CS323 Operating Systems Multimedia

Yuanyuan Zhou
Lecture 29
4/9/2003

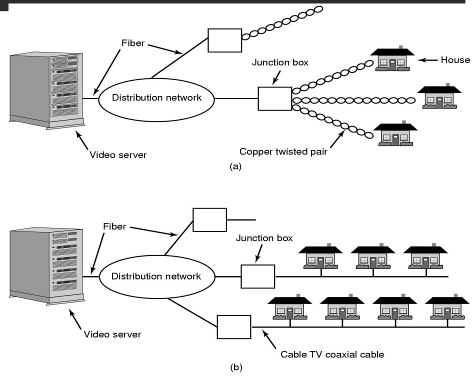
Content of this lecture

- Introduction on Multimedia
- Audio encoding
- Video encoding
- Compression

Introduction to Multimedia

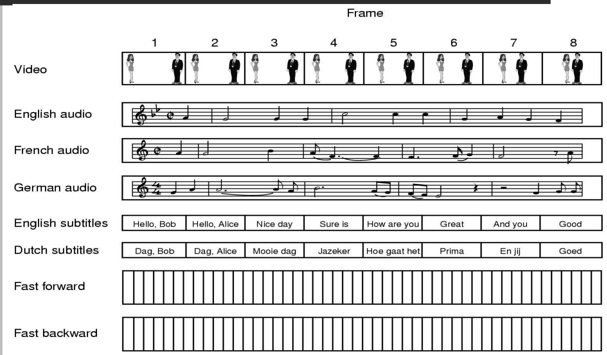
- Video
  - Hostage: BMW shortfilm
- Audio
  - Song from Chicago

Video on Demand



Video On Demand: (a) ADSL vs. (b) cable

Multimedia Files



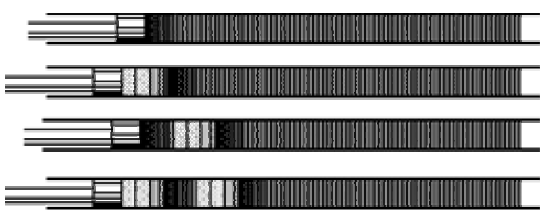
A movie may consist of several files

Multimedia Issues

- Analog to digital
- Problem: need to be acceptable by ears or eyes
  - Jitter
- Require high data rate
  - Large storage
  - Compression
- Require real-time playback
  - Scheduling
  - Quality of service
  - Resource reservation

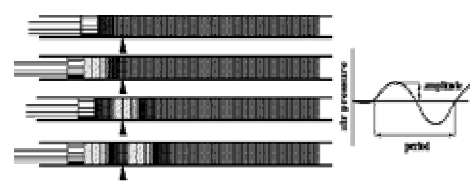
Audio

- Sound is a continuous wave that travels through the air.
- The wave is made up of pressure differences.



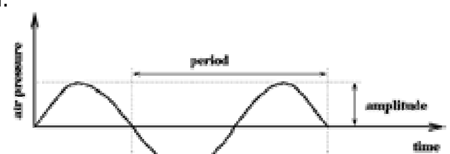
How do we hear sound?

- Sound is detected by measuring the pressure level at a point
- When an acoustic signal reaches the otter- ear (Pinna), the generated wave will be transformed into energy and filtered through the middle-ear. The inner-ear (Cochlea) transforms the energy into nerve activity.
- In similar way, when an acoustic wave strikes a microphone, the microphone generates an electrical signal, representing the sound amplitude as a function of time.



Basic Sound Concepts

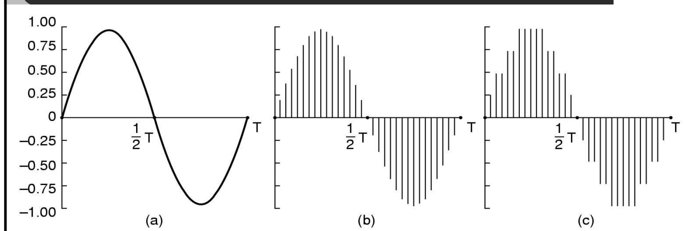
- Frequency represents the number of periods in a second (measured in hertz, cycles/second)
- Human hearing frequency range: 20 Hz - 20 kHz (audio), voice is about 500 Hz to 2 kHz.
- Amplitude of a sound is the measure of displacement of the air pressure wave from its mean.



Computer Representation of Audio

- Speech is analog in nature and it is converted to digital form by an analog-to-digital converter (ADC).
- A transducer converts pressure to voltage levels.
- Convert analog signal into a digital stream by discrete sampling
- Discretization both in time and amplitude (quantization)

Audio Encoding (1)



Audio Waves Converted to Digital

- electrical voltage input
- sample voltage levels at intervals to get a vector of values: (0, 0.2, 0.5, 1.1, 1.5, 2.3, 2.5, 3.1, 3.0, 2.4,...)
- A computer measures the amplitude of the waveform at regular time intervals to produce a series of numbers (samples).
- The ADC process is governed by four factors: *sampling rate, quantization, linearity, and conversion speed*. binary number as output

Audio Encoding (2)

- Sampling Rate: rate at which a continuous wave is sampled (measured in Hertz)
- Examples: CD standard - 44100 Hz, Telephone quality - 8000 Hz
- The audio industry uses 5.0125 kHz, 11.025 kHz, 22.05 kHz, and 44.1 kHz as the standard sampling frequencies. These frequencies are supported by most sound cards.
- Question: How often do you need to sample a signal to avoid losing information?

Audio Encoding (3)

- Answer: It depends on how fast the signal is changing,. Real Answer: twice per cycle (this follows from Nyquist sampling theorem)
- **Nyquist Sampling Theorem**: If a signal f(t) is sampled at regular intervals of time and at a rate higher than twice the highest significant signal frequency, then the samples contain all the information of the original signal.
- Example: CD's actual sampling frequency is 22050 Hz, but because of Nyquist's Theorem, we need to sample the signal twice, therefore the sampling frequency is 44100Hz.

Audio Encoding (4)

- The best-known technique for voice digitization is **Pulse-Code Modulation (PCM)**.
- PCM is based on the sampling theorem.
- If voice data are limited to 4000 Hz, then PCM samples 8000 samples/second which is sufficient for the input voice signal.
- PCM provides analog samples which must be converted to digital representation. Each of these analog samples must be assigned a binary code. Each sample is approximated by being **quantized** as explained above.

Audio Encoding (5)

- *Quantization (sample precision)*: the resolution of a sample value.
  - Samples are typically stored as raw numbers (linear PCM format) or as logarithms (u-law or A-law)
  - Quantization depends on the number of bits used measuring the height of the waveform
  - Example: 16-bit CD quality quantization results in over 65536 values
- Audio Formats are described by the **sample rate** and **quantization**
  - Voice quality: 8-bit quantization, 8000 Hz u-law mono (8kBytes/s)
  - 22 kHz 8-bit linear mono (22 kBytes/second) and stereo (44 kBytes/s)
  - CD quality 16-bit quantization, 44100 Hz linear stereo (176.4 kBytes/s = 44100 samples x 16 bits/sample x 2 (two channels)/8000)

Audio Formats

- Available formats on SUN
  - **au** - Sun File Format
  - **wav** - Microsoft RIFF/waveform Format
  - **al** - Raw A-law Data Format
  - **u** - Raw u-law Data Format
  - **snd** - NeXT File Format
- Available formats on Microsoft-Windows-based systems (RIFF formats):
  - Waveform audio file format for digital audio hardware
  - MIDI file format for standard MIDI files
  - Audio Video Interleaved (AVI) Indeo file format
- RIFF (Resource Interchange File Format) forms the basis of a number of file formats. RIFF (similarly to TIFF - Tagged Image File Format) is a tagged file format. Tags allow applications capable of reading RIFF files to read RIFF files by another application, hence the word interchange in RIFF.
- Other Formats/Players - RealPlayer 7 (Windows NT) with RealAudio, MP3 (MPEG Audio Layer 3) audio, Midi players; MP3 players (MP3.com)

Video Encoding

Scanning Pattern for NTSC Video and Television

Color Encoding

- During the scanning, a camera creates three signals: RGB (red, greed and blue) signals.
- For compatibility with black-and-white video and because of the fact that the three color signals are highly correlated, a new set of signals of different space are generated.
- The new color systems correspond to the standards such as NTCS, PAL, SECAM.
- For transmission of the visual signal we use three signals: 1 luminance (brightness- basic signal) and 2 chrominance (color signals).
- In NTSC signal the luminance and chrominance signals are interleaved;
- The goal at the receiver is : (1) separate luminance from chrominance components, and (2) avoid interference between them (cross-color, cross luminance)

Basic Concepts of Image Formats

- Important Parameters for **Captured Image Formats**:
  - Spatial Resolution (pixels x pixels)
  - Color encoding (quantization level of a pixel: e.g., 8-bit, 24-bit)
  - Examples: `SunVideo' Video Digitizer Board allows pictures of 320 by 240 pixels with 8-bit gray-scale or color resolution.
  - For a precise demonstration of image basic concepts try the program **xv** which displays images and allows to show, edit and manipulate the image characteristics.
- Important Parameters for **Stored Image Formats**:
  - Images are stored as a 2D array of values where each value represents the data associated with a pixel in the image (bitmap or a color image).
  - The stored images can use flexible formats such as the RIFF (Resource Interchange File Format). RIFF includes formats such as bitmats, vector-representations, animations, audio and video.
  - Currently, most used image storage formats are GIF (Graphics Interchange Format), XBM (X11 Bitmap), Postscript, JPEG (see compression chapter), TIFF (Tagged Image File Format), PBM (Portable Bitmap), BMP (Bitmap).

Digital Transmission Bandwidth

- **Bandwidth requirements for Images**:
  - Raw Image Transmission Bandwidth:= size of the image:= spatial resolution x pixel resolution;
  - Compressed Image Transmission Bandwidth: = depends on the compression scheme(e.g., JPEG) and content of the image;
  - Symbolic Image Transmission bandwidth:= size of the instructions and variables carrying graphics primitives and attributes.
- **Bandwidth Requirements for Video**:
  - Uncompressed Video Bandwidth:= image size x frame rate;
  - Compressed Video Bandwidth:= depends on the compression scheme(e.g., Motion JPEG, MPEG) and content of the video (scene changes).
- **Example:** Assume the following video characteristics - 720,000 pixels per image(frame), 8 bits per pixel quantization, and 60 frames per second frame rate. The *Video Bandwidth*:= 720,000 pixels per frame x 8 bits per pixel x 60 fps
- which results in HDTV data rate of 43,200,000 bytes per second = 345.6 Mbps When we use MPEG compression, the bandwidth goes to 34 Mbps wit some loss in image/video quality.

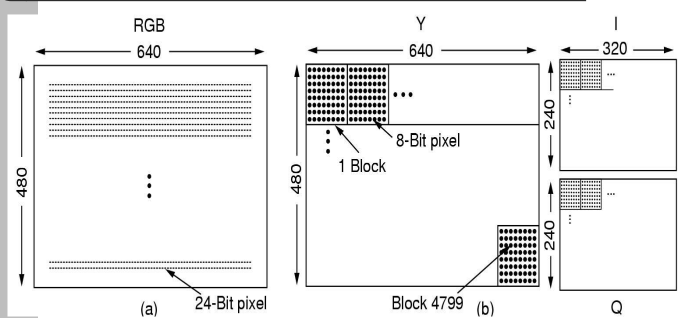
Video Compression

- Compression is important due to limited bandwidth
- All compression systems require two algorithms:
  - Encoding at the source
  - Decoding at the destination
- Two types of compression algorithms: Lossless or Lossy
- Simple lossless compression algorithm is the **Run-length Coding**, where multiple occurring bytes are grouped together as Number-OccuranceSpecial-CharacterCompressed-Byte. For example, 'AAAAAABBBBBDDDDDAAAAAAAA' can be encoded as `6!A5!B5!D8!A', where `!' is the special character. The compression ratio is 50% (12/24*100%).
- Best known lossy image compression algorithm is JPEG; best known lossy video compression algorithm is MPEG (MPEG-1, MPEG-2, MPEG-4, MPEG-7)

JPEG: Joint Photographic Experts Group

- 6 major steps to compress an image: (1) block preparation, (2) DCT (Discrete Cosine Transform) transformation, (3) quantization, (4) further compression via differential compression, (5) zig-zag scanning and run-length coding compression, (6) Huffman coding compression
- Quantization step represents the lossy step where we loose data in a non-invertible fashion.
- Differential compression means that we consider similar blocks in the image and encode only the first block and for the rest of the similar blocks, we encode only differences between the previous block and current block. The hope is that the difference is a much smaller value, hence we need less bits to represent it. Also often the differences end up close to 0 and can be very well compressed by the next compression - run-length coding.
- Huffman compression is a lossless statistical encoding algorithm which takes into account frequency of occurrence (not each byte has the same weight)

JPEG: Block Preparation



RGB input data and block preparation

Eyes responds to luminance (Y) more than chrominance (I and Q)

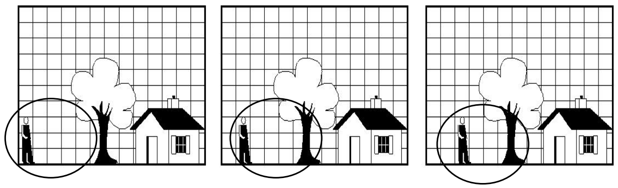
MPEG: Motion Picture Experts Group

- MPEG-1 was designed for video recorder-quality output (320x240 for NTSC) using the bit rate of 1.2 Mpbs.
- MPEG-2 is for broadcast quality video into 4-6Mbps (it fits into the NTSC or PAL broadcast channel)
- MPEG takes advantage of temporal and spatial redundancy. Temporal redundancy means that two neighboring frames are similar, almost identical.
- MPEG-2 output consists of three different kinds of frames that have to be processed:
  - I (Intracoded) frames - self-contained JPEG-encoded still pictures
  - P (Predictive) frames - Block-by-block difference with the last frame
  - B (Bidirectional) frames - Differences with the last and next frames

The MPEG Standard (2)

- I frames are self-contained, hence they are used for fast forward and rewind operations in VOD applications
- P frames code interfame differences. They are based on the idea of macroblocks, which cover 16x16 pixels in the luminance space and 8x8 pixels in the chrominance space. The algorithm searches for similar macroblocks in the current and previous frame, and if they are only slightly different, it encodes only the difference and the motion vector in order to find the position of the macroblock for decoding.
- B frames can be encoded if three frames are available at once: the past one, the current one and the future one. Similar to P frame, the algorithm takes a macroblock in the current frame and looks for similar macroblocks in the past and future frames.
- MPEG is suitable for stored video because it is an asymmetric lossy compression. The encoding takes long time, but the decoding is very fast.
- The frames are delivered at the receiver in the dependency order rather than display order, hence we need buffering to reorder the frames.

The MPEG Standard (3)



Consecutive Video Frames

Video Format

- **Video Digitizer** - is an analog-to-digital converter
- Important parameters resulting from a digitizer:
  - digital image resolution
  - quantization
  - frame rate
- Examples: Parallax XVideo The camera takes NTSC signal and the Parallax video board digitizes it. The resulting video has the following parameters:
  - 640x480 pixels spatial resolution;
  - 24 bits per pixel resolution (16777216 shades of gray or color)
  - 20 fps; if the image resolution is 320x240 pixels then this video board can provide even 30 fps.
- Output of digital video goes mostly to raster displays which have large video RAM memories. These displays use for presentation of color systems the **Color Look Up Table** (lut).