4th Class

2016-2017

Digital Video & Audio

الفديو والصوت الرقمي

أستاذ المادة : م. د. إقباس عز الدين

Introduction to Multimedia

Multimedia is content that uses a combination of different content forms such as text, audio, images, animation, video and interactive content. Multimedia contrasts with media that use only rudimentary بدائي computer displays such as text-only or traditional forms of printed or hand-produced material.

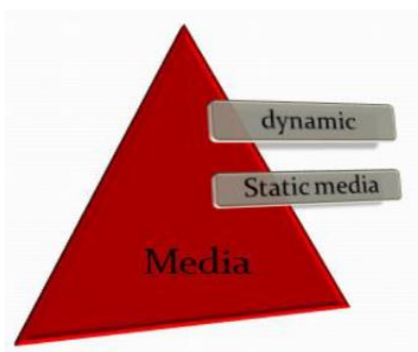
Multimedia Definition

Digital multimedia is any combination of two or more media, represented in a digital form, sufficiently well integrated to be presented via a single interface, or manipulated by a single computer program.

The perception of new information depends on the diversity of views and the presentation of this information to the learner; so that the desire for education increases, when added visual and audio effects to the education system (scientific research indicates that human receives more than 80% of knowledge through the sense of hearing and sight and about 13-20 % of those during the hearing, followed by the other senses, ranging from 1-5%, which senses touch, taste and smell) - for all this has been to focus on the selection and use of multi-media presentation of information technologies.

The terms **static** and **dynamic** media refer to the way in which a medium is presented.

- Static media, such as the newspaper, is printed in black and white and cannot be altered once it has been written. The static data are those in which the temporal component is not present and do not need a synchronization among the data.
- Dynamic media is constantly updated and is interactive with player control. Dynamic media is developing in the modern world and the classic example is a website. In the dynamic data the temporal variable is present.



Dynamic Media

1- Video
2- Animation
3- sound

Static Media

1-Image
2- Text

Delivery of multimedia

1- Online delivery uses a network to send information from one computer to another. The network may be LAN or WAN (internet).
2- Offline delivery used removable storage, CDs, etc.

Disadvantages of Multimedia

1- High monetary cost

- May require reformatting of information.
- Requires high amount of disk space.
- Requires powerful computer.
- Large amount of varied hardware is necessary like: CDROM drive, Sound card, Speakers and Scanner.

2- Technical expertise needed to set up multimedia system.

3- Encourages reliance (الأتكال) on technology.

- Users may begin to rely on the presence of technology
- Can be seen as only source of information.

Multimedia Systems

A Multimedia System is a system capable of processing multimedia data and applications. It is characterized by the processing, storage, generation, manipulation and rendition of Multimedia information.

Multimedia system can be classified into:

1- Playback Systems: is responsible for the operation of multimedia.

2- Authoring System: Multimedia authoring tools provide an integrated environment for binding together the content and functions of your project. Authoring systems typically include the ability to create, organize, and edit the elements of multimedia like graphics, sounds, animations and video clips

Components of a Multimedia System

Now let us consider the Components (Hardware and Software) required for a multimedia system:

1- Capture devices — Video Camera, Video Recorder, Audio Microphone, Keyboards, mice, graphics tablets, 3D input devices, tactile sensors, VR devices. Digitizing Hardware.

2- Storage Devices — Hard disks, CD-ROMs, DVD-ROM, etc.

3- Communication Networks — Local Networks, Intranets, Internet, Multimedia or other special high speed networks.

4- Computer Systems — Multimedia Desktop machines, Workstations, MPEG/VIDEO/DSP Hardware.

5- Display Devices — CD-quality speakers, High-definition television (HDTY), Super Video Graphics Array(SVGA), Hi-Resolution monitors, and Color printers etc.

Characteristics of a Multimedia System

A Multimedia system has four basic characteristics:

1. Multimedia systems must be computer controlled.
2. Multimedia systems are **integrated**.
3. The information they handle must be represented **digitally**.
4. The interface to the final presentation of media is usually **interactive**.

Challenges for Multimedia Systems

1. Supporting multimedia applications over a computer network.
2. Multimedia systems may have to render a variety of media at the same instant -- a distinction from normal applications. There is a temporal relationship between many forms of media (***e.g.*** Video and Audio. There are two forms of problems here
    - Sequencing within the media playing frames in correct order/time frame in video
    - Synchronization — inter-media scheduling (e.g. Video and Audio) — Lip synchronization is clearly important for humans to watch playback of video and audio and even animation and audio.

Desirable Features for a Multimedia System

Given the above challenges the following feature a desirable for a Multimedia System:

1- **Very High Processing Power** — needed to deal with large data processing and **real time** delivery of media.

2- **Multimedia Capable File System** —needed to deliver real-time media **e.g**. Video/Audio Streaming. Special Hardware/Software needed. ***e.g*** RAID technology.

3- **Data Representations** — File Formats that support multimedia should be easy to handle yet allow for compression/decompression in real-time.

4- **Efficient and High I/O** —input and output to the file subsystem needs to be efficient and fast. Need to allow for real-time recording as well as playback of data. E.g. Direct to Disk recording systems.

5- **Special Operating System** —to allow access to file system and process data efficiently and quickly. Needs to support direct transfers to disk, real-time scheduling, fast interrupt processing, I/O streaming etc.

6- **Storage and Memory** — large storage units and large memory (several GB or more). Large Caches also required and high speed buses for efficient management.

7- **Network Support** — Client-server systems common as distributed systems common.

8- **Software Tools** — user friendly tools needed to handle media, design and develop applications, deliver media.

Digitization Definition

When we have a continuously varying signal, both the value we measure and the intervals at which we can measure it, can vary infinitesimally.

In contrast, if we were to convert it to a digital signal, we would have to restrict both of these to a set of discrete values that could be represented in some fixed number of bits.

Therefore,

Digitization is the process of converting a signal from analogue to digital form. It consists of two steps:

1- **Sampling**, when we measure the signal's value at discrete intervals.

2- **Quantization**, when we restrict the value to a fixed set of levels.

Sampling rate is the number of samples in a fixed amount of time or space, ex: the spatial resolution (number of pixels) of the digitized image.

Quantization level is the equally spaced levels to which a signal is quantized, ex: the number of grey levels (number of bits) in the digitized image.

Original 8-bit image,
256 gray levels

Quantized to 6 bits,
64 gray levels

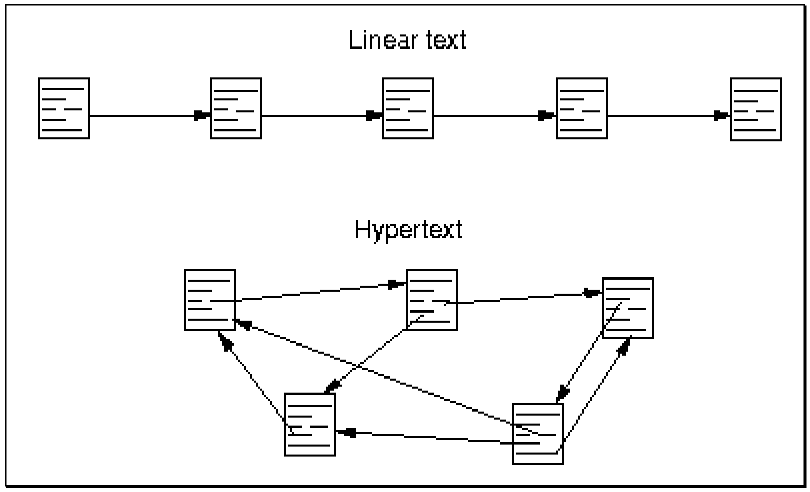
Quantized to 3 bits,

Quantized to 1 bits,

Multimedia may be broadly divided into linear and non-linear categories.

1)Linear "or Passive" Multimedia is a type of a multimedia that is designed to be presented in a sequential manner. It has a distinct beginning and end. It goes on a logical flow from a starting point to a conclusion without any navigation control for the viewer such as a cinema presentation or a YouTube video.

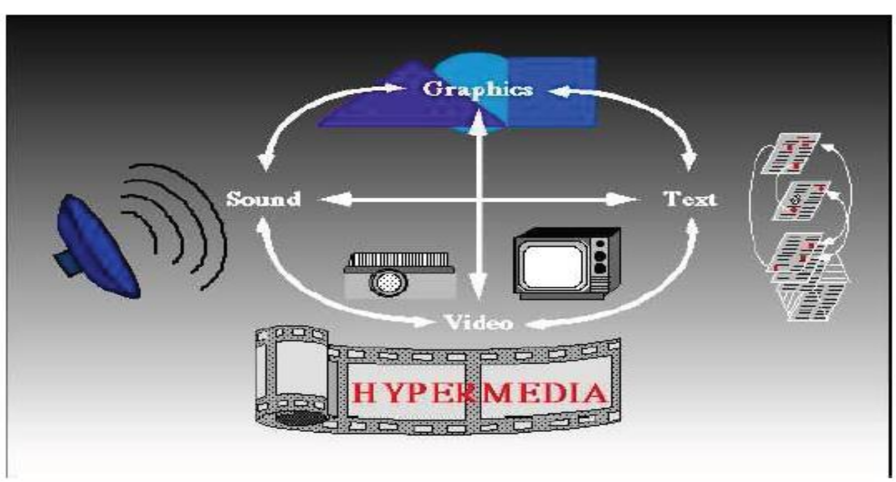
2)Non-linear is a non-sequential type of multimedia where the person needs to interact with a computer program, thus making him in control of the experience. With the presence of an interface, the person and the computer interacts with each other. From a starting point, the person using a nonlinear multimedia is given a range of options that, according to his own preferences, will lead him to a new information, as used with a computer game and Website .



Hypertext and Hypermedia

Unlike the typical printed book, which is read sequentially from beginning to end, **hypertext** is inherently nonlinear, where individual pages can be combined using links, allowing the user to jump from one page to another by clicking on a representation of the link.

When **other media** can be embedded in the linked pages the combination is called **Hypermedia**. **So, Hypermedia** is the generalization of hypertext to include other kinds of media: images, audio clips and video clips are typically supported in addition to text.



Example Hypermedia Applications?

1- The World Wide Web (WWW) is the best example of a hypermedia application.

2- PowerPoint

3-Adobe Acrobat (or PDF software).

4- Adobe Flash.

Multimedia Applications

Multimedia finds its application in various areas including, advertisements, art, education, entertainment, engineering,medicine, disabilities, and spatial-temporal applications. Several examples are as follows:

1-World Wide Web

2- Multimedia Authoring, e.g. Adobe/Macromedia Director

3- Hypermedia courseware

4- Video-on-Demand VOD: are systems which allow users to select and watch/listen to video or audio content when they choose to, rather than having to watch at a specific broadcast time, ex: YouTube.

5-Interactive TV: Interactivity degrees are ranging from :

a- low (TV on/off, volume, changing channels).

b-moderate interactivity (simple movies on demand without player controls).

c-high interactivity in which, for example, an audience member affects the program being watched. The most obvious example of this would be any kind of real-time voting on the screen, in which audience votes create decisions that are reflected in how the show continues.

6- Computer Games.

7- Virtual reality: refers to computer technologies that use software to generate realistic images, sounds and other sensations that replicate a real environment (or create an imaginary setting), and simulate a user's physical presence in this environment, by enabling the user to interact with this space and any objects depicted therein using specialized display screens or projectors and other devices-- for example, in simulations for pilot or combat training, which depict realistic images and sounds of the world, where the normal laws of physics apply (e.g., in flight simulators), or it can differ significantly from reality, such as in VR video games that take place in fantasy (الخيال) settings, where gamers can use fictional magic and telekinesis powers.

8- Digital video editing and production systems: The process of manipulating video images. Video editing includes cutting segments (trimming), re-sequencing clips, and adding transitions and other Special Effects.

9-Multimedia Database systems: is a collection of related multimedia data. The multimedia data include one or more primary media data types such as text, images, graphic objects (including drawings, sketches and illustrations) animation sequences, audio and video.

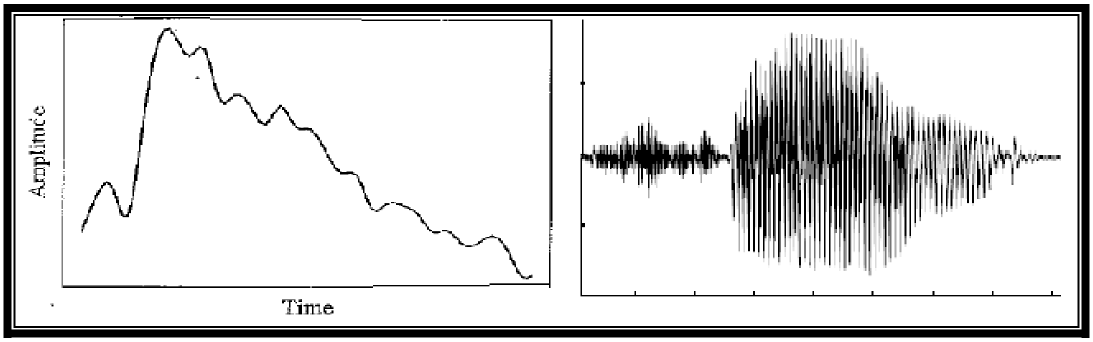
What is sound

Sound is a wave phenomenon like light, but is macroscopic and involves molecules of air being compressed and expanded under the action of some physical device.

When someone speaks their vocal chords vibrate. That sends sound waves through the air. It's just like dropping a rock in the water and watching the waves move out in a circle. For every different sound, there's a different wave.

The sound waves are generated by a sound source, such as the vibrating diaphragm of a stereo speaker. The sound source creates vibrations in the surrounding medium. As the source continues to vibrate the medium, the vibrations propagate away from the source at the speed of sound, thus forming the sound wave.

- Since sound consists of measurable pressures, we can detect it by measuring the pressure level at a location, using a transducer to convert pressure to voltage levels.
- Sound is a continuous wave that travels through the air.
- There are three ways that the wave phenomenon can vary : it varies in amplitude, frequency, and phase



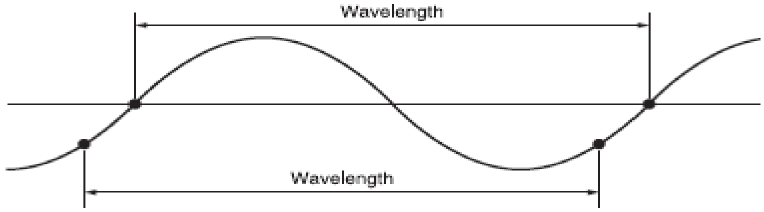
An analog signal: continues measurement of amplitude or energy of wave

- Amplitude is a general indicator for the amount of energy contained in a wave. For example, the amplitude of a sound wave is a measure of its volume.

- Sound waves have normal wave properties such as reflection, refraction (change the angle when entering a medium with different density) and diffraction (bending around an obstacle).

Wave length

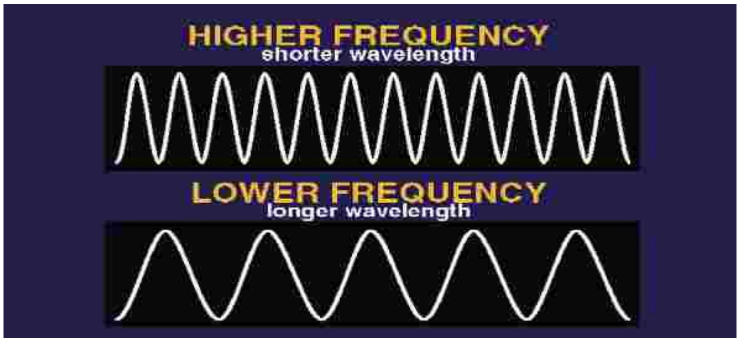
- Wave length : is the distance between one part of a wave and the same part of the next wave.

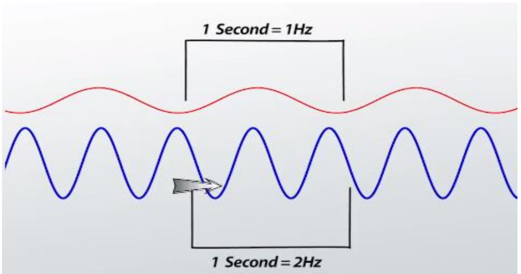


Wavelength is defined as the distance between two points at the same place on adjacent cycles.

Frequency

Is the number of occurrences of a repeating event per unit of time (number of cycle per second of the wave). The unit of frequency is the Hertz (Hz). The hertz is equivalent to cycles per second, i.e., "1/second". A note that is vibrating at 256 Hz will be caused by sound waves that vibrate at 256 times a second. The generally accepted standard range of audible frequencies is 20 Hz to 20 kHz. Frequencies below 20 Hz are generally felt rather than heard. High frequencies are the first to be affected by hearing loss due to age and/or prolonged exposure to very loud noises.





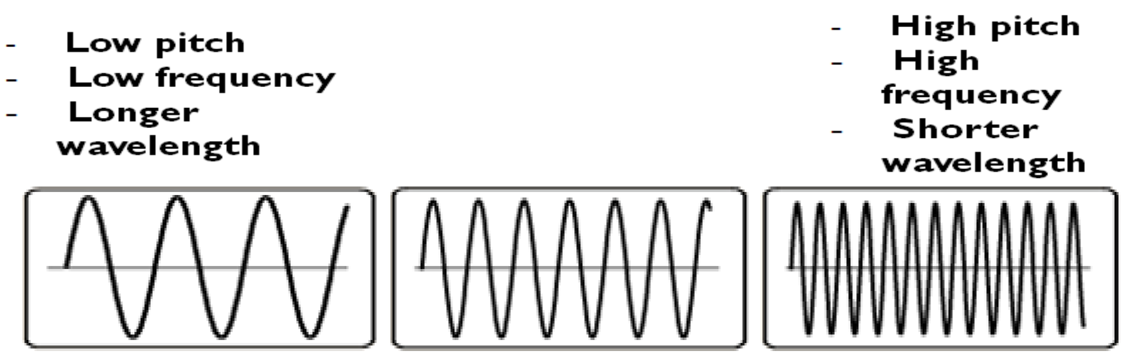
Sound can be classified depending on the frequency into the types:

1. Infrasound, which is less than 20 Hz and not audible to the human ear, where the frequency is too low, and the human ear cannot sense it. The most important source is a vibrating and sliding movement of the layers of the earth's crust and the resulting earthquakes and volcanoes and therefore it's very important to monitor earthquakes and volcanic activity tracking. Some animals can sense earthquakes before they occur.
2. Range of hearing, which extends from about 20 Hz to 20,000 Hz, which sounds audible to humans,
3. Ultrasound, greater than 20,000 Hz, which is not audible to humans and fall outside the scope of the sense of the human ear. This type of wave is still under consideration and attention given intensive task for applications that affect many areas in industry, medicine and others.

Hertz is the property of sound that most determines pitch.

Pitch

The quality that allows us to classify a sound as relatively high or low. It represents the cyclic, repetitive nature of the vibrations that make up sound. For simple sounds, pitch relates to the frequency of the slowest vibration in the sound (called the fundamental harmonic). Sometimes individuals identify different pitches for the same sound, based on their personal experience of particular sound patterns.



Sound output from the man's throat for that output of the Throat women varies. The fundamental frequency of male voice lies in the range of 85 to 180 Hz while that of the female in the range of 150 to 300 Hz, and for the voice of women is characterized by smooth, and sharpness, while the characterized of male voice rough-and heavy-handed.

Recording Sound

There are two types of recording such as:-

1) Stereo recordings are made by recording on two channels, and are lifelike and realistic.

2) Mono sounds are less realistic, flat, and not as dramatic, but they have a smaller file size.

Stereo sounds require twice the space as compared to mono recordings.

To calculate the storage space required, the following formula are used:

Mono Recording:

File size = Sampling rate × duration of recording in seconds × (bits per sample/8) × 1

Stereo Recording:

File size = Sampling rate × duration of recording in seconds × (bits per sample/8) × 2

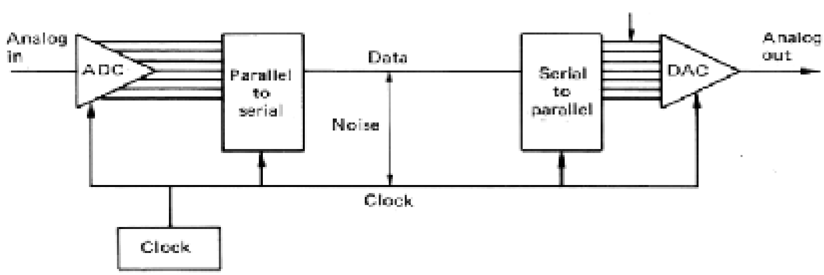
What is an audio signal?

An analog audio signal is an electrical waveform which is a representation of the velocity of a microphone diaphragm. Such a signal is two-dimensional in that it carries a voltage changing with respect to time.

In computers, audio is the sound system that comes with or can be added to a computer. An audio card contains a special built-in processor and memory for processing audio files and sending them to speakers in the computer.

An audio file is a record of captured sound that can be played back. Sound is a sequence of naturally analog signals that are converted to digital signals by the audio card, using a microchip called an Analog-to-Digital Converter (ADC).

When sound is played, the digital signals are sent to the speakers where they are converted back to analog signals that generate varied sound.



Digital audio

Sound is the terminology used in the analog form, and the digitized form of sound is called as audio. Digital audio is created when you represent the characteristics of a sound wave using numbers. One can digitized sound from a microphone, synthesizer, existing recordings, live radio and television broadcasts and popular CD and DVDs. Digitized sound is sampled sound.

Every *n*th fraction of a second, a **sample** of sound is taken and stored as digital information in bits and bytes.

The quality of this digital recording depends upon how often the samples are taken (**sampling rate)**, and how many numbers are used to represent the value of each sample (**bit depth**).

Why binary?

Arithmetically, the binary system is the simplest numbering scheme possible. Figure bellow shows that there are only two symbols: 1 and 0. Each symbol is a binary digit, abbreviated to bit. Logically, binary allows a system of thought in which statements can only be true or false.

What is binary?

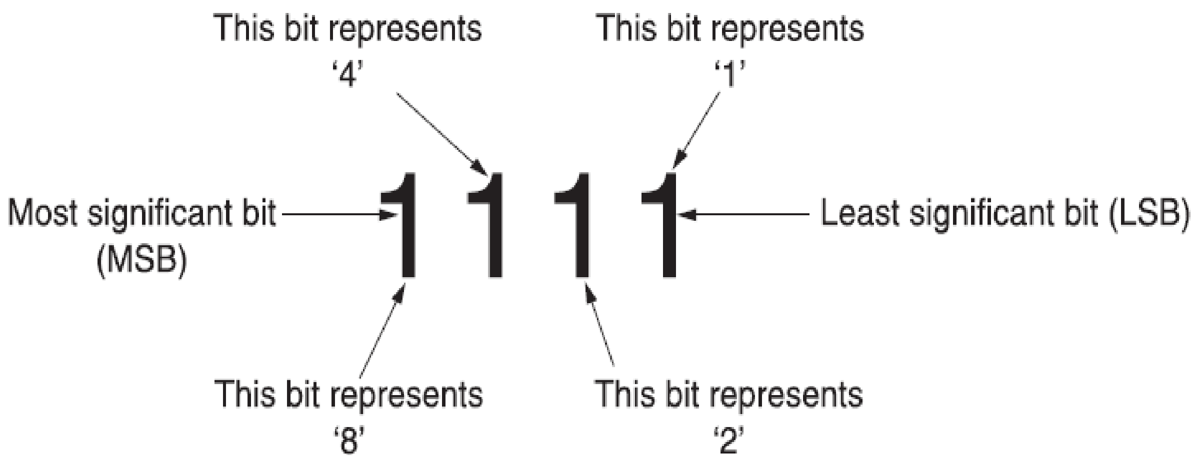
(a) Mathematically:
The simplest numbering scheme possible, there are only two symbols:

1 and **0**

Logically:
A system of thought in which there are only two states:

True and **False**

The robustness of binary signals means that bits can be packed more densely onto storage media, increasing the performance.

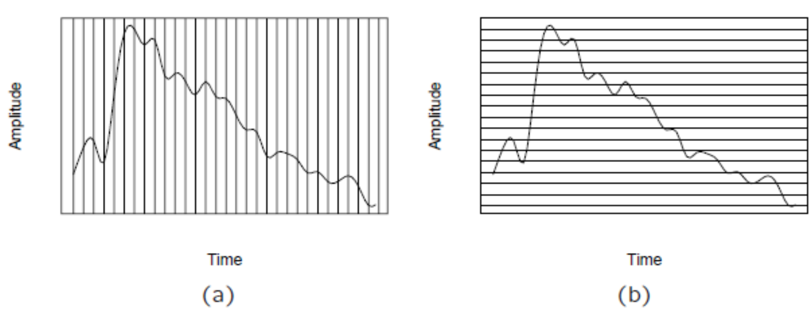


Examples

$0\,1\,0\,1_2 = 5_{10}$
$1\,1\,1\,1_2 = 15_{10}$
$1\,0\,0\,1_2 = 9_{10}$

Digitization

The amplitude value is a continuous quantity. Since we are interested in working with such data in computer storage, we must *digitize* the *analog signals* (i.e., continuous-valued voltages) produced by microphones.

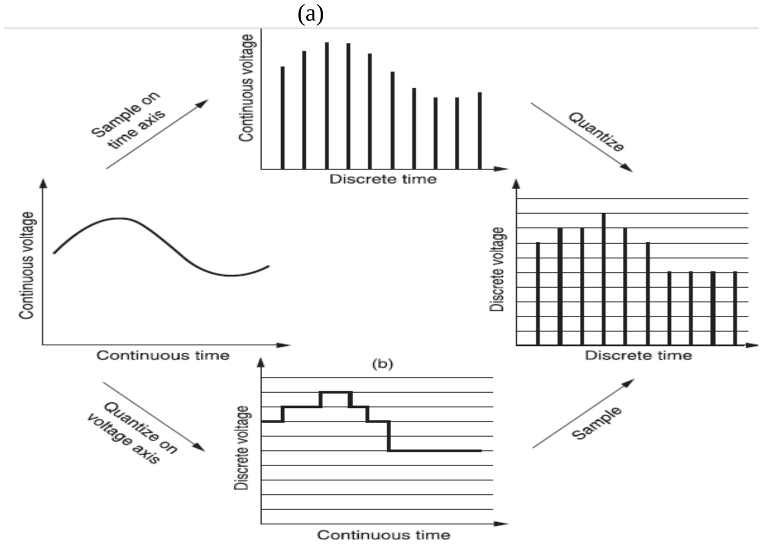


Sampling and quantization: (a) sampling the analog signal in the time dimension;
(b) quantization is sampling the analog signal in the amplitude dimension.

- To digitize, the signal must be sampled in the two dimensions: in time, and in amplitude.

- The first kind of sampling, using measurements only at evenly spaced time intervals, is simply called, sampling.
- The rate at which sampling is performed is called the sampling rate.

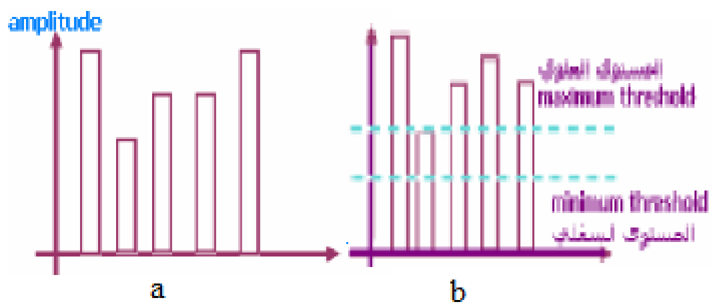
(a)



(a)Sampling (b) quantized

To decide how to digitize audio data we need to answer the following questions:

- What is the sampling rate?
- How finely is the data to be quantized?
- How is audio data formatted? (file format)



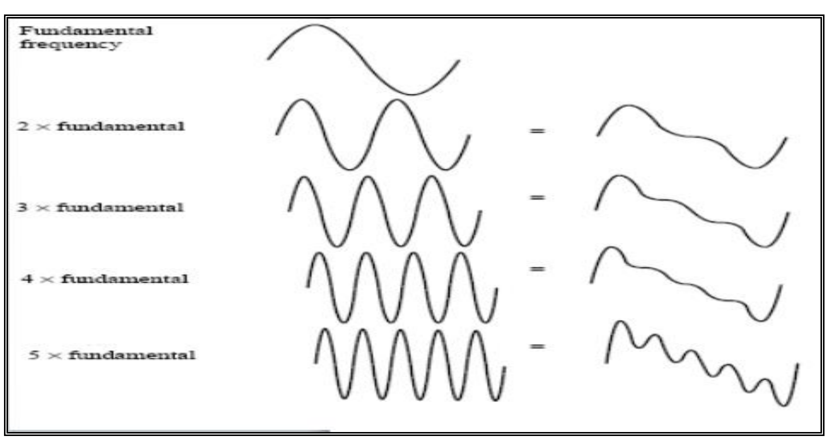
(a)Sampling (b) Quantization

Sampling rate

The choice of sampling rate (the rate at which the signal voltage must be examined to convey the information in a changing signal) is important in any system; if it is too low, the signal will be degraded, and if it is too high, the number of samples to be recorded will rise unnecessarily, as will the cost of the system.

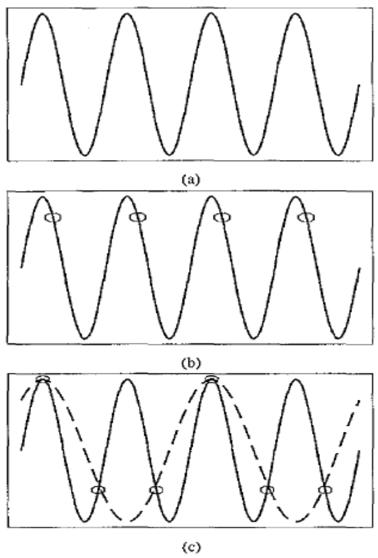
Nyquist Theorem

- Signals can be decomposed into a sum of sinusoids, if we are willing to use enough sinusoids. So, weighted sinusoids can be used to build up complex signal.



Building up a complex signal by superposing sinusoids.

- The Nyquist Theorem, also known as the sampling theorem, is a principle that engineers follow in the digitization of analog signals. For analog-to-digital conversion (ADC) to result in a faithful reproduction of the signal, slices, called samples, of the analog waveform must be taken frequently.
- The Nyquist theorem states how frequently we must sample in time to be able to recover the original sound.
-



Aliasing. (a): A single sinusoid.
(b): Sampling rate just equals the actual frequency, produces a constant.
(c): Sampling at 1.5 times per cycle produces an alias perceived frequency.

- If sampling rate just equals the actual frequency, as shown in Fig.(b), a false signal is detected: it is simply a constant, with zero frequency.
- If sample at 1.5 times the actual frequency, as shown in Fig.(c), an incorrect (alias) frequency is obtained.

- For correct sampling, we must use a sampling rate equal to at least twice the maximum frequency content in the signal. This rate is called the Nyquist rate.

- ***Nyquist Sampling Rate*****:** is the lowest sampling rate that can be used without having aliasing. The sampling rate for an analog signal must be at least two times the maximum frequency content in the signal.

- Aliasing occurs when a system is measured at an insufficient sampling rate.

To avoid the aliasing there are two approaches:

1. raise the sampling frequency to satisfy the sampling theorem,
2. The other is to filter off the unnecessary high-frequency component from the continuous-time signal. We limit the signal frequency by an effective low-pass filter, called anti-aliasing pre filter, so that the remained highest frequency is less than half of the intended sampling rate.

Nyquist frequency: half of the Nyquist rate.

- Since it would be impossible to recover frequencies higher than Nyquist frequency in any event, most systems have a low-pass antialiasing filter that restricts the frequency content in the input to the sampler to a range at or below Nyquist frequency.

relationship among the sampling frequency, true frequency, and the alias frequency:

1) f_{alias} = fsampling − f_{true}, for f_{true} < fsampling < 2× f_{true}

For example: if the true frequency is 5.5 kHz and the sampling frequency is 8 kHz, then the alias frequency

f_{alias} = 8 – 5.5 = 2.5 kHz;

2) falias = *n* × fsampling − f_{true} for fsampling < f_{true} < 2× f_{true}

where *n* is the lowest integer that makes *n* × fsampling > f_{true}

For example: if the true frequency is 7.5 kHz and the sampling frequency is 5 kHz, then the alias frequency is

f_{alias} = 2 × 5 – 7.5 = 2.5 kHz; where *n* =2

Bit rate

In telecommunications and computing, bit rate (R) is the number of bits that are conveyed or processed per unit of time. The bit rate is quantified using the bits per second unit (symbol: bit/s or bps)

Bit Rate (bits/second) = Qunatization * Sampling

= (bits/sample) * (sample/second)

The Bandwidth of the Sound (in Hertz) that the Electronic Device Sampling the Sound Signals affects Sound Quality.

Table 1 shows the various Sampling Rates and their Applications.

Sampling Rates	Applications
8,000 Hz	Telephone Quality
11,025 Hz	Low-end Radio Quality (good for voice)
22,050 Hz	Radio Quality (good for music and voice)
44,100 Hz	CD Quality (high quality)

The bandwidth of a speech signal is from 50 Hz through to 10 kHz and that of a music signal is from 15 Hz through to 20 kHz. You want to digitize these signals using the Nyquist criterion.

Q1/ What is the bit rate produced for the speech signal if 12 bits are used per sample?

Solution:
The bandwidth of a speech signal is from 50 Hz through to 10 kHz
Here we take 2 times as per Nyquist Theorem

Bit Rate for Speech = (Bits/Sample) * (Sample/Second)

= 12 * (2 * 10) = 240 Kbps

Q2/ What is the bit rate produced for the music signal when 16 bits per sample are used?

Solution:
The bandwidth of a music signal is from 15Hz through to 20 KHz.
Here we take 2 times as per Nyquist Theorem

Bit Rate for Music = (Bits/Sample) * (Sample/Second)

= 16 * (2 * 20) = 640 Kbps

Q3/ How many mega bytes of storage do you need for 10 minutes of stereophonic music?

solution

Space for 10 min music = Bit rate * time =10*60*640K

= 600 sec * 640 Kbps

= 384000 Kbits = (384 * 10^6)/8

= 48 Mbytes

Thus, space for mono sound = 48 Mbytes &

space for Stereophonic music = 2 * 48 = 96 Mbytes

Intelligible speech :

In speech communication, intelligibility is a measure of how comprehensible speech is in given conditions. Intelligibility is affected by the quality of the speech signal, the type and level of background noise, reverberation, and, for speech over communication devices, the properties of the communication system.

The high-frequency limit of human hearing extends to approximately 20 kHz, but studies have shown that intelligible speech requires frequencies only up to 4kHz.

Q4/ What is the Nyquist rate for reliable speech communications?

solution

The reliable sampling rate for speech communication is 2*4kHz=8kHz.

Q5/ Justify why the sampling rate for an audio Compact Disc (CD) is 44.1 kHz.

Solution:

The ear is the ultimate receptor of the rendered digital audio. Since it can hear frequencies no more than 20KHz, the original signal must be sampled at least 40 KHz to capture all the 20KHz frequencies. With some additional room (10%) for higher ranges, the CD audio is sampled at 44.1 KHz.

Q6/ Why do you think people sound different on the phone from in person?

Solution:

People's voice are sampled and reconstructed through the telephone lines, which cause loss of high frequencies. That's why it will be sound different from in person.

Q7/ Suppose intelligible speech requires 7 bits per sample. If the phone system is designed to precisely meet the requirements for speech (which is the case), what is the maximum bit rate allowable over telephone lines?

Solution:

Bit Rate (bits/second) = Qunatization * Sampling
= (bits/sample) * (sample/second)

Max bit rate over telephone lines = 7 x 2 x 4 = 56Kbps.

Q8/ CDs use 16 bits per sample. What is the bit rate of music coming off a CD?

Is a modem connection fast enough to support streamed CD quality audio?

Solution:

Bit Rate (bits/second) = Quantization * Sampling
= (bits/sample) * (sample/second)

Bit rate of CD sound = 16 x 44.1 = 705.6 Kbps for mono,

and its 705.6 * 2 = 1.411 Mbps (for stereo).

Broadbands at most homes are getting to a point where this can be supported but still for practical efficiency, CD audio streams are compressed using audio compression schemes like MP3 or AAC.

Q9/ You are asked to design anti-aliasing filters for:
i) FM Radio,
ii) CD Recording and
ii) Telephone System.
What will be typical filter design requirements for these applications?
Solution :

Sampling Rates	Applications
8,000 Hz	Telephone Quality
11,025 Hz	Low-end Radio Quality (good for voice)
22,050 Hz	Radio Quality (good for music and voice)
44,100 Hz	CD Quality (high quality)

For those devices, the fs (sampling frequency) is fixed.
If $fs < 2f_M$, aliasing will occur.
So the pre-filter should cut off the high frequency of the original signal to control f_M, and make $f_M <= fs /2$.
Thus,
for FM Radio, the pre-filter should cut off frequencies higher than 11kHz;
CD 22.05kHz;
Telephone 4kHz (for speech)

Signal-to-Noise Ratio (SNR)

In analog and digital communications, signal-to-noise ratio, often written S/N or SNR, is a measure of signal strength relative to background noise. The ratio is usually measured in decibels (dB).

If the incoming signal strength in microvolts is Vs, and the noise level is Vn, then: **$S/N = 20 \log_{10} (Vs/Vn)$**

If $V_s = V_n$, then S/N = 0. Where log(1)=0

In this situation, the signal borders on unreadable, because the noise level severely competes with it. In digital communications, this will probably cause a reduction in data speed because of frequent errors that require the source (transmitting) computer or terminal to resend some packets of data.

Ideally, V_s is greater than V_n, so a high signal-to-noise ratio is positive. As an example, suppose that V_s = 10.0 microvolts and V_n = 1.00 microvolt. Then:

$S/N = 20 \log_{10}(10.0) = 20.0$ dB

where log(10)=1

This results in the signal being clearly readable.

If the signal is much weaker but still above the noise -- say, 1.30 microvolts -- then:

$S/N = 20 \log_{10}(1.30) = 2.28$ dB

where $\log_{10}(1.30) = 0.113943352$

This is a marginal situation. There might be some reduction in data speed under these conditions.

If V_s is less than V_n, then S/N is negative, representing a low signal-to-noise ratio. In this type of situation, reliable communication is generally not possible unless steps are taken to increase the signal level and/or decrease the noise level at the destination (receiving) computer or terminal.

Signal-to-Quantization-Noise Ratio (SQNR)

For digital signals, we must take into account the fact that only quantized values are stored.

For a digital audio signal, the precision of each sample is determined by the number of bits per sample, typically 8 or 16.

_ Aside from any noise that may have been present in the original analog signal, there is also an additional error that results from quantization.

(a) If voltages are actually in 0 to 1 but we have only 8 bits in which to store values, then effectively we force all continuous values of voltage into only 256 different values.

(b) This introduces a round off error. It is not really "noise". Nevertheless it is called quantization noise (or quantization error).

The quality of the quantization is characterized by the *signal-to-quantization-noise ratio(SQNR).*

Quantization noise is defined as: the difference between the actual value of the analog signal, for the particular sampling time, and the nearest quantization interval value.

Audio Filtering

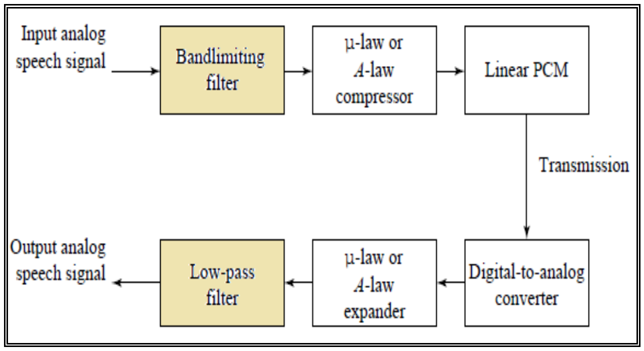
(a) Prior to sampling and AD conversion, the audio signal is usually filtered to remove unwanted frequencies by a ***band pass filter***. The frequencies kept depend on the application:

- For speech, typically from 50Hz to 10kHz is retained, and other frequencies are blocked by the use of a band-pass filter that screens out lower and higher frequencies.

- An audio music signal will typically contain from about 20Hz up to 20kHz.

(b) At the DA converter end, high frequencies may reappear in the output because of sampling and then quantization, smooth input signal is replaced by a series of step functions containing all possible frequencies. So at the decoder side, a lowpass filter is used after the DA circuit.

The sampled signal can be returned to the continuous-time domain simply by passing it into a ***low-pass filter***. This filter can be called a reconstruction filter.



Modulation

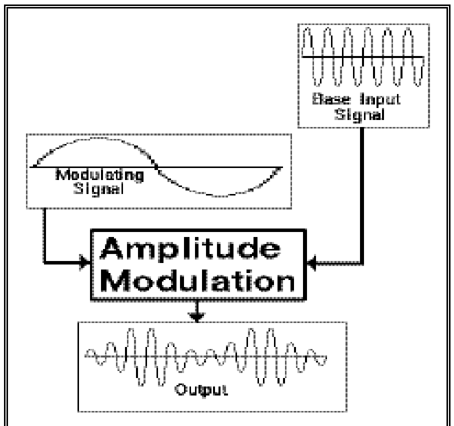
In electronics and telecommunications, modulation is the process of varying some characteristic (line amplitude, frequency, phase) of a high frequency signal wave (carrier wave), with a modulating signal that typically contains information to be transmitted.

Modulation is the addition of information to an electronic carrier signal. A carrier signal is one with a steady waveform -- constant height (amplitude) and frequency. Information can be added to the carrier by varying its properties like amplitude or frequency.

e.g.: the speech and music are used to modulate the carrier, instead of a switch or key. Two types of modulation are listed below :

a-<u>Amplitute modulation</u>

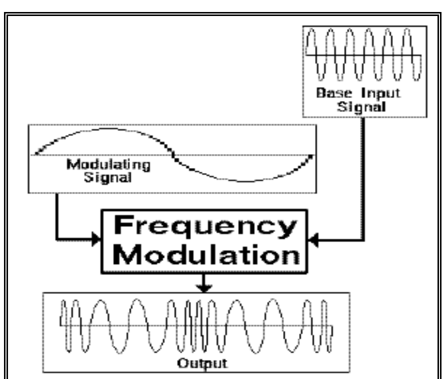
The encoding of information in a carrier wave by changing the amplitude of the carrier wave according to the waveform being transmitted, while the frequency remains constant.



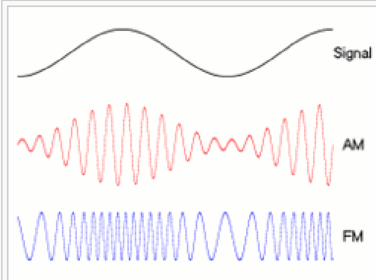
Amplitude Modulation

b- <u>frequency modulation</u>

The encoding of information in a carrier wave by varying the frequency of the wave.



Frequency Modulation



A signal may be carried by an AM or FM radio wave.

Why do we modulate a signal during transmission and demodulation during receiving ? Why don't we send a signal directly without modulation ?

1- Modulation allow the use of a smaller antenna: Theory shows that in order to transmit a wave effectively, the length of the transmitting antenna should be approximately equal to the wavelength of the wave. For low frequency the wavelength becomes very high. So physical length of antenna becomes impractically large.

2- Also the energy of a wave depends upon its frequency. The greater the frequency, the greater the distance of transmission. For low frequency, it is hard to transmit the signal at large distance. The efficiency of the signal can also decrease by the interference. The problem can be overcome by using the low frequency signal to modulate a much higher frequency signal in term of carrier wave.

Examples: Think about your car radio. There are more than a dozen (or so) channels on the radio at any time, each with a given frequency:

100.1MHz, 102.5MHz etc... Each channel gets a certain range (usually about 0.22MHz), and the entire station gets transmitted over that range. Modulation makes it all possible, because it allows us to send voice and music (which are essential baseband signals) over a bandpass channel.

Synthetic Sounds

- Can we create sound wave rather than record and play it back ??

In fact, we can create it on the computer and skip the recording. We call this process s*ound synthesis*.

A synthesizer was a stand-alone sound generator that can vary pitch, loudness, and tone.

Synthesizer: electronic musical instrument operated by a keyboard, producing a wide variety of sounds by generating and combining signals of different frequencies

Synthesizers are computer programs that create sound waves from scratch, it can shaping sound waves — for taking one wave and twisting it into another. Music synthesis can produce almost limitless sounds, including many that we can’t find in nature.

There are two fundamentally different approaches to handling stored sampled audio.

1. Frequency Modulation FM synthesis: more interesting sound is created by changing the argument of the main sinusoid term.

2. Wave Table synthesis: A more accurate way of generating sounds from digital signals. In this technique, the actual digital samples of sounds from real instruments are stored.

Since wave tables are stored in memory on the sound card, they can be manipulated by software so that sounds can be combined, edited, and enhanced.

Wave table synthesis is more accurate and expensive than FM synthesis, partly because the data storage needed is much larger.

MIDI: Musical Instrument Digital Interface

MIDI forms a protocol adopted by the electronic music industry that enables computers, synthesizers, keyboards, and other musical devices to communicate with each other. It is a scripting language that codes "events or messages" that stand for the production of sounds. E.g., a MIDI event might include values for the pitch of a single note, its duration, and its volume.

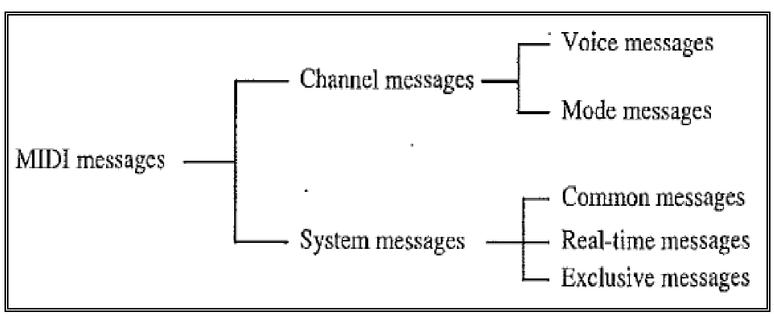
A *MIDI keyboard* produces no sound, instead generating sequences of MID instructions (messages), and since MIDI files don't contain sampled audio like MP3 or WAV files, they're comparatively much smaller than audio files. A minute of compressed audio adds up to around 10Mb of data, while a minute of sound translated into MIDI only takes up 10Kb. This makes MIDI a great choice for memory-starved devices like cell phones and video games.

MIDI channels are used to separate messages. There are 16 channels, numbered from 0 to 15. The idea is that each channel is associated with a particular instrument – for example, channel 1 is the piano, channel 10 is the drums.

When you record music onto a computer using MIDI, the software saves this list of messages and instructions as a .MID file. If you play the .MID file back on an electronic keyboard, the keyboard's internal synthesizer software follows the instructions to play back the song. The keyboard will play a certain key with a certain velocity and hold it for a specified amount of time before moving on to the next note.

MIDI messages can be classified into two types:

1- Channel messages: special message to an instrument's channel.
2- system messages: general message for all instruments indicating a change in tuning or timing such as timing signals for synchronization, positioning information in pre-recorded, MIDI sequences, and detailed setup information for the destination device..



MIDI message taxonomy.

Here are a few examples of typical MIDI messages:

1- Note On: signals that a key has been pressed or a note on another instrument (like a MIDI guitar or clarinet) has been played. The Note On message includes instructions for what key was pressed, what channel, what pitch, and at what volume.

2- Note Off: signals that the key has been released or the note is done playing.

3- Polyphonic Key Pressure is a measurement of how hard a key is pressed. On some keyboards, this adds vibrato or other effects to the note.

4- Control Change indicates that a controller has been pressed or turned.

5- Pitch Wheel Change signals.

Voice message	Status byte	Data byte1	Data byte2
Note Off	&H8n	Key number	Note Off velocity
Note On	&H9n	Key number	Note On velocity
Polyphonic Key Pressure	&HAn	Key number	Amount
Control Change	&HBn	Controller number	Controller value
Program Change	&HCn	Program number	None
Channel Pressure	&HDn	Pressure value	None
Pitch Bend	&HEn	MSB	LSB

Quantization and Transmission of Audio

Quantization and transformation of data are collectively known as coding of the data.

For audio, the µ-law technique for companding audio signals is usually combined with an algorithm that exploits the temporal redundancy present in audio signals.

Where companding algorithms in digital domain reduce the quantization error (hence increasing signal to quantization noise ratio).

Differences in signals between the present and a past time can reduce the size of signal values and also concentrate the histogram of pixel values into a much smaller range.

The result of reducing the variance of values is that lossless compression methods produce a bitstream with shorter bit lengths for more likely values.

In general, producing quantized sampled output for audio is called

1- PCM (Pulse Code Modulation).

3- The differences version is called **DPCM**

4- crude but efficient variant is called **DM**

5- The adaptive version is called **ADPCM**.

Pulse Code Modulation PCM

_ The basic techniques for creating digital signals from analog signals are sampling and quantization.

Sampling is invariably done uniformly – we select a sampling rate and produce one value for each sampling time.

In the magnitude direction, we digitize by quantization, selecting breakpoints in magnitude and remapping any value within an interval to one representative output level.

Assuming a bandwidth for speech from about 50 Hz to about 10 kHz, the Nyquist rate would dictate a sampling rate of20 kHz. Using uniform quantization without companding, the minimum sample size we could get away with would likely be about 12 bits. Hence, for mono speech transmission the bitrate would be 240 kbps. With companding, we can

safely reduce the sample size to 8 bits with the same perceived level of quality and thus reduce the bitrate to 160 kbps. However, the standard approach to telephony assumes that the highest-frequency audio signal we want to reproduce is about 4 kHz. Therefore, the sampling rate is only 8 kHz, and the companded bitrate thus reduces to only 64 kbps.

Differential Coding of Audio

Audio is often stored not in simple PCM but in a form that exploits differences. For a start, differences will generally be smaller numbers and hence offer the possibility of using fewer bits to store.

Streaming Audio

- Streaming audio means sound is delivering over a network and played as it arrives without having to be stored on the computer.

- Because of lower bandwidth required by audio, streaming is more successful for sound than it is for video.
- Example for streaming audio format: Real Network, RealAudio, streaming QuickTime.
- Software required for playing streaming audio: RealPlayer, and Windows media Player.

Audio File Type

An audio file format is a container format for storing audio data on a computer system. There are many formats of audio:

1) WAV: WAV is a sound file developed by Microsoft for use on windows based machines. WAV stands for Waveform Audio File Format. The WAV file uses interesting algorithms to compress raw sound without a loss in quality.

2) AIFF: AIFF stands for Audio Interchange File Format. Similar to how Microsoft developed WAV for Windows, AIFF is a format that was developed by Apple for Macintosh systems back in 1988.

3) RealAudio: Developed by Progressive Networks, RealAudio was the first format to allow for real time streaming of music and sound over the web. Listeners are required to download the Realplayer to

enjoy sound in RealAudio Format. The Realplayer can also stream video and is currently in use by millions of internet users worldwide.

4) MP3: stands for MPEG-1 Audio Layer 3. It was released back in 1993 and quickly exploded in popularity, eventually becoming the most popular audio format in the world for music files.

MP3 is a type of compression that can dramatically reduce file size without drastically reducing sound quality. MP3 works by, among other things, chopping off all sounds that are outside of the normal human range of hearing.

MP3 file format

An MP3 file is made up of multiple MP3 frames, which consist of a header and a data block. This sequence of frames is called an elementary stream. Frames are not independent items ("byte reservoir") and therefore cannot be extracted on arbitrary frame boundaries. The MP3 Data blocks contain the (compressed) audio information in terms of frequencies and amplitudes. The diagram shows that the MP3 Header consists of a sync word, which is used to identify the beginning of a valid frame. This is followed by a bit indicating that this is the MPEG standard and two bits that indicate that layer 3 is used; hence MPEG-1 Audio Layer 3 or MP3. After this, the values will differ, depending on the MP3 file.; as noted in the diagram.

Here is a presentation of the header content. Characters from A to M are used to indicate different fields. In the table below, you can see details about the content of each field.

AAAAAAAA AAABBCCD EEEEFFGH IIJJKLMMIIJJKLMM

Sign	Length (bits)	Position (bits)	Description
A	11	(31-21)	Frame sync (all bits set)
B	2	(20,19)	MPEG Audio version ID 00 - MPEG Version 2.5 (unofficial) 01 - reserved 10 - MPEG Version 2 (ISO/IEC 13818-3) 11 - MPEG Version 1 (ISO/IEC 11172-3)
C	2	(18,17)	Layer description 00 - reserved 01 - Layer III 10 - Layer II 11 - Layer I
D	1	(16)	Protection bit 0 - Protected by CRC (16bit crc follows header) 1 - Not protected
E	4	(15,12)	Bitrate index

bits	V1,L1	V1,L2	V1,L3	V2,L1	V2, L2 & L3
0000	free	free	free	free	free
0001	32	32	32	32	8
0010	64	48	40	48	16
0011	96	56	48	56	24
0100	128	64	56	64	32
0101	160	80	64	80	40
0110	192	96	80	96	48
0111	224	112	96	112	56

1000	256	128	112	128	64
1001	288	160	128	144	80
1010	320	192	160	160	96
1011	352	224	192	176	112
1100	384	256	224	192	128
1101	416	320	256	224	144
1110	448	384	320	256	160
1111	bad	bad	bad	bad	bad

NOTES: All values are in kbps
V1 - MPEG Version 1
V2 - MPEG Version 2 and Version 2.5
L1 - Layer I
L2 - Layer II
L3 - Layer III

bitrate	single channel	stereo	intensity stereo	dual channel
free	yes	yes	yes	yes
32	yes	no	no	no
48	yes	no	no	no
56	yes	no	no	no
64	yes	yes	yes	yes
80	yes	no	no	no
96	yes	yes	yes	yes
112	yes	yes	yes	yes
128	yes	yes	yes	yes
160	yes	yes	yes	yes
192	yes	yes	yes	yes
224	no	yes	yes	yes
256	no	yes	yes	yes
320	no	yes	yes	yes
384	no	yes	yes	yes

F 2 (11,10) Sampling rate frequency index (values are in Hz)

bits	MPEG1	MPEG2	MPEG2.5
00	44100	22050	11025
01	48000	24000	12000
10	32000	16000	8000
11	reserv.	reserv.	reserv.

G 1 (9) Padding bit
0 - frame is not padded

H	1	(8)	Private bit. It may be freely used for specific needs of an application.
I	2	(7,6)	Channel Mode 00 - Stereo 01 - Joint stereo (Stereo) 10 - Dual channel (2 mono channels) 11 - Single channel (Mono)

Graphics and Image data types

a) 1-Bit Images

Images consist of *pixels,* or picture elements in digital images. A 1-bit image consists of on and off bits only and thus it is the simplest type of image. Each pixel is stored as a single bit (0 or 1). Such an image is also referred to as a *binary image or* 1-bit *monochrome* image, since it contains no color.

A 640 x 480 monochrome image requires :

640 x 480/8 = 38400 bytes of storage

Monochrome 1-bit images can be satisfactory for pictures containing only simple graphics and text.

Monochrome I-bit Lena image

b)8-Bit Gray-Level Images

One for which each pixel has a *gray value* between 0 and 255. Each pixel is represented by a single byte - for example, a dark pixel might have a value of 10, and a bright one might be 230.

Image resolution refers to the number of pixels in a digital image (higher resolution always yields better quality).

High resolution for such an image might be 1,600 x 1,200, whereas lower resolution might be 640 x 480. Notice that here we are using an aspect ratio of 4:3 it has been found to look natural.

Grayscale image of Lena

Such an array must be stored in hardware; we call this hardware a frame buffer. Special hardware called a " graphics/video" card is used for this purpose.

Each pixel is usually stored as a byte (a value between 0 to 255), so a 640 x 480 grayscale image requires :

640 x 480 = 307,200 bytes

307,200 / 1024 = 300 kilobytes of storage

c) 24 -Bit Color Images

In a color 24-bit image, each pixel is represented by three bytes, usually representing RGB. Since each value is in the range 0-255, this format supports :

256 x 256 x 256 = 16,777,216 possible combined colors.

However, such flexibility does result in a storage penalty:

640 x 480 x 24-bit = 7372800 bits

7372800 / 8 = 921600 bytes of storage without any compression.

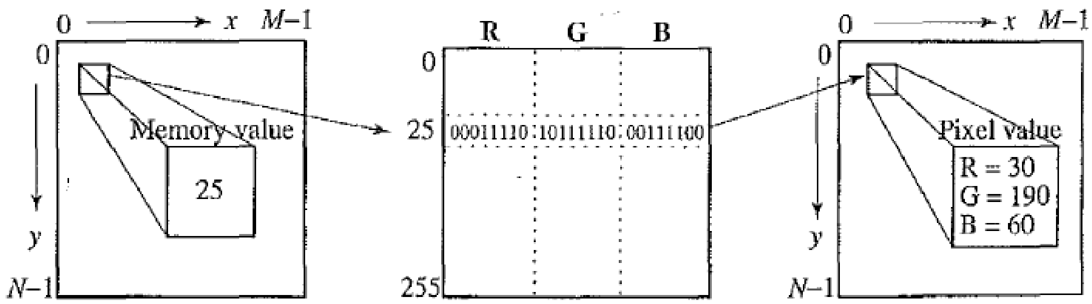
d) 8 -Bit Color Images

If space is a concern, reasonably accurate color images can be obtained by quantizing the color information to collapse it. Many systems can make use of only 8 bits of color information in producing a screen image.

Such image files use the concept of a *lookup table* to store color information. Basically, the image stores not color but instead just a set of bytes, each of which *is* an index into a table with 3-byte values that specify the color for a pixel with that lookup table index.

The idea used in 8-bit color images is to store only the index, or code value, for each pixel. Then, if a pixel stores, say, the value 25, the meaning is to go to row 25 in a color lookup table (LUT). For an 8-bit image, the image file can store in the file header information just what 8-bit values for R, G, and B correspond to each index.

Note the great savings in space for 8-bit images over 24-bit ones



Color LUT for 8-bit color images

Image File Formats

Why do we have so many different types of image file format?

1• there are many different types of images and application with varying requirements.

2• market share proprietary information, and a lack of coordination within the imaging industry.

Many image types can be converted to one of other type by easily available image conversion software.

Some popular file formats for information exchange are described below.

a) GIF

Graphics Interchange Format (GIF) is one of the most important image format, because of its historical connection to the WW'V and HTML markup language as the first image type recognized by net browsers. It was devised initially for transmitting graphical images over phone lines via modems. The GIF standard uses the Lempel-Ziv-Welch algorithm (form of compression).

The GIF standard is limited to 8-bit (256) color images only. It is best suited for images with few distinctive colors (e.g., graphics or drawing).

b) JPEG

The most important current standard for image compression is JPEG. This standard was created by a working group of the International Organization for Standardization (ISO) that was informally called the Joint Photographic Experts Group and is therefore so named.

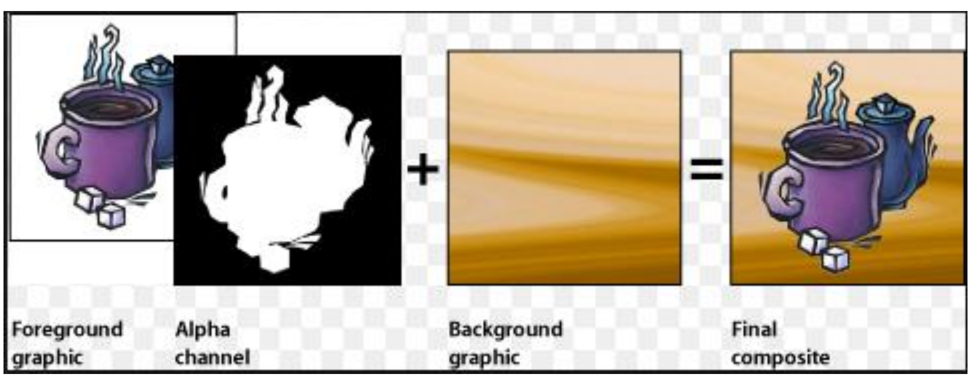
JPEG is a commonly used method of lossy compression for digital images, particularly for those images produced by digital photography. The degree of compression can be adjusted, allowing a selectable tradeoff between storage size and image quality. JPEG typically achieves 10:1 compression with little perceptible loss in image quality.

The human vision system has some specific limitations, which IPEG takes advantage of to achieve high rates of compression. The eye-brain system cannot see extremely fine detail. If many changes occur within a few pixels, we refer to that image segment as having *high spatial frequency* - that is, a great deal of change in *(x, y)* space. The color information in IPEG is *decimated* (partially dropped, or averaged) and then small blocks of an image are represented in the spatial frequency domain *(u, v).*

c) PNG

One interesting development stemming from the popularity of the Internet is efforts toward more system-independent image formats. One such format is *Portable Network Graphics* (PNG). This standard is meant to supersede the GIF standard and extends it in important ways. Special features of PNG files include support for up to 48 bits of color information. Files may also contain gamma-correction information for correct display of color images and alpha-channel information for such uses as control of transparency.

alpha-channel: In graphics, a portion of each pixel's data that is reserved for transparency information. 32-bit graphics systems contain four channels -- three 8-bit channels for red, green, and blue (RGB) and one 8-bit alpha channel. The alpha channel is really a mask-- it specifies how the pixel's colors should be merged with another pixel when the two are overlaid, one on top of the other. This is especially important for animation, where the background changes from one frame to the next.



d) TIFF

Tagged Image File Format (TIFF) is another popular image file format. Developed by the Aldus Corporation in the 1980s, it was later supported by Microsoft. TIFF can store many different types of images: 1-bit, grayscale, 8-bit, 24-bit RGB, and so on. TIFF was originally a loss1ess format, but a new JPEG tag allows you to opt for JPEG compression. Since TIFF is not as user-controllable as JPEG, it does not provide any major advantages over the latter.

e) **EXIF**

Exchange Image File (EXIF) is an image format for digital cameras. Initially developed in 1995. Compressed EXIF files use the baseline JPEG format.

f) Graphics Animation Files

A few dominant formats are aimed at storing graphics animations (i.e., series of drawings or graphic illustrations) as opposed to video (i.e., series of images). The difference is that animations are considerably less demanding of resources than video files.

Animation vs Video

• A video is created using a camcorder, mobile, or a movie camera and no preparation is needed and one can just pick up the camera and start shooting any object, still or moving, with the camera.

• Animation is created by a cartoonist or an artist who draws a series of illustrations in different angles that are fed into a computer to convert them into a video mode adding music or voices.

• Making animation is more difficult than creating a video but once converted into a video; there is virtually no difference between the two.

Many formats are used for animation, such as:

Apple QuickTime, GIF89, DL, FLC, FLI, GL (better quality moving pictures that can handle larger file sizes)

g) PS and PDF

PostScript is an important language for typesetting, and many high-end printers have a PostScript interpreter built into them. PostScript is a vector-based, rather than pixel-based, picture language.

However, the PostScript page description language itself does not provide compression; in fact, PostScript files are just stored as ASCII, therefore files are often large.

Another text + figures language has begun to supersede PostScript: Adobe Systems Inc. includes LZW compression in its *Portable Document Format* (PDF) file format.

The Adobe Acrobat PDF reader can also be configured to read documents structured as linked elements, with clickable content and handy summary tree-structured link diagrams provided.

h) Windows Metafile WMF

Windows Metafile (WMF) is the native vector file format for the Microsoft Windows operating environment.

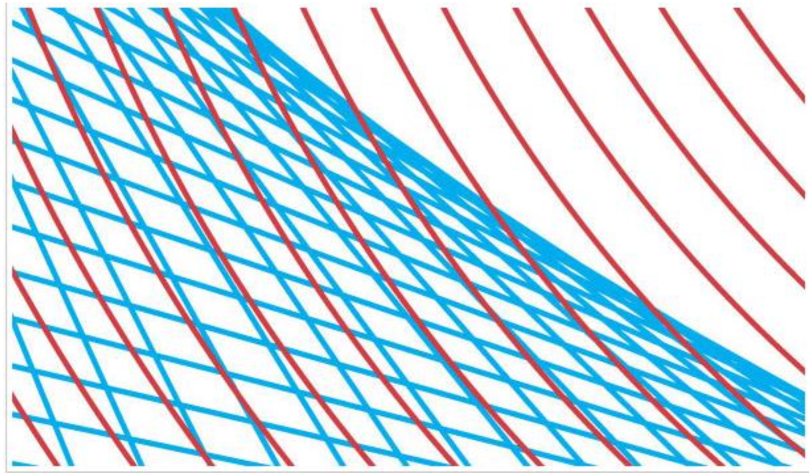
i) Windows BMP

BitMap (BMP) is the major system standard graphics file format for Microsoft Windows, used in Microsoft Paint and other programs. It makes use of run-length encoding compression and can fairly efficiently store 24-bit bitmap images.

The Difference between Vectors and Bitmaps Vector

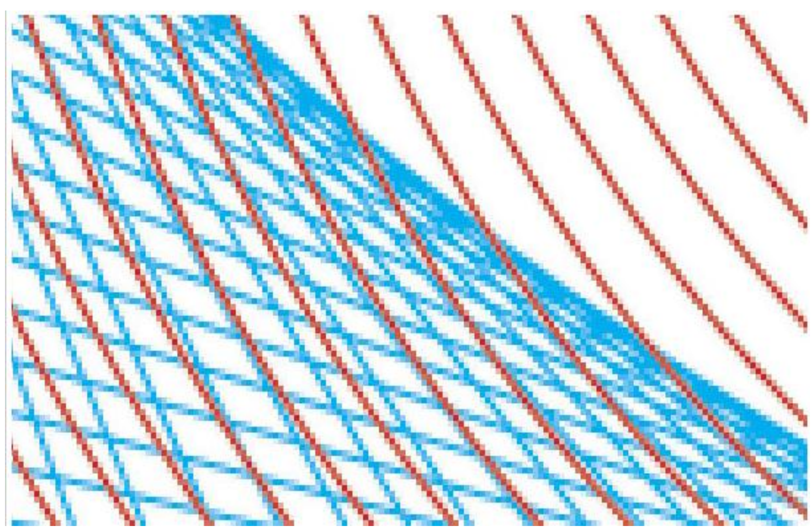
A vector file is a file that's constructed from shapes that are defined by mathematical equations. As a result, when you zoom into a vector image, the quality remains sharp, like the image below scaled to 400%.

The fact that vectors are constructed using mathematical equations is their strength. It keeps the details in the image as high as possible, and means that if you want to change how large or how small you would like the image to be printed, you can do so without worrying about quality.



Vector image at 400% (on screen)

A **bitmap** file (also known as a **raster**) is an image that's constructed from pixels. If you zoom into a bitmap file (like the image below scaled to 400%) you start to see the individual pixels and, consequently, it's important that bitmap files for of printing are used with a resolution that is high enough — typically, 300 dpi (dots per inch) or more**.**



Bitmap image at 400% (on screen)

The strength of using a bitmap file format lies in the tools you can use to create and edit them (like Photoshop). There are certain techniques that you simply can’t do within a vector graphics editor. This gives you greater flexibility when creating or modifying artwork.

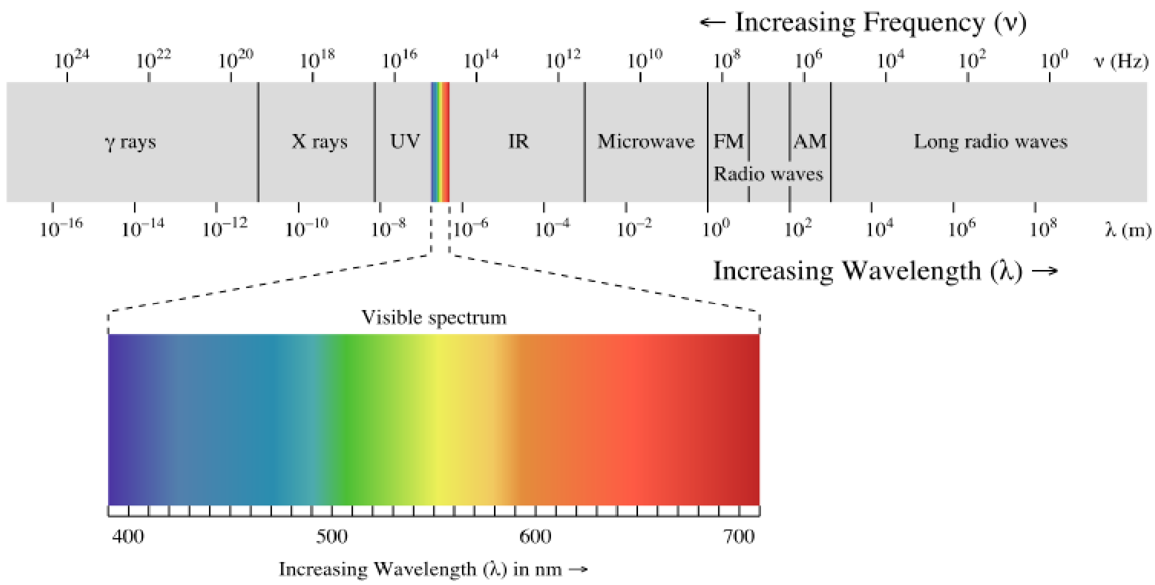
Another strength of bitmaps is linked with their use in photography. Photos are created as bitmaps within a typical digital camera, normally as JPGs or RAW files. So if you’re working with photographic-quality images, you’ll typically need to use a bitmap file.

<u>Color in Image and Video</u>

a-Basics of Color

i-Light and Spectra

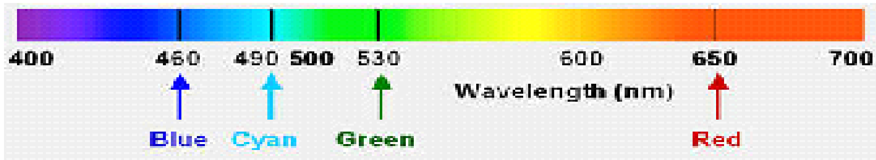
most light sources produce contributions over many wavelengths. Humans cannot detect all light ~ just contributions that fall in the visible wavelength.



Electromagnetic spectrum (visible-light range highlighted).

The human eye is sensitive to electromagnetic radiation with wavelengths between about 380 and 700 nanometers. This radiation is known as *light*. The eye has three classes of color-sensitive light receptors , which respond roughly to red, green and blue light (around 650, 530 and 460 nm, respectively).

Short wavelengths produce a blue sensation, and long wavelengths produce a red one.



Some of unvisible Electromagnetic spectrum are:

a) **X-rays** is a form of electromagnetic radiation. X-ray wavelengths are shorter than those of UV rays. X-rays can identify bone structures and have been used for medical imaging.

b) Ultraviolet (UV) is an electromagnetic radiation with a wavelength from 10 nm to 400 nm, shorter than that of visible light but longer than X-rays. UV radiation is present in sunlight. It is also produced by specialized lights.

Suntan, freckling and sunburn are familiar effects of over-exposure, along with higher risk of skin cancer. Ultraviolet is also responsible for the formation of bone-strengthening vitamin D in most land vertebrates, including humans. The UV spectrum thus has effects both beneficial and harmful to human health.

UV light is used to kill or inactivate microorganisms by destroying nucleic acids and disrupting their DNA, leaving them unable to perform vital cellular functions. UV is used in a variety of applications, such as food, air, and water purification.

c) **Infrared** (**IR**) is an invisible electromagnetic radiation with longer wavelengths than those of visible light, extending from the nominal red edge of the visible spectrum at 700 nanometers.

IR application: Infrared radiation is used in industrial, scientific, and medical applications. Night-vision devices using active near-infrared illumination allow people or animals to be observed without the observer being detected. Infrared astronomy uses sensor-equipped telescopes to penetrate dusty regions of space such as molecular clouds, detect objects such as planets. Infrared thermal-imaging cameras are used to detect heat loss in insulated systems, to observe changing blood flow in the skin, and to detect overheating of electrical equipment.

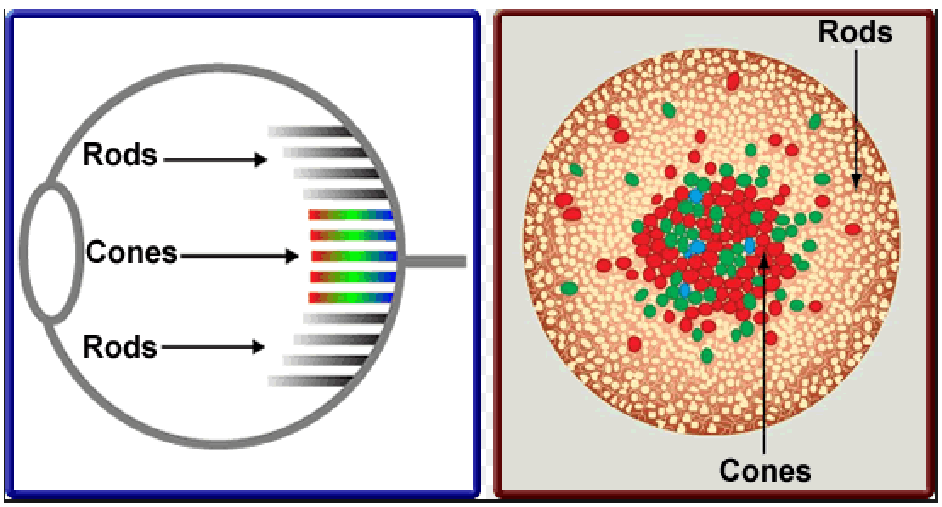
d) **Microwave**

Are a form of electromagnetic radiation with wavelength between radio waves and infrared.

Microwave technology is extensively used for telecommunications. Microwaves are used in spacecraft communication, and much of the world's data, TV, and telephone communications are transmitted long distances by microwaves between ground stations and communications satellites. Microwaves are also employed in microwave ovens and in radar technology.

ii- Human Vision

The eye works like a camera, with the lens focusing an image onto the retina. The retina consists of an array of *rods* and three kinds of *cones*, so named because of their shape. The rods come into play when light levels are low and produce an image in shades of gray. For higher light levels, the cones each produce a signal. Because of their differing pigments, the three kinds of cones are most sensitive to red *(R)*, green (*G*), and blue *(B)* light.

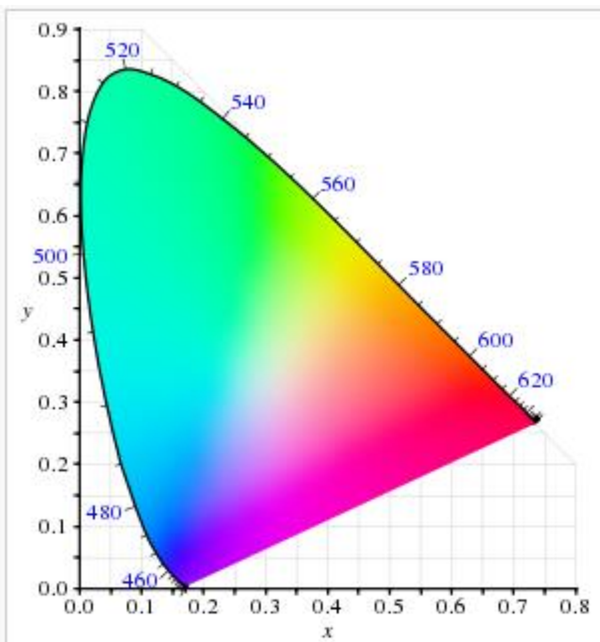


iii-Cones and perception

The response in each color channel in the eye is proportional to the number of neurons firing. For the red channel, any light falling anywhere in the nonzero part of the red cone will generate some response. So the total response of the red channel is the sum over all the light falling on the retina to which the red cone is sensitive, weighted by the sensitivity at that wavelength.

b-CIE Chromaticity Diagram

Since the human eye has three types of color sensors that respond to different ranges of wavelengths, a full plot of all visible colors is a three-dimensional figure. However, the concept of color can be divided into two parts: brightness and chromaticity. For example, the color white is a bright color, while the color grey is considered to be a less bright version of that same white. In other words, the chromaticity of white and grey are the same while their brightness differs.



The CIE 1931 color space chromaticity diagram.

i-Camera Systems

Humans develop camera systems in a similar fashion. A good camera has three signals produced at each pixel location (corresponding to a retinal position). Analog signals are converted to digital, truncated to integers, and stored. If the precision used is 8·bit, the maximum value for any of *R*, G, *B* is 255, and the minimum is 0.

ii-CRT display and Gamma Correction

The cathode ray tube (CRT) is a vacuum tube that contains one or more electron guns and a phosphorescent screen, and is used to display images.

Our eyes do not perceive light the way cameras do. With a digital camera, when twice the number of photons hit the sensor, it receives twice the signal (a "linear" relationship). That's not how our eyes work. Instead, we perceive twice the light as being only a fraction brighter — and increasingly so for higher light intensities (a "nonlinear" relationship).

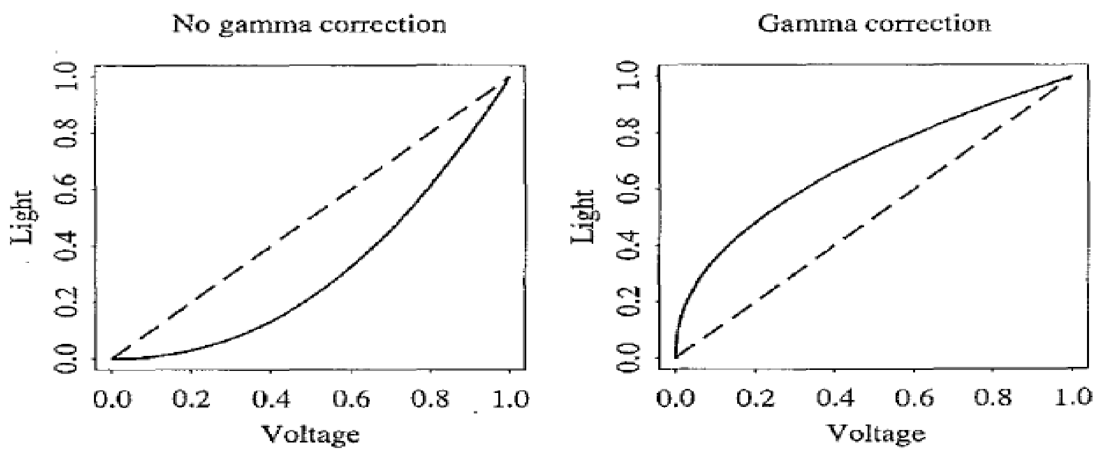
Compared to a camera, we are much more sensitive to changes in dark tones than we are to similar changes in bright tones. There's a biological reason for this peculiarity: it enables our vision to operate over a broader range of luminance. Otherwise the typical range in brightness we encounter outdoors would be too overwhelming.

But how does all of this relate to gamma? In this case, gamma is what translates between our eye's light sensitivity and that of the camera. When a digital image is saved, it's therefore "gamma encoded" — so that twice the value in a file more closely corresponds to what we would perceive as being twice as bright.

On the left is the image as it might appear on an un-corrected monitor.
The center image should look right on a monitor with a gamma of around 1.8

The RGB numbers in an image file are converted back to analog and drive the electron guns in the cathode ray tube (CRT). Electrons are emitted proportional to the driving voltage, and we would like to have the CRT system produce light linearly related to the voltage. Unfortunately, it turns out that this is not the case. The light emitted is actually roughly proportional to the voltage raised to a power; this power is called "gamma", with symbol y.

Thus, if the file value in the red channel is R, the screen emits light proportional to R^Y, with the red phosphor paint on the screen that is the target of the red-channel electron gun.



(a) (b)

FIGURE 4: Effect of gamma correction: (a) no gamma correction (b) gamma correction of signal.

Figure 4(a) shows the light output with no gamma correction applied. We see that darker values are displayed too dark. This is also shown in Figure 5, which display a linear ramp from left to right. Figure 4(b) shows the effect of precorrecting signals by applying the power law $R^{1/Y}$,

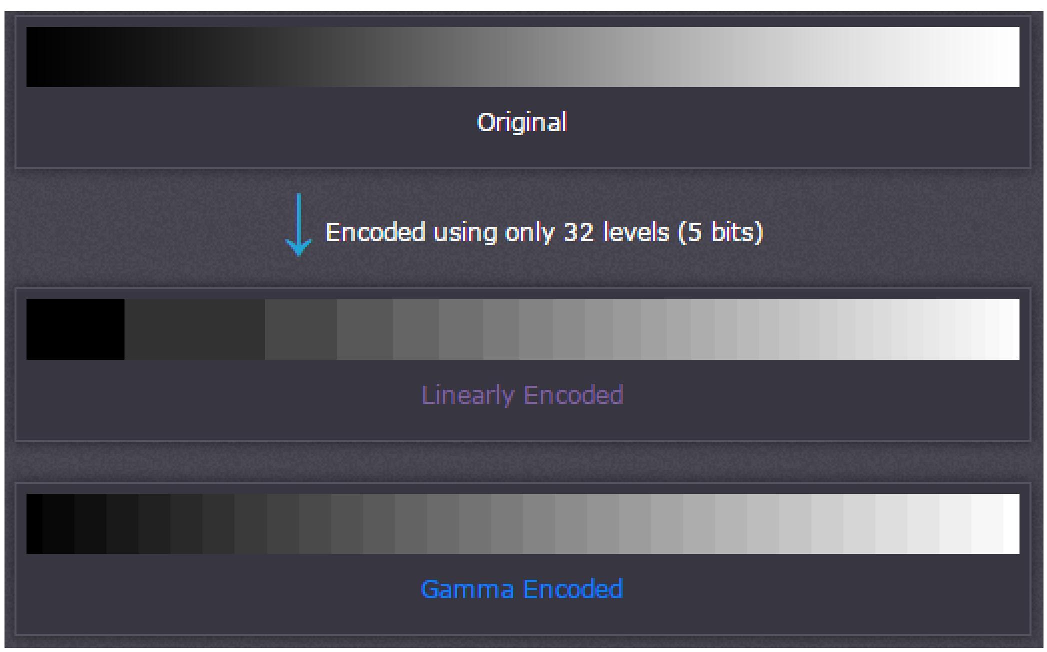


Figure 5

With the advent of CRT-based computer systems, the situation has become even more interesting. The camera may or may not have inserted gamma correction; software may write the image file using some gamma; software may decode expecting some (other) gamma; the image is stored in a frame buffer, and it is common to provide a lookup table for gamma in the frame buffer. After all, if we generate images using computer graphics, no gamma is applied, but a gamma is still necessary to precompensate for the display. Adobe Photoshop allows us to try different gamma values.

For WWW publishing, it is important to know that a Macintosh does gamma correction in its graphics card, with a gamma of 1.8.

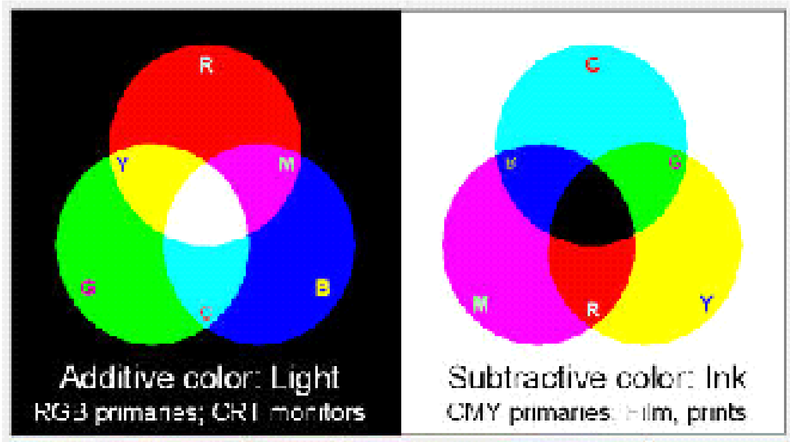
SGI machines expect a gamma of *1.4*, and most PCs do no extra gamma correction and likely have a display gamma of about 2.2.

Color image and video representations

A color model is a mathematical model describing the way colors can be represented as tuples of numbers, typically as three or four values. We will focus on some models:

The RGB (CMY) Color Model

RGB and its subset CMY form the most basic and well-known color model. This model bears closest resemblance to how we perceive color. It also corresponds to the principles of additive and subtractive colors.



The objects may either transmit light (transparencies) or reflect light (paper, for example).

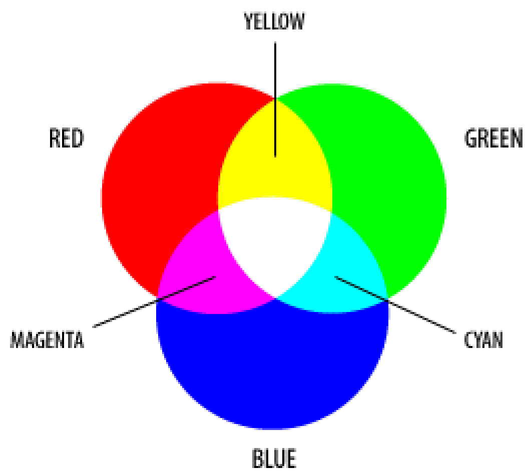
A range of colors can be reproduced by one of two complimentary approaches:

1)The RGB (Additive Colors)

Additive colors are created by mixing spectral light in varying combinations. The most common examples of this are television screens and computer monitors, which produce colored pixels by firing red, green, and blue electron guns at phosphors on the television or monitor screen.

It is used for Web graphics, but it cannot be used for print production.

The *additive primary* colors are red (R), green (G), and blue (B). Adding R and G light makes yellow (Y). Similarly, G + B = cyan (C) and R + B = magenta (M). Combining all three additive primaries makes white.

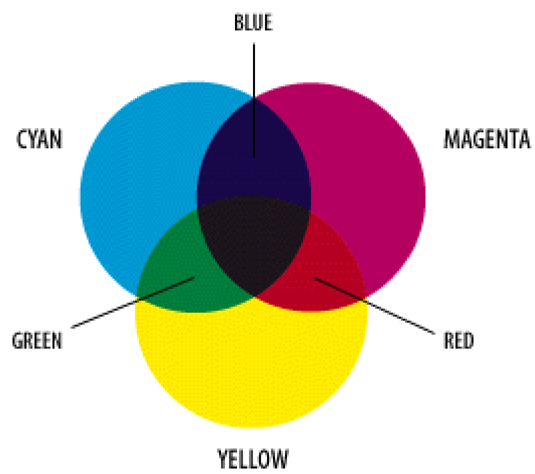


- The importance of RGB as a color model is that it relates very closely to the way we perceive color with the r g b receptors in our retinas. RGB is the basic color model used in television or any other medium that projects the color. It is the basic color model on computers and

2)CMY(K) ***(Subtractive color)***

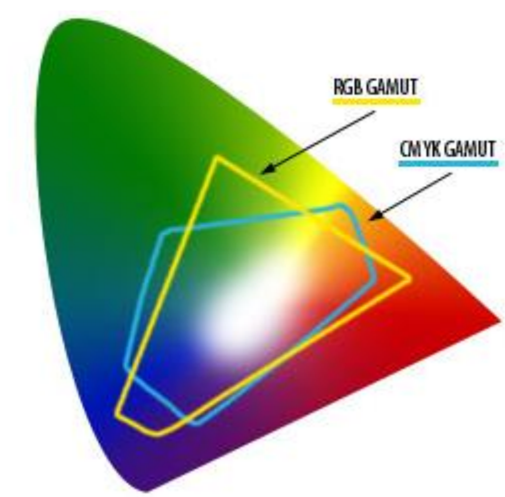
The *subtractive primaries* are C, M and Y. Cyan absorbs red; hence C is sometimes called "minus red" (-R). Similarly, M is -G and Y is -B.

Cyan, magenta, and yellow correspond roughly to the primary colors in art production: red, blue, and yellow. In the illustration below, you can see the CMY counterpart to the RGB model shown above:



Just as the primary colors of CMY are the secondary colors of RGB, the primary colors of RGB are the secondary colors of CMY. But as the illustrations show, the colors created by the subtractive model of CMY don't look exactly like the colors created in the additive model of RGB. Particularly, CMY cannot reproduce the brightness of RGB colors.

The illustration below shows the representative RGB and CMY gamuts over the 1931 CIE Chromaticity Diagram (representing the whole gamut of human color perception):

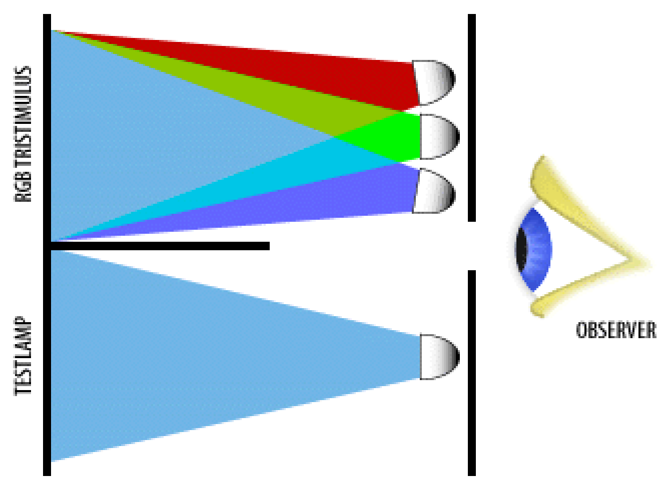


Both models fall short of reproducing all the colors we can see. Furthermore, they differ to such an extent that there are many RGB colors that cannot be produced using CMY(K), and similarly, there are some CMY colors that cannot be produced using RGB.

3)The CIE Color Models

The CIE color model was developed to be completely independent of any device and is based as closely as possible on how humans perceive color.

The observer viewed a split screen. On one half a test lamp cast a pure spectral color on the screen. On the other half, three lamps emitting varying amounts of red, green, and blue light attempted to match the spectral light of the test lamp. The observer viewed the screen through an aperture and determined when the two halves of the split screen were identical. The RGB tristimulus values for each distinct color could be obtained this way.

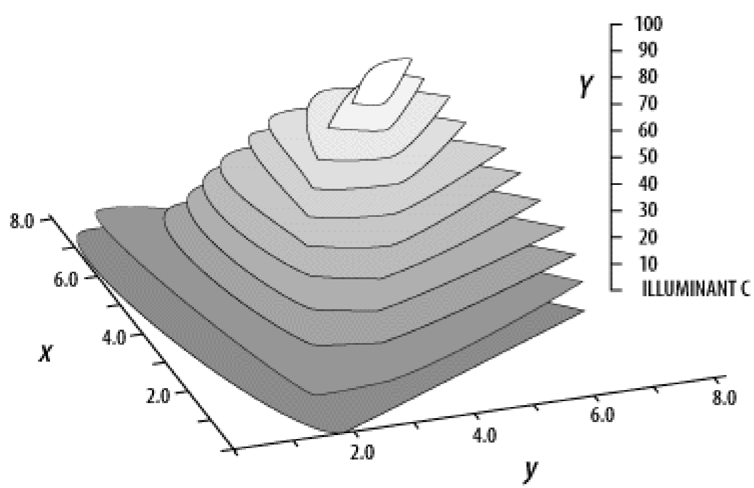


The significant difference between the 1931 and 1964 standard observers was the field of vision used to view the screens. The 1931 observer had a 2° field of vision. This was later considered inadequate in many cases since it did not take in enough of the observer's peripheral vision. The 1964 specification widened the observer's field of vision to 10° in order to get tristimulus values that reflect a wider retinal sensitivity. The following is a summary of some CIE models:

i-CIEXYZ

The original CIE model using the chromaticity diagram adopted in 1931.

The CIE XYZ color space was designed so that the Y parameter was a measure of the luminance of a color. The chromaticity of a color was then specified by the two derived parameters x and y.

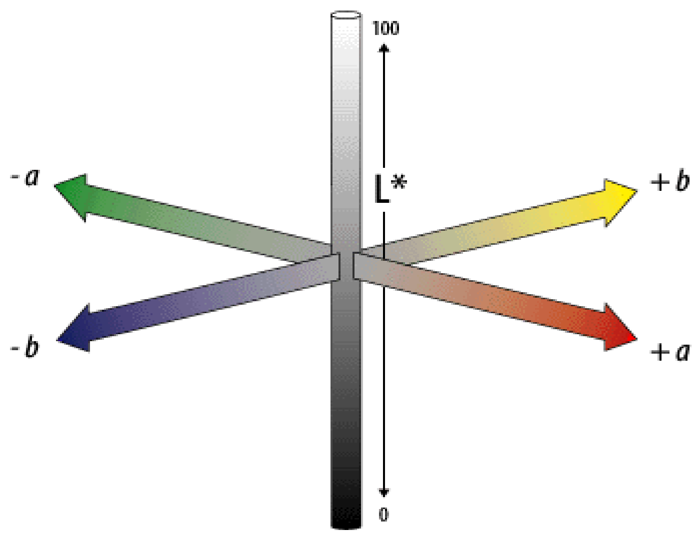


ii-CIELUV

iii-CIELAB

A different approach that defines colors along two polar axes for color (a) and (b) and a third for lightness (L). Roughly, the maximum and minimum of value *a* correspond to red and green, while *b* ranges from yellow to blue.

The L a b color model contains all perceivable colors, which means that its gamut exceeds those of the CMY and the RGB. where color spaces like the RGB contains about 90% of all perceivable colors.



Therefore, values are only needed for two color axes and for the lightness or grayscale axis (L), which is separate (unlike in RGB, CMY or XYZ where lightness depends on relative amounts of the three color channels).

CIELAB, Like all CIE models, is device independent (unlike RGB and CMY).

Designation	*Name*	*Notes*
RGB	Red, Green, Blue	The native format
CYMK	Cyan, Yellow, Magenta, Black	For color printing
HSB	Hue, Saturation, Brightness	Related to human perception
L-a-b	Luminance, a (green to red) and b (blue to yellow)	The CIE model

Color television is most similar to the CIE (Commission International d'Eclairage) model of color with a luminance and two color-related components.

Video Visual Effect of Motion

The visual effect of motion is due to biological phenomenon of Persistence of vision where an object seen by the human eye remains mapped on the eye's retina for a brief time after viewing (approximately 25 ms).

Due to this phenomena of our vision system, a discrete sequence of individual pictures can be perceived as a continuous sequence

Basics of Video

Definition: It is the technology of electronically capturing, recording, processing, storing, transmitting, and display of moving visual sequence of still images representing scenes in motion. Video systems vary greatly in the resolution of the display and refresh rate. Video can be carried on a variety of media, including broadcast, tapes, DVDs, computer files etc.

Types of color video signals

Video signals can be organized in three different ways: Component video, Composite video, and S-video.

i- Component Video (3 wires)

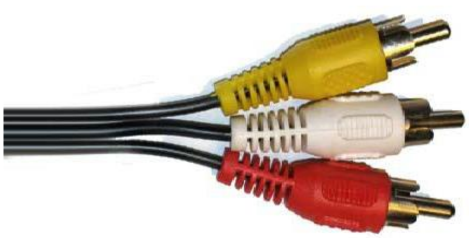
One way of maintaining signal clarity is by separating the components of a video signal so that they do not interfere with each other.

Component video: Higher-end video systems make use of three separate video signals for the red, green, and blue image planes. Each color channel is sent as a separate video signal.

(a) Most computer systems use Component Video, with separate signals for R, G, and B signals.

(b) For any color separation scheme, Component Video gives the best color reproduction since there is no "crosstalk" between the three channels.

(c) This is not the case for S-Video or Composite Video, discussed next. Component video, however, requires more bandwidth and good synchronization of the three components.



Composite Video- (1 wires)

In contrast to component video, all video information, chrominance and luminance are mixed together into a single carrier wave.

a) Chrominance is a composition of two color components.

b) The chrominance and luminance components can be separated at the receiver end and then the two color components can be further recovered.

c) When connecting to TVs Composite Video uses only one wire and video color signals are mixed, not sent separately. The audio and sync signals are additions to this one signal.

This mixing causes interference between the luminance and chrominance signals.

Dot-Crawl is a defect that results from crosstalk due to the intermodulation of the chrominance and luminance components of the signal, where Dot-crawl affects the edges of color and manifests itself as moving dots of colour along these edges. Dot-Crawl can be eliminated by using an S-Video, or component video connection. Like component video, composite-video cables do not carry audio and are often paired with audio cables.

This is the type of signal used by broadcast color TVs.

Enlarged detail from a video source exhibiting dot crawl.
Note the distinctive checkerboard pattern on the vertical edges between yellow and blue areas.

ii-Separate Video (S-Video)

S-Video: e.g., in S-VHS uses two wires, one for luminance and another for a composite chrominance signal.

• As a result, there is less crosstalk between the color information and the crucial gray-scale information.

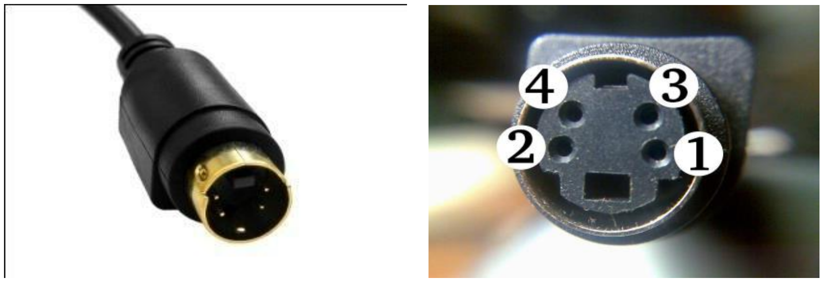
• The reason for placing luminance into its own part of the signal is that black-and-white information is most crucial for visual perception.

– In fact, humans are able to differentiate spatial resolution in grayscale images with a much higher acuity than for the color part of color images.

– As a result, we can send less accurate color information than must be sent for intensity information — we can only see fairly large blobs of color, so it makes sense to send less color detail.

It does not carry audio on the same cable.

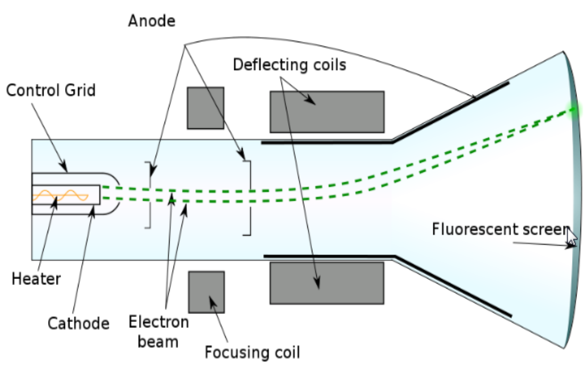
The infamous dot crawl is eliminated.

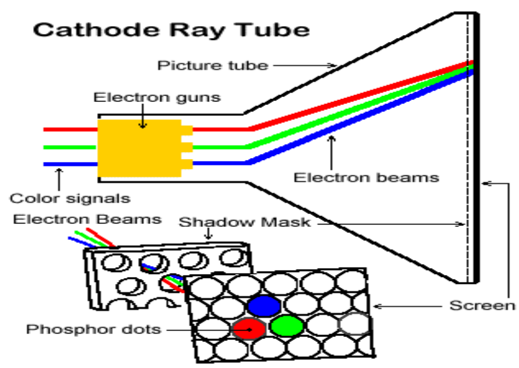


A standard 4-pin S-Video cable connector, with each signal pin (3, 4) paired with its own ground pin (1,2)

Video Display

In conventional TV sets or monitors, the video signal is displayed using a CRT (Cathode Ray Tube).An electron beam sweeps the screen from top to bottom beam carrying the corresponding pattern information, such as intensity in a viewed scene.



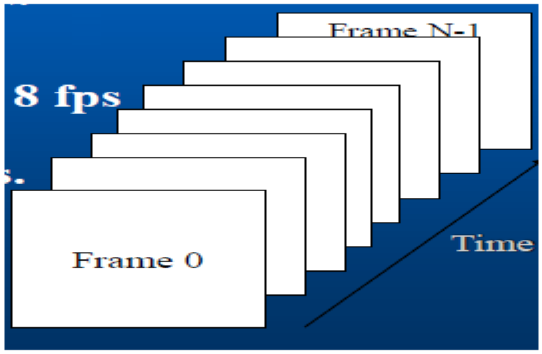


Characteristics of video streams

It is possible to describe the characteristics of the video by many parameters, some of them are: Number of frames per second, Aspect ratio, Color space and bits per pixel, video quality, formats (analog or digital), and video compression method (for digital only).

1- Number of frames per second (Frame rate)

- Frame rate, the number of still pictures per unit of time of video, ranges from 6 or 8 frames per second (fps) for old mechanical cameras to 120 or more frames per second for new professional cameras.



- PAL standards (Europe, Asia, Australia, etc.) and SECAM (France, Russia, parts of Africa etc.) specify 25 frame/s,

- while NTSC standards (USA, Canada, Japan, etc.) specify 29.97 frames.

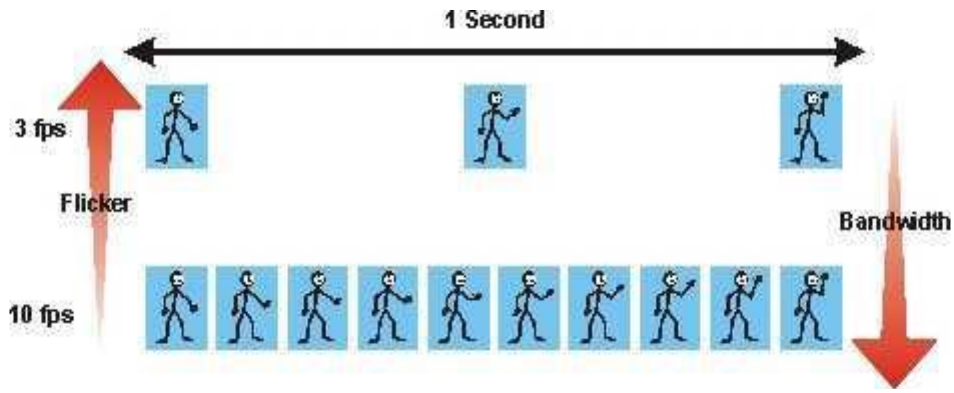
- The minimum frame rate to achieve a comfortable illusion of a moving image is about 16 frames per second.

2- Aspect ratio

The **aspect ratio** describes the dimensions of video screens or the proportional relationship between its width and its height. It is commonly expressed as two numbers separated by a colon (W:H), for example; the screen aspect ratio of a traditional television screen is 4:3 and high definition televisions use an aspect ratio of 16:9.

3- Interlaced vs progressive

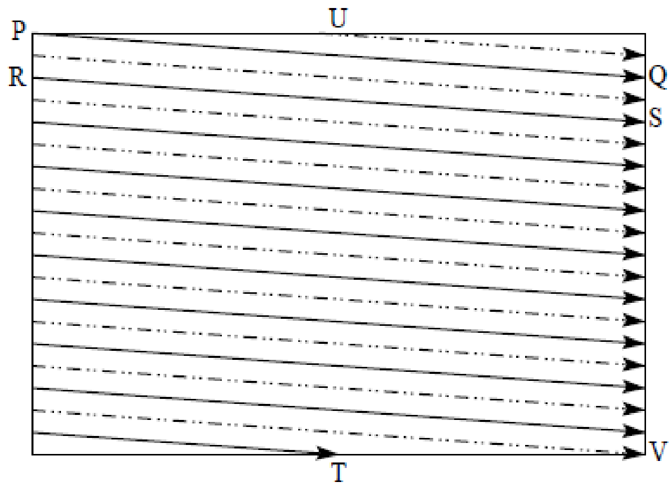
Flicker: is a visible fading between cycles displayed on video displays, especially the refresh interval on cathode ray tube (CRT) based computer screens.



There are two common methods for "painting" a video image on an electronic display:

a- Interlaced video is a technique for doubling the perceived frame rate of a video display without consuming extra bandwidth. The interlaced signal contains two fields to create a frame. One field contains all odd-numbered lines in the image; the other contains all even-numbered lines. This enhances motion perception to the viewer, and reduces flicker.

Interlacing was invented because, when standards were being defined, it was difficult to transmit the amount of information in a full frame quickly enough to avoid flicker. The double number of fields presented to the eye reduces perceived flicker.



Interlaced raster scan

Because of interlacing, the odd and even lines are displaced in time from each other - generally not noticeable except when very fast action is taking place on screen, when blurring may occur.

Initially the odd-numbered lines are scanned and then the process is repeated for even-numbered lines -this time starting at the second row.

- For example, in the video in Fig. 1, the moving helicopter is blurred more than is the still background.

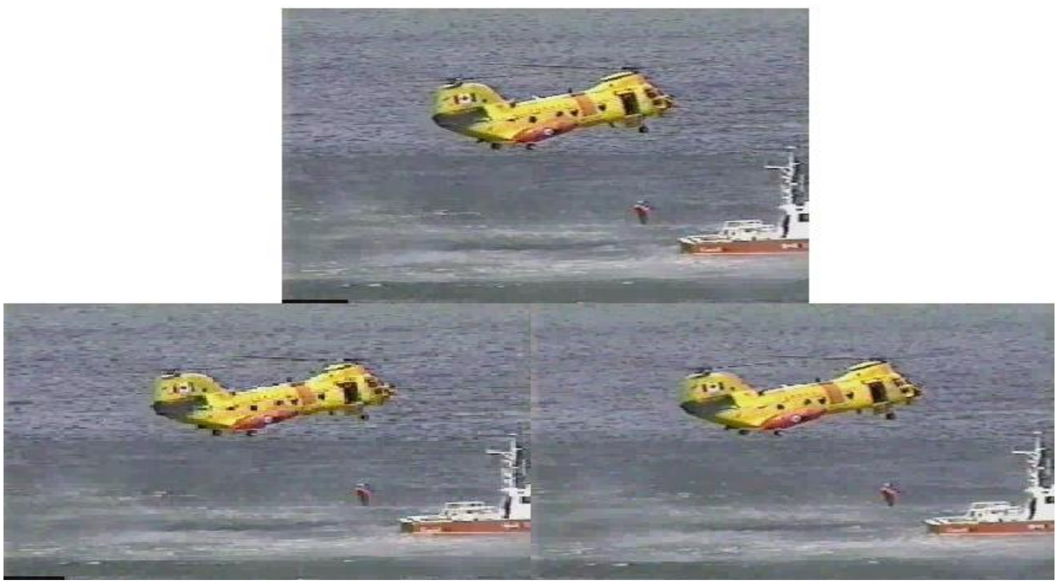


Fig. 1: Interlaced scan produces two fields for each frame.
At top: The video frame, at left: odd field, at right: even field,

A Phase Alternating Line (PAL)-based television set display, for example, scans 50 fields every second (25 odd and 25 even). The two sets of 25 fields work together to create a full frame

NTSC, PAL and SECAM are interlaced formats. Abbreviated video resolution specifications often include an i to indicate interlacing. For example, PAL video format is often specified as 576i50, where 576 indicates the total number of horizontal scan lines, i indicates interlacing, and 50 indicates 50 fields (half-frames) per second.

b- Progressive scanning (alternatively referred to as noninterlaced scanning, is a way of displaying, storing, or transmitting moving images in which all the lines of each frame are drawn in sequence.

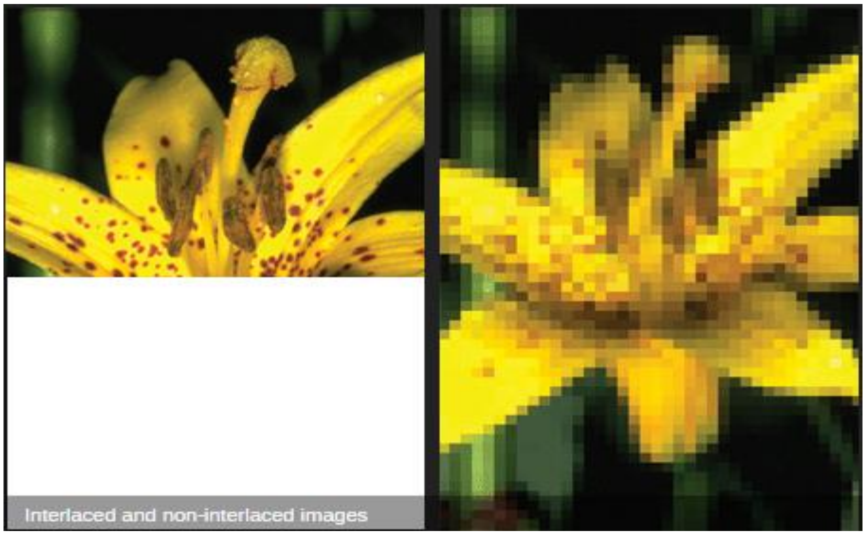
progressive scan interlace

So interlacing is in fact a clever way to compress a movie when one cannot use digital compression methods. Interlacing reduces the bandwidth (= storage space) by half, without losing vertical resolution in quiet areas.

THE DIFFERENCE BETWEEN INTERLACED AND NON-INTERLACED IMAGES

These images look and function almost the same way as a non-interlaced image with one exception–how it appears to load to your visitor. If you have a large image on your site and someone with a slower Internet connection comes to view that image, a non-interlaced image will simply be blank until the data transfers and then slowly it will appear from top to bottom.

An interlaced image will appear completely, but it will be highly pixelated. As the data transfers, the picture will begin to get clearer and clearer until the full resolution becomes apparent. A perfect example of the difference between the two can be seen below:



(a) non-interlaced (b) interlaced
Half of figure downloaded

Deinterlacing is the process of converting interlaced video, such as common analog television signals into a non-interlaced form.

To display interlaced video on a progressive scan display requires a process called deinterlacing. This is an imperfect technique, and causes various artifacts—particularly in areas with objects in motion. Providing the best picture quality for interlaced video signals requires expensive and complex devices and algorithms.

Method 1: Capturing one field and combining it with the next field, Problem: "combing" effect.

Method 2: Line doubler

The most basic and literal way to double lines is to repeat each scanline. Most line doublers use digital interpolation to recreate the missing lines in an interlaced signal, and the resulting quality depends on the technique used. Generally a line doubler will only interpolate within a single field, rather than merging information from adjacent fields, to preserve the smoothness of motion.

When interlaced video is watched on a progressive monitor with very poor deinterlacing, it exhibits combing when there is movement between two fields of one frame.

4- Video formats:

a- Analog Video: Analog video is represented as a continuous (time-varying) signal.

i- NTSC Video

- NTSC (National Television System Committee) TV standard is mostly used in North America and Japan. It uses the familiar 4:3 aspect ratio (i.e., the ratio of picture width to its height) and uses 525 scan lines per frame at 30 frames per second (fps). NTSC follows the interlaced scanning system, and each frame is divided into two fields, with 262.5 lines/field**.**

ii- PAL Video

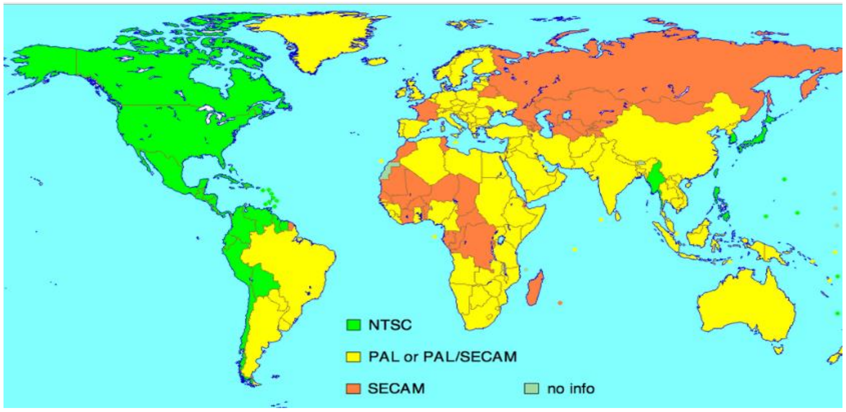
- PAL (Phase Alternating Line) is a TV standard widely used in Western Europe, China, India, and many other parts of the world.
- PAL uses 625 scan lines per frame, at 25 frames/second, with a 4:3 aspect ratio and interlaced fields.

iii- SECAM Video

SECAM stands for Systeme Electronique Couleur Avec Memoire,the third major broadcast TV standard.

- SECAM also uses 625 scan lines per frame, at 25 frames per second, with a 4:3 aspect ratio and interlaced fields.
- SECAM and PAL are very similar. They differ slightly in their color coding scheme.

	Line duration (µS)	*Picture height (line)*	*Line rate (Hz)*	*Frame rate (Hz)*	*Active picture area*	
					Width (µS)	*Height (lines)*
RS-170		525	15,750	30		486
NTSC	63.49	525	15,734.26	29.96	720	486
PAL, SECAM	64	625	15,625	25	52	576



b- Digital Video

Digital video is represented as a sequence of digital images.

The advantages of digital representation for video are many. For example:

(a) Video can be stored on digital devices or in memory, ready to be processed (noise removal, cut and paste, etc.), and integrated to various multimedia applications;

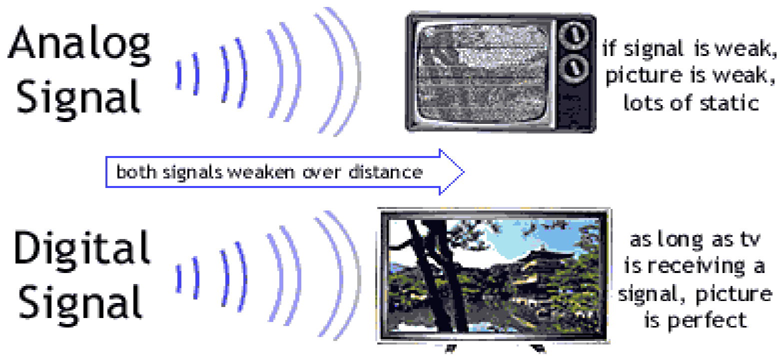
(b) Direct random access is possible, which makes nonlinear video editing achievable as a simple, rather than a complex, task;

(c) Repeated recording does not degrade image quality;

(d) Ease of encryption and better tolerance to channel noise.

(e) An advantage digital has over analog is that analog signals can't be compressed as well as a digital signal can.

Almost all digital video uses component video.



Homework: why can't analog video be compressed like digital?

Chroma Subsampling

Digital signals are often compressed to save transmission time and reduce file size. Since the human visual system is much more sensitive to variations in brightness than color, a video system can be optimized by devoting more bandwidth to the luma component (usually denoted Y'), than to the color difference components Cb and Cr.

This became a central motivation behind early forms of analog and digital compression. Video signals would be separated into a lightness or “luma” component and two color or “chroma” components, similar to how images can be separated into three red, green and blue (RGB) components. The luma and chroma components would then be referred to as YUV (with analog) or YCbCr (with digital).

Once separated, the chroma resolution would then be reduced by half or more through a process called "chroma subsampling." The end result is a video signal that appears more detailed at the same broadcast bandwidth, since the luma component occupies a greater fraction of the video signal:

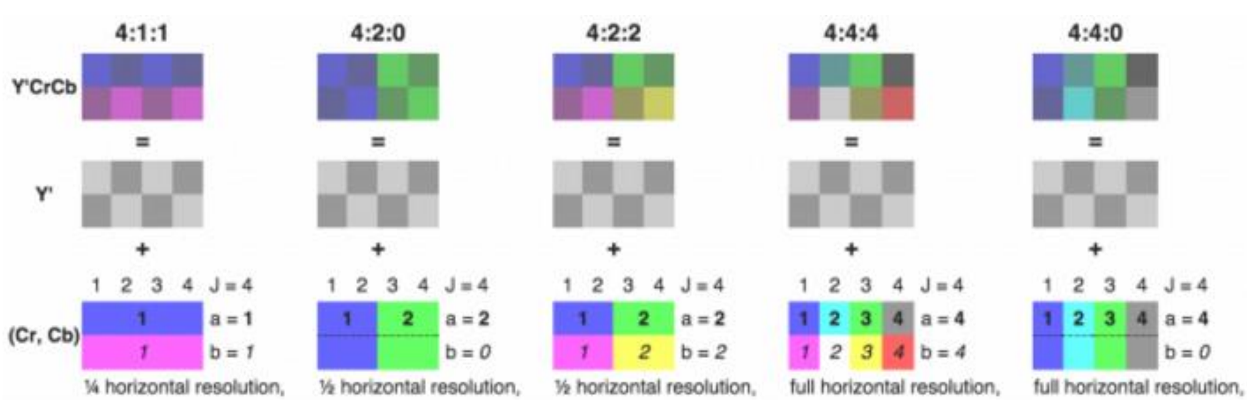
In compressed images, for example, the 4:2:2 Y'CbCr scheme requires two-thirds the bandwidth of (4:4:4) R'G'B'. This reduction results in almost no visual difference as perceived by the viewer.

Sampling systems and ratios

The subsampling scheme is commonly expressed as a three part ratio J:a:b (e.g. 4:2:2) that describe the number of luminance and chrominance samples in a conceptual region that is J pixels wide, and 2 pixels high.

The scheme"4:2:2" indicates horizontal subsampling of the Cb and C*r* signals by a factor of 2. That is, of four pixels horizontally labeled 0 to 3, all four *Y* s are sent, and every two Cbs and two *Crs* are sent.

The chroma subsampling scheme "4:4:4" indicates that no chroma subsampling is used: each pixel's Y, Cb and Cr values are transmitted, 4 for each of Y, Cb, Cr.



mapping examples

Scheme 4:2:0, along with others, is commonly used in JPEG and MPEG.

High Definition TV (HDTV)

The introduction of wide-screen movies brought the discovery that viewers seated near the screen enjoyed a level of participation not experienced with conventional movies. Apparently the exposure to a greater field of view, especially the involvement of peripheral vision, contributes to the sense of "being there". The main thrust of High Definition TV (HDTV) is not to increase the "definition" in each unit area, but rather to increase the visual field, especially its width.

The salient difference between conventional TV and HDTV:

(a) HDTV has a much wider aspect ratio of 16:9 instead of 4:3 where 16:9 is closer to aspect ratio of the human eye sight.

(b) HDTV moves toward progressive (non-interlaced) scan. The rationale is that interlacing introduces serrated edges to moving objects and flickers along horizontal edges.

(c) HDTV has higher resolution 1280 ×720 or 1920 ×1080.

Video quality

Video quality is a characteristic of a video passed through a video transmission or processing system, representing a measure of perceived degradation with respect to the original source video. Video processing systems may introduce some amount of distortion or artifacts in the video signal, which negatively impacts the user's perception of a system. For many stakeholders such as content providers, service providers and network operators, the assurance of video quality is an important task.

Compression Loss, Artifacts, and Visual Quality:

Compression artifacts are noticeable distortions in compressed video, when it is subsequently decompressed and presented to a viewer. Such distortions can be present in compressed signals other than video as well. These distortions are caused by the lossy compression techniques involved. One of the goals of compression algorithms is to minimize the distortion while maximizing the amount of compression. However, depending on the algorithm and the amount of compression, the output has varying levels of diminishing quality or introduction of artifacts.

Compression loss is manifested in many different ways and results in some sort of visual impairment. Quantization is the process of mapping a large set of input values to a smaller set—for example, rounding the input values to some unit of precision. The round-off error introduced by the process is referred to as *quantization error* or the *quantization noise*. In other words, the difference between the input signal and the quantized signal is the quantization error.

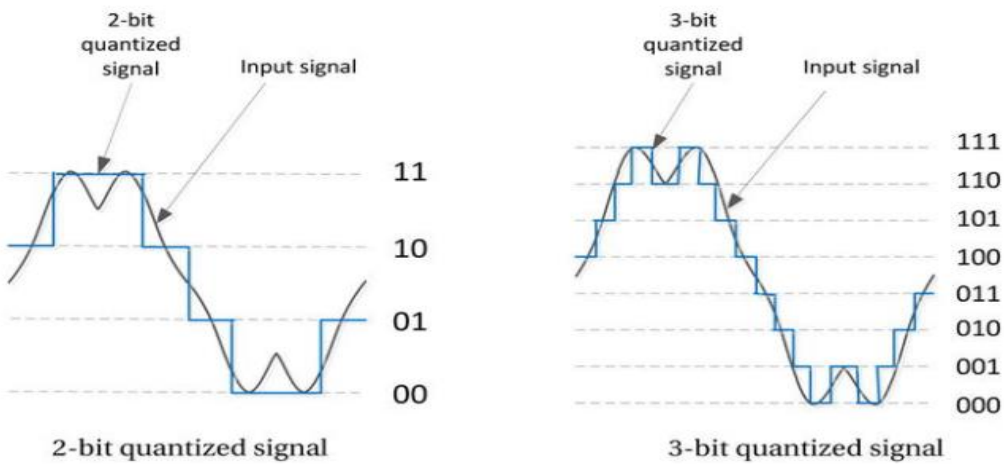
There are two major sources of quantization noise in video applications:

1) when an analog signal is converted to digital format;

2) when high-frequency components are discarded during a lossy compression of the digital signal.

1-Quantization of Samples:

The digitization process of an image converts the continuous-valued brightness information of each sample at the sensor to a discrete set of integers representing distinct gray levels—that is, the sampled image is quantized to these levels.



Quantized signals with different bit resolution

2-Frequency Quantization

In frequency quantization, an image or a video frame undergoes a transform, such as the discrete cosine transform, to convert the image into the frequency domain.

The human eye is fairly good at seeing small differences in brightness over a relatively large area, but not so good at distinguishing the exact strength of a high frequency (rapidly varying) brightness variation. This fact allows one to reduce the amount of information required by ignoring the high frequency components. This is done by simply dividing each component in the frequency domain by a constant for that component, and then rounding to the nearest integer. This is the main lossy operation in the whole process. As a result of this, it is typically the case that many of the higher frequency components are rounded to zero, and many of the rest become small positive or negative numbers.

$$\begin{bmatrix} 1260 & -1 & -12 & -5 & 2 & -2 & -3 & 1 \\ -23 & -17 & -6 & -3 & -3 & 0 & 0 & -1 \\ -11 & -9 & -2 & 2 & 0 & -1 & -1 & 0 \\ -7 & -2 & 0 & 1 & 1 & 0 & 0 & 0 \\ -1 & -1 & 1 & 2 & 0 & -1 & 1 & 1 \\ 2 & 0 & 2 & 0 & -1 & 1 & 1 & -1 \\ -1 & 0 & 0 & -1 & 0 & 2 & 1 & -1 \\ -3 & 2 & -4 & -2 & 2 & 1 & -1 & 0 \end{bmatrix}$$

Transform coefficients

$$\begin{bmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 112 & 100 & 103 & 99 \end{bmatrix}$$

Quantization matrix

$$\begin{bmatrix} 79 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ -2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Quantized coefficients

$$\begin{bmatrix} 1264 & 0 & -10 & 0 & 0 & 0 & 0 & 0 \\ -24 & -12 & 0 & 0 & 0 & 0 & 0 & 0 \\ -14 & -13 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Reconstructed coefficients

Figure 1: Quantization of a block of transform coefficients

The quantization matrix is the same size as the block of transform coefficients, which is input to the quantizer. To obtain quantized coefficients, an element-by-element division operation is performed, followed by a rounding to the nearest integer. For example, in Figure 1, quantization of the coefficient (the upper left element) by doing round (1260/16) gives the quantized coefficient 79. Notice that after quantization, mainly low-frequency coefficients are retained, while high-frequency coefficients have become zero and are discarded before transmission. Reconstruction is performed by multiplying the quantized coefficients by the same quantization matrix elements. However, the resultant reconstruction contains the quantization error.

It should be noted that the large number of zeros that appear in the quantized coefficients matrix is not by accident; the quantization matrix is designed in such a way that the high-frequency components–which are not very noticeable to the HVS–are removed from the signal. This allows greater compression of the video signal with little or no perceptual degradation in quality.

Example2: This is an example of DCT coefficient matrix:

$$\begin{bmatrix} -415 & -33 & -58 & 35 & 58 & -51 & -15 & -12 \\ 5 & -34 & 49 & 18 & 27 & 1 & -5 & 3 \\ -46 & 14 & 80 & -35 & -50 & 19 & 7 & -18 \\ -53 & 21 & 34 & -20 & 2 & 34 & 36 & 12 \\ 9 & -2 & 9 & -5 & -32 & -15 & 45 & 37 \\ -8 & 15 & -16 & 7 & -8 & 11 & 4 & 7 \\ 19 & -28 & -2 & -26 & -2 & 7 & -44 & -21 \\ 18 & 25 & -12 & -44 & 35 & 48 & -37 & -3 \end{bmatrix}$$

A common quantization matrix is:

$$\begin{bmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 112 & 100 & 103 & 99 \end{bmatrix}$$

Dividing the DCT coefficient matrix element-wise with this quantization matrix, and rounding to integers results in:

$$\begin{bmatrix} -26 & -3 & -6 & 2 & 2 & -1 & 0 & 0 \\ 0 & -3 & 4 & 1 & 1 & 0 & 0 & 0 \\ -3 & 1 & 5 & -1 & -1 & 0 & 0 & 0 \\ -4 & 1 & 2 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Typically this process will result in matrices with values primarily in the upper left (low frequency) corner. By using a zig-zag ordering to group the non-zero entries and run length encoding, the quantized matrix can be much more efficiently stored than the non-quantized version.

3- Color Quantization

Color quantization is a method to reduce the number of colors in an image. As the HVS is less sensitive to loss in color information, this is an efficient compression technique.

Common Artifacts

some common artifacts that are typically found in various image and video compression applications.

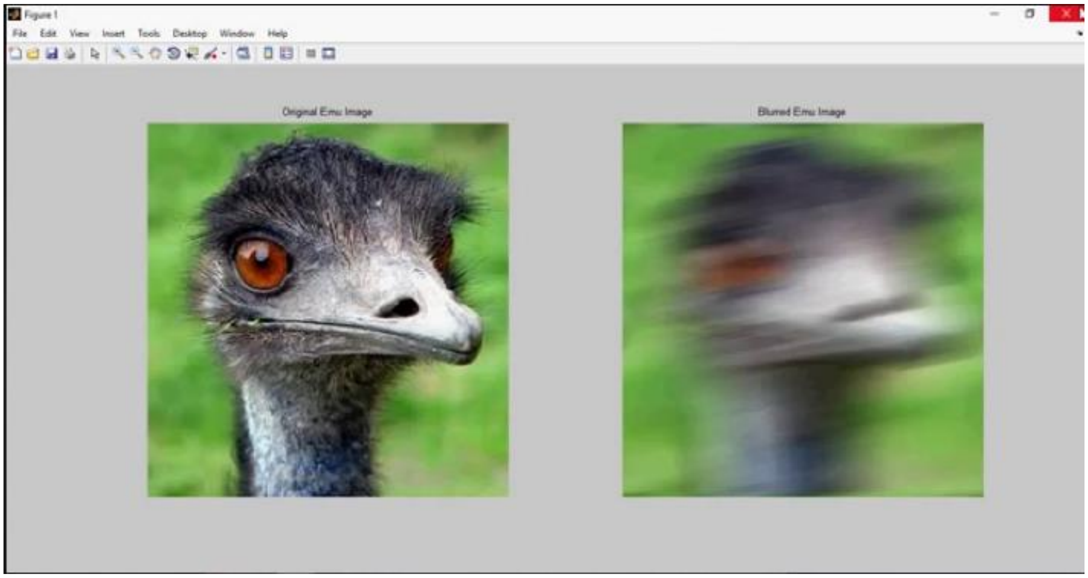
a- Blurring Artifact

Blurring of an image refers to a smoothing of its details and edges, and it results from direct or indirect low-pass filter effects of various processing. Blurring of an object appears as though the object is out of focus. Generally speaking, blurring is an artifact the viewer would like to avoid, as clearer, crisper images are more desirable.

But sometimes, blurring is intentionally introduced by using a Gaussian function:

1) To reduce image noise or to enhance image structures at different scales.
2) Typically, this is done as a pre-processing step before compression algorithms may be applied, attenuating high-frequency signals and resulting in more efficient compression.

3) This is also useful in edge-detection algorithms, which are sensitive to noisy environments



4) Graphics, image, or video editing tools may also generate the motion blur effect for artistic reasons; the most frequent synthetic motion blur is found when computer-generated imagery (CGI) is added to a scene in order to match existing real-world blur or to convey a sense of speed for the objects in motion.

Also blur can appear in:

5) Motion blur appears in the direction of motion corresponding to rapidly moving objects in a still image or a video. It happens when the image being recorded changes position (or the camera moves) during the recording of a single frame, because of rapid movement of objects. For example, motion blur is often an artifact in sports content with fast motion.

6) Deinterlacing by the display

7) Compression artifacts present in digital video streams can contribute additional blur during fast motion.

8) Motion blur has been a more severe problem for LCD displays, owing to their sample-and-hold nature, where a continuous signal is sampled and the sample values are held for a certain time to eliminate input signal variations. In these displays, the impact of motion blur can be reduced by controlling the backlight.

b-**Block Boundary Artifact**

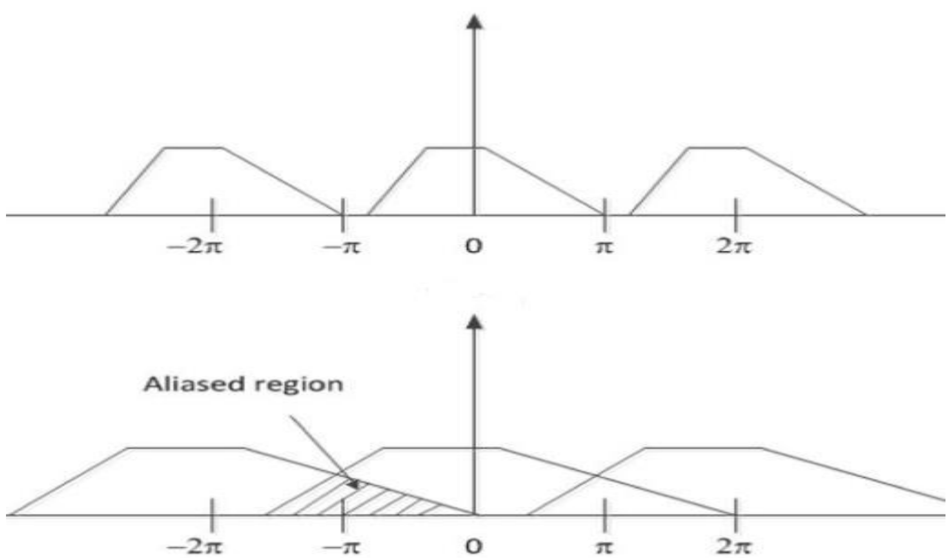
The *block boundary artifact* is the result of independently quantizing the blocks of transform coefficients, leading to discontinuities in the reconstructed block boundaries. These block-boundary discontinuities are usually visible, especially in the flat color regions such as the sky, faces, and so on, where there are little details to mask the discontinuity.

Original video frame

Reconstructed video frame with visible block boundaries

b- Aliasing Artifacts

If the transformed signal overlap with its shifted replicas. In case of such overlap, the original signal cannot be unambiguously recovered from its downsampled version, as the overlapped region represents two copies of the transformed signal at the same time. One of these copies is an *alias,* or replica of the other. This overlapping effect is called *aliasing*.



Transform domain effect of downsampling, causing aliasing

aliasing is an effect that causes different signals to become indistinguishable (or aliases of one another).

Aliasing is generally avoided by applying low pass filters anti-aliasing filters to the analog signal before sampling.

When a digital image is viewed, a reconstruction is performed by a display or printer device, and by the eyes and the brain. If the image data is processed in some ways during sampling or reconstruction, the reconstructed image will differ from the original image, and an alias is seen.

Aliasing can **occur** in :

1- signals sampled in time, for instance digital audio, and is referred to as temporal aliasing.

2- in spatially sampled signals, for instance moiré patterns in digital images. Aliasing in spatially sampled signals is called spatial aliasing.

An example of spatial aliasing is the moiré pattern observed in a poorly pixelized image of a brick wall. Spatial anti-aliasing techniques avoid such poor pixelizations.

Aliasing can be caused either by:

i- The sampling stage this may be distinguished by calling sampling aliasing prealiasing

ii- The reconstruction stage; which may be distinguished by calling reconstruction aliasing postaliasing.

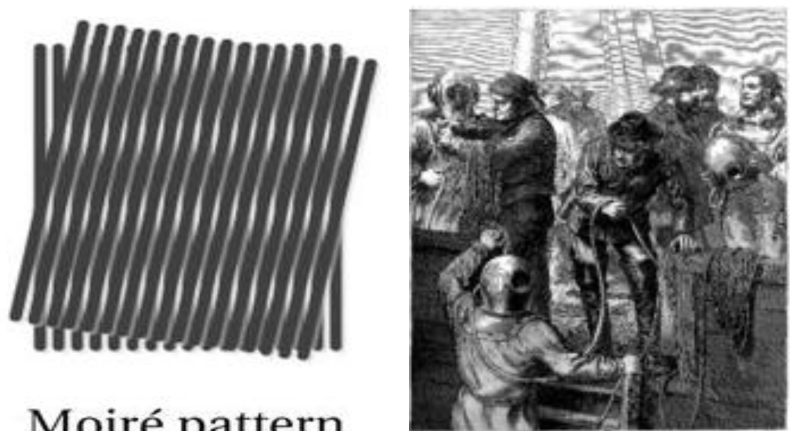
Temporal aliasing is a major concern in the sampling of video and audio signals. Music, for instance, may contain high-frequency components that are inaudible to humans. If a piece of music is sampled at 32000 samples per second (Hz), any frequency components above 16000 Hz (the Nyquist frequency for this sampling rate) will cause aliasing when the music is reproduced by a digital to analog converter (DAC). To prevent this an anti-aliasing filter is used to remove components above the Nyquist frequency prior to sampling.

i- Jaggies

Is a common form of aliasing artifact produces visible stair like lines where there should be smooth straight lines or curves in a digital image. These stairs or steps are a consequence of the regular, square layout of a pixel. With increasing image resolution, this artifact becomes less visible. Also, anti-aliasing filters are useful in reducing the visibility of the aliased edges.

ii- Moiré Pattern

Due to undersampling of a fine regular pattern, a special case of aliasing occurs in the form of *moiré patterns*. It is an undesired artifact of images produced by various digital imaging and computer graphics techniques—for example, ray tracing a checkered surface. The moiré effect is the visual perception of a distinctly different third pattern, which is caused by inexact superposition of two similar patterns. In as in the following figure, moiré effect can be seen as an undulating pattern, while the original pattern comprises a closely spaced grid of straight lines.



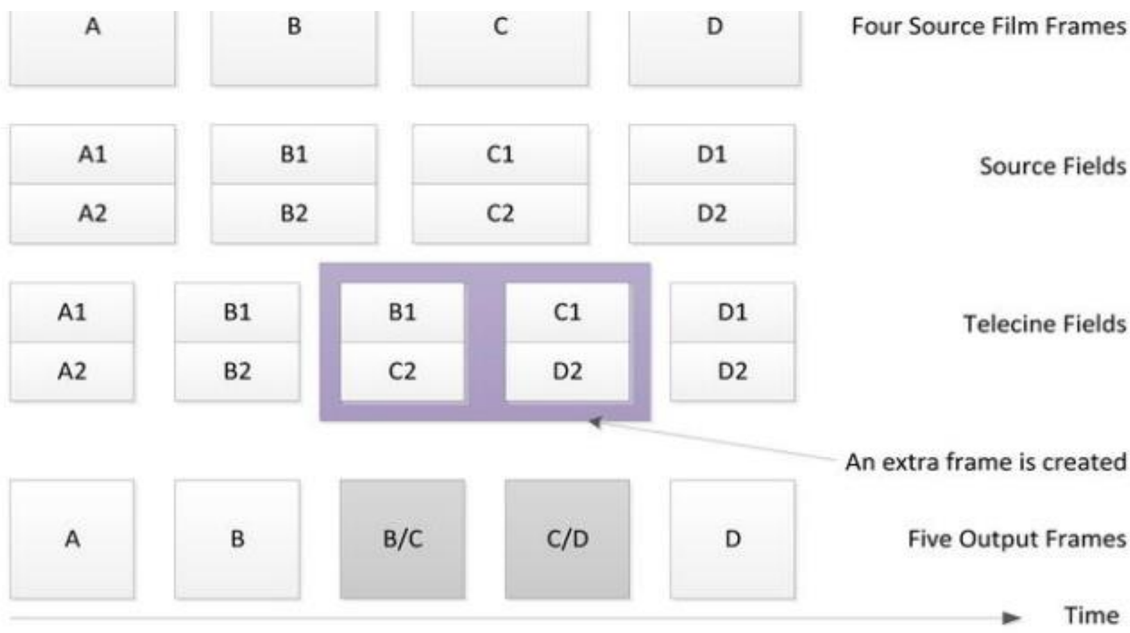
Moiré pattern

Flickering Artifacts

Flicker is perceivable interruption in brightness for a sufficiently long time (e.g., around 100 milliseconds) during display of a video. It is a flashing effect that is displeasing to the eye. Flicker occurs on old displays such as cathode ray tubes (CRT) when they are driven at a low refresh rate. Since the shutters used in liquid crystal displays (LCD) for each pixel stay at a steady opacity, they do not flicker, even when the image is refreshed.

Telecine Judder

A flicker-like artifact is the *telecine judder*. In order to convert the 24 fps film content to 30 fps video, a process called telecine is commonly applied. The process converts every four frames of films to five frames of interlaced video. Some DVD or Blu-ray players, line doublers, or video recorders can detect telecine and apply a reverse telecine process to reconstruct the original 24 fps video content. As in next figure that shows the telecine process.



The telecine process

Notice that by the telecine process two new frames B/C and C/D are created, that were not part of the original set of source frames. Thus, the telecine process creates a slight error in the video signal compared to the original film frames. This used to create the problem for films viewed on NTSC television that they would not appear as smooth as when viewed in a cinema. This problem was particularly visible during slow, steady camera movements that would appear slightly jerky when telecined.

Factors Affecting Visual Quality

Visual artifacts resulting from loss of information due to processing of digital video signals usually degrade the perceived visual quality.

The following are important contributing factors that affect visual quality:

a- Sensor noise: Sensor noise, is an undesirable by-product of image capture that affects visual quality.

b- Characteristics of video: Visual quality is affected by digital video characteristics including:

1- bit depth: Typical video frames use 8 bits for each pixel component, while premium quality videos allocate 10 to 16 bits.
2- Resolution: high-definition video frames are four to six times as large as standard.
3- frame rate: Frame rate is another important factor; although the HVS can perceive slow motion at 10 frames per second (fps) and smooth motion at 24 fps, higher frame rates imply smoother motion, especially for fast-moving objects. For example, a moving ball may be blurry at 30 fps, but would be clearer at 120 fps. Very fast motion is more demanding–wing movements of a hummingbird would be blurry at 30 fps, or even at 120 fps; for clear view of such fast motion, 1000 fps may be necessary.
4- frame complexity: One measure of the complexity of a frame is the amount of details or *spatial business* of the frame. Artifacts in frames with low complexity and low details are generally more noticeable than frames with higher complexity.
   The spatial information (detail) and temporal information (motion) of the video are critical parameters. These play a crucial role in determining the amount of video compression.

c- Amount of compression: Highly compressed video has lower visual quality than lightly compressed video.

d- Methods of compression: Lossless compression retains all the information present in the video signal, so it does not introduce quality degradation. On the other hand, lossy compression aims to control the loss of quality by performing a careful tradeoff between visual quality and amount of compression.

e- Multiple generations of compression: Some video applications may employ multiple generations of compression, where a compressed video signal is decompressed before compressing again with possibly different parameters. This may result in quality degradation owing to the use of different quantization maps for each generation. To avoid such quality loss, robust design of quantization parameters is necessary.

Video Quality Evaluation Methods and Metrics

Video quality evaluation is performed to describe the quality of a set of video sequences under study. Video quality can be evaluated with two approaches:

1- Subjectively: the actual visual quality of the image or video content is determined based on subjective evaluation done by humans.

2- Objectively: Objective video models are mathematical models that can automatically evaluate the quality of the multimedia content, predicting the subjective judgment. In this context, the term model may refer to a simple statistical model which can be implemented in software or hardware.

In general, the aforementioned models are based on criteria that can be measured objectively – that is, free from human interpretation. They can be automatically evaluated by a computer program. Unlike a panel of human observers, an objective model will always output the same quality score for a given set of input parameters.

Classification of objective video quality metrics:

Objective metrics can be classified by the amount of information available about the original signal, the received signal, or whether there is a signal present at all:

1- Full Reference Methods (FR): FR metrics compute the quality difference by comparing the original video signal against the received video signal. Typically, every pixel from the source is compared against the corresponding pixel at the received video, with no knowledge about the encoding or transmission process in between. FR metrics are usually the most accurate at the expense of higher computational effort.

2- Reduced Reference Methods (RR): RR metrics extract some features of both videos and compare them to give a quality score. They are used when all the original video is not available, or when it would be practically impossible to do so, e.g. in a transmission with a limited bandwidth. This makes them more efficient than FR metrics.

3- No-Reference Methods (NR): NR metrics try to assess the quality of a distorted video without any reference to the original signal. Due to the absence of an original signal, they may be less accurate than FR or RR approaches, but are more efficient to compute. NR can be divided into:

a- Pixel-Based Methods (NR-P): Pixel-based metrics use a decoded representation of the signal and analyze the quality based on the pixel information. Some of these evaluate specific degradation types only, such as blurring or other coding artifacts.

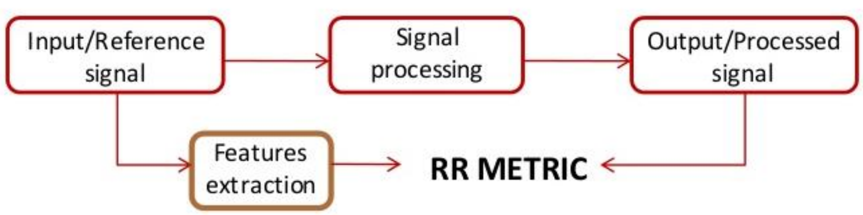
b- Parametric/ Bitstream Methods (NR-B): These metrics make use of features extracted from the transmission container and/or video bitstream, e.g. MPEG-TS packet headers, motion vectors and quantization parameters. They do not have access to the original signal.

c- Hybrid Methods (Hybrid NR-P-B): Hybrid metrics combine parameters extracted from the bitstream with a decoded video signal. They are therefore a mix between NR-P and NR-B models.

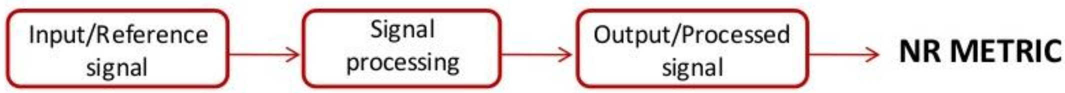
FR, RR and NR Scenarios

- **Full Reference** approach:

- **Reduced Reference** approach:



- **No-Reference** approach:



Simple full-reference metrics

The most traditional ways of evaluating quality of digital video processing system (e.g. a video codec) are FR-based. Among the oldest FR metrics are signal-to-noise ratio (SNR) and peak signal-to-noise ratio (PSNR), which are calculated between every frame of the original video signal and the video passed through a system (e.g., an encoder or a transmission channel). PSNR is the most widely used objective image quality metric, and the average PSNR over all frames can be considered a video quality metric. However, PSNR values do not correlate well with perceived picture quality due to the complex, highly non-linear behavior of the human visual system HVS.

Where PSNR is an expression for the ratio between the maximum possible power of a signal and the power of distorting noise that affects the quality of its representation after compression, processing, or transmission. The proposal is that the higher the PSNR, the better degraded image has been reconstructed to match the original image.

More complex full-reference metrics

With the success of digital video, a large number of more precise metrics were developed. These metrics are inherently more complex than PSNR, and need more computational effort to calculate predictions of video quality. Among those metrics are for example structural similarity

(SSIM) and the MOtion-tuned Video Integrity Evaluation (MOVIE) Index.

These metrics are deliver better perceptual (similarity or motion) picture quality predictions than do traditional methods such as the peak signal-to-noise ratio (PSNR) and mean squared error (MSE), which have proven to be inconsistent with human visual perception.

Subjective video quality

The main goal of many objective video quality metrics is to automatically estimate the average user's (viewer's) opinion on the quality of a video processed by a system. Their main idea is: video sequences are shown to a group of viewers and then their opinion is recorded and averaged to evaluate the quality of each video sequence. However, the testing procedure may vary depending on what kind of system is tested.

Data Compression

data compression involves encoding information using fewer bits than the original representation. There are two basic categories of compression:

1- Lossless compression:

Lossless compression algorithms usually exploit statistical redundancy to represent data without losing any information, so that the process is reversible. For example, an image may have areas of color that do not change over several pixels; instead of coding "red pixel, red pixel, ..." the data may be encoded as "279 red pixels". This is a basic example of run-length encoding; there are many schemes to reduce file size by eliminating redundancy.

2- Lossy compression:

Lossy compression is the converse of lossless data compression. In these schemes, some loss of information is acceptable. Dropping nonessential detail from the data source can save storage space. Lossy data compression schemes are designed by research on how people perceive the data in question.

For example, the human eye is more sensitive to subtle variations in luminance than it is to the variations in color.

JPEG image compression works in part by rounding off nonessential bits of information.

A number of popular compression formats exploit these perceptual differences, including those used in music files, images, and video.

Lossy image compression can be used in digital cameras, to increase storage capacities with minimal degradation of picture quality.

Similarly, DVDs use the lossy MPEG-2 video coding format for video compression.

In lossy audio compression, methods of psychoacoustics are used to remove non-audible components of the audio signal.

The two basic compression standards are JPEG and MPEG. In broad terms, JPEG is associated with still digital pictures, whilst MPEG is dedicated to digital video sequences.

The group of MPEG standards that include the MPEG 1, MPEG-2, MPEG-4 and H.264 formats have some similarities, as well as some notable differences.

The basics of compression

Compression basically means reducing image data. To reduce the media overheads for distributing these sequences, the following techniques are commonly employed to achieve desirable reductions in image data:

- Reduce color nuances within the image
- Reduce the color resolution with respect to the prevailing light intensity
- Remove small, invisible parts, of the picture

- Compare adjacent images and remove details that are unchanged between two images

The first three are image based compression techniques, where only one frame is evaluated and compressed at a time. The last one is for video compression technique where different adjacent frames are compared as a way to further reduced the image data. All of these techniques are based on an accurate understanding of how the human brain and eyes work together to form a complex visual system.

Video compression

A video clip is a sequence of pictures or frames. Each picture can be processed in isolation, much like a still image. Compression algorithms aim at lowering the total number of parameters required to represent the signal, while delivering a reasonable quality picture to the player. There are four main redundancies present in the video signal:

a) _ Spatial

b) _ Temporal

c) _ Perceptual

d) _ Statistical

Spatial redundancy occurs where neighboring pixels in a frame of a video signal are related; it could be an object of a single color.

If consecutive pictures also are related there is temporal redundancy. The human visual system has psychovisual redundancy; not all the visual information is treated with the same relevance. An example is lower acuity to color detail than luminance.

Finally, not all parameters occur with the same probability in an image. This statistical redundancy can be used in the coding of the image parameters. For example, frequently occurring parameters can be coded with fewer bits (Huffman coding).

1- Intraframe (Spatial) compression

Spatial redundancy occurs where neighboring pixels in a frame of a video signal are related. The compression of a single image in the spatial domain is called intraframe. This compression can be done by:

a- Transform coding

Transform coding converts the spatial array of pixel blocks, a bit map, to the frequency domain. Different transforms have been proposed, like:

i- The Discrete Cosine Transform codecs DCT .

ii- Discrete Fourier (DFT),

iii- Karhunen Lo'eve (KLT),

iv- Walsh–Hadamard (WHT).

The DFT was ruled out because it suffers discontinuities at the block boundaries. For typical real-world images, the DCT outperforms the WHT and DFT in the energy compaction. The KLT has the optimal decomposition but requires a very large number of operations, so it has the longest processing time. The DCT generally has been adopted because it approaches the performance of the KLT, but with a lower processing overhead. Any encoder that is to be used for video-conferencing or other live applications must have a short processing delay. So the DCT was chosen as the best compromise.

v- Wavelet compression: Although the DCT has been very successful, there has long been a demand for other compression schemes that do not suffer from the blocking artifacts that often are visible in DCT compression. Wavelet compression often has been cited as an alternative. It provides a time-frequency representation of the image and can achieve the same image quality as DCT at much higher compression ratios. It has been employed with some success in desktop video-editing systems as an alternative to motion JPEG. It is now becoming a more mainstream technology since it has been adopted as the core technology of the JPEG2000 standard for still image coding.

b- Model-based compression :

This is an alternative to waveform coding. The codec attempts to model the scene and then transmit descriptors, rather than a representation, of the spatial image.

i- Fractal

The fractal compression technique relies on the fact that, in certain images, parts of the image resemble other parts of the same image. Similar sections of an image are located, and then the fractal algorithm is applied.

ii- Object-based coding

Object-based coding has been adopted as the basis for MPEG-4 coding. A scene is represented by a number of video objects. Each object is described by its shape, texture, and motion.

2- *Temporal or interframe compression*

A sequence of video images has little change from one picture to the next, except at scene changes. This redundancy can be exploited by transmitting only the difference between successive pictures; by this means, a large reduction in the data rate can be achieved. This temporal or interframe compression allows the data rate for video sequences to be reduced much more than a sequence of unrelated still images. It typically allows a further 3:1 reduction over any initial spatial (intraframe) compression.

3- Visual perception

The human retina process the visual scene in a way that detects edges and lines. This allows objects to be rapidly separated from the background. A consequence of this is that codecs that destroy or create edges will be viewed as creating perceptible distortions. Another feature is that fine detail near the edges of objects is not perceived with great acuity. So there are opportunities for perceptual redundancies where distortions are easily noticed. A good compression architecture has to exploit the mechanisms of visual perception.

Still images – JPEG compression

The JPEG format is widely used to compress continuous-tone grayscale and color images. JPEG compresses the file size by selectively discarding picture data. This is called lossy compression. The final picture is a representation of the original, but the original can never be restored from the JPEG file. The standard also supports a lossless compression with about a 3:1 reduction in data, but it is more commonly used in the lossy mode with ratios of 20:1 or greater.

JPEG compression is based on a technique called the discrete cosine transform.

A lower compression ratio results in less data being discarded, but the JPEG compression algorithm will degrade any fine detail in an image. As the compression is increased, more artifacts become apparent. Wave-like patterns and blocky areas become visible, and there is ringing on sharp edges. The levels of artifacts that are considered acceptable depend upon the application. A picture reproduced in a magazine should have no visible artifacts, whereas minor distortions are to be expected with a thumbnail on a web page. It is more important to have a small file size with a short download time.

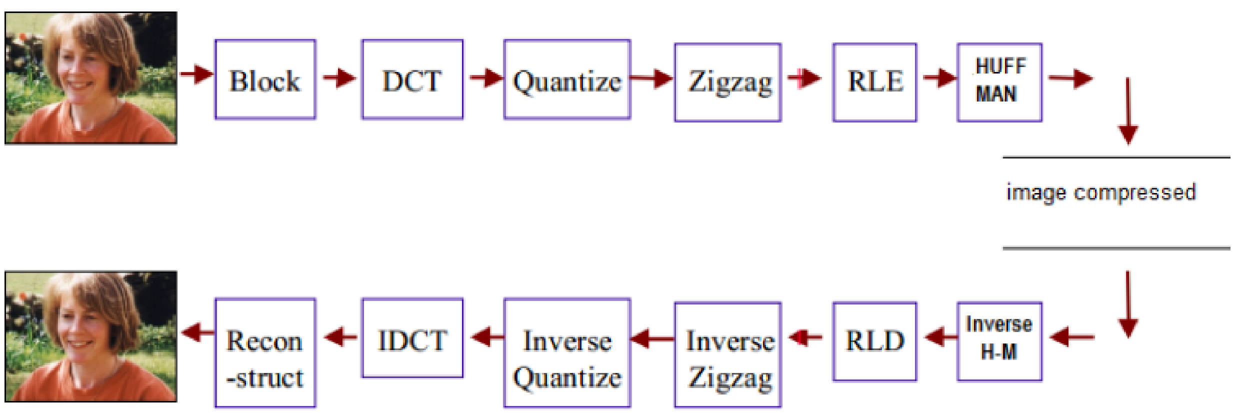


Figure.1 The steps of JPEG algorithm in lossy sequential mode.

MPEG compression

The first public standard of the MPEG committee was the MPEG-1 which first parts were released in 1993. The name **MPEG** is an acronym for **Moving Pictures Experts Group**. MPEG-1 video compression is based upon the same technique that is used in JPEG. In addition to that it also includes techniques for efficient coding of a video sequence.
Consider the video sequence displayed in Figure 2. The picture to the left is the first picture in the sequence followed by the picture in the middle and then the picture to the right. When displayed, the video

sequence shows a man running from right to left with a house that stands still.

Figure 2: *three-picture JPEG video sequence.*

In JPEG each picture in the sequence is coded as a separate unique picture resulting in the same sequence as the original one.

In MPEG video only the new parts of the video sequence is included together with information of the moving parts. The video sequence of Figure 2 will then appear as in Figure 3. But this is only true during the transmission of the video sequence to limit the bandwidth consumption. When displayed it appears as the original video sequence again.

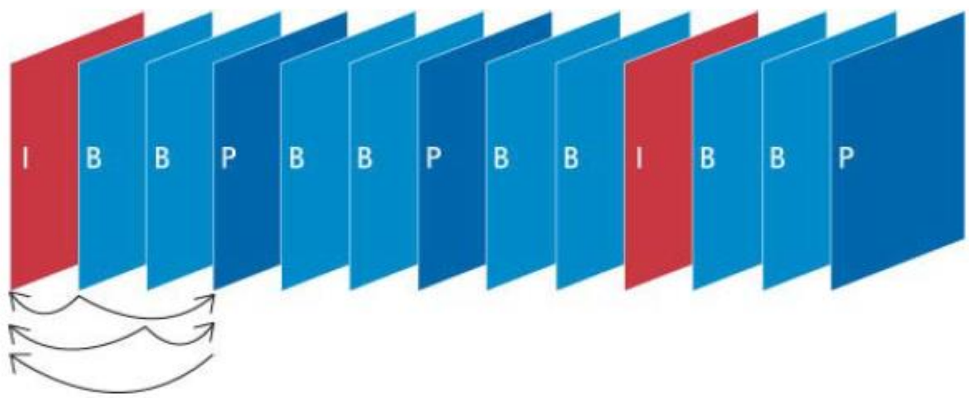
Figure 3. A three-picture MPEG video sequence.

More on MPEG compression

MPEG-4 is a fairly complex and comprehensive standard that has some characteristics that are important to understand. They are outlined below.

- **Frame types**

The basic principle for video compression is the image-to-image prediction. The first image is called an I-frame and is self-contained, having no dependency outside of that image. The following frames may use part of the first image as a reference. An image that is predicted from one reference image is called a P-frame and an image that is bidirectionally predicted from two reference images is called a B-frame.



The temporal compression is arranged in short sequence of frames called a group of pictures (GOP). MPEG defines three types of frame within the group:

1- **Intraframe or I-frame** These are coded spatially, solely from information contained within the frame.

2- **Predicted frame or P-frame** These are coded from previous I- or P-frame pictures. The decoder uses motion vectors to predict the content from the previous frames.

3- **Bidirectional frame or B-frame** These pictures use past and future I and P pictures as a reference, effectively interpolating an intermediate picture.

The video decoder restores the video by decoding the bit stream frame by frame. Decoding must always start with an I-frame, which can be decoded independently, while P- and B-frames must be decoded together with current reference image(s).

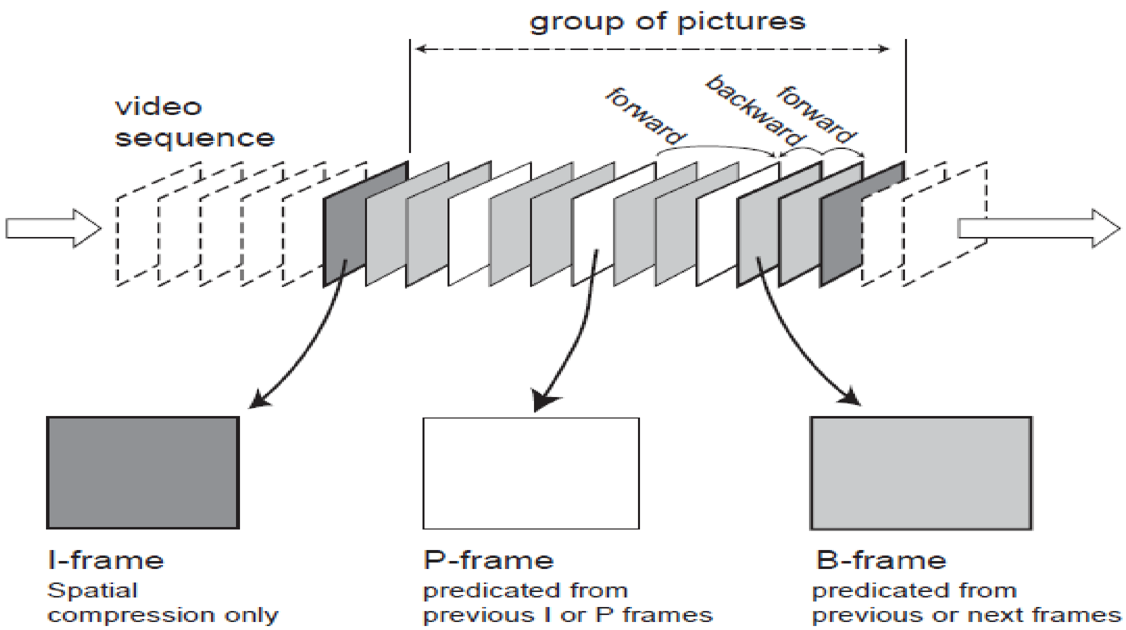


Figure5 I, P, and B frames. *Note that a P-frame may only reference a preceding I- or P-frame, while a B-frame may reference both preceding and succeeding I- and P-frames.*

- **Group of Pictures**

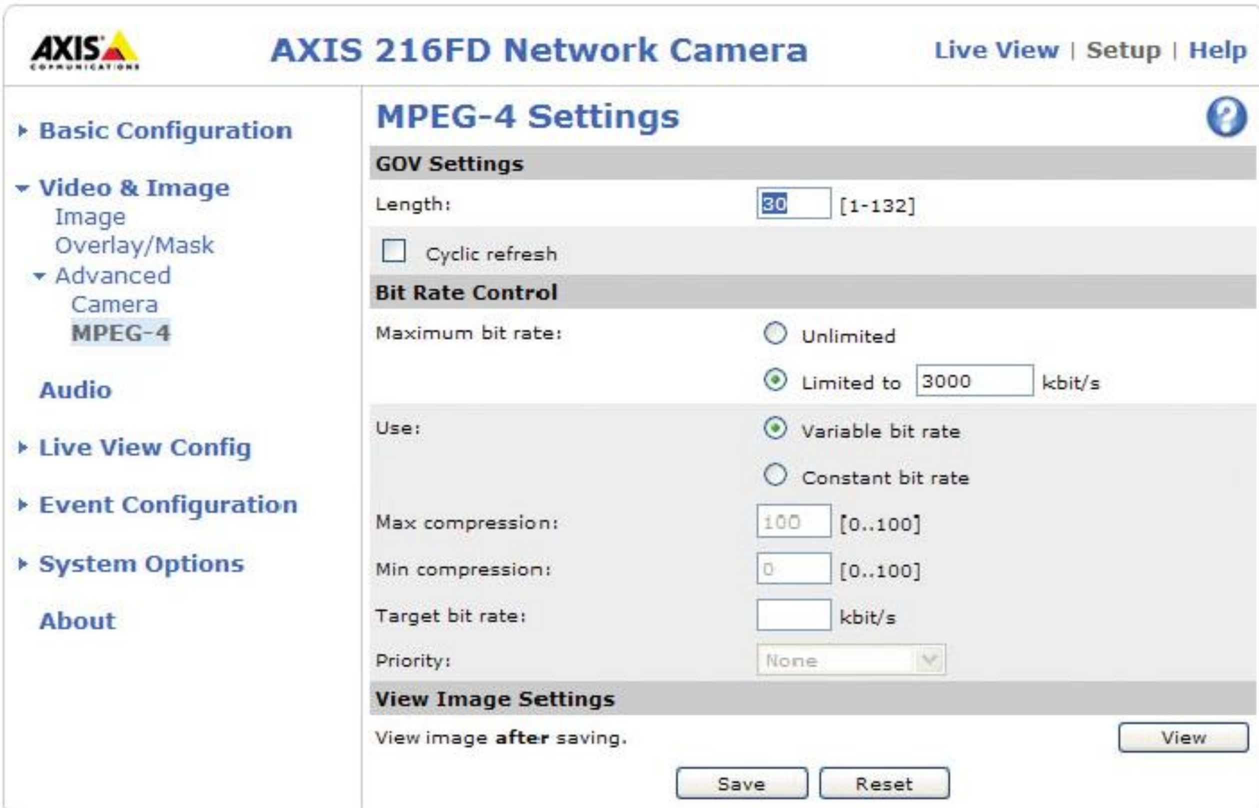
One parameter that can be adjusted in MPEG-4 is the Group of Pictures (GOP) length and structure, also referred to as Group of Video (GOV) in some MPEG standards. It is normally repeated in a fixed pattern, for example:

> GOV = 4, e.g. IPPP IPPP ...

> GOV = 15, e.g. IPPPPPPPPPPPPPP IPPPPPPPPPPPPPP ...

> GOV = 8, e.g. IBPBPBPB IBPBPBPB ...

The appropriate GOP depends on the application. By decreasing the frequency of I-frames, the bit rate can be reduced. By removing the B-frames, latency can be reduced, where latency is the delay time that needed to interpret (decompress) the data and view it on the monitor.



An interface in a network camera where the length of the Group of Video (GOV), i.e. the number of frames between two I-frames, can be adjusted to fit the application.

- Video objects

Consider the cartoon that produced by cel animation, for example, consider a downhill skier. The animators deconstruct the storyboard image into components: the background and the characters.

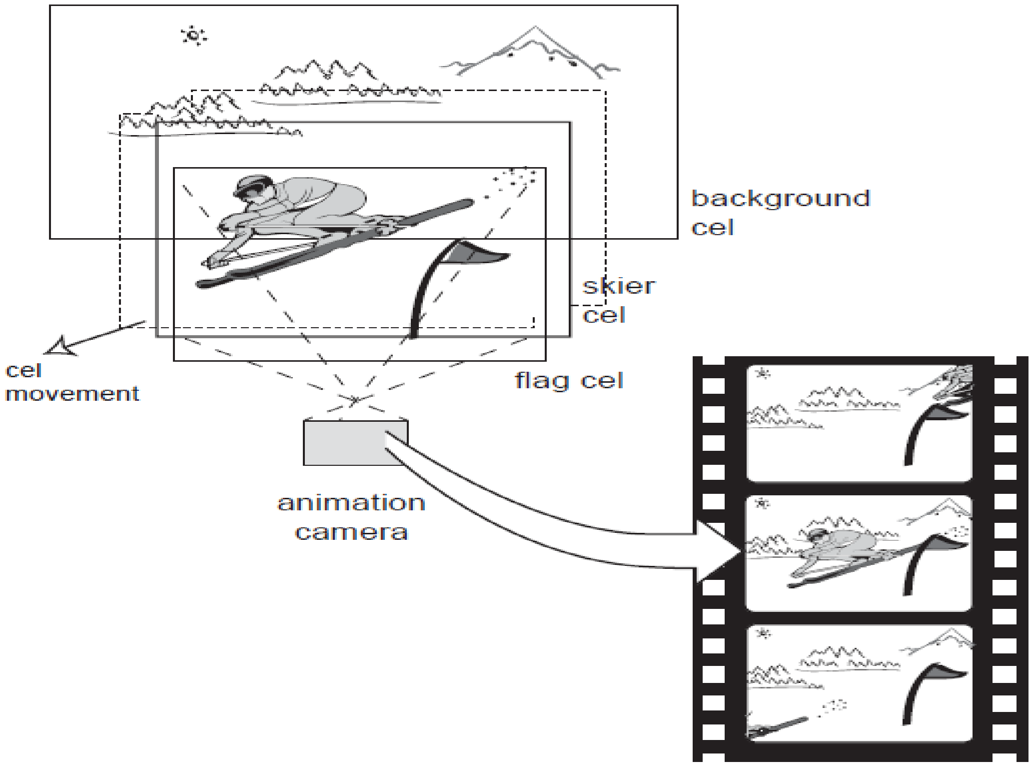
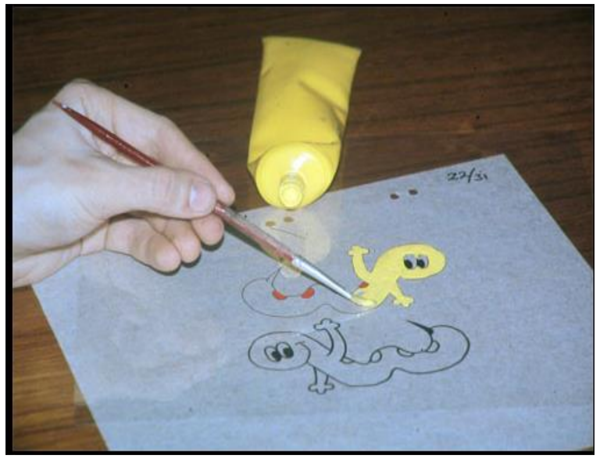


Figure7 Cel animation for animated cartoons

Each scene consists of a single background drawing, with a number of characters and foreground objects. Each is drawn separately on a clear sheet called a cel. Then each frame of a scene is photographed, with the cels stacked in the appropriate order. The skier cel can be moved for each exposure or more likely a sequence of cels will represent the movement of the skier. So a great deal of drawing effort has been saved. Only one background has been drawn, and only one flag for the foreground. The savings by drawing each frame in its entirety is immense.



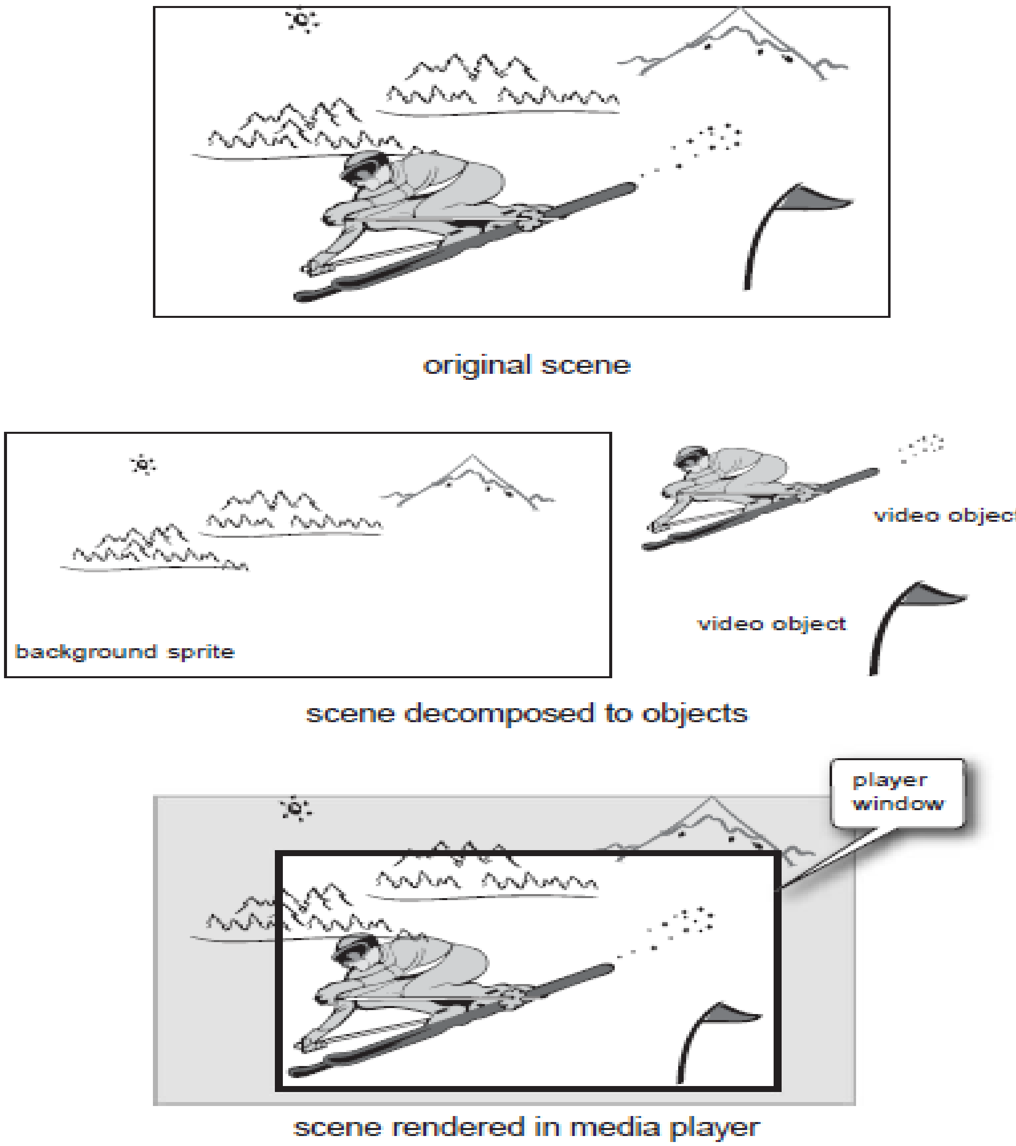
Painting on cel (see: https://www.youtube.com/watch?v=ARXwDSkY4CE)

This deconstruction of a video sequence into different components or objects is the basis of MPEG-4 video coding,

MPEG-4 breaks away from the cinematic representation and moves

toward the virtual reality world of video games.

A sprite is defined as a large video object that is persistent over time. The media player can crop and spatially transform (warp) the sprite.



MPEG-4 video objects.

.

Audio compression

Audio compression is a form of data compression designed to reduce the size of audio data files and as a result reduce the transmission bandwidth and storage requirements of audio data.

Lossless audio compression produces a representation of digital data that decompress to an exact digital duplicate of the original audio stream, unlike playback from lossy compression techniques such as MP3. Compression ratios are around 50–60% of original size, which is similar to those for generic lossless data compression. Lossless compression is unable to attain high compression ratios due to the complexity of waveforms and the rapid changes in sound forms.

Lossy audio compression

The innovation of lossy audio compression was to use psychoacoustics to recognize that not all data in an audio stream can be perceived by the human auditory system. Most lossy compression reduces perceptual redundancy by first identifying perceptually irrelevant sounds, that is, sounds that are very hard to hear. Typical examples include high frequencies or sounds that occur at the same time as louder sounds. Those sounds are coded with decreased accuracy or not at all.

Audio vs. Speech Compression Techniques:

Speech encoding is an important category of audio data compression. The perceptual models used to estimate what a human ear can hear are generally somewhat different from those used for music. The range of frequencies needed to convey the sounds of a human voice are normally far narrower than that needed for music, and the sound is normally less complex. As a result, speech can be encoded at high quality using a relatively low bit rate.

- Psycho-acoustic Model

Psychoacoustics – study of how sounds are perceived by humans

Uses perceptual coding and eliminate information from audio signal that is inaudible to the ear.

- Goals of audio compression

1- Reduced required storage space

2- Reduced required transmission bandwidth.

If the data to be compressed is analog (such as a voltage that varies with time), quantization is employed to digitize it into numbers (normally integers). This is referred to as analog-to-digital (A/D) conversion. If the integers generated by quantization are 8 bits each, then the entire range of the analog signal is divided into 256 intervals and all the signal values within an interval are quantized to the same number. If 16-bit integers are generated, then the range of the analog signal is divided into 65,536 intervals.

This relation illustrates the compromise between high resolution (a large number of analog intervals) and high compression (small integers generated). This application of quantization is used by several speech compression methods. This is accomplished, in general, by some combination of two approaches:

- Only encoding sounds that could be made by a single human voice.

- Throwing away more of the data in the signal—keeping just enough to reconstruct an "intelligible" voice rather than the full frequency range of human hearing.

Digital Rights Management

Digital Rights Management is the use of computer technology to regulate the authorized use of digital media content, and to manage the consequences of such use; for example, a payment.

A Digital Rights Management (DRM) system encrypts the content so that distribution can be controlled in accordance with the agreed rights and their terms and conditions. To this end, it wraps prices and business rules around the content to enable the payment transaction.

The digital information could be electronic books, research reports, or graphic images like still photographs.