

Digital Audio

Multimedia Systems (Module 1 Lesson 1)

Summary:

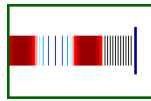
- ❑ Basic concepts underlying sound
- ❑ Facts about human perception of sound
- ❑ Computer representation of sound (Audio)
- ❑ A brief introduction to MIDI

Sources:

- ❑ My research notes
- ❑ Dr. Ze-Nian Li's course material at:
<http://www.cs.sfu.ca/CourseCentral/365/li/>

Sound Facts

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

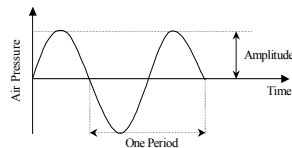


The human Ear detecting Sound

Sound Facts

Wave Characteristics

- ❑ **Frequency:** Represents the number of periods in a second and is measured in *hertz (Hz)* or *cycles per second*.
Human hearing frequency range: 20Hz to 20kHz (audio)
- ❑ **Amplitude:** The measure of displacement of the air pressure wave from its mean. Related to but not the same as loudness

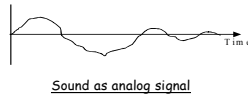


One particular frequency component

Principles of Digitization

□ Why Digitize?

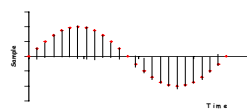
- Microphones, video cameras produce *analog signals* (continuous-valued voltages)
- To store audio or video data into a computer, we must *digitize* it by converting it into a stream of numbers.



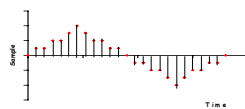
Principles of Digitization

- **Sampling:** Divide the horizontal axis (time) into discrete pieces
- **Quantization:** Divide the vertical axis (signal strength - voltage) into pieces. For example, 8-bit quantization divides the vertical axis into 256 levels. 16 bit gives you 65536 levels. Lower the quantization, lower the quality of the sound
 - Linear vs. Non-Linear quantization:
 - If the scale used for the vertical axis is linear we say its linear quantization;
 - If its logarithmic then we call it non-linear (μ -law or A-law in Europe). The non-linear scale is used because small amplitude signals are more likely to occur than large amplitude signals, and they are less likely to mask any noise.

Sampling and Quantization



Sampling



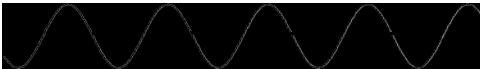
3-bit quantization

- Sampling rate: Number of samples per second (measured in Hz)
- E.g., CD standard audio uses a sampling rate of 44,100 Hz (44100 samples per second)
- 3-bit quantization gives 8 possible sample values
- E.g., CD standard audio uses 16-bit quantization giving 65536 values.
- Why Quantize?
 - To Digitize!

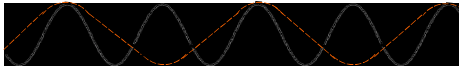
Nyquist Theorem



Consider a sine wave



Sampling once a cycle
Appears as a constant signal



Sampling 1.5 times each cycle
Appears as a low frequency
sine signal

- For Lossless digitization, the sampling rate should be **at least** twice the maximum frequency responses

Application of Nyquist Theorem

- Nyquist theorem is used to calculate the optimum sampling rate in order to obtain good audio quality.
- The CD standard sampling rate of 44100 Hz means that the waveform is sampled 44100 times per sec.
- Digitally sampled **audio** has a bandwidth of (20 Hz - 20 KHz). By sampling at twice the maximum frequency (40 KHz) we would achieve good audio quality.
- CD audio slightly exceeds this, resulting in an ability to represent a bandwidth of around 22050 Hz.

Quantization (Quality -> SNR)

- In any analog system, some of the voltage is what you want to measure (*signal*), and some of it is random fluctuations (*noise*).
- *SNR*: Signal to Noise ratio captures the quality of a signal (dB)
- *Signal to Quantization Noise Ratio (SQNR)*
- The quantization error (or quantization noise) is the difference between the actual value of the analog signal at the sampling time and the nearest quantization interval value.
- *The largest (worst) quantization error is half of the interval?*

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

SQNR Calculation (WC)

- If we use N bits per sample, the range of the digital signal is: -2^{N-1} to 2^{N-1}
- The worst-case signal to quantization noise ratio is given by:

$$SQNR = 20 \log \frac{V_{\text{signal}}}{V_{\text{quant-noise}}} = 20 \log \frac{2^{N-1}}{1/2} = N \times 20 \log 2 = 6.02N \text{ (dB)}$$

- Each bit adds about 6 dB of resolution, so 16 bits enable a maximum SQNR = 96 dB.

Miscellaneous Audio Facts

Typical Audio Formats

- Popular audio file formats include .au (Unix), .aiff (MAC, SGI), .wav (PC, DEC)
- A simple and widely used audio compression method is Adaptive Delta Pulse Code Modulation (ADPCM). Based on past samples, it predicts the next sample and encodes the difference between the actual value and the predicted value.

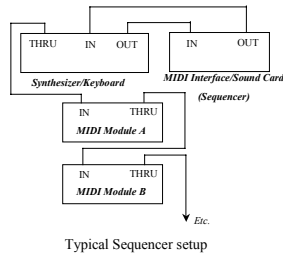
Audio Quality vs. Data Rate

Quality	Sample Rate (KHz)	Bits per Sample	Mono/ Stereo	Data Rate (kBytes/sec) (uncompressed)	Frequency Band
Telephone	8	8	Mono	8	200-3400 Hz
AM Radio	11.025	8	Mono	11.0	540-1700 KHz
FM Radio	22.050	16	Stereo	88.2	
CD	44.1	16	Stereo	176.4	20-20000 Hz
DAT	48	16	Stereo	192.0	20-20000 Hz

MIDI

❑ Musical Instrument Digital Interface

a protocol that enables computer, synthesizers, keyboards, and other musical devices to communicate with each other.



❑ Setup:

- ❑ MIDI OUT of synthesizer is connected to MIDI IN of sequencer.
- ❑ MIDI OUT of sequencer is connected to MIDI IN of synthesizer and "through" to each of the additional sound modules.

❑ Working:

- ❑ During recording, the keyboard-equipped synthesizer is used to send MIDI message to the sequencer, which records them.
- ❑ During play back, messages are sent out from the sequencer to the sound modules and the synthesizer which will play back the music.

MIDI: Data Format

- ❑ Information traveling through the hardware is encoded in MIDI data format.
- ❑ The encoding includes *note* information like beginning of note, frequency and sound volume; upto 128 notes
- ❑ The MIDI data format is digital
- ❑ The data are grouped into MIDI messages
- ❑ Each MIDI message communicates one *musical event* between machines. An event might be pressing keys, moving slider controls, setting switches and adjusting foot pedals.
- ❑ 10 mins of music encoded in MIDI data format is about 200 Kbytes of data. (compare against CD-audio!)