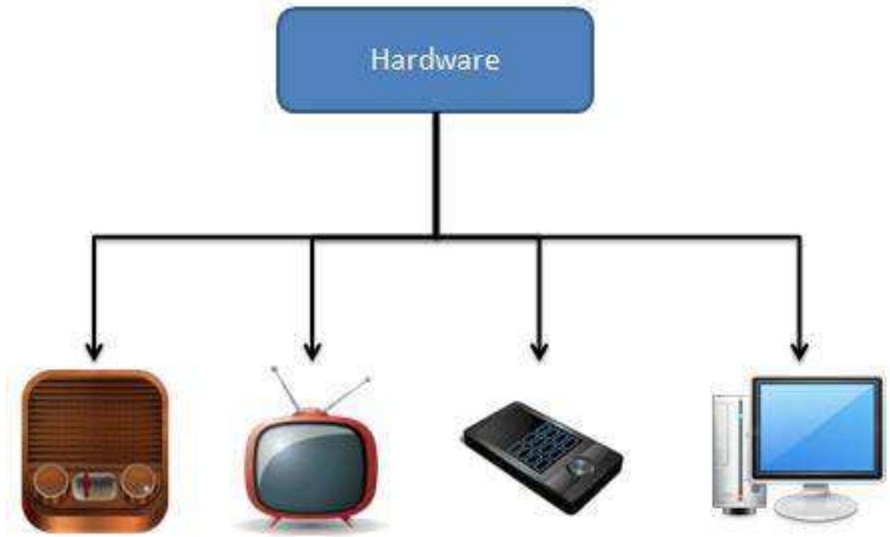
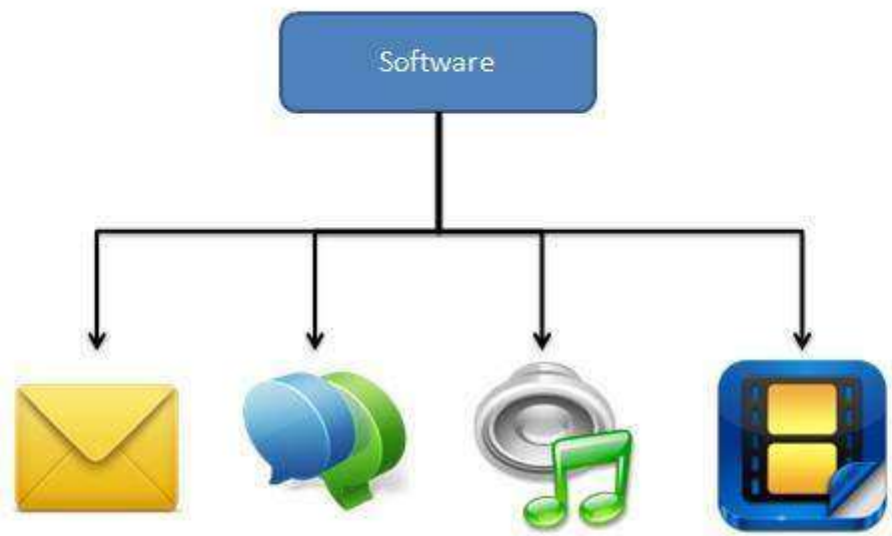
Multimedia Introduction

- Multimedia is an interactive media and provides multiple ways to represent information to the user in a powerful manner.
- It provides an interaction between users and digital information. It is a medium of communication.
- Some of the sectors where multimedias is used extensively are education, training, reference material, business presentations, advertising and documentaries.

Definition of Multimedia

- By definition Multimedia is a representation of information in an attractive and interactive manner with the use of a combination of text, audio, video, graphics and animation.
- In other words we can say that Multimedia is a computerized method of presenting information combining textual data, audio, visuals (video), graphics and animations.
- For examples: E-Mail, Yahoo Messenger, Video Conferencing, and Multimedia Message Service (MMS).
- Multimedia as name suggests is the combination of Multi and Media that is many types of media (hardware/software) used for communication of information.


Hardware


Software

Components of Multimedia

- Following are the common components of multimedia:
- **Text**- All multimedia productions contain some amount of text. The text can have various types of fonts and sizes to suit the profession presentation of the multimedia software.
- **Graphics**- Graphics make the multimedia application attractive. In many cases people do not like reading large amount of textual matter on the screen.
  - **Bitmap images**- Bitmap images are real images that can be captured from devices such as digital cameras or scanners. Generally bitmap images are not editable. Bitmap images require a large amount of memory.
  - **Vector Graphics**- Vector graphics are drawn on the computer and only require a small amount of memory. These graphics are editable.
- **Audio**- A multimedia application may require the use of speech, music and sound effects. These are called audio or sound element of multimedia. Speech is also a perfect way for teaching.
- **Video**- The term video refers to the moving picture, accompanied by sound such as a picture in television. Video element of multimedia application gives a lot of information in small duration of time. Digital video is useful in multimedia application for showing real life objects.
- **Animation**- Animation is a process of making a static image look like it is moving. An animation is just a continuous series of still images that are displayed in a sequence.

Applications of Multimedia

- Following are the common areas of applications of multimedia.
- **Multimedia in Business**- Multimedia can be used in many applications in a business.
- The multimedia technology along with communication technology has opened the door for information of global wok groups.
- Today the team members may be working anywhere and can work for various companies. Thus the work place will become global.
- The multimedia network should support the following facilities:
  - Voice Mail
  - Electronic Mail
  - Multimedia based FAX
  - Office Needs
  - Employee Training
  - Sales and Other types of Group Presentation
  - Records Management

- **Multimedia in Marketing and Advertising**- By using multimedia marketing of new products can be greatly enhanced. Multimedia boost communication on an affordable cost opened the way for the marketing and advertising personnel.
- **Multimedia in Entertainment**- By using multimedia marketing of new products can be greatly enhanced. Multimedia boost communication on an affordable cost opened the way for the marketing and advertising personnel. Presentation that have flying banners, video transitions, animations, and sound effects are some of the elements used in composing a multimedia based advertisement to appeal to the consumer in a way never used before and promote the sale of the products.
- **Multimedia in Education**- Many computer games with focus on education are now available. Consider an example of an educational game which plays various rhymes for kids. The child can paint the pictures, increase reduce size of various objects etc apart from just playing the rhymes. Several other multimedia packages are available in the market which provide a lot of detailed information and playing capabilities to kids.
- **Multimedia in Bank**- Bank is another public place where multimedia is finding more and more application in recent times. People go to bank to open saving/current accounts, deposit funds, withdraw money, know various financial schemes of the bank, obtain loans etc. Every bank has a lot of information which it wants to impart to in customers. For this purpose, it can use multimedia in many ways. Bank also displays information about its various schemes on a PC monitor placed in the rest area for customers.

- **Multimedia in Hospital**- Multimedia best use in hospitals is for real time monitoring of conditions of patients in critical illness or accident. The conditions are displayed continuously on a computer screen and can alert the doctor/nurse on duty if any changes are observed on the screen. Multimedia makes it possible to consult a surgeon or an expert who can watch an ongoing surgery line on his PC monitor and give online advice at any crucial juncture.
- **Multimedia Pedagogues**- Pedagogues are useful teaching aids only if they stimulate and motivate the students. The audio-visual support to a pedagogue can actually help in doing so. A multimedia tutor can provide multiple numbers of challenges to the student to stimulate his interest in a topic.
- **Communication Technology and Multimedia Services**- The advancement of high computing abilities, communication ways and relevant standards has started the beginning of an era where you will be provided with multimedia facilities at home. These services may include:
- Basic Television Services
- Interactive entertainment
- Digital Audio
- Video on demand
- Home shopping
- Financial Transactions
- Interactive multiplayer or single player games
- Digital multimedia libraries
- E-Newspapers, e-magazines

Hypermedia

-Is nothing but multimedia.

-an extension to hypertext providing multimedia facilities, such as those handling sound and video.

-The term can be used as a noun (a medium with multiple content forms) or as an adjective describing a medium as having multiple content forms.

-Can be example as one of the multimedia application.

-Is not to be constrained to be text based. It can include other media.

EXAMPLES

The World Wide Web (WWW) is the best example of hypermedia applications Through WWW it is possible to deliver hypertext, graphics, animation and sound between different computer environments .

Most educational IT applications are hypermedia and these include:

T**utorial software packages :**A computer tutorial is an interactive software program created as a learning tool.

Knowledge webpage : Are web pages from the internet that contain a myriad of information regarding different topics based on what the user needs

Simulation instructional games :Educational games and simulations, unlike direct forms of instruction, are experiential exercises.

The presentation of information-learning activities in hypermedia is said to be sequenced in a non-linear manner This fact makes it therefore important to understand hypermedia in the educational context in order to ensure their successful integration in the teaching-learning process.

Characteristics of hypermedia applications

1.Learner Control.
2.Learner wide range of navigation routes.
3.Variety of media

Learner Control-

The learners makes his own decisions on the path, flow or events of instruction.

Learner wide range of navigation routes-

The learner controls the sequence and pace of his path depending on his ability and motivation.

Variety of media-

Hypermedia includes more than one media (text, graphics, audio, animation and video clip) but does not necessarily use all types of media in one presentation.

World Wide Web(WWW)

The World Wide Web(abbreviated WW Worthe Web) is an information space where documents and other web resources are identified by Uniform Resource Locators (URLs), inter linked by Hypertext links, and can be accessed via the internet.

CONCEPT OF WWW

URL

Uniform resource locator: The last part of the puzzle required to allow the web to work is a URL. This is the address which indicates where any given document lives on the web

HTML

Hypertext markup language: A standardized system for tagging text files to achieve font, color, graphic & hyperlink effects on WWW pages.

HTTP

Hypertext Transfer Protocol: This protocol requests the 'HTML' document from the server and serves it to the browser.

HYPERLINK: A hyperlink is a word, phrase, or image that you can click on to jump to a new document or a new section within the current document. Hyperlinks are found in nearly all Web pages, allowing users to click their way from page to page. Text hyperlinks are often blue and underlined .

HYPERTEXT: Hypertext is text that links to other information. By clicking on a link in a hypertext document, a user can quickly jump to different content. Hypertext is usually associated with Web pages. Today nearly every web page includes links to other pages and both text and images can be used as links to more content.

WEBPAGE: A web page or webpage is are source of information that is suitable for the world wide web and can be accessed through a web browser. This information is usually in HTML or XHTML format and may provide navigation to other web pages via hyper text links.

WEBSITES :A website is a collection of web pages, images, videos or other digital assets that is hosted on one or more web servers, usually accessible via the internet. All publicly accessible web sites are seen collectively as constituting the world wide web.

Multimedia Software

- Multimedia software tells the hardware what to do. For example, multimedia software tells the hardware to display the color blue, play the sound of cymbals crashing etc.
- To produce these media elements(movies, sound, text, animation, graphics etc.) there are various software available in the market such as Paint Brush, Photo Finish, Animator, Photo Shop, 3D Studio, Corel Draw, Sound Blaster, IMAGINET, Apple Hyper Card, Photo Magic, Picture Publisher.

Multimedia Software Categories

- Following are the various categories of Multimedia software
- **Device Driver Software**- These software's are used to install and configure the multimedia peripherals.
- **Media Players**- Media players are applications that can play one or more kind of multimedia file format.
- **Media Conversion Tools**- These tools are used for encoding / decoding multimedia contexts and for converting one file format to another.
- **Multimedia Editing Tools**- These tools are used for creating and editing digital multimedia data.
- **Multimedia Authoring Tools**- These tools are used for combing different kinds of media formats and deliver them as multimedia contents.

Multimedia Software Tools

- Multimedia applications are created with the help of following mentioned tools and packages.
- The sound, text, graphics, animation and video are the integral part of multimedia software. To produce and edit these media elements, there are various software tools available in the market. The categories of basic software tools are:
- **Text Editing Tools**- These tools are used to create letters, resumes, invoices, purchase orders, user manual for a project and other documents. MS-Word is a good example of text tool. It has following features:
  - Creating new file, opening existing file, saving file and printing it.
  - Insert symbol, formula and equation in the file.
  - Correct spelling mistakes and grammatical errors.
  - Align text within margins.
  - Insert page numbers on the top or bottom of the page.
  - Mail-merge the document and making letters and envelopes.
  - Making tables with variable number of columns and rows.

- **Painting and Drawing Tools**- These tools generally come with a graphical user interface with pull down menus for quick selection. You can create almost all kinds of possible shapes and resize them using these tools.
- Drawing file can be imported or exported in many image formats like .gif, .tif, .jpg, .bmp, etc. Some examples of drawing software are Corel Draw, Freehand, Designer, Photoshop, Fireworks, Point etc. These software have following features:
- Tools to draw a straight line, rectangular area, circle etc.
- Different color selection option.
- Pencil tool to draw a shape freehand.
- Eraser tool to erase part of the image.
- Zooming for magnified pixel editing.
- **Video Editing Tools**- These tools are used to edit, cut, copy, and paste your video and audio files. Video editing used to require expensive, specialized equipment and a great deal of knowledge.
- The artistic process of video editing consists of deciding what elements to retain, delete or combine from various sources so that they come together in an organized, logical and visually planning manner.

- **Sound Editing Tools**- These tools are used to integrate sound into multimedia project very easily. You can cut, copy, paste and edit segments of a sound file by using these tools. The presence of sound greatly enhances the effect of a mostly graphic presentation, especially in a video. Examples of sound editing software tools are: Cool Edit Pro, Sound Forge and Pro Tools. These software have following features:
    - Record your own music, voice or any other audio.
    - Record sound from CD, DVD, Radio or any other sound player.
    - You can edit, mix the sound with any other audio.
    - Apply special effects such as fade, equalizer, echo, reverse and more.
- **Animation and Modeling Tools**- An animation is to show the still images at a certain rate to give it visual effect with the help of Animation and modeling tools. These tools have features like multiple windows that allow you to view your model in each dimension, ability to drag and drop primitive shapes into a scene, color and texture mapping, ability to add realistic effects such as transparency, shadowing and fog etc. Examples of Animations and modeling tools are 3D studio max and Maya.

Graphics and Image Data Representation

- An image consists of a rectangular array of dots called pixels. The size of the image is specified in terms of width X height, in numbers of the pixels. The physical size of the image, in inches or centimeters, depends on the resolution of the device on which the image is displayed. The resolution is usually measured in DPI (Dots Per Inch). An image will appear smaller on a device with a higher resolution than on one with a lower resolution. For color images, one needs enough bits per pixel to represent all the colors in the image. The number of the bits per pixel is called the depth of the image.

Graphics Image data types

- Images can be created by using different techniques of representation of data called data type like monochrome and colored images. Monochrome image is created by using single color whereas colored image is created by using multiple colors. Some important data types of images are following:
- **1-bit images**- An image is a set of pixels. Note that a pixel is a picture element in digital image. In 1-bit images, each pixel is stored as a single bit (0 or 1). A bit has only two states either on or off, white or black, true or false. Therefore, such an image is also referred to as a binary image, since only two states are available. 1-bit image is also known as 1-bit monochrome images because it contains one color that is black for off state and white for on state.
- A 1-bit image with resolution 640*480 needs a storage space of 640*480 bits.
- 640 x 480 bits. = (640 x 480) / 8 bytes = (640 x 480) / (8 x 1024) KB= 37.5KB.
- The clarity or quality of 1-bit image is very low.

- **8-bit Gray level images**- Each pixel of 8-bit gray level image is represented by a single byte (8 bits). Therefore each pixel of such image can hold 2^8=256 values between 0 and 255. Therefore each pixel has a brightness value on a scale from black (0 for no brightness or intensity) to white (255 for full brightness or intensity). For example, a dark pixel might have a value of 15 and a bright one might be 240.
- A 8-bit image with resolution 640 x 480 needs a storage space of 640 x 480 bytes=(640 x 480)/1024 KB= 300KB. Therefore an 8-bit image needs 8 times more storage space than 1-bit image.
- **24-bit color images** - In 24-bit color image, each pixel is represented by three bytes, usually representing RGB (Red, Green and Blue). Usually true color is defined to mean 256 shades of RGB (Red, Green and Blue) for a total of 16777216 color variations. It provides a method of representing and storing graphical image information an RGB color space such that a colors, shades and hues in large number of variations can be displayed in an image such as in high quality photo graphic images or complex graphics.
- Many 24-bit color images are stored as 32-bit images, and an extra byte for each pixel used to store an alpha value representing special effect information.

- A 24-bit color image with resolution 640 x 480 needs a storage space of 640 x 480 x 3 bytes = (640 x 480 x 3) / 1024=900KB without any compression. Also 32-bit color image with resolution 640 x 480 needs a storage space of 640 x 480 x 4 bytes= 1200KB without any compression.
- Disadvantages
- Require large storage space
- Many monitors can display only 256 different colors at any one time. Therefore, in this case it is wasteful to store more than 256 different colors in an image.
- **8-bit color images** - 8-bit color graphics is a method of storing image information in a computer's memory or in an image file, where one byte (8 bits) represents each pixel. The maximum number of colors that can be displayed at once is 256. 8-bit color graphics are of two forms. The first form is where the image stores not color but an 8-bit index into the color map for each pixel, instead of storing the full 24-bit color value. Therefore, 8-bit image formats consists of two parts: a color map describing what colors are present in the image and the array of index values for each pixel in the image. In most color maps each color is usually chosen from a palette of 16,777,216 colors (24 bits: 8 red, 8green, 8 blue).

Image file formats

- **GIF- Graphics Interchange Formats**- The GIF format was created by CompuServe. It supports 256 colors. GIF format is the most popular on the Internet because of its compact size. It is ideal for small icons used for navigational purpose and simple diagrams. GIF creates a table of up to 256 colors from a pool of 16 million. If the image has less than 256 colors, GIF can easily render the image without any loss of quality. When the image contains more colors, GIF uses algorithms to match the colors of the image with the palette of optimum set of 256 colors available. Better algorithms search the image to find and the optimum set of 256 colors
- **PNG- Portable Network Graphics**- PNG is the only lossless format that web browsers support. PNG supports 8 bit, 24 bits, 32 bits and 48 bits data types. One version of the format PNG-8 is similar to the GIF format. But PNG is the superior to the GIF. It produces smaller files and with more options for colors. It supports partial transparency also. PNG-24 is another flavor of PNG, with 24-bit color supports, allowing ranges of color akin to high color JPEG. PNG-24 is in no way a replacement format for JPEG because it is a lossless compression format. This means that file size can be rather big against a comparable JPEG. Also PNG supports for up to 48 bits of color information.

- **TIFF- Tagged Image File Format**- The TIFF format was developed by the Aldus Corporation in the 1980 and was later supported by Microsoft. TIFF file format is widely used bitmapped file format. It is supported by many image editing applications, software used by scanners and photo retouching programs.
- **BMP- Bitmap**- The bitmap file format (BMP) is a very basic format supported by most Windows applications. BMP can store many different type of image: 1 bit image, grayscale image, 8 bit color image, 24 bit RGB image etc. BMP files are uncompressed. Therefore, these are not suitable for the internet. BMP files can be compressed using lossless data compression algorithms.
- **EPS- Encapsulated Postscript**- The EPS format is a vector based graphic. EPS is popular for saving image files because it can be imported into nearly any kind of application. This file format is suitable for printed documents. Main disadvantage of this format is that it requires more storage as compare to other formats.
- **PDF- Portable Document Format**- PDF format is vector graphics with embedded pixel graphics with many compression options. When your document is ready to be shared with others or for publication. This is only format that is platform independent. If you have Adobe Acrobat you can print from any document to a PDF file. From illustrator you can save as .PDF.

- **EXIF- Exchange Image File**- Exif is an image format for digital cameras. A variety of tage are available to facilitate higher quality printing, since information about the camera and picture - taking condition can be stored and used by printers for possible color correction algorithms.it also includes specification of file format for audio that accompanies digital images.

- **WMF- Windows Meta File**- WMF is the vector file format for the MS-Windows operating environment. It consists of a collection of graphics device interface function calls to the MS-Windows graphic drawing library. Metafiles are both small and flexible, hese images can be displayed properly by their proprietary software's only.
- **PICT**- PICT images are useful in Macintosh software development, but you should avoid them in desktop publishing. Avoid using PICT format in electronic publishing-PICT images are prone to corruption.
- **Photoshop**- This is the native Photoshop file format created by Adobe. You can import this format directly into most desktop publishing applications.

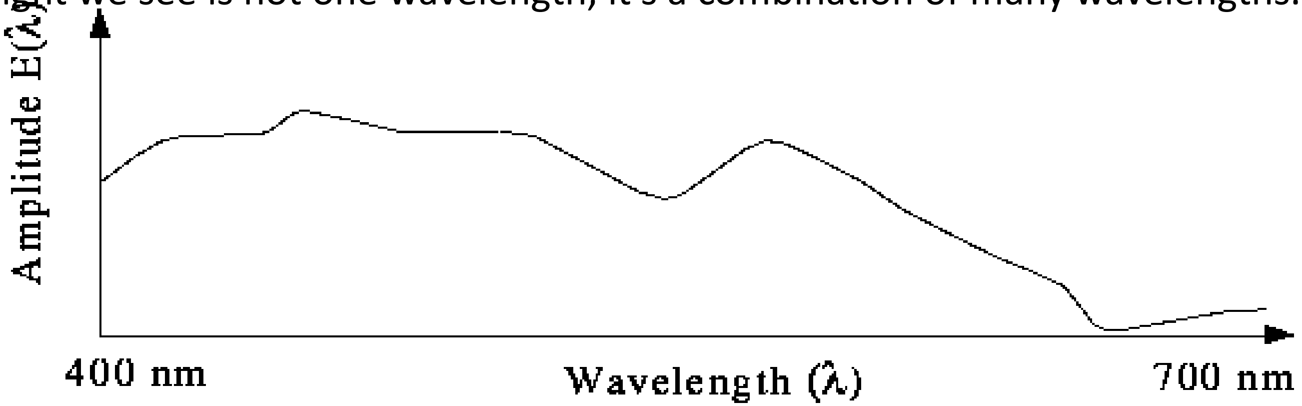
Color in Image and Video

Color Science

Light and Spectra

Visible light is an electromagnetic wave in the 400 nm - 700 nm range.

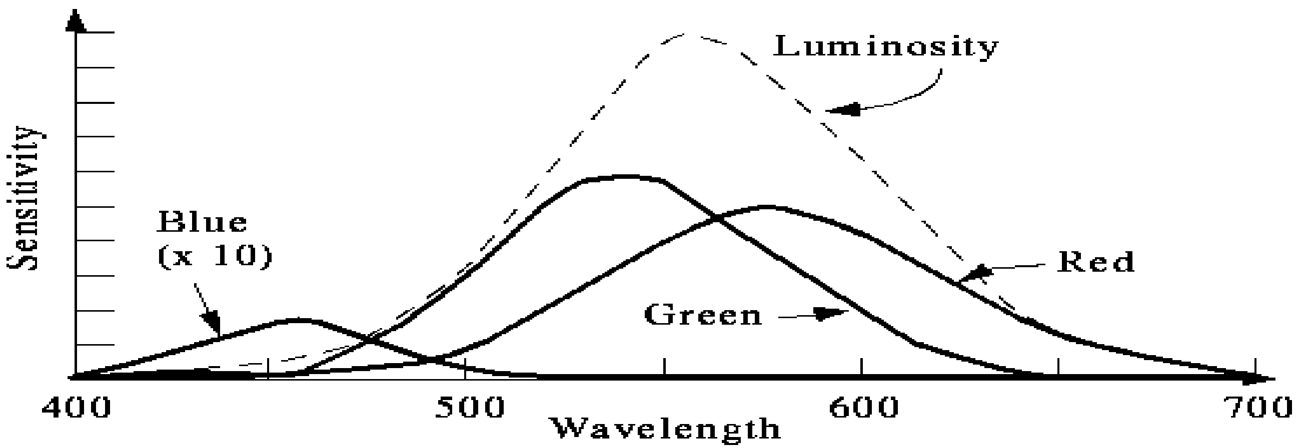
Most light we see is not one wavelength, it's a combination of many wavelengths.



The profile above is called a *spectral power distribution* or *spectrum*.

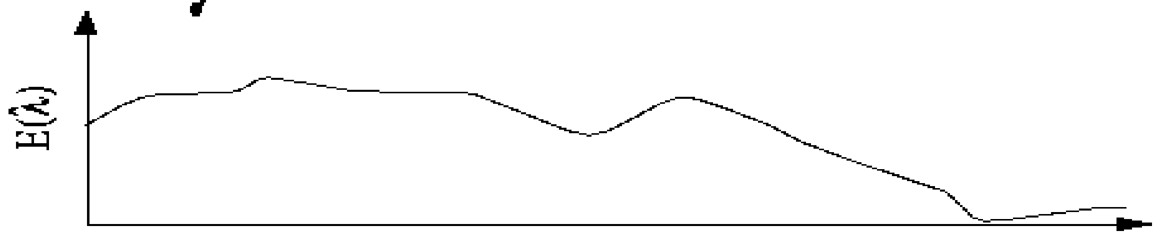
The Human Retina
The eye is basically just a camera

Each neuron is either a *rod* or a *cone*. Rods are not sensitive to color.
Cones and Perception
Cones come in 3 types: red, green and blue. Each responds differently to various frequencies of light. The following figure shows the spectral sensitivity functions of the cones and the luminous-efficiency function of the human eye.

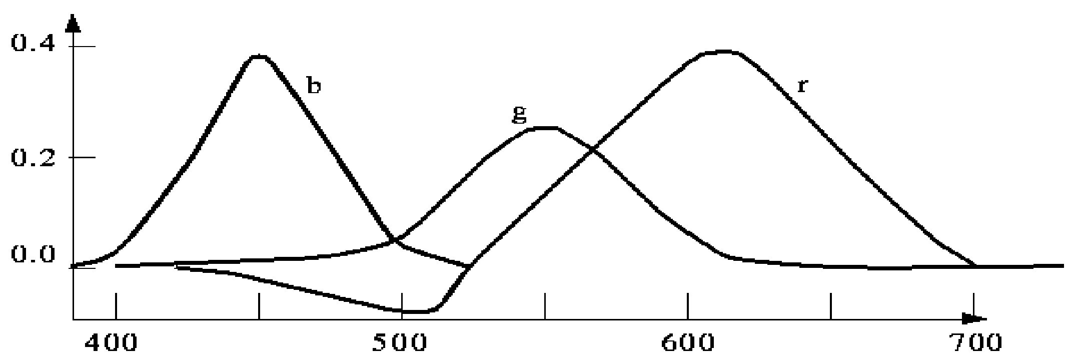


The color signal to the brain comes from the response of the 3 cones to the spectra being observed. That is, the signal consists of 3 numbers:

$$G = \int E(\lambda)\, S_G(\lambda)\, d\lambda$$



where *E* is the light (spectral power distribution) and *S* are the spectral sensitivity functions.
A color can be specified as the sum of three colors. So colors form a 3 dimensional vector space.
The following figure shows the amounts of three primaries needed to match all the wavelengths of the visible spectrum.

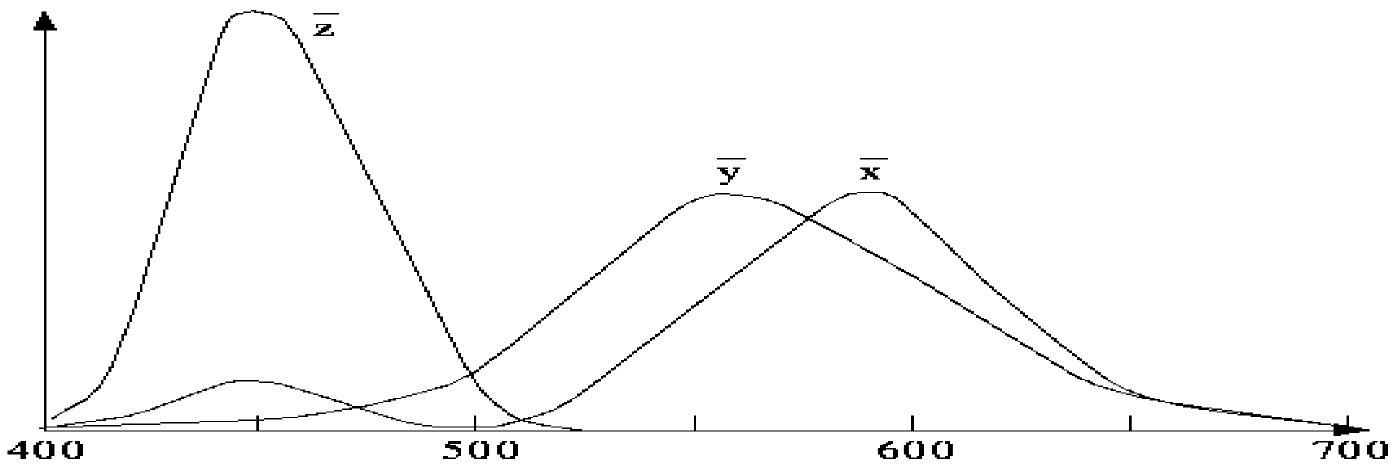


The negative value indicates that some colors cannot be exactly produced by adding up the primaries.

CIE Chromaticity Diagram

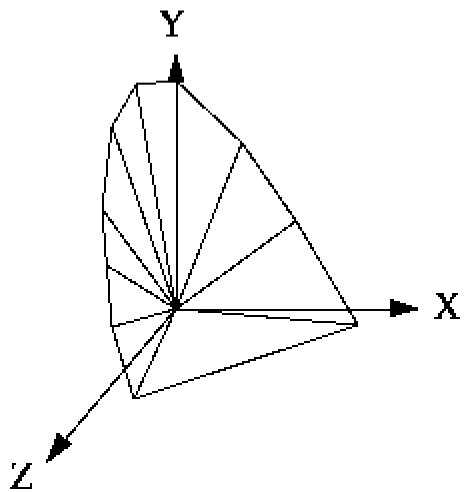
Q: Does a set of primaries exist that span the space with only positive coefficients?
A: Yes, but no pure colors.

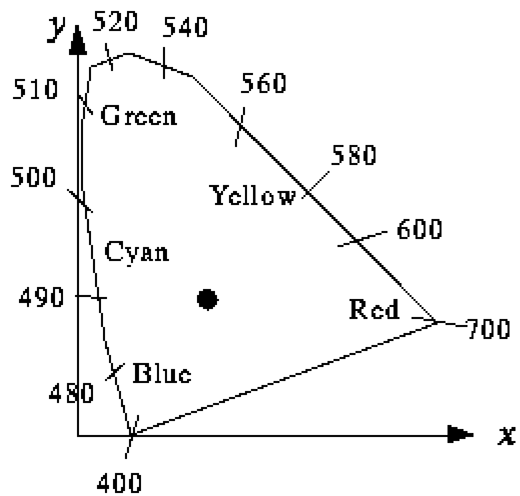
In 1931, the CIE (Commission International de L'Eclairage, or International Commission on Illumination) defined three standard primaries **(X, Y, Z)**. The **Y** primary was intentionally chosen to be identical to the luminous-efficiency function of human eyes.



The above figure shows the amounts of X, Y, Z needed to exactly reproduce any visible color.

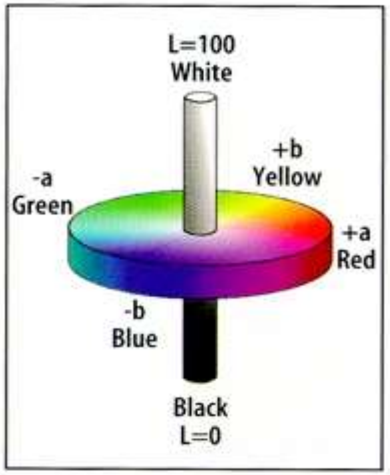
All visible colors are in a "horseshoe" shaped cone in the X-Y-Z space. Consider the plane *X+Y+Z=1* and project it onto the X-Y plane, we get the *CIE chromaticity diagram* as below.





The edges represent the "pure" colors (sine waves at the appropriate frequency)
White (a blackbody radiating at 6447 kelvin) is at the "dot"
When added, any two colors (points on the CIE diagram) produce a point on the line between them.

L*a*b (Lab) Color Model



Lab model

A refined CIE model, named CIE L*a*b in 1976
Luminance: L

Chrominance: a -- ranges from green to red, b -- ranges from blue to yellow Used by *Photoshop*

Color Models in Images

A color image is a 2-D array of (R,G,B) integer triplets. These triplets encode how much the corresponding phosphor should be excited in devices such as a monitor.

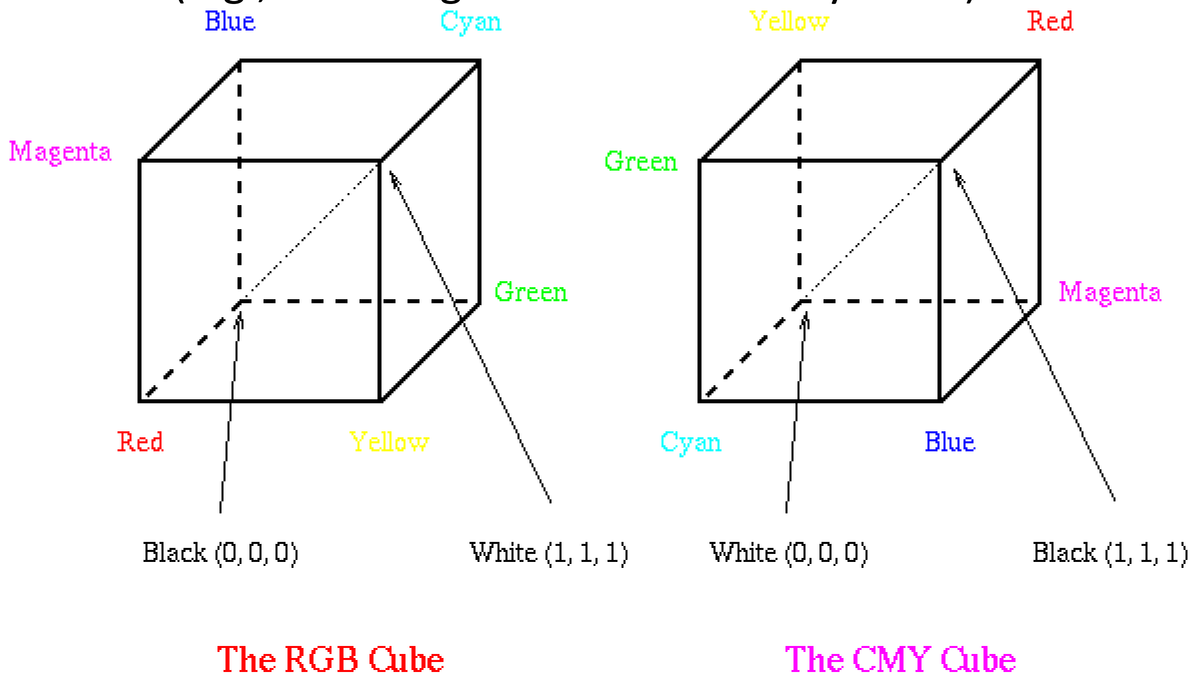
RGB Color Model for CRT Displays

CRT displays have three phosphors (RGB) which produce a combination of wavelengths when excited with electrons.

CMY Color Model

Cyan, Magenta, and Yellow (CMY) are complementary colors of RGB. They can be used as *Subtractive Primaries*.

CMY model is mostly used in printing devices where the color pigments on the paper absorb certain colors (e.g., no red light reflected from cyan ink).



The RGB and CMY Cubes

Conversion between RGB and CMY:

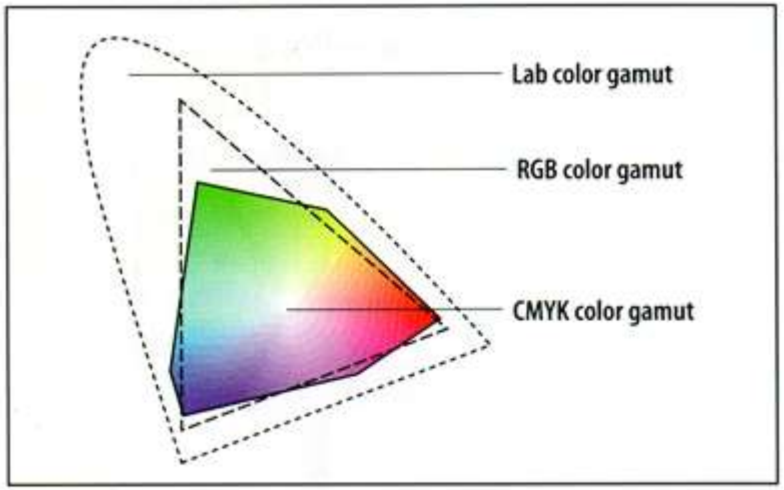
-- e.g., convert White from (1, 1, 1) in RGB to (0, 0, 0) in CMY.

$$\begin{bmatrix} C \\ M \\ Y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Sometimes, an alternative CMYK model (K stands for *Black*) is used in color printing (e.g., to produce darker black than simply mixing CMY).

K := min (C, M, Y), C := C - K, M := M - K, Y := Y - K.

Comparison of Three Color Gamuts



•The Lab gamut covers all colors in visible spectrum
•The RGB gamut is smaller, hence certain visible colors (e.g. pure yellow, pure cyan) cannot be seen on monitors
•The CMYK gamut is the smallest (but not a straight subset of the RGB gamut)

•The *gamut* of colors is all colors that can be reproduced using the three primaries

Color Models in Video

YIQ and YUV are the two commonly used color models in video

YUV Color Model

Initially, for PAL analog video, it is now also used in CCIR 601 standard for digital video

Y (luminance) is the CIE Y primary.

$$Y = 0.299R + 0.587G + 0.114B$$

Chrominance is defined as the difference between a color and a reference white at the same luminance. It can be represented by U and V -- the *color differences*.

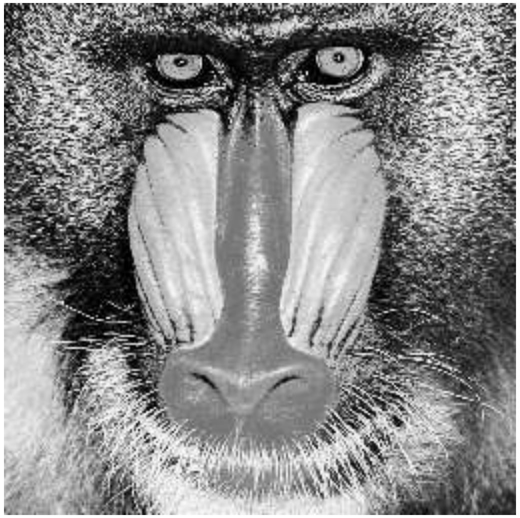
$$U = B - Y$$
$$V = R - Y$$

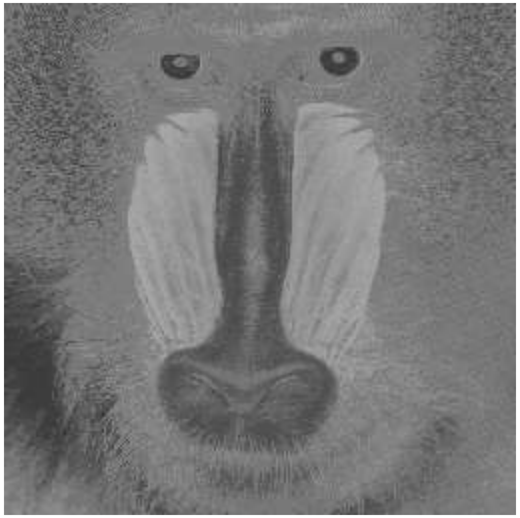
If b/w image, then U = V = 0. --> No chrominance!

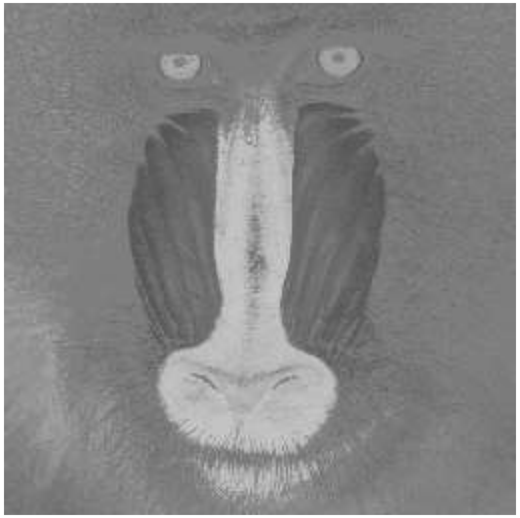
** In actual PAL implementation:

$$U = 0.492\,(B - Y)$$
$$V = 0.877\,(R - Y)$$

Sample YUV Decomposition:







Eye is most sensitive to Y. In PAL, 5 (or 5.5) MHz is allocated to Y, 1.3 MHz to U and V.

YCbCr Color Model

The YCbCr model is closely related to the YUV, it is a scaled and shifted YUV.

$Cb = (B - Y) / 1.772 + 0.5$
$Cr = (R - Y) / 1.402 + 0.5$
The chrominance values in YCbCr are always in the range of 0 to 1.
YCbCr is used in JPEG and MPEG.

YIQ Color Model

YIQ is used in NTSC color TV broadcasting, it is downward compatible with B/W TV where only Y is used.
Although U and V nicely define the color differences, they do not align with the desired human perceptual color sensitivities. In NTSC, I and Q are used instead.

I is the orange-blue axis, Q is the purple-green axis.
I and Q axes are scaled and rotated R - Y and B - Y (by 33 degrees clockwise).
$I = 0.877(R - Y) \cos 33 - 0.492(B - Y) \sin 33$
$Q = 0.877(R - Y) \sin 33 + 0.492(B - Y) \cos 33$
Namely,
$I = 0.736(R - Y) - 0.268(B - Y) = 0.596R - 0.275G - 0.321B$
$Q = 0.478(R - Y) + 0.413(B - Y) = 0.212R - 0.523G + 0.311B$
The YIQ transform:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.275 & -0.321 \\ 0.212 & -0.523 & 0.311 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Eye is most sensitive to Y, next to I, next to Q.

In NTSC broadcast TV, 4.2 MHz is allocated to Y, 1.5 MHz to I and 0.55 MHz to Q. For VCR, Y is cut down to 3.2 MHz and I to 0.63 MHz.

Fundamental Concepts in Video and Digital Audio

- Types of Video Signals
- Analog Video
- Digital Video
- Digitization of Sound
- MIDI
- Quantization and Transmission of Audio

Types of video signals

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) However, requires more bandwidth and good synchronization of the three components.

Composite Video — 1 Signal

Color (“chrominance”) and intensity (“luminance”) signals are mixed into a single carrier wave.

a) Chrominance is a composition of two color components (I and Q, or U and V).

b) In NTSC TV, e.g., I and Q are combined into a Chroma signal, and a color subcarrier is then employed to put the Chroma signal at the high-frequency end of the signal shared with the luminance signal.

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

d) When connecting to TVs or VCRs, 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.

- Since color and intensity are wrapped into the same signal, some interference between the luminance and chrominance signals is inevitable.

S-Video — 2 Signals

As a compromise, (separated video, or Super-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 gray scale 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.

Analog Video

- An analog signal f(t) samples a time-varying image. So-called “progressive” scanning traces through a complete picture (a frame) row-wise for each time interval.
- In TV, and in some monitors and multimedia standards as well, another system, called “interlaced” scanning is used:

a) The odd-numbered lines are traced first, and then the even-numbered lines are traced. This results in “odd” and “even” fields — two fields make up one frame.

b) In fact, the odd lines (starting from 1) end up at the middle of a line at the end of the odd field, and the even scan starts at a half-way point.

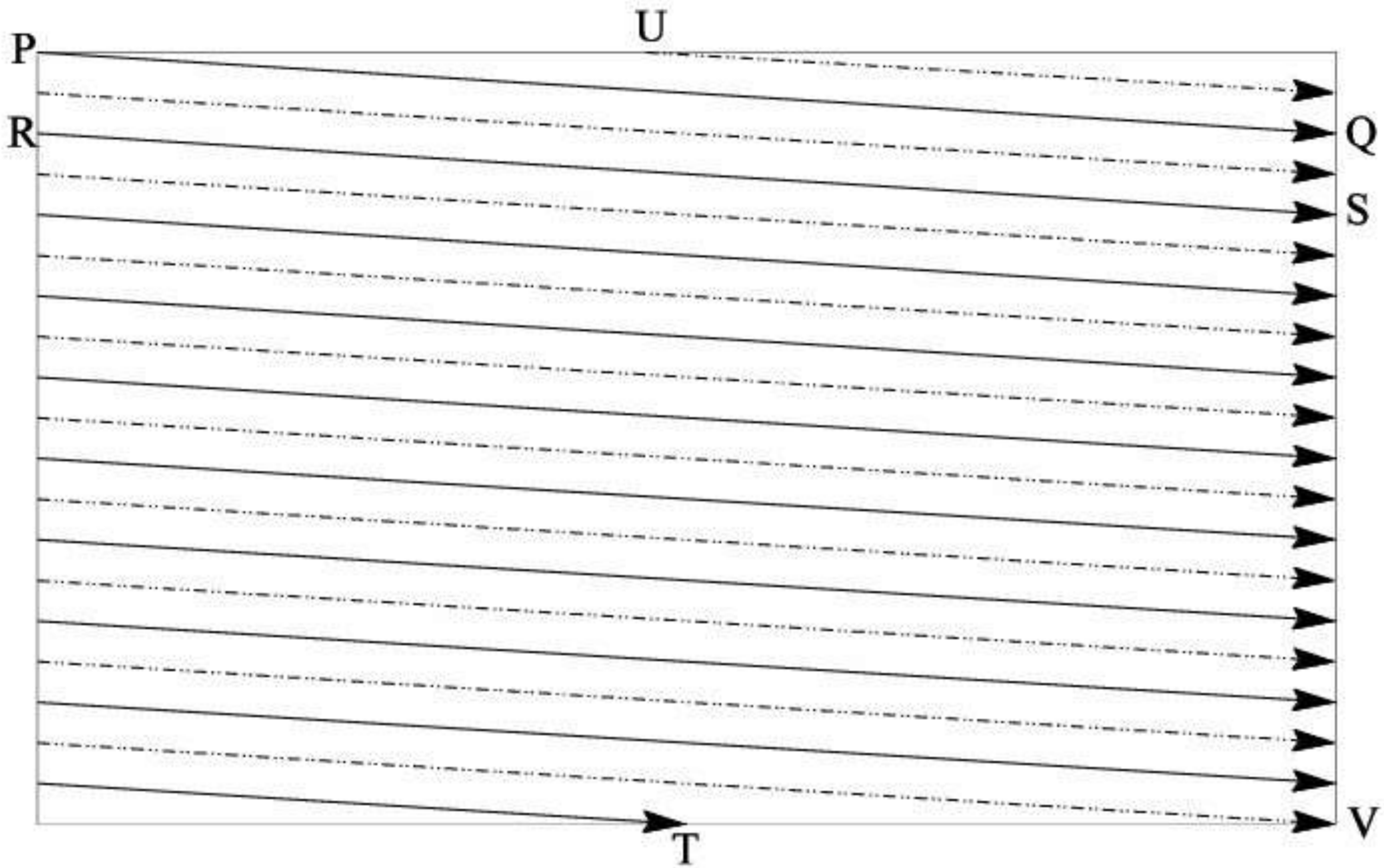


Fig. 3.1: Interlaced raster scan

c) Figure 3.1 shows the scheme used. First the solid (odd) lines are traced, P to Q, then R to S, etc., ending at T; then the even field starts at U and ends at V.

d) The jump from Q to R, etc. in Figure 3.1 is called the horizontal retrace, during which the electronic beam in the CRT is blanked. The jump from T to U or V to P is called the vertical retrace.

- 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

NTSC Video

NTSC(National Television System Committee)

- Mostly used in North America and Japan.
- Uses the familiar 4:3 aspect ratio (i.e., the ratio of picture width to its height)
- Uses 525 scan lines per frame at 30 fps.
- Follows the interlaced scanning system, and each frame is divided into two fields, with 262.5 lines/field. Uses the YIQ color model

- Vertical retrace takes place during 20 lines reserved for control information at the beginning of each field. Hence, the number of active video lines per frame is only 485.
- Similarly, almost 1/6 of the raster at the left side is blanked for horizontal retrace and sync. The non-blanking pixels are called active pixels.
- The horizontal retrace takes 10.9 μ sec

- Fig. 3.3 shows the effect of "vertical retrace & sync" and "horizontal retrace & sync" on the NTSC video raster.

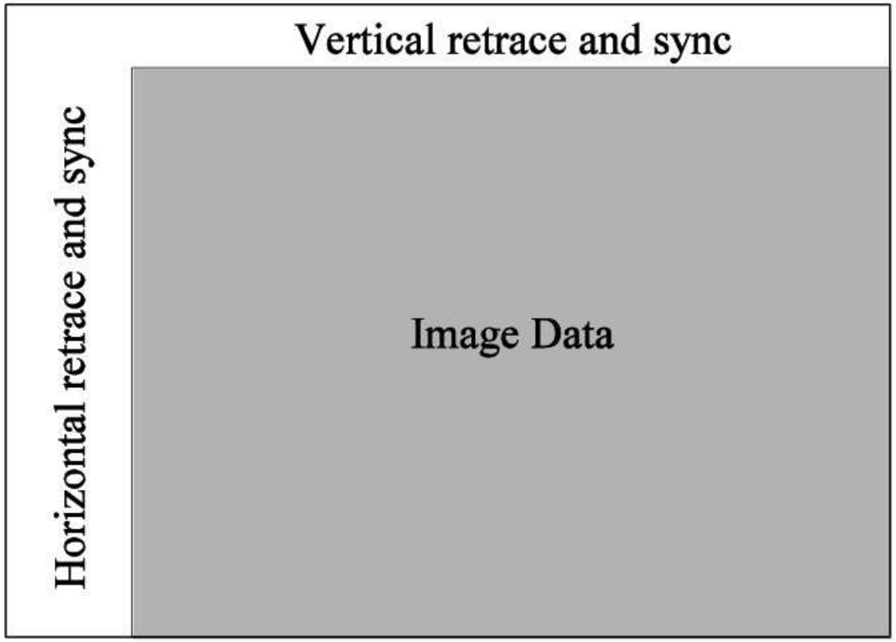


Fig. 3.3: Video raster, including retrace and sync data

- NTSC video is an analog signal with no fixed horizontal resolution. Therefore one must decide how many times to sample the signal for display: each sample corresponds to one pixel output.
- A “pixel clock” is used to divide each horizontal line of video into samples. The higher the frequency of the pixel clock, the more samples per line there are.
- Different video formats provide different numbers of samples per line, as listed in Table 3.1.

Table 3.1: Samples per line for various video formats

Format	Samples per line
VHS	240
S-VHS	400-425
Betamax	500
Standard 8 m	300
Hi-8 mm	425

PAL Video

PAL(Phase Alternating Line)

- Widely used in Western Europe, China, India, and many other parts of the world.
- Uses 625 scan lines per frame, at 25 fps Hasa4:3 aspect ratio and interlaced fields.
- YUV color model is used

SECAM Video

Uses625scan lines per frame, at 25 fps

With a 4:3 aspect ratio and interlaced fields.

SECAM and PAL are very similar. They differ slightly in their color coding scheme:

(a) In SECAM, U and V signals are modulated using separate color subcarriers at 4.25 MHz and 4.41 MHz respectively.

(b) They are sent in alternate lines, i.e., only one of the U or V signals will be sent on each scan line.

- Table 3.2 gives a comparison of the three major analog broadcast TV systems.

Table 3.2: Comparison of Analog Broadcast TV Systems

TV System	Frame Rate (fps)	# of Scan Lines	Total Channel Width (MHz)	Bandwidth Allocation (MHz)		
				Y	I or U	Q or V
NTSC	29.97	525	6.0	4.2	1.6	0.6
PAL	25	625	8.0	5.5	1.8	1.8
SECAM	25	625	8.0	6.0	2.0	2.0

Digital Video

- 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 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.

Chroma Subsampling

- Chroma subsampling is the practice of encoding images by implementing less resolution for Chroma information than for luma information, taking advantage of the human visual system's lower acuity for color differences than for luminance.
- It is used in many video encoding schemes– both analog and digital– and also in JPEG encoding.

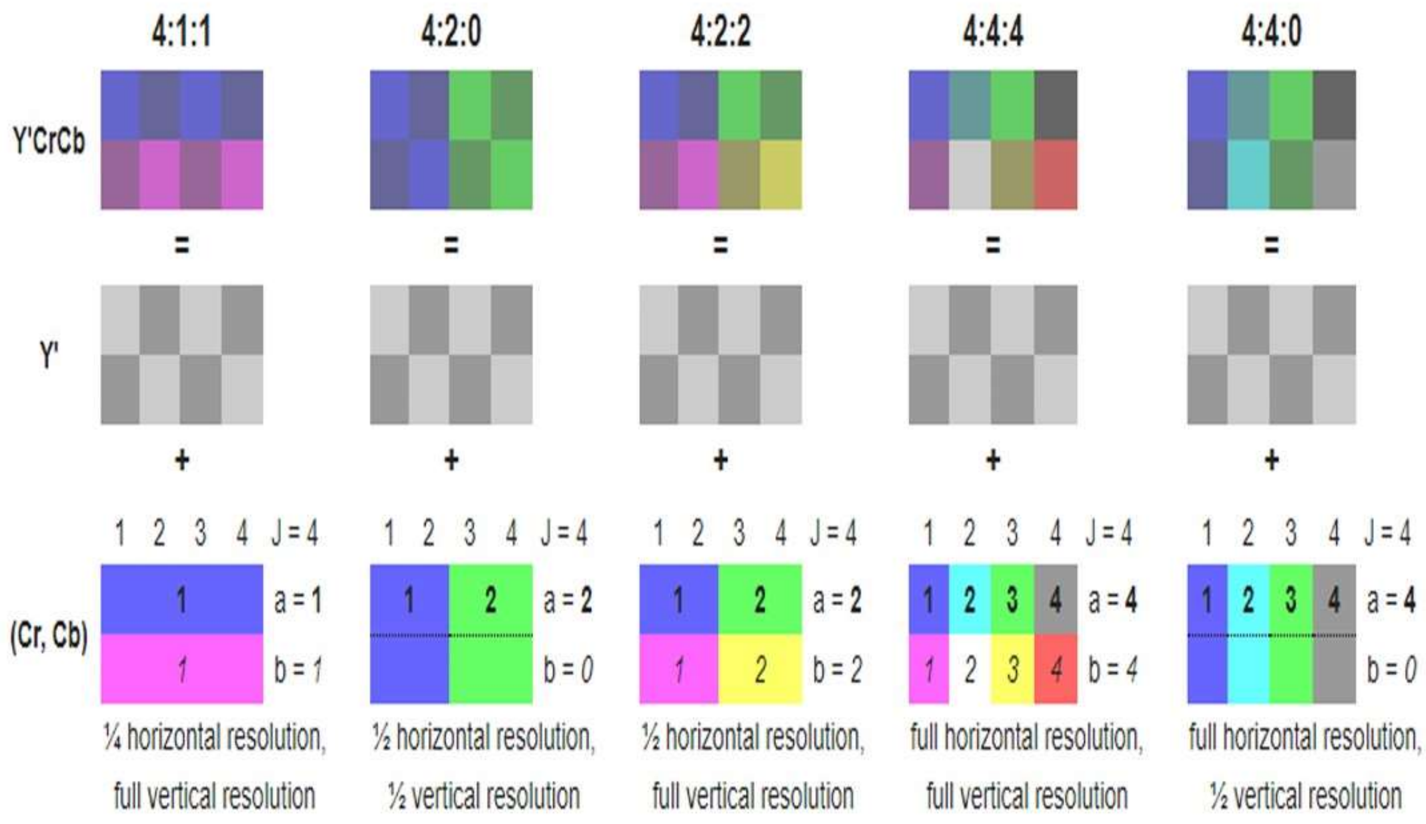
(a) The Chroma subsampling scheme “4:4:4” indicates that no Chroma subsampling is used: each pixel’s Y, Cb and Cr values are transmitted, 4 for each of Y, Cb, Cr.

(b) The scheme “4:2:2” indicates horizontal subsampling of the Cb, Cr signals by a factor of 2. That is, of four pixels horizontally labeled as 0 to 3, all four Ys are sent, and every two Cb’s and two Cr’s are sent, as (Cb0, Y0)(Cr0, Y1)(Cb2, Y2)(Cr2, Y3)(Cb4, Y4), and so on (or averaging is used).

(c) The scheme “4:1:1” subsamples horizontally by a factor of 4.

(d) The scheme “4:2:0” subsamples in both the horizontal and vertical dimensions by a factor of 2. Theoretically, an average Chroma pixel is positioned between the rows and columns as shown Fig.3.4.

- Sampling systems and ratios The subsampling scheme is commonly expressed as a three part ratio J:a:b (e.g. 4:2:2) or four parts if alpha channel is present (e.g. 4:2:2:4), that describe the number of luminance and chrominance samples in a conceptual region that is J pixels wide, and 2 pixels high. The parts are (in their respective order): J: horizontal sampling reference (width of the conceptual region). Usually, 4.
- a: number of chrominance samples (Cr, Cb) in the first row of J pixels.
- b: number of changes of chrominance samples (Cr, Cb) between first and second row of J pixels.
- Alpha: horizontal factor (relative to first digit). May be omitted if alpha component is not present, and is equal to J when present


4:1:1
4:2:0
4:2:2
4:4:4
4:4:0
Y'CrCb
=
Y'
+
1 2 3 4 J = 4
(Cr, Cb)
a = 1
b = 1
a = 2
b = 0
a = 2
b = 2
a = 4
b = 4
a = 4
b = 0
¼ horizontal resolution, full vertical resolution
½ horizontal resolution, ½ vertical resolution
½ horizontal resolution, full vertical resolution
full horizontal resolution, full vertical resolution
full horizontal resolution, ½ vertical resolution

CCIR Standards for Digital Video

Consultative Committee for International Radio produced CCIR-601, for component digital video.

- This standard has since become standard ITU-R-601, an international standard for professional video applications
- adopted by certain digital video formats including the popular DV video.

• Table 3.3 shows some of the digital video specifications, all with an aspect ratio of 4:3. The CCIR 601 standard uses an interlaced scan, so each field has only half as much vertical resolution (e.g., 240 lines in NTSC).

- CIF stands for Common Intermediate Format specified by the CCITT (International Telegraph and Telephone Consultative Committee).

(a) The idea of CIF is to specify a format for lower bitrate.

(b) CIF is about the same as VHS quality. It uses a progressive (non-interlaced) scan.

(c) Note, CIF is a compromise of NTSC and PAL in that it adopts the ‘NTSC frame rate and half of the number of active lines as in PAL

Table 3.3: Digital video specifications

	CCIR 601 525/60 NTSC	CCIR 601 625/50 PAL/SECAM	CIF	QCIF
Luminance resolution	720 x 480	720 x 576	352 x 288	176 x 144
Chrominance resolution	360 x 480	360 x 576	176 x 144	88 x 72
Color Subsampling	4:2:2	4:2:2	4:2:0	4:2:0
Fields/sec	60	50	30	30
Interlaced	Yes	Yes	No	No

HDTV (High Definition TV)

• The main thrust of HDTV (High Definition TV) is not to increase the "definition" in each unit area, but rather to increase the visual field especially in its width.

• HDTV is a television system providing an image resolution that that is of substantially higher resolution than that of standard-definition television.

(a) The first generation of HDTV was based on an analog technology

developed by Sony and NHK in Japan in the late 1970s.

(b) MUSE (Multiple sub-Nyquist Sampling Encoding) was an improved NHK HDTV with hybrid analog/digital technologies that was put in use in the 1990s. It has 1,125 scan lines, interlaced (60 fields per second), and 16:9 aspect ratio.

(c) Since uncompressed HDTV will easily demand more than 20 MHz bandwidth, which will not fit in the current 6 MHz or 8 MHz channels, various compression techniques are being investigated.

(d) It is also anticipated that high quality HDTV signals will be transmitted using more than one channel even after compression.

A brief history of HDTV evolution:

(a) In 1987, the FCC decided that HDTV standards must be compatible with the existing NTSC standard and be confined to the existing VHF (Very High Frequency) and UHF (Ultra High Frequency) bands.

(b) In 1990, the FCC announced a very different initiative, i.e., its preference for a full-resolution HDTV, and it was decided that HDTV would be simultaneously broadcast with the existing NTSC TV and eventually replace it.

(c) Witnessing a boom of proposals for digital HDTV, the FCC made a key decision to go all-digital in 1993. A "grand alliance" was formed that included four main proposals, by General Instruments, MIT, Zenith, and AT&T, and by Thomson, Philips, Sarnoff and others.

(d) This eventually led to the formation of the ATSC (Advanced Television Systems Committee) — responsible for the standard for TV broadcasting of HDTV.

(e) In 1995 the U.S. FCC Advisory Committee on Advanced Television Service recommended that the ATSC Digital Television Standard be adopted.

- The standard supports video scanning formats shown in Table 3.4. In the table, “I” mean interlaced scan and “P” means progressive (non-interlaced) scan.

Table 3.4: Advanced Digital TV formats supported by ATSC

# of Active Pixels per line	# of Active Lines	Aspect Ratio	Picture Rate
1,920	1,080	16:9	60I 30P 24P
1,280	720	16:9	60P 30P 24P
704	480	16:9 & 4:3	60I 60P 30P 24P
640	480	4:3	60I 60P 30P 24P

- For video, MPEG-2 is chosen as the compression standard. For audio, AC-3 is the standard. It supports the so-called 5.1 channel Dolby surround sound, i.e., five surround channels plus a subwoofer channel.
- The salient difference between conventional TV and HDTV:

(a)HDTV has a much wider aspect ratio of 16:9 instead of 4:3.

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

• The FCC has planned to replace all analog broadcast services with digital TV broadcasting by the year 2009. The services provided will include:

– SDTV (Standard Definition TV): the current NTSC TVor higher.

– EDTV(Enhanced Definition TV): 480 active lines or higher, i.e., the third and fourth rows in Table 3.4.

– HDTV (High Definition TV): 720 active lines or higher.

Digitization of Sound

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.

(a) For example, a speaker in an audio system vibrates back and forth and produces a longitudinal pressure wave that we perceive as sound.

(b) Since sound is a pressure wave, it takes on continuous values, as opposed to digitized ones.

(c) Even though such pressure waves are longitudinal, they still have ordinary wave properties and behaviors, such as reflection (bouncing), refraction (change of angle when entering a medium with a different density) and diffraction (bending around an obstacle).

(d) If we wish to use a digital version of sound waves we must form digitized representations of audio information.

Digitization

- Digitization means conversion to a stream of numbers, and preferably these numbers should be integers for efficiency.
- Fig. 3.5 shows the 1-dimensional nature of sound: amplitude values depend on a 1D variable, time. (And note that images depend instead on a 2D set of variables, x and y).

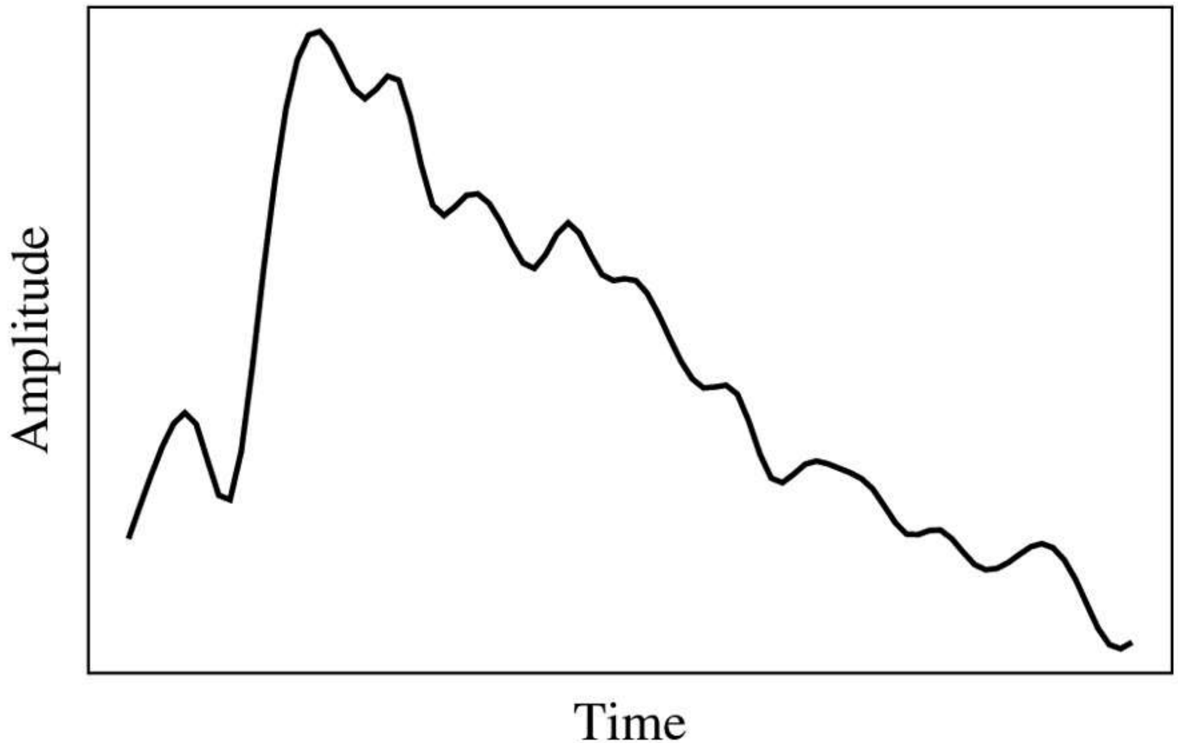


Fig. 3.5: An analog signal: continuous measurement of pressure wave.

• The graph in Fig. 3.5 has to be made digital in both time and amplitude. To digitize, the signal must be sampled in each dimension: in time, and in amplitude.

(a) Sampling means measuring the quantity we are interested in, usually at evenly-spaced intervals.

(b) The first kind of sampling, using measurements only at evenly spaced time intervals, is simply called, sampling. The rate at which it is performed is called the sampling frequency (see Fig. 3.6(a)).

(c) For audio, typical sampling rates are from 8 kHz (8,000 samples per second) to 48 kHz. This range is determined by the Nyquist theorem, discussed later.

(d) Sampling in the amplitude or voltage dimension is called quantization. Fig. 3.6(b) shows this kind of sampling.

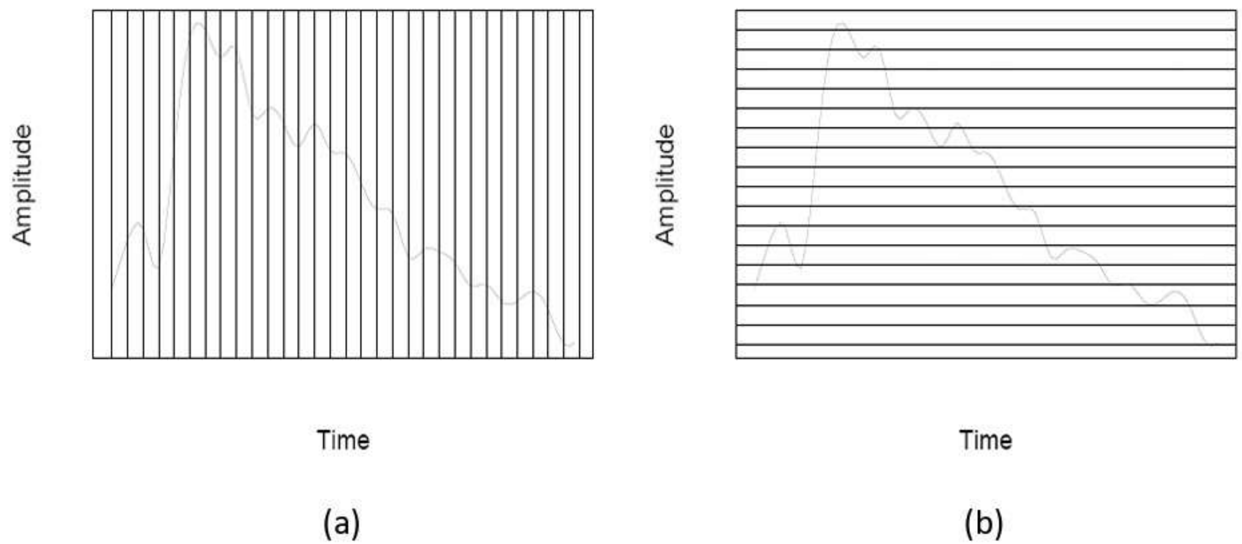


Fig. 3.6: Sampling and Quantization. (a): Sampling the analog signal in the time dimension. (b): Quantization is sampling the analog signal in the amplitude dimension.

• Thus to decide how to digitize audio data we need to answer the following questions: 1. What is the sampling rate?
2. How finely is the data to be quantized, and is quantization uniform?
3. How is audio data formatted? (file format)

• **Nyquist Theorem**: If a signal is band-limited, i.e., there is a lower limit f1 and an upper limit f2 of frequency components in the signal, then the sampling rate should be at least 2(f2 − f1).

• **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 an antialiasing filter that restricts the frequency content in the input to the sampler to a range at or below Nyquist frequency.

• The relationship among the Sampling Frequency, True Frequency, and the Alias Frequency is as follows: f alias = f sampling − f true, for f true < f sampling < 2 × f true

Signal to Noise Ratio (SNR)

- The ratio of the power of the correct signal and the noise is called the signal to noise ratio (SNR) — a measure of the quality of the signal.
- The SNR is usually measured in decibels (dB), where 1 dB is a tenth of a bel. The SNR value, in units of dB, is defined in terms of base-10 logarithms of squared voltages, as follows:

$$SNR = 10\log_{10}\frac{V_{signal}^2}{V_{noise}^2} = 20\log_{10}\frac{V_{signal}}{V_{noise}}$$

a) The power in a signal is proportional to the square of the voltage. For example, if the signal voltage V signal is 10 times the noise, then the SNR is $20 * \log10(10) = 20dB$.

b) In terms of power, if the power from ten violins is ten times that from one violin playing, then the ratio of power is 10dB, or 1B.

- The usual levels of sound we hear around us are described in terms of decibels, as a ratio to the quietest sound we are capable of hearing. Table 3.5 shows approximate levels for these sounds.

Table 3.5: Magnitude levels of common sounds, in decibels

Threshold of hearing	0
Rustle of leaves	10
Very quiet room	20
Average room	40
Conversation	60
Busy street	70
Loud radio	80
Train through station	90
Riveter	100
Threshold of discomfort	120
Threshold of pain	140
Damage to ear drum	160

Signal to Quantization Noise Ratio (SQNR)

- 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).

(a) Quantization noise: the difference between the actual value of the analog signal, for the particular sampling time, and the nearest quantization interval value.

(b) At most, this error can be as much as half of the interval.

(c) For a quantization accuracy of N bits per sample, the SQNR can be simply expressed:

$$SQNR = 20\log_{10}\frac{V_{signal}}{V_{quan\_noise}} = 20\log_{10}\frac{2^{N-1}}{\frac{1}{2}}$$

$$= 20 \times N \times \log 2 = 6.02\, N(\mathrm{dB})$$

- Notes:

(a) We map the maximum signal to $2^{N-1} - 1$ ($\simeq 2^{N-1}$) and the most negative signal to -2^{N-1}.

(b) The above equation is the *Peak* signal-to-noise ratio, PSQNR: peak signal and peak noise.

Audio Filtering

- Prior to sampling and AD conversion, the audio signal is also usually filtered to remove unwanted frequencies. The frequencies kept depend on the application:

(a) 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.

(b) An audio music signal will typically contain from about 20Hz up to 20kHz.

(c) 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.

(d) So at the decoder side, a low pass filter is used after the DAcircuit.

Audio Quality vs. Data Rate

- The uncompressed data rate increases as more bits are used for quantization.

Stereo: double the bandwidth to transmit a digital audio

signal. Table3.6:Data rate and bandwidth in sample audio applications.

Quality	Sample Rate (Khz)	Bits per Sample	Mono / Stereo	Data Rate (uncompressed) (kB/sec)	Frequency Band (KHz)
Telephone	8	8	Mono	8	0.200-3.4
AM Radio	11.025	8	Mono	11.0	0.1-5.5
FM Radio	22.05	16	Stereo	88.2	0.02-11
CD	44.1	16	Stereo	176.4	0.005-20
DAT	48	16	Stereo	192.0	0.005-20
DVD Audio	192 (max)	24(max)	6 channels	1,200 (max)	0-96 (max)

Synthetic Sounds

Digitized sound must still be converted to analog, for us to hear it. There are two fundamentally different approaches to handling stored sampled audio.

1.FM (Frequency Modulation): one approach to generating synthetic sound:

2. Wave Table synthesis: A more accurate way of generating sounds from digital signals. Also known, simply, as sampling.

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.

MIDI: Musical Instrument Digital Interface

- A protocol that enables computer, synthesizers, keyboards, and other musical device to communicate with each other.
- This protocol is a language that allows interworking between instruments from different manufacturers by providing a link that is capable of transmitting and receiving digital data.
- Transmits only commands, it does not transmit an audio signal.
- It was created in 1982.

Components of a MIDI System

1. **Synthesizer:**

- It is a sound generator (various pitch, loudness, tone color).
- A good (musician's) synthesizer often has a microprocessor, keyboard, control panels, memory, etc.

2. **Sequencer:**

- It can be a stand-alone unit or a software program for a personal computer. (It used to be a storage server for MIDI data).
- Now a days it is more a software music editor on the computer.)
- It has one or more MIDIINs and MIDIOUTs

Basic MIDI Concepts

Track:

Track in sequencer is used to organize the recordings.

Tracks can be turned on or off on recording or playing back.

Channel:

Channels are used to separate information in a MIDI system. There are 16 MIDI channels in one cable.

Channel numbers are coded into each MIDI message.

Timbre:

The quality of the sound, e.g., flute sound, cello sound, etc. Multi timbral– capable of playing many different sounds at the same time (e.g., piano, brass, drums, etc.)

Pitch:

The musical note that the instrument plays

Voice:

Voice is the portion of the synthesizer that produces sound.

Synthesizers can have many (12, 20, 24, 36, etc.) voices.

Each voice works independently and simultaneously to produce sounds of different timbre and pitch.

Patch:

The control settings that define a particular timbre.

MIDI to WAV Conversion

- Some programs, such as early versions of Premiere, cannot include .mid files — instead, they insist on .wav format files.

a) Various share ware programs exist for approximating a reasonable conversion between MIDI and WAV formats.

b) These programs essentially consist of large lookup files that try to substitute pre-defined or shifted WAV output for MIDI messages, with inconsistent success.

Quantization and Transmission of Audio

• Coding of Audio: Quantization and transformation of data are collectively known as coding of the data.

a) 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.

b) 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 (differences, now) into a much smaller range.

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

• In general, producing quantized sampled output for audio is called PCM (Pulse Code Modulation). The differences version is called DPCM (and a crude but efficient variant is called DM). The adaptive version is called ADPCM.

Pulse Code Modulation

- The basic techniques for creating digital signals from analog signals are sampling and quantization.
- Quantization consists of selecting breakpoints in magnitude, and then re-mapping any value within an interval to one of the representative output levels.

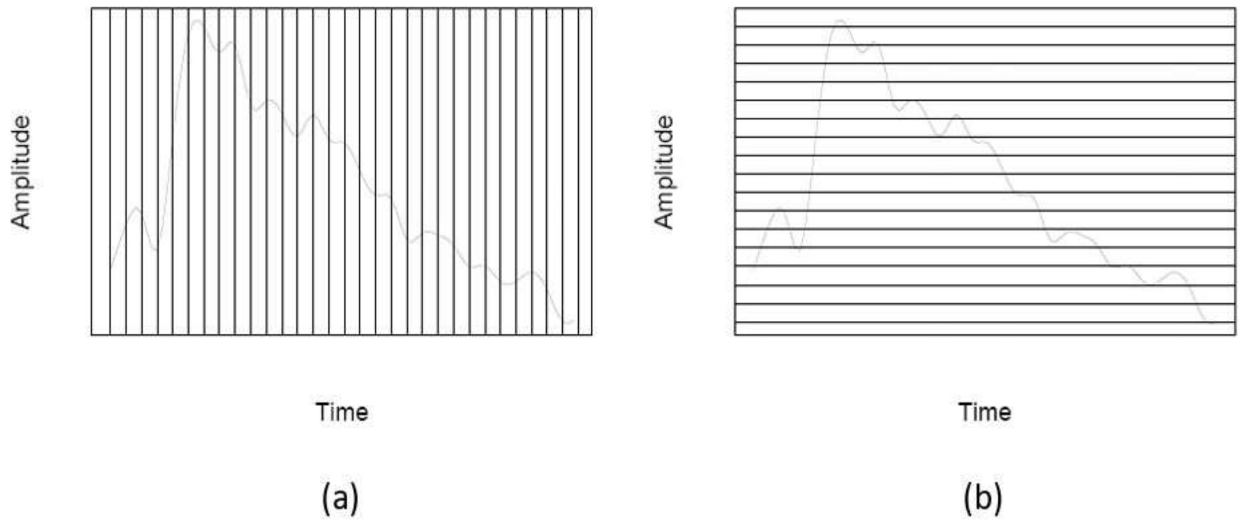


Fig. 3.2: Sampling and Quantization.

a)The set of interval boundaries are called decision boundaries, and the representative values are called reconstruction levels.

b) The boundaries for quantizer input intervals that will all be mapped into the same output level form a coder mapping.

c) The representative values that are the output values from a quantizer are a decoder mapping.

d) Finally, we may wish to compress the data, by assigning a bit stream that uses fewer bits for the most prevalent signal values.

- Every compression scheme has three stages:

A. The input data is transformed to a new representation that is easier or more efficient to compress.

B. We may introduce loss of information. Quantization is the main lossy step ⇒ we use a limited number of reconstruction levels, fewer than in the original signal.

C. Coding. Assign a code word (thus forming a binary bit stream) to each output level or symbol. This could be a fixed-length code, or a variable length code such as Huffman coding.

PCM in Speech Compression

- Assuming a bandwidth for speech from about 50 Hz to about 10 kHz, the Nyquist rate would dictate a sampling rate of 20 kHz.

a) 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 bit-rate would be 240 kbps.

(b) With companding, we can reduce the sample size down to about 8 bits with the same perceived level of quality, and thus reduce the bit-rate to 160 kbps.

(c) However, the standard approach to telephony in fact assumes that the highest-frequency audio signal we want to reproduce is only about 4 kHz. Therefore the sampling rate is only 8 kHz, and the companded bit rate thus reduces this to 64 kbps.

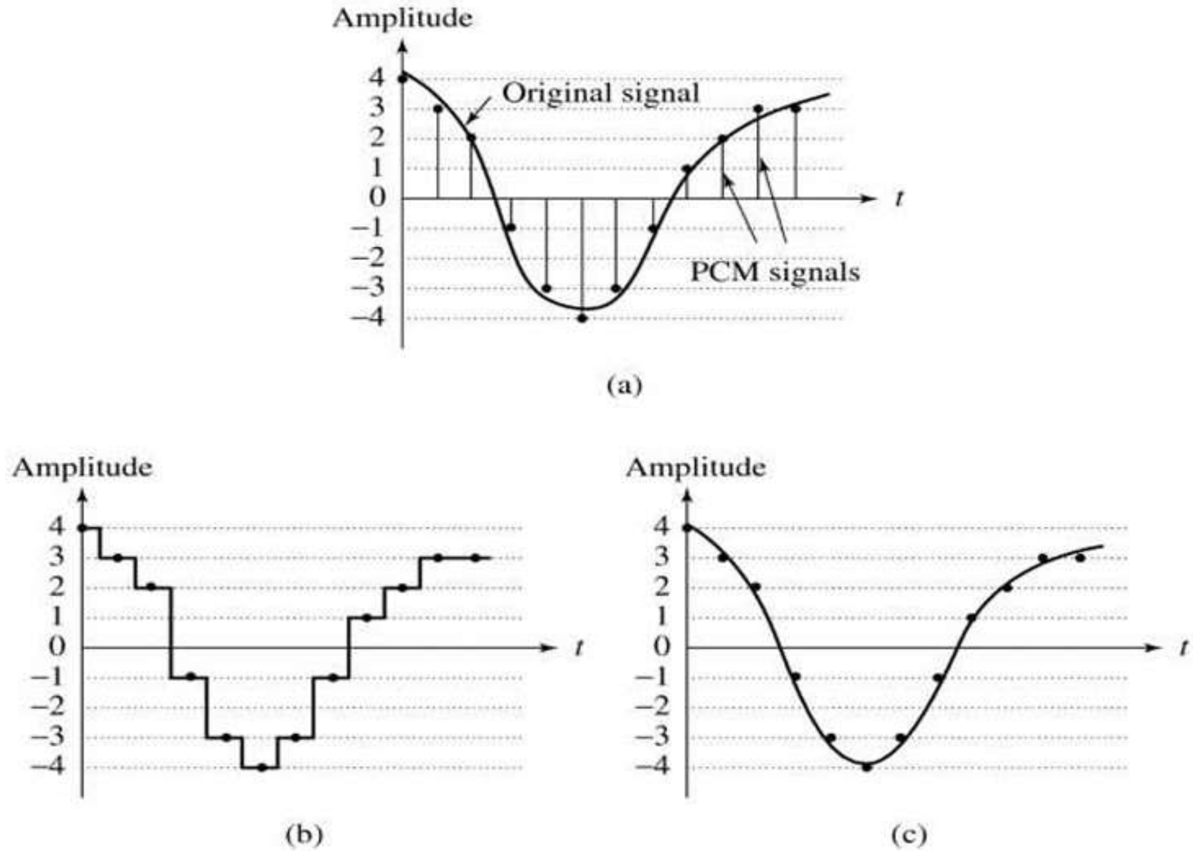


Fig. 3.7: Pulse Code Modulation (PCM). (a) Original analog signal and its corresponding PCM signals. (b) Decoded staircase signal. (c) Reconstructed signal after low-pass filtering.

• The complete scheme for encoding and decoding telephony signals is shown as a schematic in Fig. 3.8. As a result of the low-pass filtering, the output becomes smoothed and Fig. 3.7(c) above showed this effect.

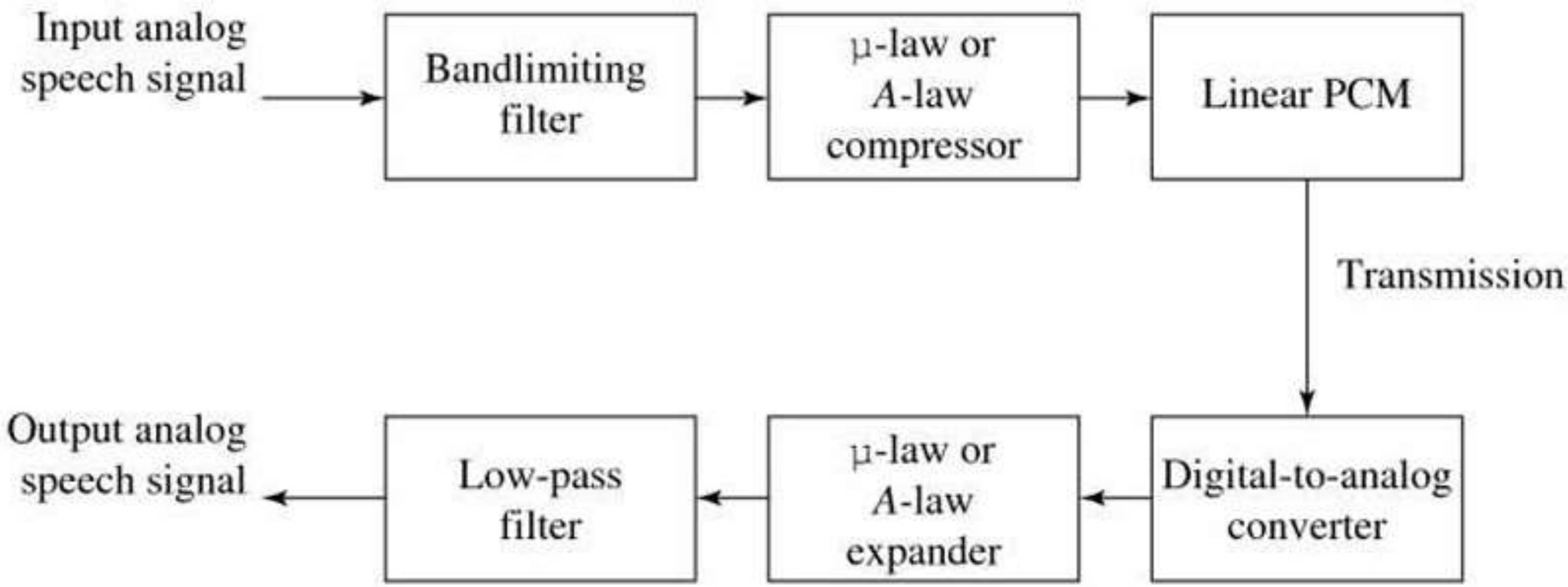


Fig. 3.8: PCM signal encoding and decoding.