CSc 461/561
Multimedia Systems
Audio representation

Jianping Pan

Spring 2015

First things first

- Course website alive
  - on http://connex.csc.uvic.ca
  - see “CSC 461: 201501 A01” tab (or adjust prefs)
  - if you are not officially “in the system” yet, please send me an email and I can add you in manually
  - course lectures embedded in connex *
    - also publicly available http://www.cs.uvic.ca/~pan/csc461
  - chat room and web forums available too
    - discussion group: get help and help others

get your A0 to me by Friday---help me help yourself---project hints next week!

* now https embedding should be working without any tweaks in your browser

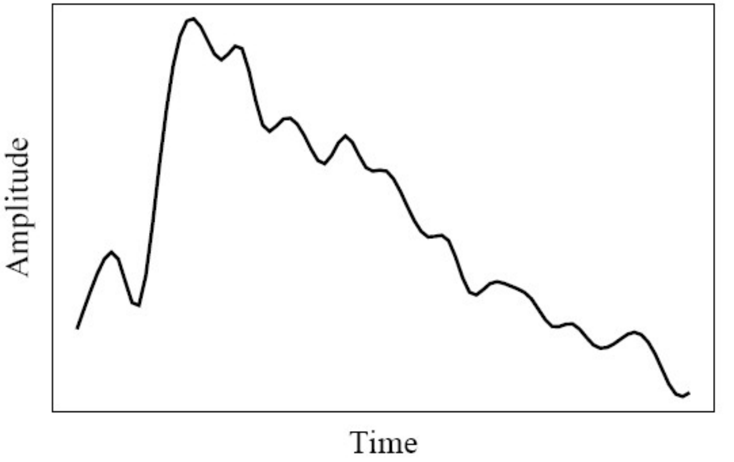
Research opportunities

- Undergraduate Research Award (USRA)
  - UVic CSc awarded 6+ for May’15 to Apr’16
  - Testdrive research for a term before a few years
  - Dept deadline: March 6, 2015
- Grad research opportunities
  - UVic CSc: theory, systems/networking, apps
  - UVic ECE: DSP, comm, power, materials, etc
  - UVic Fellowship deadline: January 15, 2015

* some examples: http://www.cs.uvic.ca/~pan/usra

Sound is a wave

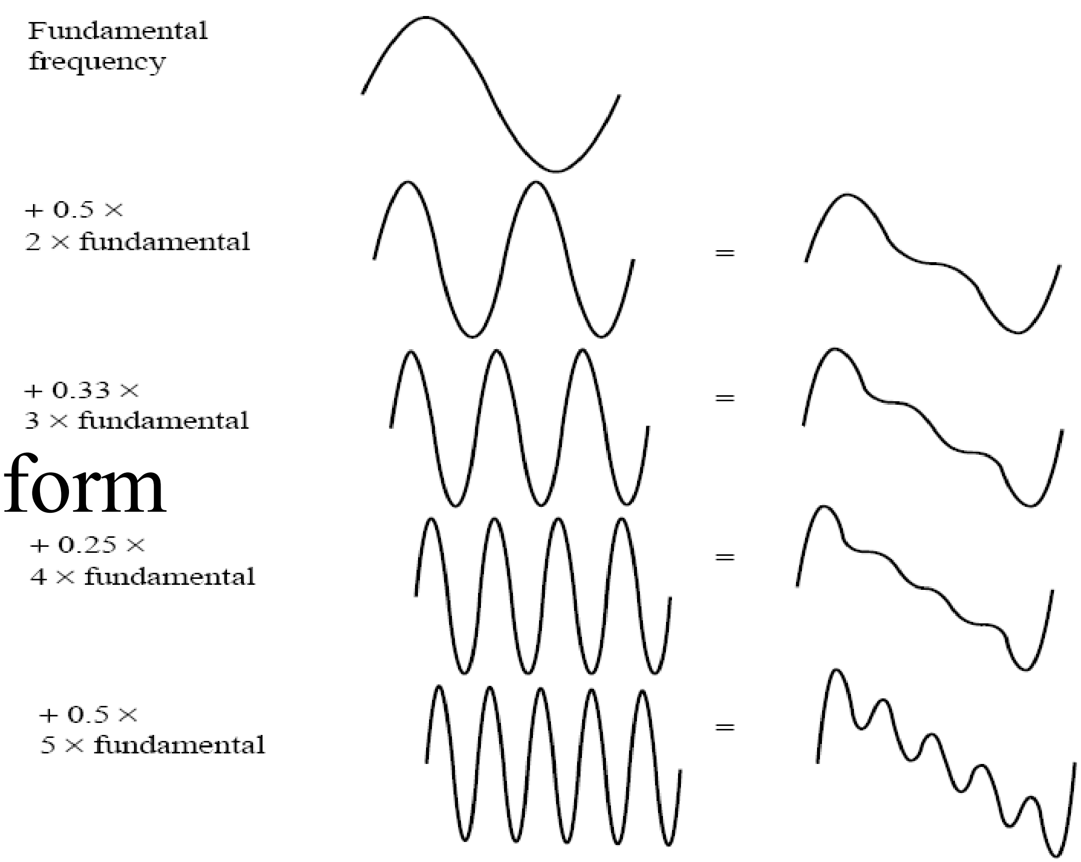
- How we hear
  - mouth / transmitter
  - air / cord
  - ear / receiver
- Continuous in both
  - time
  - amplitude
  - mechanic waves



* human perception: 20 to 20KHz (17m to 17mm), about 330m/s in air

A wave of many *waves*

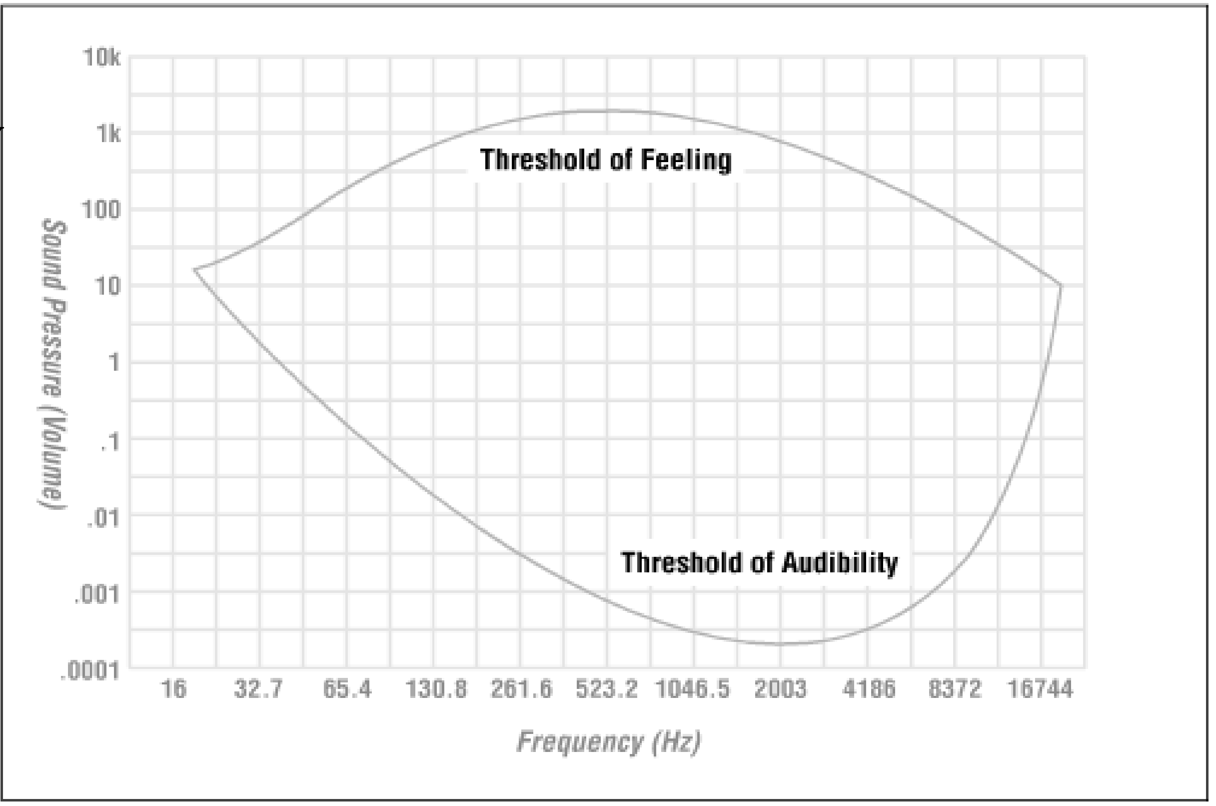
- Frequency
  - pitch
- Amplitude
  - loudness
- Fourier Transform



* any (periodic) waves: a sum of (co)sine waves of different frequencies

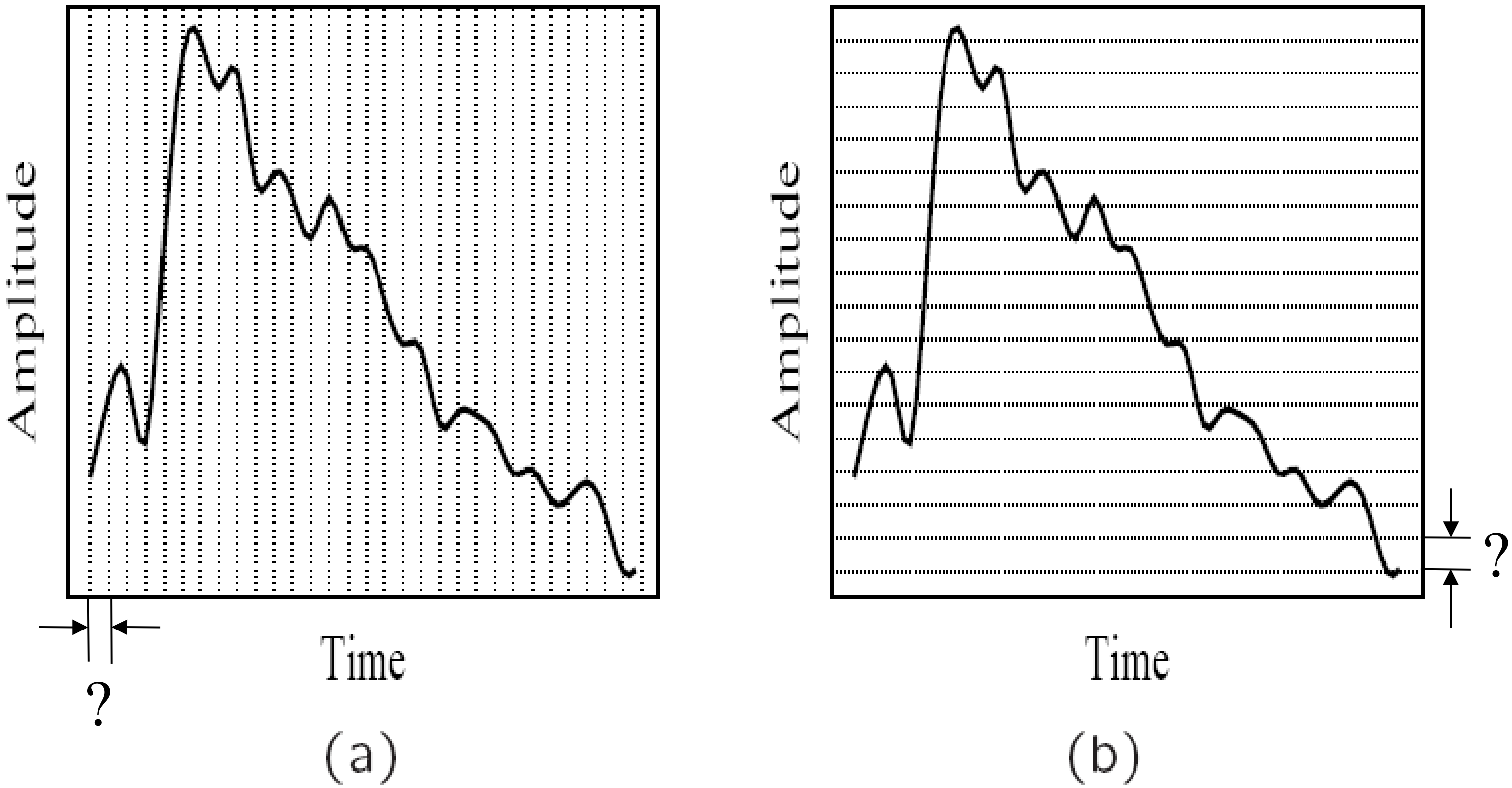
Not all waves are equal

- For human
  - frequency
  - volume



* cannot hear or too loud so to hurt?

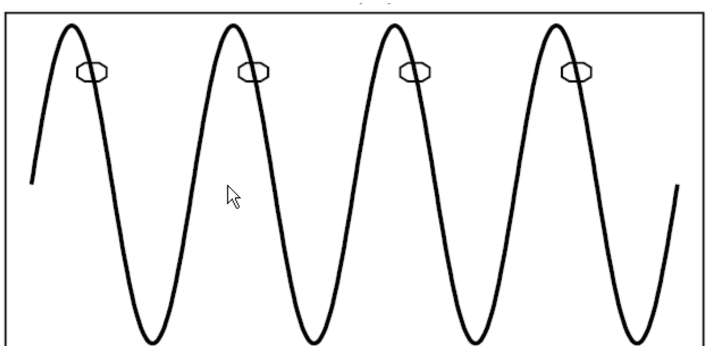
Sampling and quantization

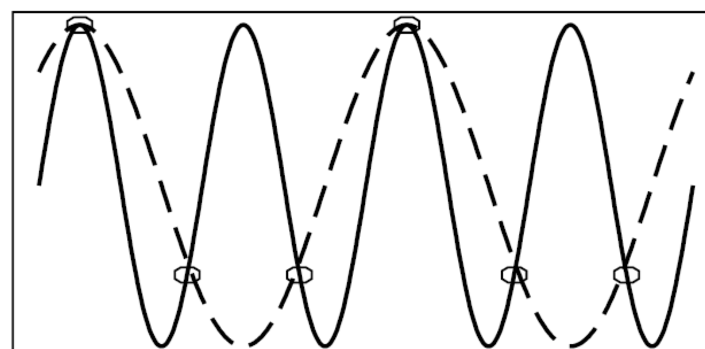


* in time and amplitude domains

Sampling rate

- If sampling_freq = true_freq
  - in this example, a constant signal
- If sampling_freq = 1.5x true_freq
  - an alias_freq of 0.5x true_freq





* how often to sample (is sufficient)?

Nyquist Theorem

- If a signal is bounded by frequency (f1,f2), then the sampling rate should be *at least* 2(f2-f1) to <u>reconstruct</u> the signal
- Human can hear 20Hz - 20KHz
  - CD sampling rate: 44.1KHz
- Human voice 300Hz - 4KHz
  - telephone systems sampling rate: 8KHz

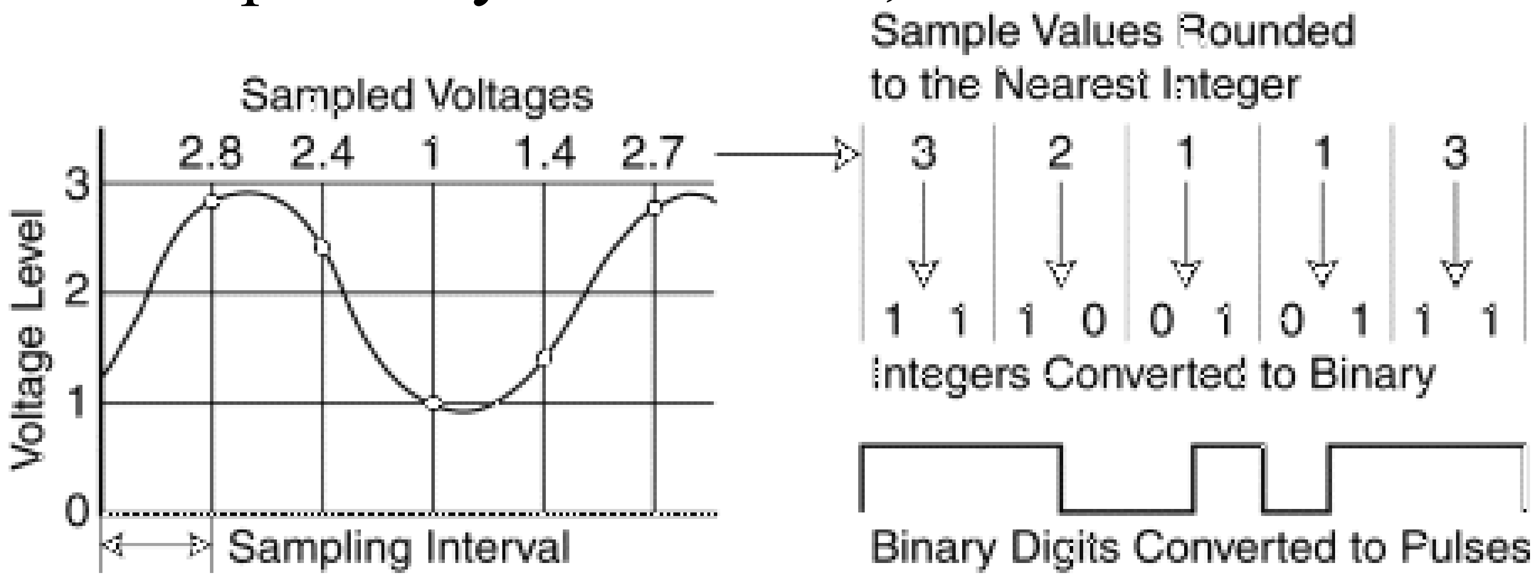
* too much or too less is suboptimal

Alias frequency

- If sampling_freq >= 2x true_freq
  - true_freq is recovered
- If sampling_freq in [true_freq, 2x true_freq)
  - alias_freq = sampling_freq - true_freq
- If sampling_freq < true_freq
  - true_freq' = true_freq mod sampling_freq
  - follow the first two approaches

Quantization depth

- How many bits per sample
  - In this example: 2 bits
  - telephone systems: 8 bits; CD: 16 bits



* telephone: 8KHz * 8bits/sample = 64Kbps

Quantization noise

- Quantization is lossy
- More bits, less information loss
- Signal-to-quantization-noise ratio
  - defined as a (log) ratio of signal/noise power
  - one more bit can add 6 db to SQNR

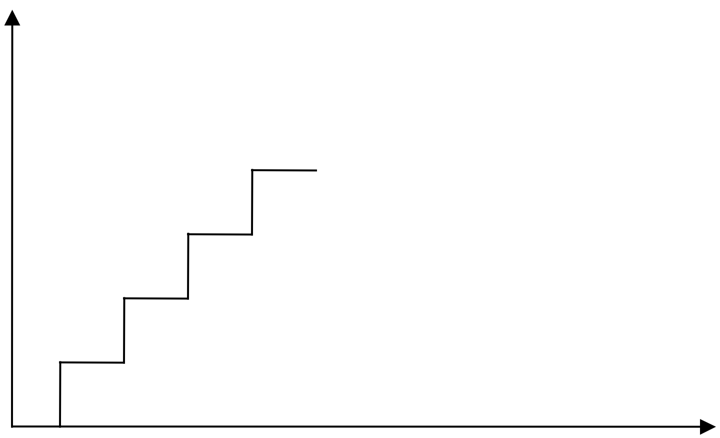
$$SQNR = 20 \log_{10} \frac{V_{signal}}{V_{quan\_noise}} = 20 \log_{10} \frac{2^{N-1}}{\frac{1}{2}}$$

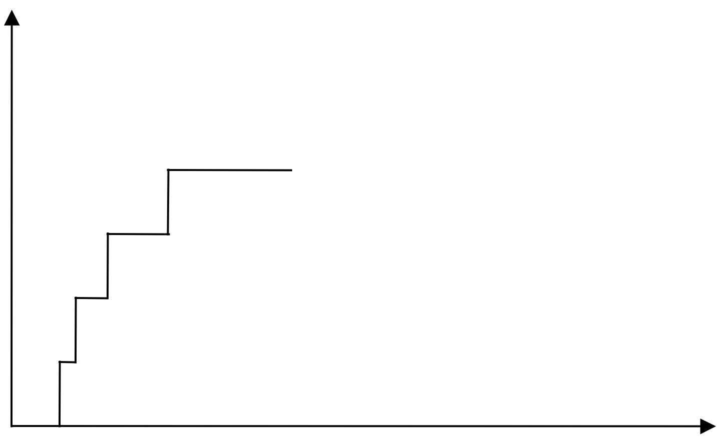
$$= 20 \times N \times \log 2 = 6.02\, N (\mathrm{dB}) \qquad (6.3)$$

* dB: decibel; 10 dB = 1 B; in honor of Alexander Graham Bell (phone)

Not all quantizer are linear

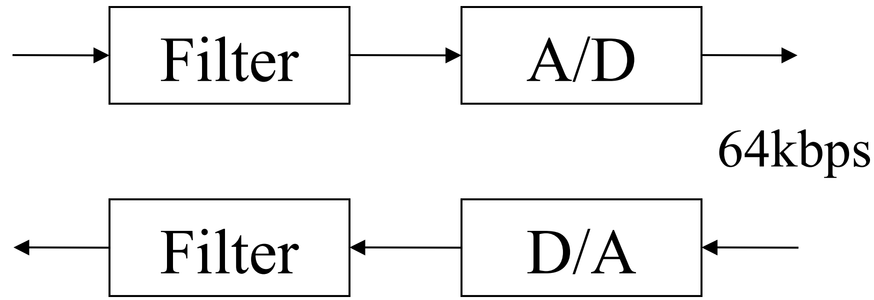
- Nonlinear/non-uniform quantization
  - humans are more sensitive to *small* changes from *small* values
- u-law or A-law (more bits for small values)





* also known as bit (budget) allocation

Pulse Code Modulation (PCM)

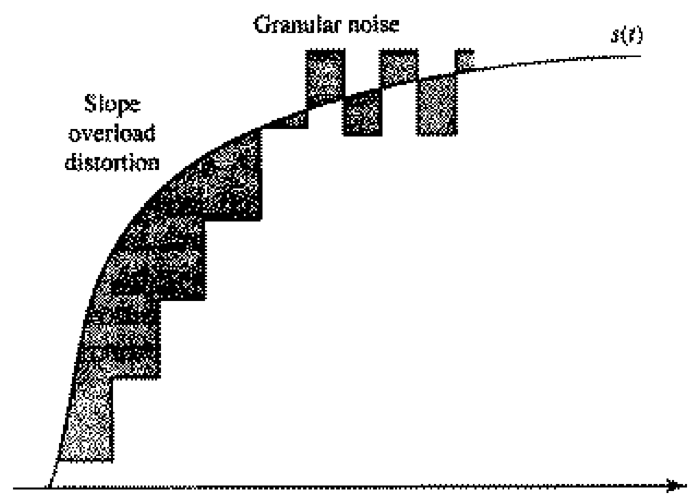


- Band-limit filter
  - sampling rate: 8KHz
- Quantizer
  - sample size: 8-bit (output) with u-law or A-law
- Data rate: 64Kbps (output)
- Reconstruction: low-pass filter
- ITU-T G.711

* recall dialup->56Kbps->DSL in CSc361?

Delta Modulation (DM)

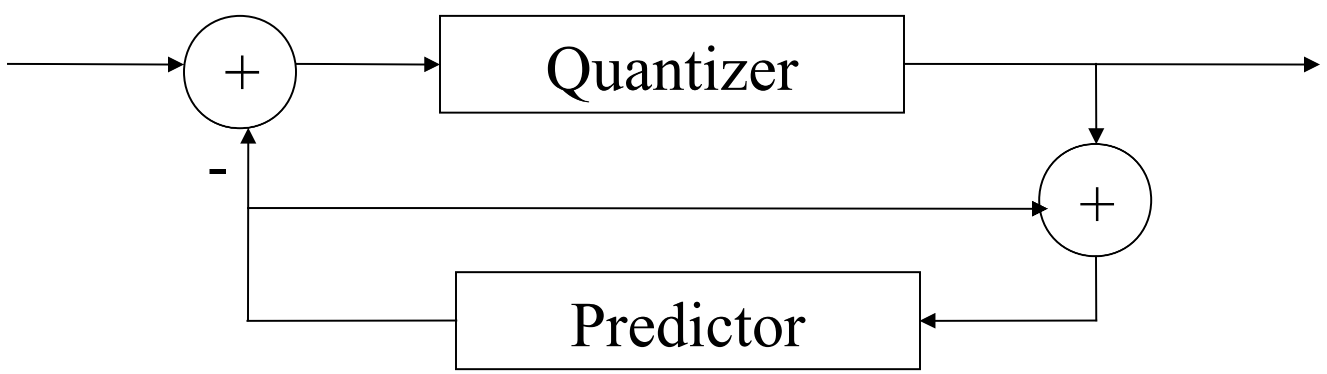
- Prediction
  - If current > last, output 1 (to increase)
  - if current < last, output 0 (to decrease)
- slope overload
  - change too fast
- granular noise
  - change too slow



* recall TCP/IP header compression in CSc361?

Differential PCM (DPCM)

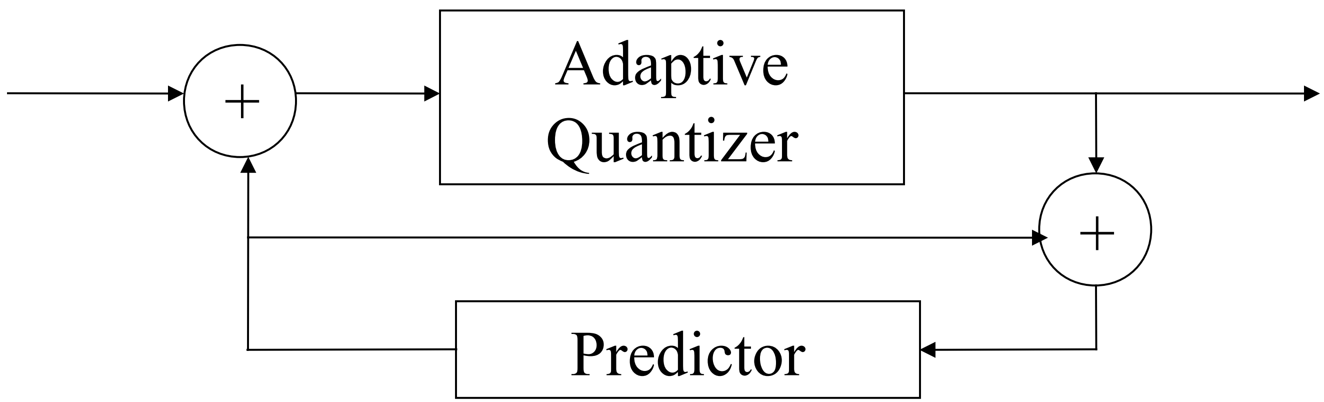
- Predictor
  - Predict the current based on the history
- Quantizer
  - quantize the difference



* essentially how to use fewer bits more efficiently

Adaptive DPCM (ADPCM)

- Adaptive quantizer
  - adapt quantization step to the difference



- ITU-T G.723: 8KHz, 4-bit, 32Kbps

* half of G.711 (64Kbps)

This lecture

- Audio presentation
  - sampling rate, alias frequency
  - quantization depth, SQNR
  - PCM, DPCM, ADPCM
- Explore further
  - MIDI [Li&Drew 6.2]

Next lecture

- Multimedia representation
  - image [Ref: Li&Drew Chap 3-4]
    - bitmap vs vector
    - gray-scale vs color [3.1.1/2, 3.1.4-6]
    - picture resolution, pixel depth
    - color space (RGB, CMY, YUV) [4.2.1-3, 4.3.2]
    - gamma correction [4.1.6]