CS 6570: Multimedia Computing System

Dr. *Fenn-Huei Simon Sheu*

Department of Computer Science
National Tsing Hua University
Tel: (03) 574-2959
Office: Room 638, EE/CS Blg.
Email: sheu@cs.nthu.edu.tw
URL: http://www.cs.nthu.edu.tw/~sheu/

Outline

- System Support for Multimedia Data
- Compression Techniques
- Video Delivery and Reliability
- Multimedia Database Technologies
- Multimedia Applications

Fundamentals of MM Systems

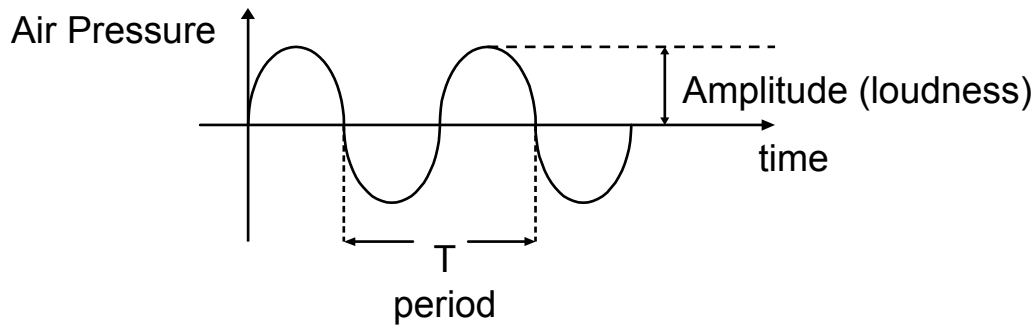
- What
- Why
- Basic Natures
- Synchronization
- Desirable Features
- Components
- Operation Models
- Related Researches
  - Database
  - Presentation
  - Conferencing
  - WWW, Internet
  - MBone
  - various fields
- Challenges

What is & Why Multimedia?

- Multi + media
  - Simulating sound, touch (feel), scene (visual), taste, smell.
  - Combination of 2+ media: audio (voice, music), image (photo, graphics), video (animation), data and the like (slide).
  - Complete coverage of human I/O's.
- Basic Human Traits
  - perception
  - retention
  - reasoning
  - presentation
- Application:
  - Replacement
    - manufacturing sys.
  - Facilitation
    - teleconferencing

Basic Natures

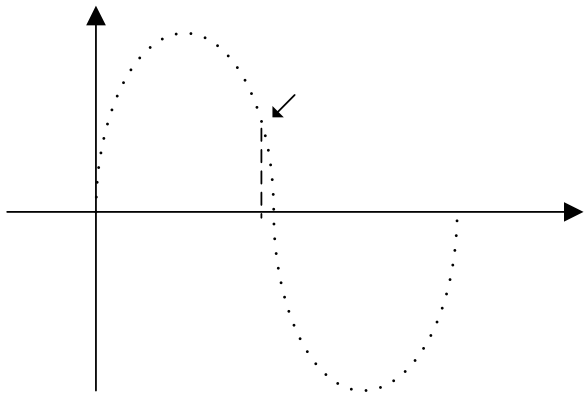
- Each media type has its own property and can be stored using respective bit (0 & 1) sequences. [encoding]
- Audio:

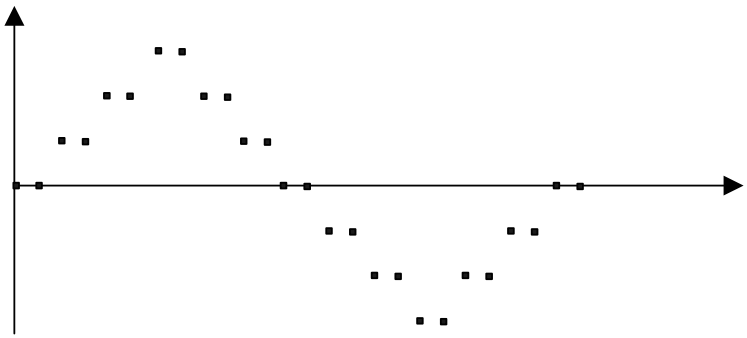


  - The ***frequency*** of a sound is the reciprocal value of the period.
  - The frequency range is divided into:
    - Human hearing frequency: 20 Hz to 20 KHz
    - Infra-sound: 0 to 20 Hz
    - Ultrasound: 20 KHz to 1 GHz
    - Hypersound: 1 GHz to 10 THz
  - We call sound within the human hearing range ***audio***, and the waves in this frequency band ***acoustic signals***.
  - Besides speech and music, we denote any other audio signal as ***noise***.

Digitized Audio

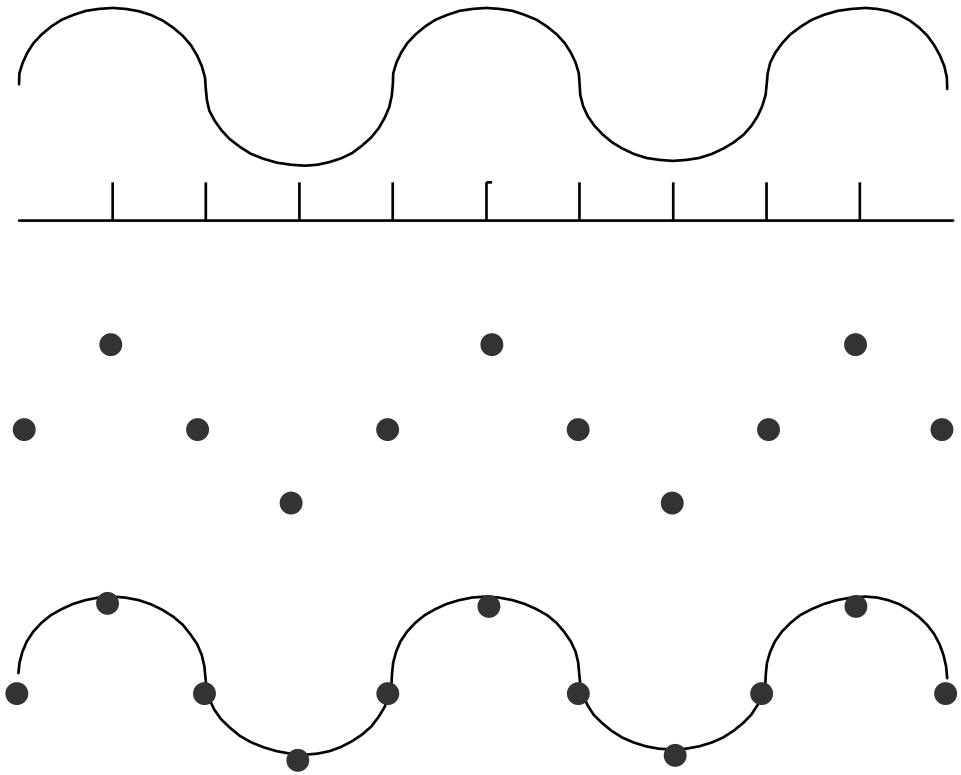
- Computer Representation of Sound:
  - Measure the wave amplitude at regular time intervals to yield a series of numbers, each called a ***sample***.





  - Sampling rate - the rate at which a continuous waveform is sampled.
    - Example: *The CD standard sampling rate of 44,100 Hz means that the waveform is sampled 44,100 times per second.*
    - In telecommunication networks, the analogue speech signal is band limited to 300 ~ 3400 Hz, and sampled at 8 kHz.
  - Quantization - the process of representing a large, possibly infinite, set of values with a much smaller set (***codewords***.)
    - Example: *An 8-bit quantization yields 256 possible values.*

Sound Hardware

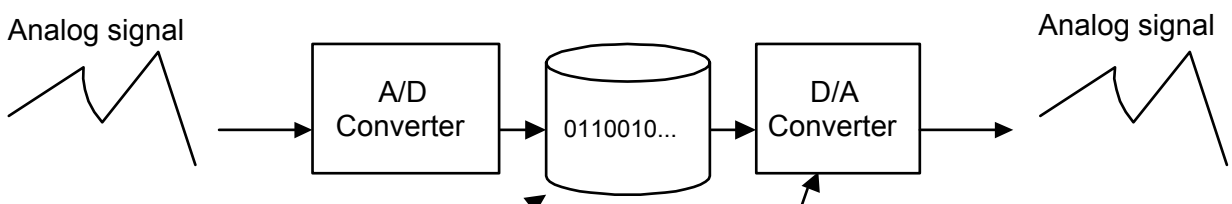


original waveform

Sampling frequency
Nyquist limit

Samples

Reconstructed waveform



All multimedia information is internally represented in digital format

Since humans only react to physical sensory stimuli, a digital-to-analog conversion necessarily takes place in any presentation of multimedia information.

The Physics of Color

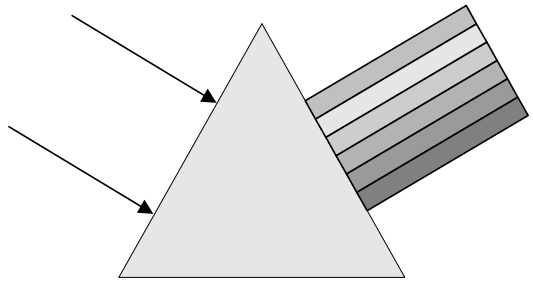
- The **radiant energy spectrum** contains audio frequencies, radio frequencies, infrared, visible light, ultraviolet rays, x-rays, and gamma rays.
- Radiant energy is measured in terms of **frequency** or **wavelength**.

$$frequency = \frac{velocity\_of\_light}{wavelength}$$

- The human eye responds to visible light wavelengths between 380 and 760 nanometers.

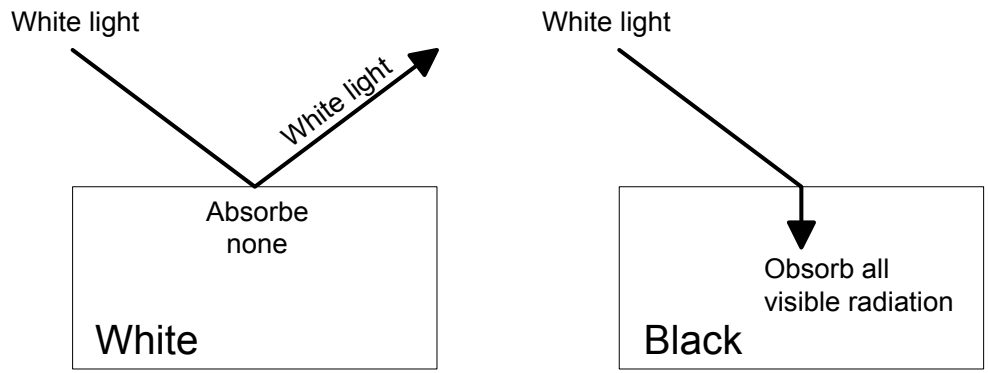
Color	Wavelength (Nanometer)	Frequency (*Hertz*)
Violet (Purple)	380 - 450	$6.6\times10^{14} - 7.9\times10^{14}$
Blue	450 - 490	$6.1\times10^{14} - 6.6\times10^{14}$
Green	490 - 560	$5.4\times10^{14} - 6.1\times10^{14}$
Yellow	560 - 590	$5.1\times10^{14} - 5.4\times10^{14}$
Orange	590 - 630	$4.8\times10^{14} - 5.1\times10^{14}$
Red	630 - 760	$3.9\times10^{14} - 4.8\times10^{14}$

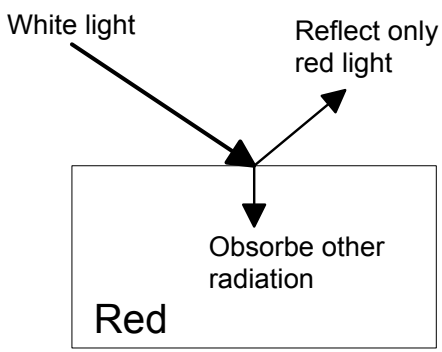
Color Physics (cont'd)



- White light consists of energy throughout the visible light spectrum.

- The color of an object depends on both the **reflectivity** of the surface of the object and the composition of the **illuminating light**.





Color Coding

- RGB Model:
  - Different intensities of **red**, **green**, and **blue** are added to generate various colors.
  - However, the eye is more sensitive to green, and least sensitive to blue.

- YUV Representation:
  - The ***luminance*** component (Y, brightness) contains the gray-scale information.
  - The ***chrominance*** component defines the **color** (U, hue) and the **intensity** (V, saturation) of the color.

    Advantage: The human eye is more susceptible to brightness than color.

    ⇒ *A compression scheme can use gray-scale information to define detail and allows loss of color information to achieve higher rates of compression (i.e., JPEG).*

    ⇒ *There is no reason to accurately reproduce details that the eye can't see.*

Images

- Images, often called *pictures*, are represented by ***bitmaps***.
- A bitmap is a spatial 2-D matrix made up of individual picture elements called ***pixels***.
  - Each pixel has a numerical value called ***amplitude***.
  - The number of bits available to code a pixel is called ***amplitude depth*** or ***pixel depth***.
  - A pixel depth may represent
    - a black or white dot in bi-tonal images
    - a level of gray in continuous-tone, monochromatic images, or
    - the color attributes of the picture element in colored pictures.
- TIFF or GIF are examples of standards for the coding of images. They are called *bitmap formats*.

Graphics

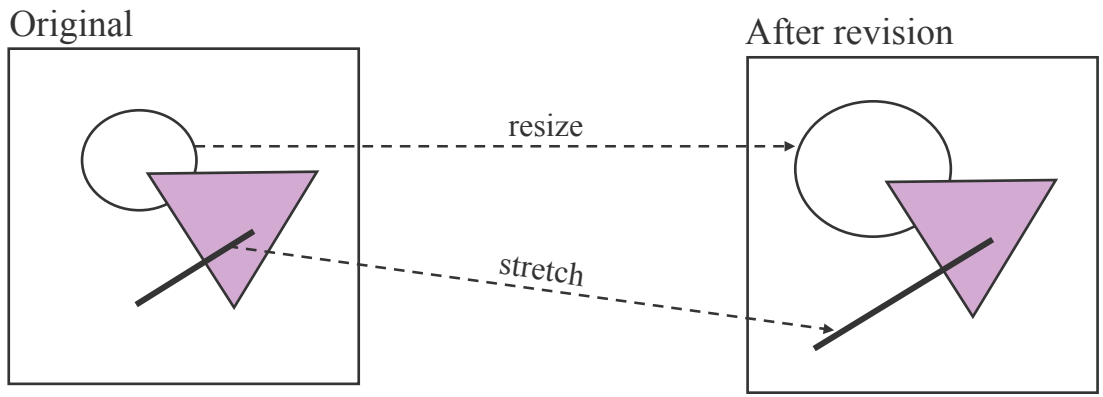
- Graphics image formats are specified through ***graphics primitives*** and their ***attributes***.
  - graphics primitives: *lines, curves, circles*
  - attributes: *thickness, gray-scale, color.*

- The semantic content of graphics is preserved in the representation.
  - Example: *A black line can be efficiently represented by a pair of spatial coordinates (a vector).*

- PHIGS, GKS, IGS are examples of graphics format standards.

Graphics vs. Images

- Graphics are **revisable** because their format retains structural information in the form of objects.



- Images are **not revisable** because their format contains no structural information.
  - Example: *If a graphic which comprises a black line is stored as a bitmap, the resulting image will not indicate that the succession of black pixels which compose the black line forms a vector.*
  - Note: *Graphics or text, once created in revisable format, may be represented and stored as images, that is, they may be converted to bitmap format.*

Video and Animation

- Both images and graphics may be displayed on a computer screen as a succession of views which create an impression of movement.

- In that case, they will be referred to as ***video*** (or motion pictures) and computer ***animation*** (or motion graphics), respectively.

- Note: *Frames of a video can be displayed directly while display of a computer animation requires real-time interpretation of the graphics frames.*

Frame Rate

- A ***frame*** is an image (or graphic) in a video (or computer animation).
- Each frame is a variant of the previous one in the video (or computer animation).
- The number of frames displayed per second is called the ***frame rate***.
  - *Between 10 and 16 fps, the viewer has an impression of motion but still feel a jerky effect.*
  - *It is above 15 or 16 fps that a smooth motion effect begins.*
- Current American TV standards use 30 fps, while European standards use 25 fps. One of the several HDTV standards operates at 60 fps.

Two classes of Media

- Discrete media: Time is not part of the semantics of the media.
  - They may be displayed according to a wide variety of timing of even sequencing, and remain meaningful.

    Examples: *text, graphics and images*.

- Continuous media: Time or more exactly time-dependency between information items, is part of the information itself.
  - If the timing is changed, or the sequencing of the items modified, the meaning is altered.

    Examples: *sound or motion video*.

Data Stream

- Transmitted information is divided into individual units (packets).
- Transmission of a sequence of individual packets forms a ***data stream***.

Examples:

+ *transmission of speech in a telephone system.*
+ *retrieval of a document from a database server.*

Characteristics of Data Streams

- **Asynchronous Transmission Mode**: provides for communication with no timely restrictions.
  - Example: Internet protocol for email transmission.

- **Synchronous Transmission Mode**: defines a maximum end-to-end delay for each packet of a data stream.
  - Needs a large temporary storage for packets arriving too early.

- **Isochronous Transmission Mode**: defines, besides a maximum end-to-end delay for each packet of a data stream, a minimum end-to-end delay.
  - The necessary storage of video data at the receiver can be significantly reduced.

Synchronization

- Temporal & Spatial Relationships
  - MM information has inter-media synchronization as an inherent property.
  - 5-minute speech:
    two video's, one showing the lecture hall and the other displaying the slides, one audio channel for speech, where the slides' images are synced with audio and video.
  - MPEG-4 is formulating such novel features.

```
group2D {
   children [
      shape {
         geometry rectangle {
            size 100.0 100.0
         }
         appearance appearance {
            material material2d {
               emissiveColor 1 0 0
               filled true
            }
         }
      }
   }
   transform2d { #tutu
      translation 20 20
      children [
         def xxx shape {
            geometry rectangle { #toto
               size 50.0 50.0
            }
...
```

Related Researches

- Database
  - * Efficiency of creation, maintenance, and retrieval of MMDBMS.
- Presentation
  - * The current trend of MM researches is to fulfil more functionality in addition to high compression rate.
  - * Allowing user interactions.
- Conferencing
  - * Proactive buffering and network management techniques are necessary to accommodate more participants in a session.
- WWW, Internet
  - * URL(Uniform Resource Locator), HTML, and beyond.
- MBone
  - * Multicast Backbone: sd (Session Directory), vat (Visual Audio Tool), wb (White Board), nv (Network Video), etc.

Limitation & Challenge

- Limitation:
  - Your imagination !!
  - The multimedia researches are relatively new areas for creative explorations.
- Challenge:
  - Various perspectives:
    - Communication People:
      - QoS (Quality of Service): data rate (mean and variation), delay characteristics, error or loss probability (bit error rate, cell or packet loss probability, error patterns.)
      - Layered Encoding (progressive) Schemes: Wavelet Encoding.
  - Paradigm shift to "?"
    - Contrivance of MIDI.
- Native Techniques:
  - Read Papers, Discussion (BBS), Ponder the Nature.