Audio Signal Processing Course Summary

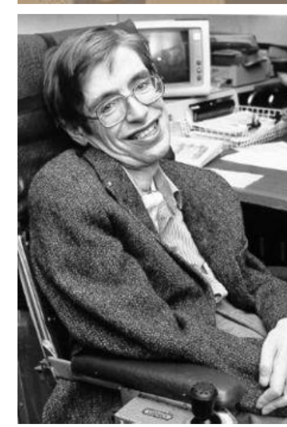
Zhiyao Duan

Spring 2020

What is Audio Signal Processing?

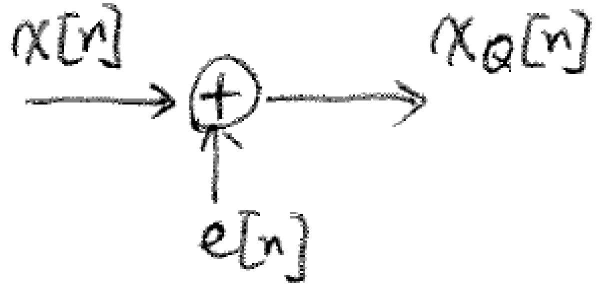
- Intentional manipulation of sound (e.g., music and speech).

- To analyze sound
  - Speaker recognition, music transcription
- To modify sound
  - Distortion, chorus, 3D audio, vocal removal
- To make new sound
  - Keyboard, speech synthesis, singing synthesis



MSR – "Penny Lane"

Quantization



- We want quantization noise $e[n]$ to have
  - Uniform PDF between [-Q/2, Q/2], so SQNR = 6.02N + 1.76 dB for full-scale sinusoidal input
  - Flat PSD, so no pitched artifacts
- Hard to achieve because $e[n]$ depends on input signal, and input signal has to satisfy unrealistic conditions: $F_X\left(\frac{k}{Q}\right) = 0$ for all $k = \pm 1, \pm 2, \ldots$

Dither

- Apply dither $d[n]$ before quantization
  - Dither: independent from input, and satisfies Schuchman's condition $F_D\left(\frac{k}{Q}\right) = 0$ for all $k = \pm 1, \pm 2, \ldots$

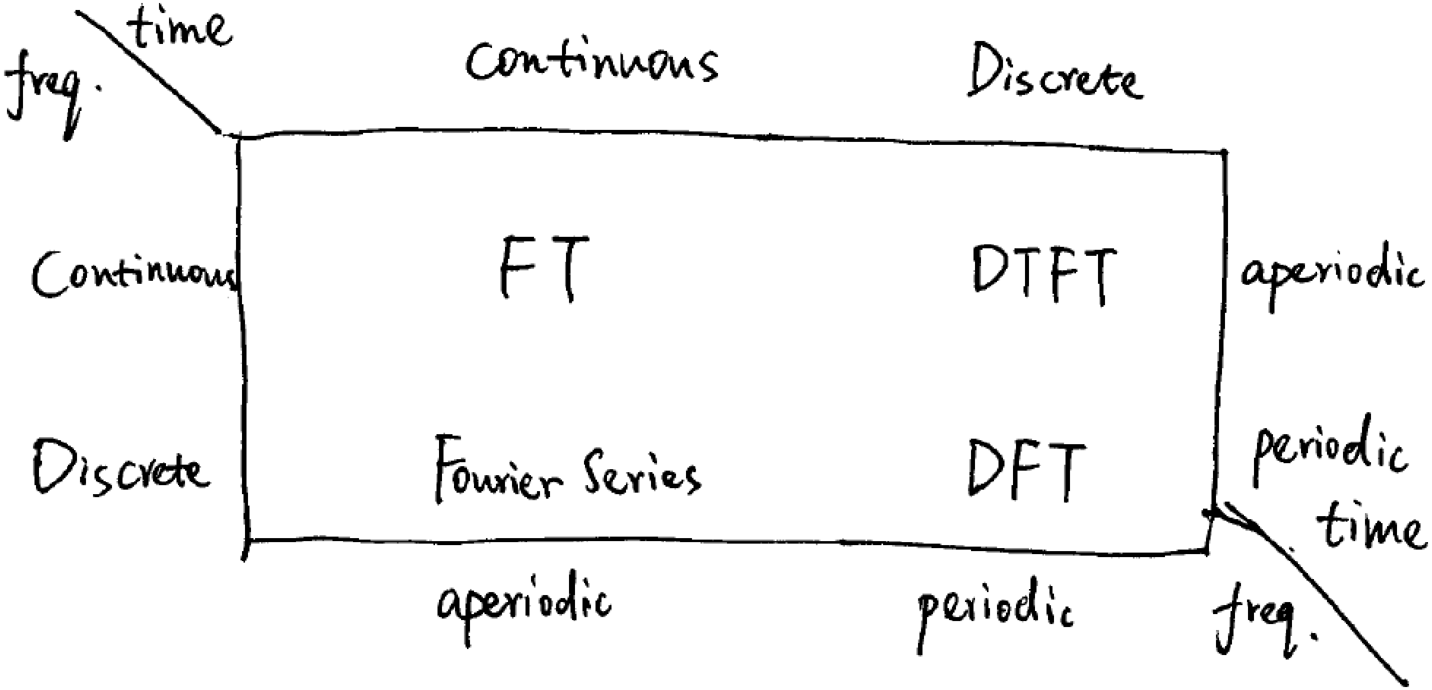
- Independent variable summation ➔ PDF convolution ➔ characteristic function multiplication
  - So now condition is satisfied as $F_{X+D}\left(\frac{k}{Q}\right) = 0$ for all $k = \pm 1, \pm 2, \ldots$

Sampling

- Sampling Theorem
  - Need to sample at least twice as fast as the highest signal frequency to prevent aliasing
  - Nyquist rate: twice of signal's maximum frequency
  - Nyquist frequency: half of sampling rate

- Benefits of oversampling
  - analog LPF ➔ digital LPF
  - SQNR increase
- Sigma-Delta Modulation
  - Shape noise to further increase SQNR

Fourier Transforms



- Frequency conversions: $f_s/2 \Leftrightarrow \pi \Leftrightarrow N/2$
- FFT algorithm

Short-time Fourier Transform (STFT)

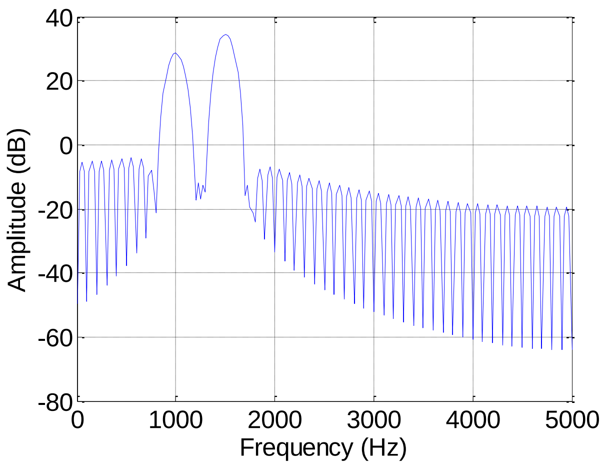
- Ideal time-frequency resolution tradeoff

$$\Delta t \cdot \Delta f = 1$$

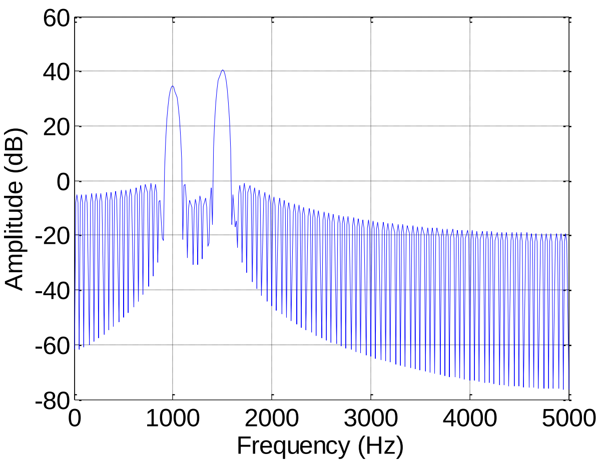
(second) (Hz)

- Zero padding does not increase freq. resolution

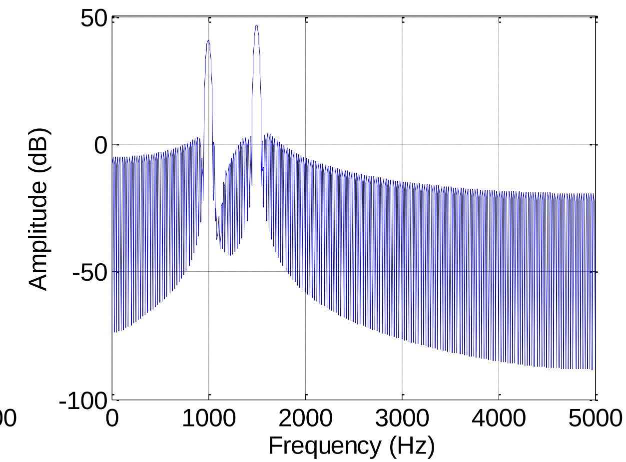
Window length: 10ms



Window length: 20ms

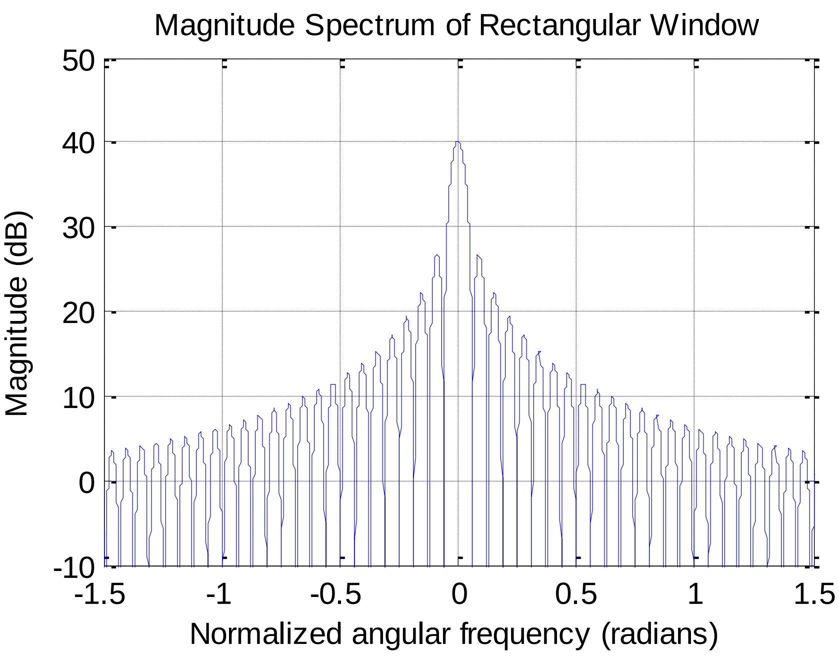


Window length: 40ms



STFT

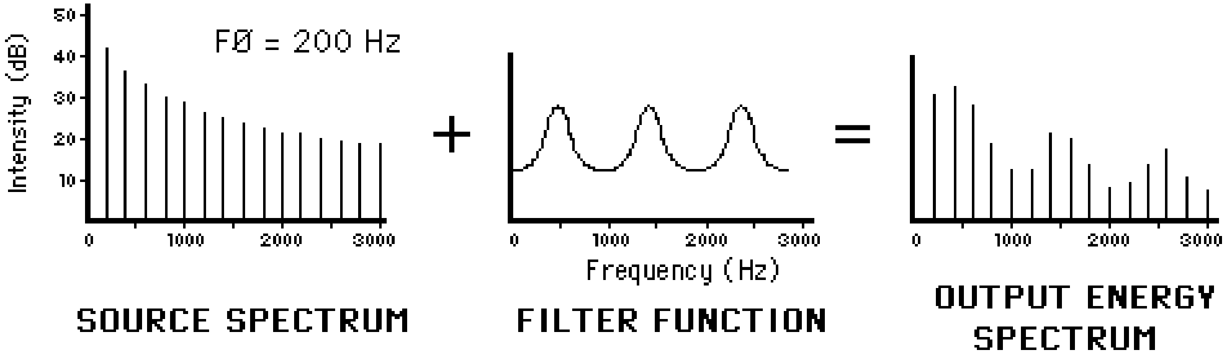
- Window effect
  - Main lobe width
  - First side lobe height
  - Side lobe decay rate



- Overlap-add technique
  - COLA, WOLA

- Phase vocoder
  - Maintain appropriate phase advance between frames

Source-filter Models

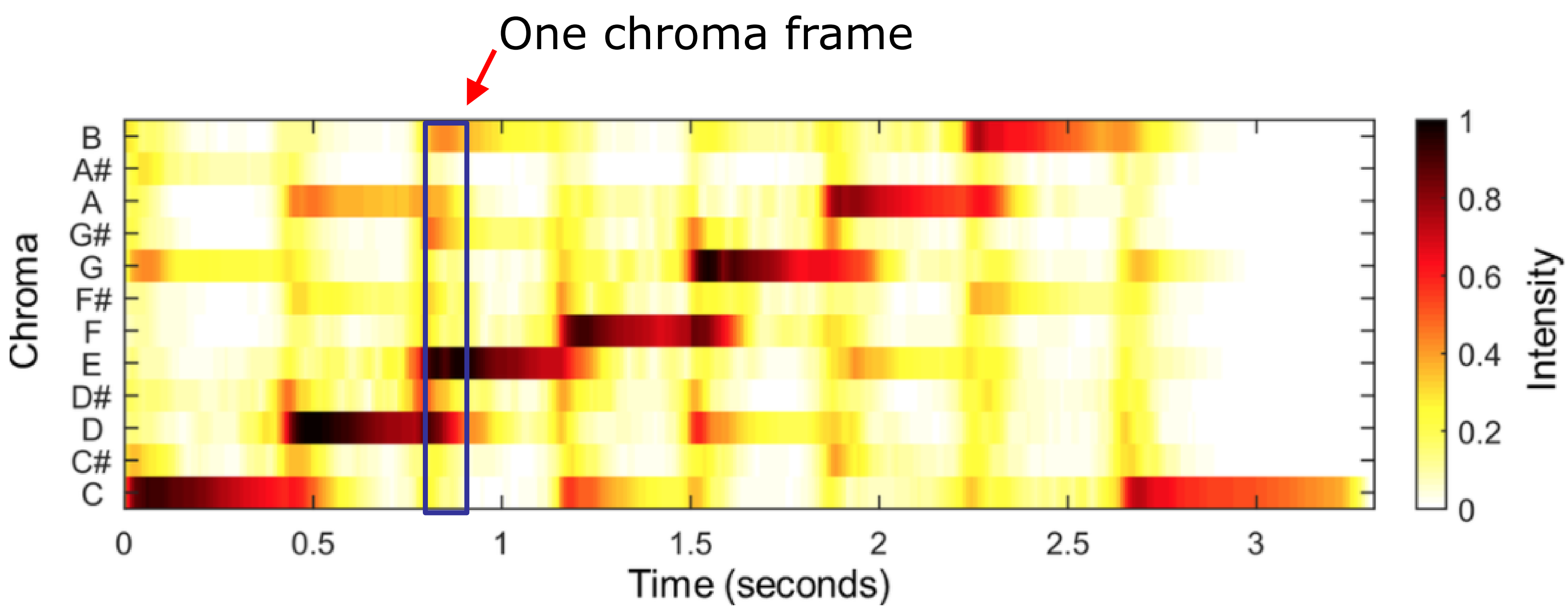


- LPC
  - time domain, autoregressive model

- Cepstrum
  - IFFT{log|FFT{x[n]}|}

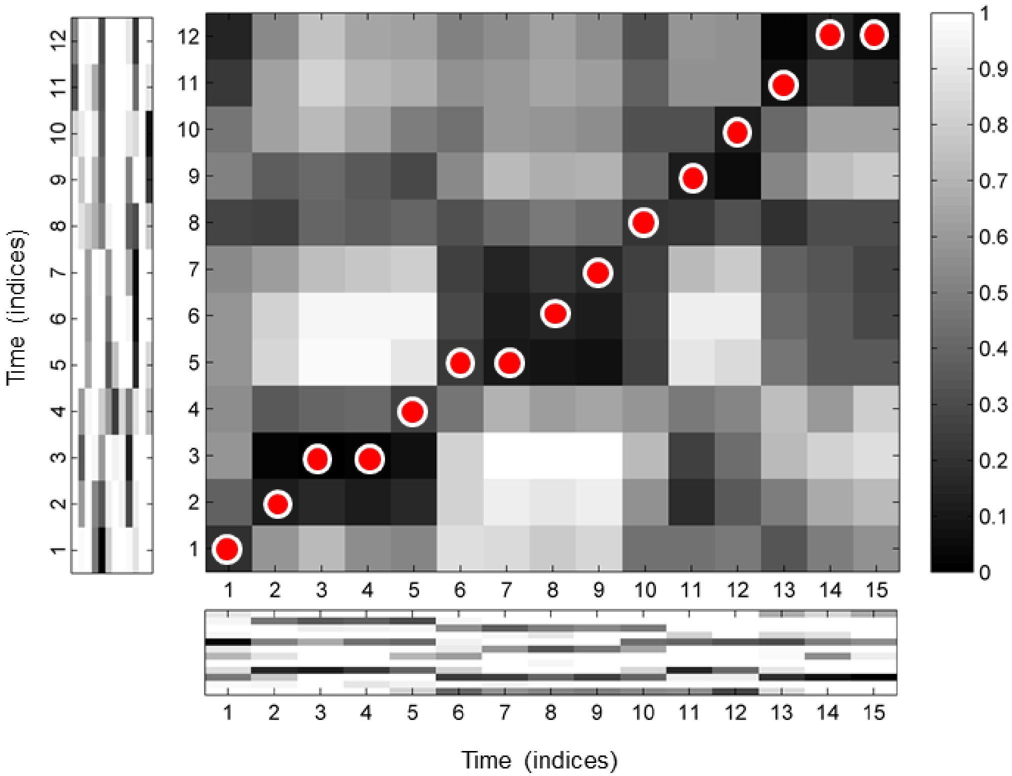
Chromagram

- Fold the magnitude spectrogram by octaves



Audio Alignment

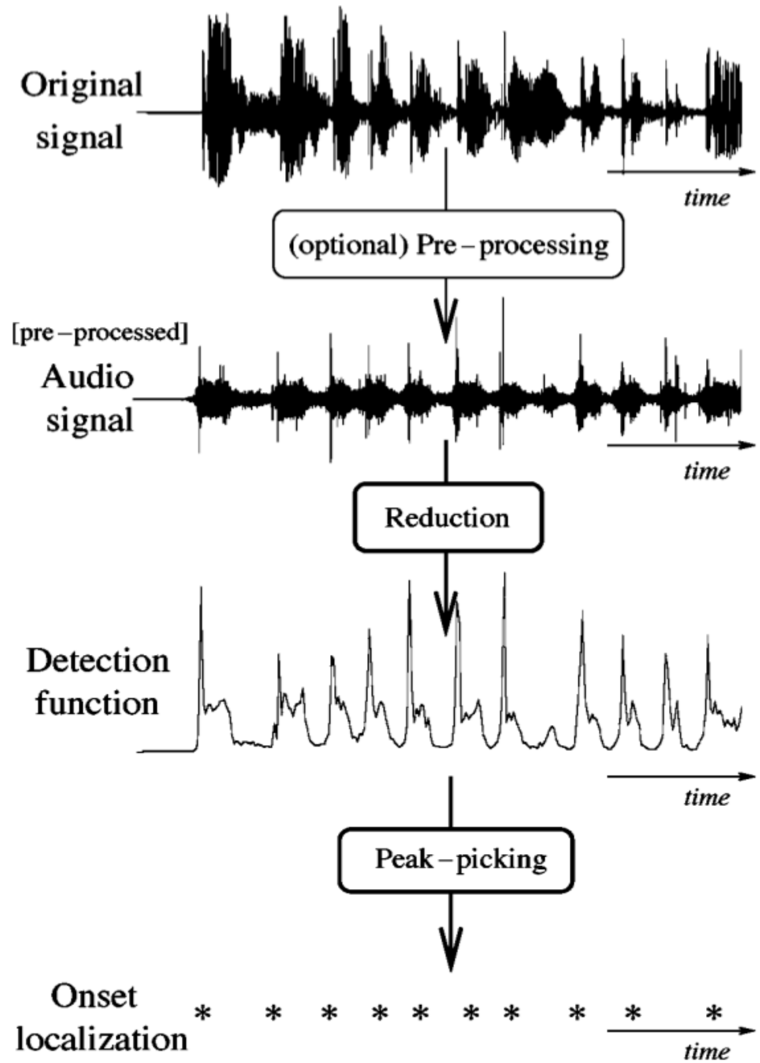
- Forward calculation of accumulated distances
- Backward tracing for alignment path



$$
\begin{aligned}
q_1 &= (N, M) \\
q_{\ell+1} &= (1, m-1) \qquad \text{if } n = 1, \\
q_{\ell+1} &= (n-1, m) \qquad \text{if } m = 1, \\
q_{\ell+1} &= \operatorname{argmin} \begin{cases} \mathbf{D}(n-1, m-1), \\ \mathbf{D}(n-1, m), \\ \mathbf{D}(n, m-1) \end{cases}
\end{aligned}
$$

Rhythm Analysis

- Onset detection
- Tempo estimation
  - E.g., tempogram
- Beat tracking
  - E.g., by dynamic programming
- Beat spectrum
  - Average of autocorrelation functions in all frequency bins



From [Bello et al., 2005]

Pitch Analysis

- Pitch perception is complicated
  - Missing fundamentals phenomenon

- Pitch detection algorithms
  - Time domain: find the period of waveform
  - Frequency domain: find the divisor of peaks
  - Cepstral domain: find the frequency gap between spectral peaks

- Pitch detection in noisy environments
- Multi-pitch detection is challenging

Digital Filters

- Linear systems
- Time-invariant systems

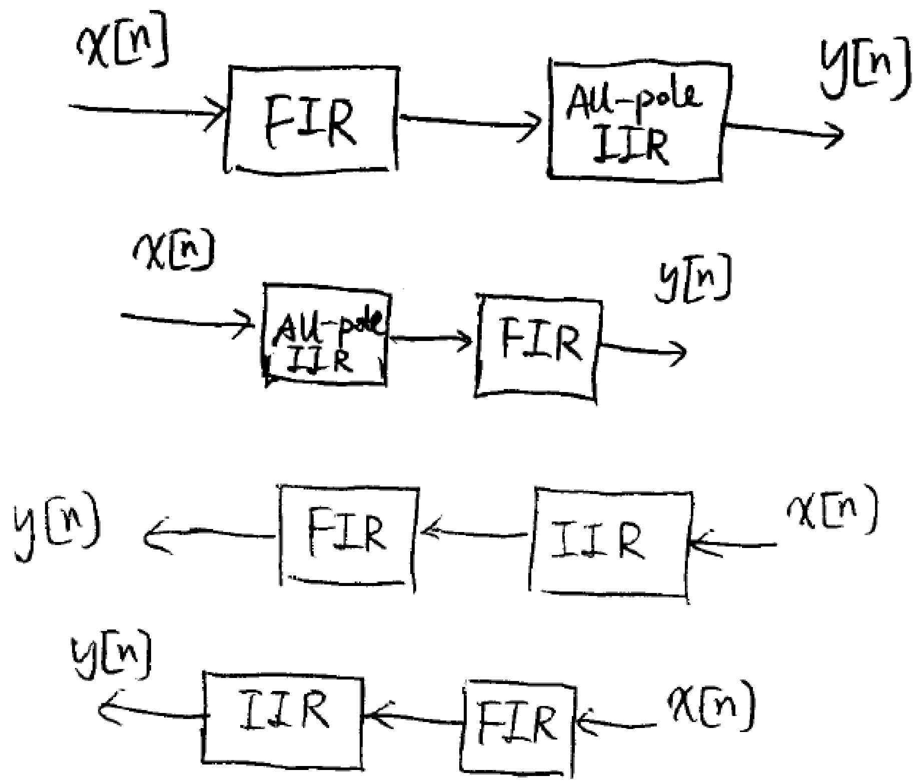
- LTI systems
  - Completely characterized by its impulse response

- Causality, stability
- FIR/IIR filters
- Z transform, zero-pole plot, frequency response

Filter Structures

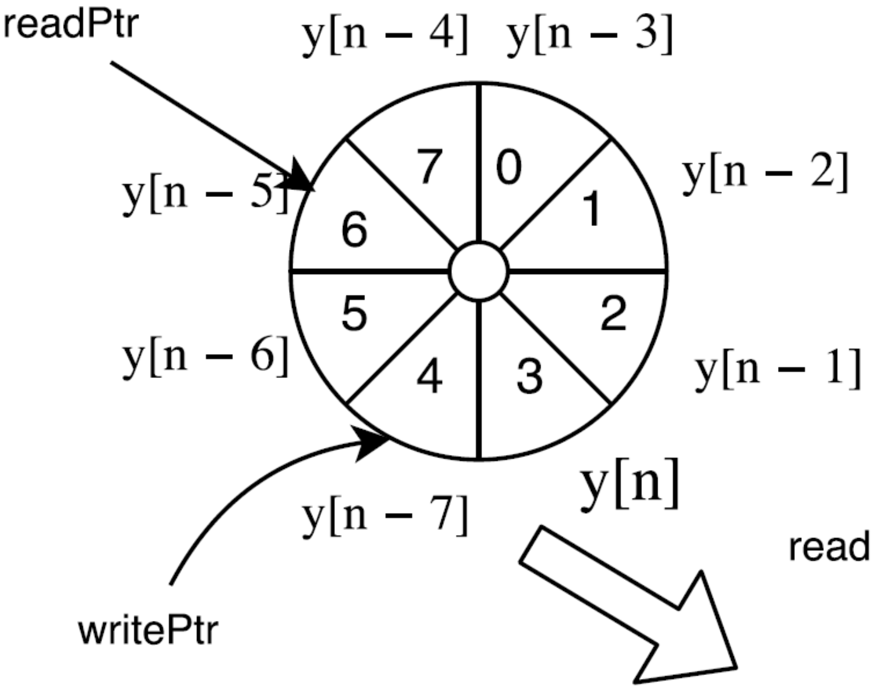
- Series combination
- Parallel combination

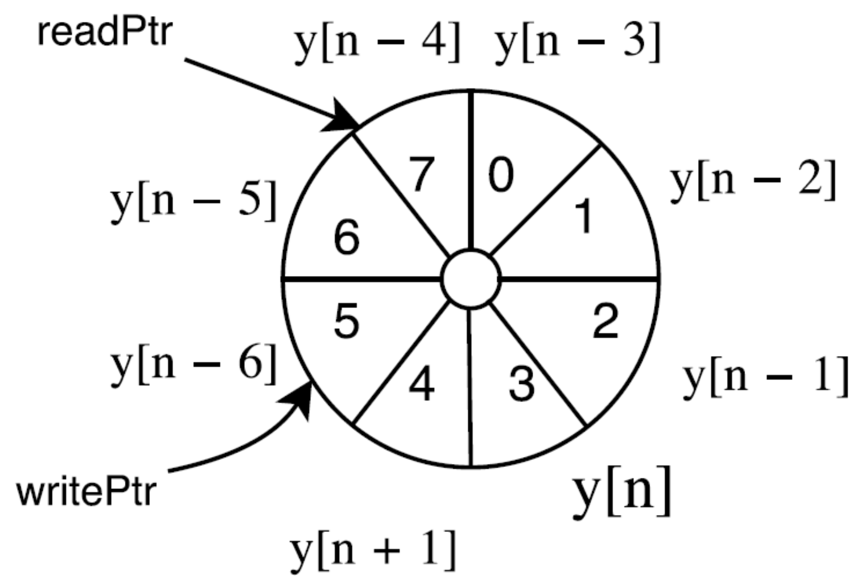
- Direct Form I
- Direct Form II
- Transposed Direct Form I
- Transposed Direct Form II



Delay Line

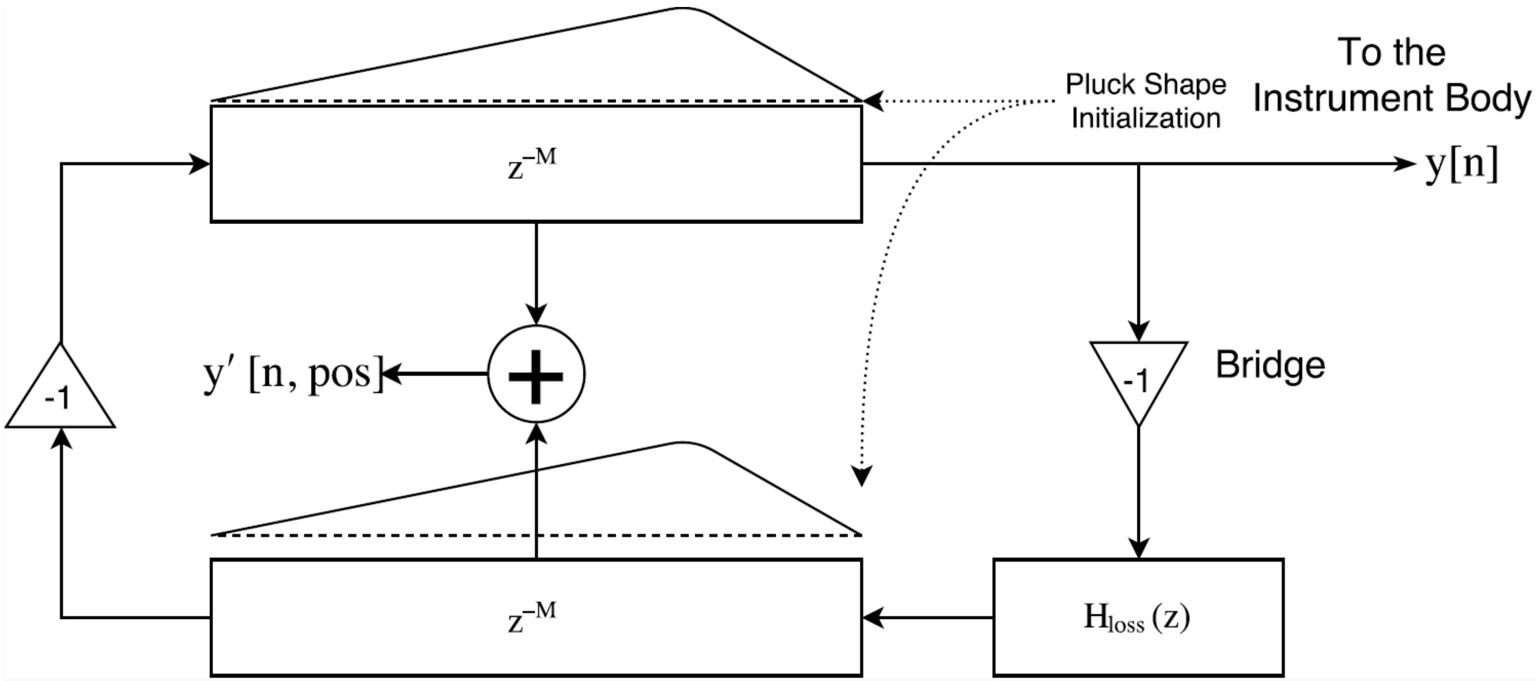
When $y[n+1]$ comes in
1. Fetch the position readPtr is looking at : $y[n-5]$
2. write $y[n+1]$
3. readPtr = (readPtr+1)%8
writePtr = (writePtr+1)%8



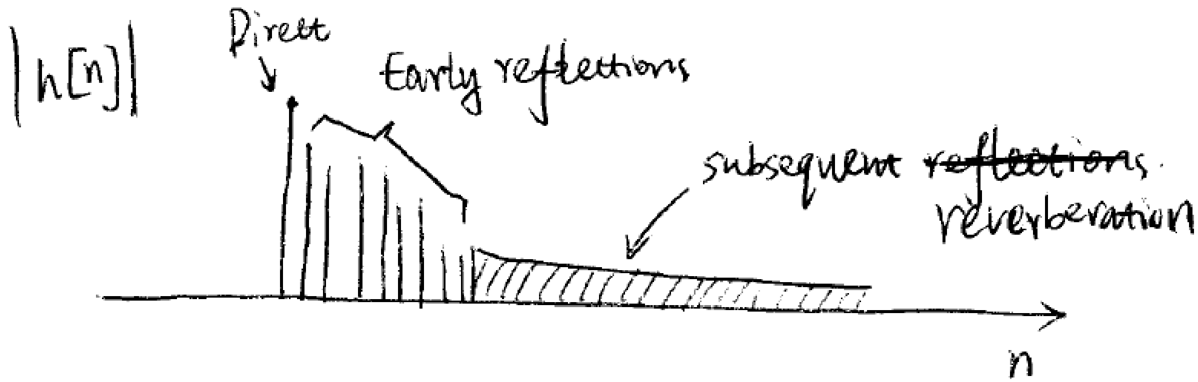


What's the delay in this case?
(writePtr-readPtr+ 8)%8

Physical Modeling of Plucked String



Room Simulation and Reverberation



- T60, room Eigen-frequencies

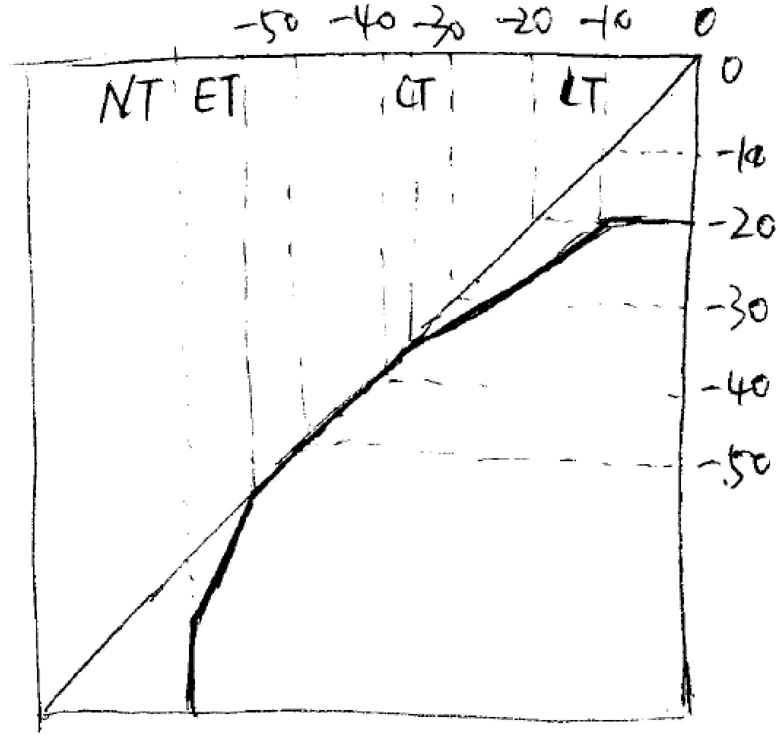
- Early reflection simulation
- Reverberation simulation
  - Schroeder algorithm: 1) parallel of comb filters to simulate room eigen-frequencies, 2) series of all-pass filters to increase echo density

Spatial Effects and Localization

- Amplitude panning
- ITD, IID
- HRTF

- Localization for separation
  - Cluster t-f bins according to ITD/IID

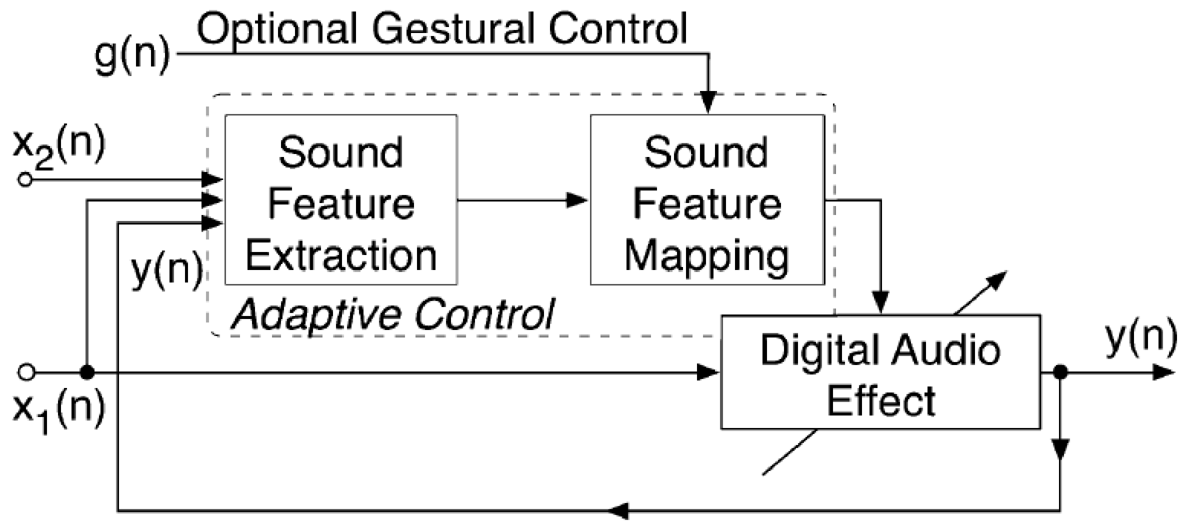
Dynamic Range Control



- Signal level estimation
  - Envelope follower
  - RMS value estimation

Adaptive Audio Effects

- Effects change over time



- Effects categorization
  - Loudness, time, pitch, space, timbre

Modulation Effects

- Tremolo (amplitude modulation)
- Vibrato (frequency modulation)
  – achieved by changing delay time
  – Fractional delay time interpolation
    - Linear interpolation: unwanted LPF effect
    - All-pass interpolation: nice!
- Flanger (phase modulation, delay < 20ms)
  – Signal + variable phase delayed copy
- Chorus (phase modulation, delay > 20ms)
  – Signal + random phase delayed copies

Course Objectives

- Good understanding on various aspects of audio signal processing
- Build intimate connections between theory and practice

- Improve implementation skills
- Gain experience in doing small-scale research projects
- Enhance capabilities of problem solving, team-working, presentation, etc.

We started here…

We are almost there!

Zoom Project Presentation

- When: May 7 (Thursday), 10-12:20 pm
- Where: Zoom (check email for meeting ID)
- Grading: faculty + TAs
- Recording: Shared within UR
- **Attendance: Required.** Students who can't attend will need to watch the recording and describe each project using own language.

- Bring your own food and drink
- Invite your friends!

Things To Do

- Final paper + presentation slides (due Thursday 5/7 night)
- Evaluate the course online with detailed feedback
- Consider taking Computer Audition in Fall 2018
  - With Dr. Andrea Cogliati
- Stay healthy!


Thank You!