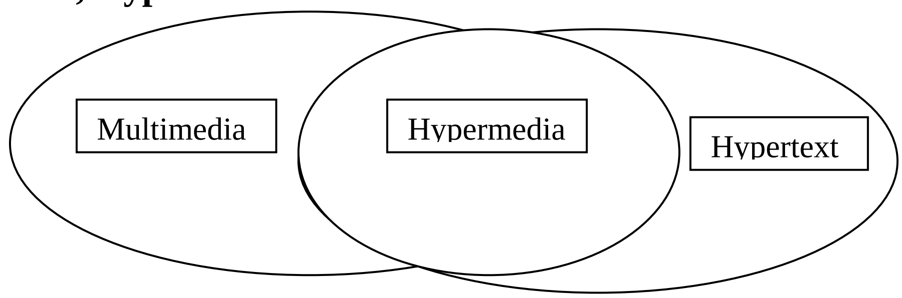


Figure: The hypertext, hypermedia and multimedia relationship.

Hypertext System

A hypertext system is mainly determined through non-linear links of information. Pointers connect the nodes. The data of different nodes can be represented with one or several media types. In a pure text system, only text parts are connected. We understand hypertext as an information object which includes links to several media.

Multimedia System

A multimedia system is characterized by computer-controlled, integrated production, manipulation, presentation, storage and communication of independent information, which is encoded at least through a continuous (time-dependent) and a discrete (time-independent) medium.

For example, if only links to text data are present, then this is not a multimedia system. It is a hypertext.

Hypermedia System

A hypermedia system includes the non-linear information links of hypertext systems and the continuous and discrete media of multimedia systems. For example, if a non-linear link consists of text and video data, then this is a hypermedia, multimedia and hypertext system.

Hypertext System Architecture

The hypertext system architecture can be divided into three layers with different functionalities:

(i) Presentation Layer

All functions connecter to the user interface are embedded. Here, nodes and pointer are mapped to the user interface. At the user interface, one or several parts of the document are visualized. This layer determines, based on the given structure and user's desired display, which data are presented and how they are presented. This layer takes over control of all inputs.

(ii) Hypertext Abstract Machine

The Hypertext Abstract Machine (HAM) is placed between the presentation and storage layers. It can expect from the underlying layer database functions for storage of multimedia data in a distributed environment. It does not have to consider input and output of the upper layer (Presentation Layer). Hypertext Abstract Machine knows the structure of the document, it has the knowledge about the pointers and its attributes. The data structure, respectively a document architecture, is constructed for the management of the document.

(iii) Storage Layer

The storage layer (also called the database layer) is the lowest layer. All functions connected with the storage of data, i.e., secondary storage management, belong to this layer. The specific properties of the different discrete and continuous media need to be considered. Functionalities from traditional database systems are expected, such as persistence (data persist through programs and processes), multi-user operations (synchronization, locks) and the restoration of data after a failure (transaction).

Document Architecture SGML

The Standard Generalized Markup Language (SGML) The Standard Generalized Markup Language, or SGML, is a language and notation for describing classes of documents. In an SGML-encoded document, the various *elements* of the document are delimited with distinguished character strings commonly called *tags*. A document may, for example, have tags that delimit elements like paragraphs, subsections, appendices, and figures. A *start-tag* of the form <X> denotes the beginning of an element, and an *end-tag* of the form </X> denotes the end of that element.

An SGML-encoded document has three parts: an *SGML declaration*, a *document type declaration*, and a *document instance*. The SGML declaration characterizes the document type declaration, and the subsequent document instance(s), in terms of character sets and optional features of SGML. The SGML declaration may be omitted if the default character set is used and no optional features are required.

In SGML, a class of documents is characterized by a grammar that indicates what markup is allowed, what markup is required, and how markup is distinguished from text. SGML defines this grammar with a *document type definition.*

SGML document is divided into two processes. Only the formatter knows the meaning of the tag and it transforms the document into a formatted document. The parser uses the tags, occurring in the document, in combination with the corresponding document type. Specification of the document structure is done with tags.

Tags are divided into different categories:

The descriptive markup (tags) describes the actual structure always in the form:

```
<start-tag> respectively also </end-tag>
```

An example is the definition of a paragraph at its beginning:

```
<paragraph> The text of the paragraph follows ......
```

The entity reference provides connection to another element. This element replaces the entity reference. This can be understood also as an abbreviation to which the actual content can be copied later at the corresponding place. The following example shows entity reference in a mathematical context:

&square x .... should be x^2

The markup declarations define the elements to which an entity reference refers. In our example of squaring a variable x, square is defined as:

```
<!ELEMENT square (....)>
```

A markup declaration can be used to define rules for the structure (the classes). The following example illustrates the construction of an article paper:

```
<!ELEMENT paper (preamble, body, postamble)>
<!ELEMENT preamble (title, author, side)>
<!ELEMENT title (#CDATA)> -- character data
<!ELEMENT body (....)>
```

Instruction for other programs in a text are entered through processing instructions. They can be meant, for example, for the formatter. Using processing instructions, different media can be inserted.

Document Architecture ODA

The Open Document Architecture (ODA) was initially called the Office Document Architecture because it supports mostly office-oriented applications. The main goal of this document architecture is to support the exchange, processing and presentation of documents in open system.

The main property of ODA is the distinction among content, logical structure and layout structure. This is in contrast to SGML where only a logical structure and the contents are defined. ODA also defines semantics. Figure shows these three aspects linked to a document.

(i) Content Portions:

The content of the document consists of Content Portions. These can be manipulated according to the corresponding medium.

A content architecture describes for each medium:

- the specification of the elements,
- the possible access functions and
- the data coding.

Individual elements are the Logical Data Units (LDUs), which are determined for each medium. The access functions serve for the manipulation of individual elements. The coding of the data determines the mapping with respect to bits and bytes.

(ii) Layout Structure:

The layout structure specifies mainly the representation of a document. It is related to a two dimensional representation with respect to a screen or paper.

(iii) Logical Structure:

The logical structure includes the partitioning of the content.

MHEG (Multimedia and Hypermedia Information Coding Expert Group)

The committee ISO/IEC JTC/SC29 (Coding of Audio, Picture, Multimedia and Hypermedia Information) works on the standardization of the exchange format for multimedia systems. The actual standards are developed at the international level in three working groups cooperating with research and industry. Figure shows that the three standards deal with the coding and compression of individual media. The results of the working groups: the Joint Photographic Expert Group (JPEG) and the Motion Picture Expert Group (MPEG) are of special importance in the area of multimedia systems.

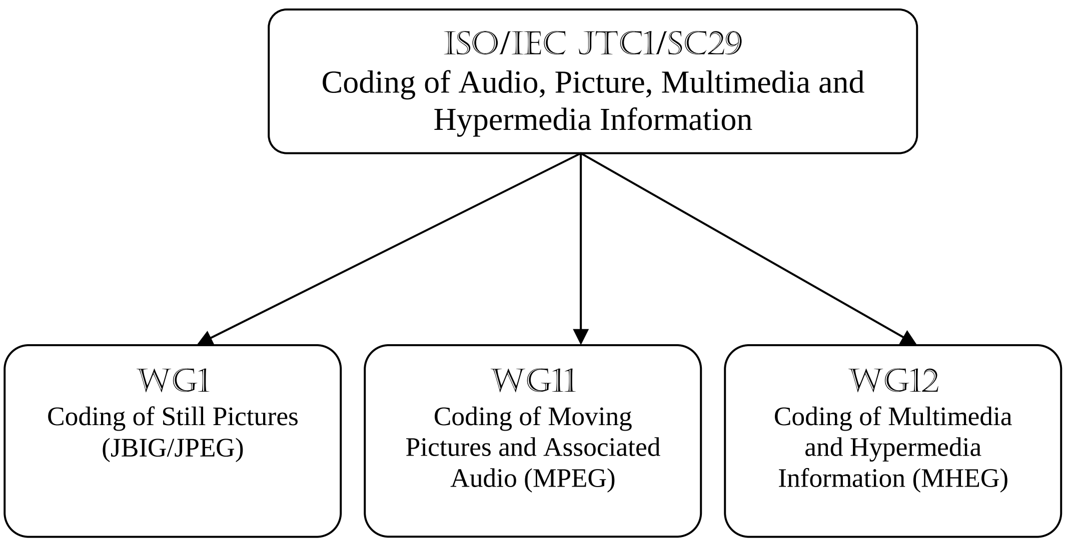


Figure: Working Groups within the ISO-SG29

The standard of this structure description is the subject of the working group WG12, which is known as the Multimedia and Hypermedia Information Coding Expert Group (MHEG). The name of the developed standard is officially called Information Technology-Coding of Multimedia and Hypermedia Information (MHEG). The final MHEG standard will be described in three documents.

Chapter 9
Multimedia Communication Systems

9.1 Definition of Multimedia Communication

The consideration of multimedia applications supports the view that local systems expand toward distributed solutions. Applications such as kiosks, multimedia mail, collaborative work systems, virtual reality applications and others require high-speed networks with a high transfer rate and communication systems with adaptive, lightweight transmission protocols on top of the networks.

From the communication perspective, we divide the higher layers of the Multimedia Communication System (MCS) into two architectural subsystems: *an application subsystem* and *a transport subsystem.*

9.2 Application Subsystem

Collaborative Computing

The current infrastructure of networked workstations and PCs, and the availability of audio and video at these end-points, makes it easier for people to cooperate and bridge space and time. In this way, network connectivity and end-point integration of multimedia provides users with a collaborative computing environment. Collaborative computing is generally known as Computer-Supported Cooperative Work (CSCW).

There are many tools for collaborative computing, such as electronic mail, bulletin boards (e.g., Usenet news), screen sharing tools (e.g., ShowMe from Sunsoft), text-based conferencing systems(e.g., Internet Relay Chat, CompuServe, American Online), telephone conference systems, conference rooms (e.g., VideoWindow from Bellcore), and video conference systems (e.g., MBone tools nv, vat). Further, there are many implemented CSCWsystems that unify several tools, such as Rapport from AT&T, MERMAID from NEC and others.

Collaborative Dimensions

Electronic collaboration can be categorized according to three main parameters: time, user scale and control. Therefore, the collaboration space can be partitioned into a three-dimensional space.

Time

With respect to time, there are two modes of cooperative work: asynchronous and synchronous. Asynchronous cooperative work specifies processing activities that do not happen at the same time; the synchronous cooperative work happens at the same time.

User Scale

The user scale parameter specifies whether a single user collaborates with another user or a group of more than two users collaborate together. Groups can be further classified as follows:

- A group may be *static* or *dynamic* during its lifetime. A group is *static* if its participating members are pre-determined and membership does not change during the activity. A group is *dynamic* if the number of group members varies during the collaborative activity, i.e., group members can join or leave the activity at any time.
- Group members may have different roles in the CSCW, e.g., a member of a group (if he or she is listed in the group definition), a participant of a group activity (if he or she successfully joins the conference), a conference initiator, a conference chairman, a token holder or an observer.
- Groups may consist of members which have *homogeneous* or *heterogeneous* characteristics and requirements of their collaborative environment.

Control

Control during collaboration can be *centralized* or *distributed*.

Centralized control means that there is a chairman (e.g., main manger) who controls the collaborative work and every group member (e.g., user agent) reports to him or her. *Distributed control* means that every group member has control over his/her own tasks in the collaborative work and distributed control protocols are in place to provide consistent collaboration.

Other partition parameter may include *locality*, and *collaboration awareness*.

Locality partition means that a collaboration can occur either in the same place (e.g., a group meeting in an officer or conference room) or among users located in different place through tele-collaboration.

Group communication systems can be further categorized into *computer-augmented collaboration systems*, where collaboration is emphasized, and *collaboration-augmented computing systems*, where the concentrations are on computing.

Group Communication Architecture

Group communication (GC) involves the communication of multiple users in a synchronous or an asynchronous mode with centralized or distributed control.

A group communication architecture consists of a *support model*, *system model* and *interface model*. The GC support model includes group communication agents that communicate via a multi-point multicast communication network as shown in following figure.

Group communication agents may use the following for their collaboration:

Group Rendezvous

Group rendezvous denotes a method which allows one to organize meetings, and to get information about the group, ongoing meetings and other static and dynamic information.

Shared Applications

Application sharing denotes techniques which allow one to replicate information to multiple users simultaneously. The remote users may point to interesting aspects (e.g., via tele-pointing) of the information and modify it so that all users can immediately see the updated information (e.g., joint editing). Shared applications mostly belong to collaboration transparent applications.

<u>Conferencing</u>

Conferencing is a simple form of collaborative computing. This service provides the management of multiple users for communicating with each other using multiple media. Conferencing applications belong to collaboration-aware applications.

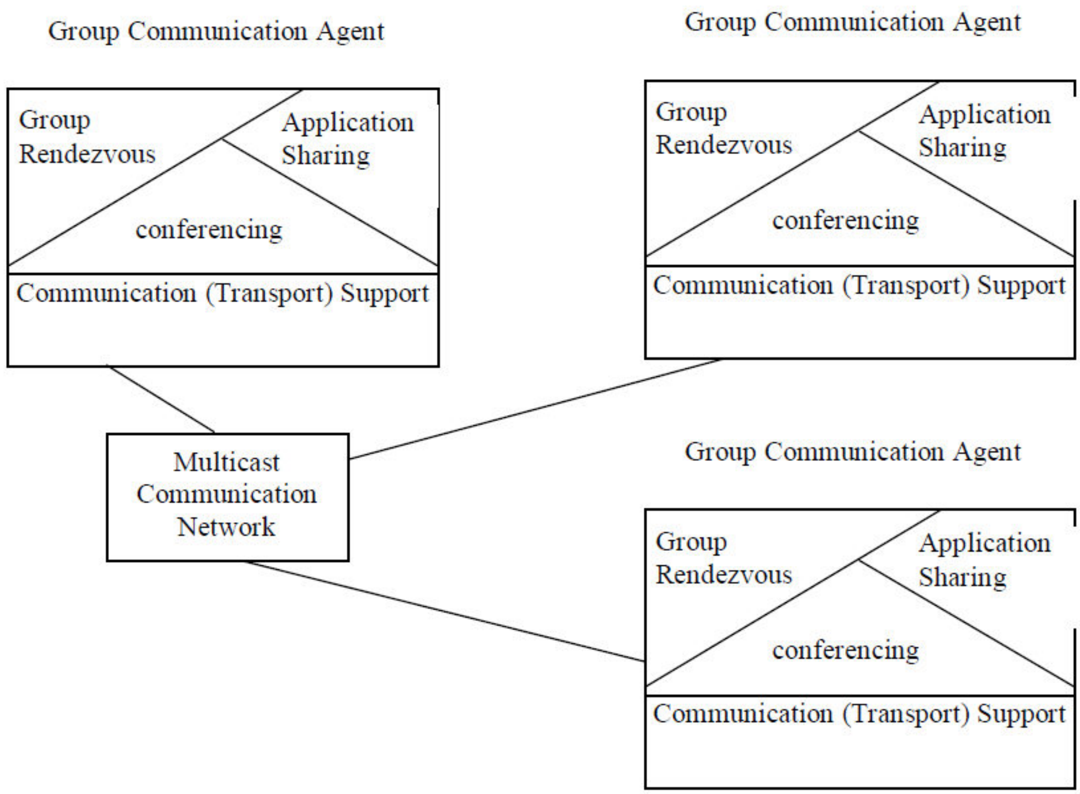


Figure 9.1: Group communication support model

The GC system model is based on a client-server model. Clients provide user interfaces for smooth interaction between group members and the system. Servers supply functions for accomplishing the group communication work, and each server specializes in its own function.

Application Sharing Approach

Sharing applications is recognized as a vital mechanism for supporting group communication activities. Sharing applications means that when a shared application program (e.g., editor) executes any input from a participant, all execution results performed on the shared object (e.g., document text) are distributed among all the participants. Shared objects are displayed, generally, in shared windows.

Application sharing is most often implemented in collaboration-transparent systems, but can also be developed through collaboration-aware, special-purpose applications. An example of a software toolkit that assists in development of shared computer applications is Bellcore's Rendezvous system (language and architecture).

Shared applications may be used as conversational props in tele-conferencing situations for collaborative document editing and collaborative software development.

An important issue in application sharing is shared control. The primary design decision in sharing applications is to determine whether they should be centralized or replicated:

Centralized Architecture

In a centralized architecture, a single copy of the shared application runs at one site. All participants' input to the application is then distributed to all sites.

The advantage of the centralized approach is easy maintenance because there is only one copy of the application that updates the shared object. The disadvantage is high network traffic because the output of the application needs to be distributed every time.

Replicated Architecture

In a replicated architecture, a copy of the shared application runs locally at each site. Input events to each application are distributed to all sites and each copy of the shared application is executed locally at each site.

The advantages of this architecture are low network traffic, because only input events are distributed among the sites, and low response times, since all participants get their output from local copies of the application. The disadvantages are the requirement of the same execution environment for the application at each site, and the difficulty in maintaining consistency.

Conferencing

Conferencing supports collaborative computing and is also called synchronous tele-collaboration.

Conferencing is a management service that controls the communication among multiple users via multiple media, such as video and audio, to achieve simultaneous face-to-face communication. More precisely, video and audio have the following purposes in a tele-conferencing system:

Video is used in technical discussions to display view-graph and to indicate how many users are still physically present at a conference. For visual support, workstations, PCs or video walls can be used.

Audio is an important component in tele-conferencing for describing and clarifying visual information. Therefore, quality audio, with true full-duplex communication and echo cancellation, and possibly enhanced with spatial queues, is necessary.

Session Management

Session management is an important part of the multimedia communication architecture. It is the core part which separates the control, needed during the transport, from the actual transport. Session management is extensively studied in the collaborative computing area; therefore we concentrate on architectural and management issues in this area.

A session management architecture is built around an entity-session manager which separates the control from the transport. By creating a reusable session manager, which is separated from the user-interface, conference-oriented tools avoid a duplication of their effort.

The session control architecture consists of the following components:

Session Manager

Session manager includes local and remote functionalities. Local functionalities may include:

- Membership control management, such as participant authentication or presentation of coordinated user interfaces;
- Control management for shared workspace, such as floor control

- Media control management, such as intercommunication among media agents or synchronization
- Configuration management, such as an exchange of interrelated QoS parameters of selection of appropriate services according to QoS; and
- Conference control management, such as an establishment, modification and a closing of a conference.

Media agents

Media agents are separate from the session manager and they are not responsible for decisions specific to each type of media. The modularity allows replacement of agents. Each agent performs its own control mechanism over the particular medium, such as mute, unmute, change video quality, start sending, stop sending, etc.

Shared Workspace Agent

The shared workspace agent transmits shared objects (e.g., telepointer coordinate, graphical or textual object) among the shared application.

Session Control

Each session is described through the session state. This state information is either private or shared among all session participants. Dependent on the functions, which an application required and a session control provides, several control mechanisms are embedded in session management:

Floor control:

In a shared workspace, the floor control is used to provide access to the shared workspace. The floor control in shared application is often used to maintain data consistency.

Conference Control:

In conferencing applications, Conference control is used.

Media control:

This control mainly includes a functionality such as the synchronization of media streams.

Configuration Control:

Configuration control includes a control of media quality, QOS handling, resource availability and other system components to provide a session according to user's requirements.

Membership control:

This may include services, for example invitation to a session, registration into a session, modification of the membership during the session etc.

9.3 Transport Subsystem

Requirements

Distributed multimedia applications put new requirements on application designers, As well as network protocol and system designers. We analyze the most important enforced requirements with respect to the multimedia transmission.

Network multimedia applications by themselves impose new requirements onto data handling in computing and communications because they need:

- substantial data throughput,
- fast date forwarding,
- service guarantees, and
- multicasting

Data Throughput

Audio and video data resemble a stream-like behavior, and they demand, even in a compressed mode, high data throughput. In a workstation or network, several of those streams may exist concurrently demanding a high throughput.

Fast Data Forwarding

Fast data forwarding imposes a problem on end-systems where different applications exist in the same end-system, and they each require data movement ranging from normal, error-free data transmission to new time-constraint traffic types. But generally, the faster a communication system can transfer a data packet, the fewer packets need to be buffered. The requirement leads to a careful spatial and temporal resource management in the end-systems and routers/switches.

Service Guarantees

Distributed multimedia applications need service guarantees; otherwise their acceptance does not come through as these systems, working with continuous media, compete against radio and television services. To achieve services guarantees, resource management must be used. Without resource management in end-systems and switches/routers, multimedia systems cannot provide reliable QoS to their users because transmission over unreserved resources leads to dropped or delayed packets.

Multicasting

Multicast is important for multimedia-distributed applications in terms of sharing resources like the network bandwidth and the communication protocol processing at end-systems.

Transport Layer

Transport protocols, to support multimedia transmission, need to have new features and provide the following function, semi-reliability, multicasting, NAK (None-Acknowledgment)-based error recovery mechanism and rate control.

First, we present transport protocols, such as TCP and UDP, which are used in the Internet protocol stack for multimedia transmission, and secondly we analyze new emerging transport protocols, such as RTP, XTP and other protocols, which are suitable for multimedia.

Transmission Control Protocol (TCP)

Early implementations of video conferencing applications were implemented on top of the TCP protocol. TCP provides a reliable, serial communication path, or virtual circuit, between processes exchanging a full-duplex stream of bytes. Each process is assumed to reside in an internet host that is identified by an IP address.

Each process has a number of logical, full-duplex ports through which it can set up and use as full-duplex TCP connections.

User Datagram Protocol (UDP)

UDP is a simple extension to the Internet network protocol IP that supports multiplexing of datagrams exchanged between pairs of Internet hosts. It offers only multiplexing and checksumming, nothing else. Higher-level protocols using UDP must provide their own retransmission, packetization, reassembly, flow control, congestion avoidance, etc.

Real-time Transport Protocol (RTP)

RTP is an end-to-end protocol providing network transport function suitable for applications transmitting real-time data, such as audio, video or simulation data over multicast or unicast network services.

Xpress Transport Protocol (XTP)

XTP was designed to be an efficient protocol, taking into account the low error ratios and higher speeds of current networks. It is still in the process of augmentation by the XTP Form to provide a better platform for the incoming variety of applications. XTP integrates transport and network protocol functionalities to have more control over the environment in which it operates.

XTP is intended to be useful in a wide variety of environments, from real-time control systems to remote procedure calls in distributed operating systems and distributed databases to bulk data transfer. It defines for this purpose six service types: connection, transaction, unacknowledged data gram, acknowledged datagram, isochronous stream and bulk data.

In XTP, the end-user is represented by a context becoming active within an XTP implementation.

Network Layer

The requirements on the network layer for multimedia transmission are a provision of high bandwidth, multicasting, resource reservation and Qos guarantees, new routing protocols with support for streaming capabilities and new higher-capacity routers with support of integrated services.

Internet Protocol (IP)

IP provides for the unreliable carriage of datagrams form source host to destination host, possibly passing through one or more gateways (routers) and networks in the process.

Internet Group Management Protocol (IGMP)

Internet Group Management protocol (ICMP) is a protocol for managing Internet multicasting groups. It is used by conferencing applications to join and leave particular multicast group. The basic service permits a source to send datagrams to all members of a multicast group. There are no guarantees of the delivery to any or all targets in the group.

Resource reservation Protocol (RSVP)

RSVP is a protocol which transfers reservations and keeps a state at the intermediate nodes. It does not have a data transfer component. RSVP messages are sent as IP datagrams, and the router keeps "soft state", which is refreshed by periodic reservation messages. In the absence of the refresh messages, the routers delete the reservation after a certain timeout.

9.4 Quality of Service and Resource Management

Every product is expected to have a quality apart from satisfying the requirements. The quality is measured by various parameters.

Parameterization of the services is defined in ISO (International Standard Organization) standards through the notion of *Quality of Service (QoS)*. The ISO standard defines QoS as a concept for specifying how "good" the offered networking services are. QoS can be characterized by a number of specific parameters.

QoS Layering

Traditional QoS (ISO standards) was provided by the network layer of the communication system. An enhancement of QoS was achieved through inducing QoS into transport services. For MCS, the QoS notion must be extended because many other services contribute to the end-to-end service quality.

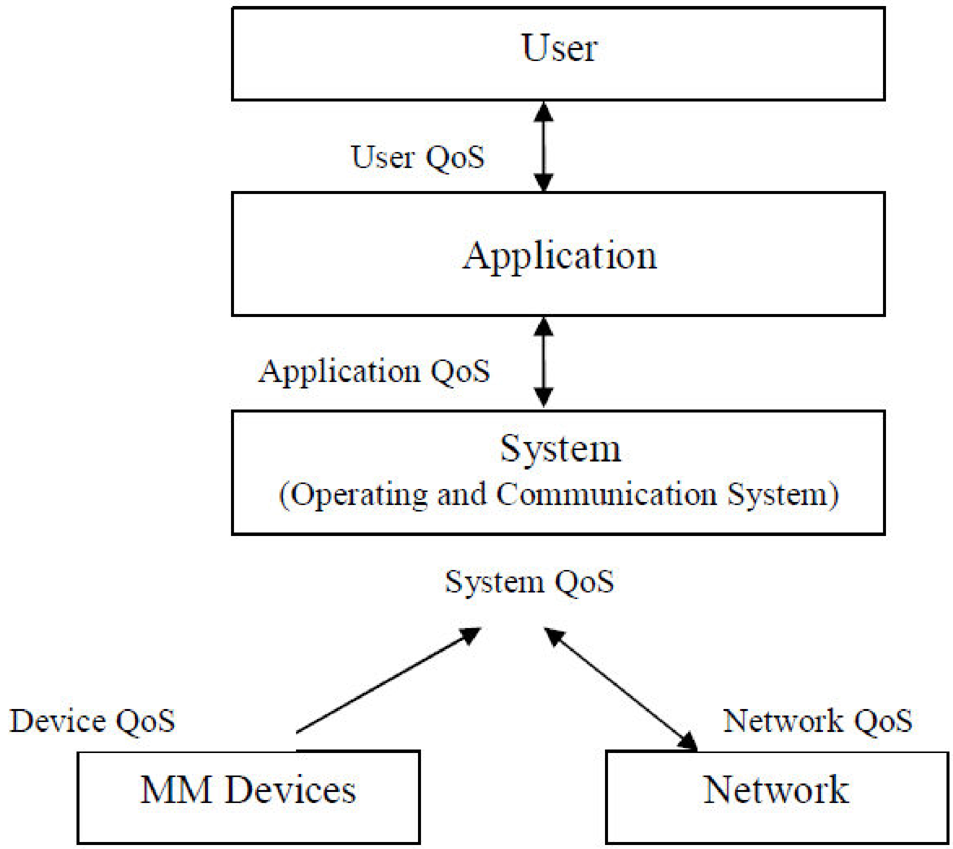


Figure 9.2: Quality of Service Layered model for the Multimedia Communication System

The MCS consists of three layers: application, system (including communication services and operating system services), and devices (network and Multimedia (MM) devices). Above the application may or may not reside a human user.

This implies the introduction of QoS in the application (application QoS), in the system (system QoS) and in the network (network QoS). In the case of having a human user, the MCS may also have a user QoS specification. We concentrate in the network layer on the network device and its QoS because it is of interest to us in the MCS. The MM devices find their representation (partially) in application QoS.

Resource Management

Multimedia systems with integrated audio and video processing are at the limit of their capacity, even with data compression and utilization of new technologies. Current computers do not allow processing of data according to their deadlines without any resource reservation and real-time process management.

Resource management in distributed multimedia systems covers several computers and the involved communication networks. It allocates all resources involved in the data transfer process between sources and sinks. In an integrated distributed multimedia system, several applications compete for system resources. This shortage of resources requires careful allocation. The system management must employ adequate scheduling algorithms to serve the requirements of the applications. Thereby, the resource is first allocated and then managed.

Resource management in distributed multimedia systems covers several computers and the involved communication networks. It allocates all resources involved in the data transfer process between sources and sinks.

Resource

A resource is a system entity required by tasks for manipulating data. Each resource has a set of distinguishing characteristics:

- There are *active* and *passive* resources. An *active* resource is, for example, the CPU or a network adapter for protocol processing; it provides a service. A *passive* resource is, for example, the main memory (buffer space) or bandwidth (link throughput); it denotes some system capabilities required by active resources.
- A resource can be either used *exclusively* by one process or *shared between various processes*. For example, a loudspeaker is an exclusive resource, whereas bandwidth is a shared resource.
- A resource that exists only once in the system is known as a *single resource*, otherwise it is a *multiple resource*. In a transputer-based multiprocessor system, the individual CPU is a multiple resource.