CTP431- Music and Audio Computing
Audio Signal Processing (Part #2)

Graduate School of Culture Technology
KAIST
Juhan Nam

Types of Audio Signal Processing

- Filter/EQ
- Compressor
- Delay-based Effects
  - Delay, reverberation
- Spatial Effect
  - HRTF
- Playback Rate Conversion
  - Resampling

Filters

- Adjust the level of a certain frequency band
  - Lowpass
  - Highpass
  - Bandpass
  - Notch
  - Resonant Filter
  - Equalizer

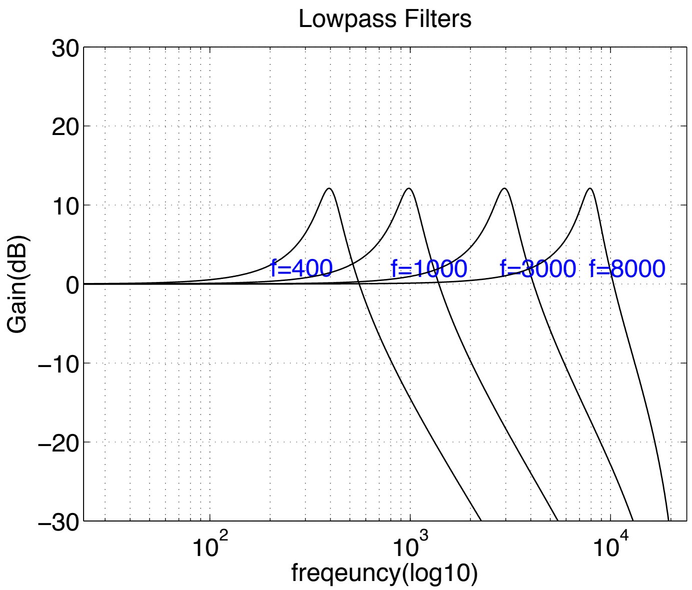
- Parameters
  - Cut-off/Center Frequency
  - Q: sharpness/resonance

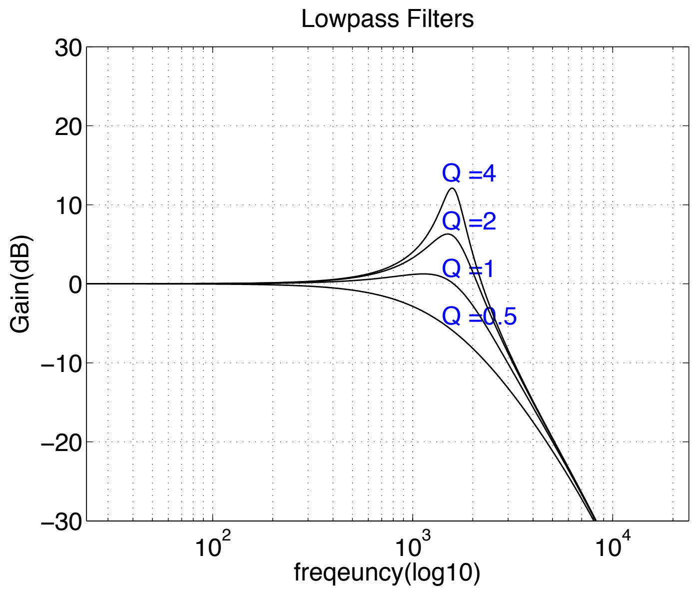
Low-pass Filter

- Transfer Function

$$H(z) = \left(\frac{1-\cos\Theta}{2}\right)\frac{1+2z^{-1}+1z^{-2}}{(1+\alpha)-2\cos\Theta z^{-1}+(1-\alpha)z^{-2}} \qquad \alpha = \frac{\sin\Theta}{2Q} \qquad \Theta = 2\pi f_c / f_s$$

  – fc : cut-off frequency, Q: resonance

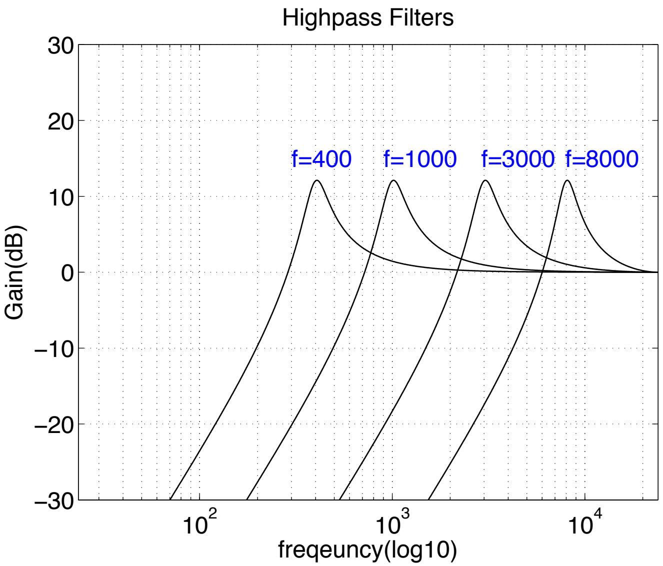


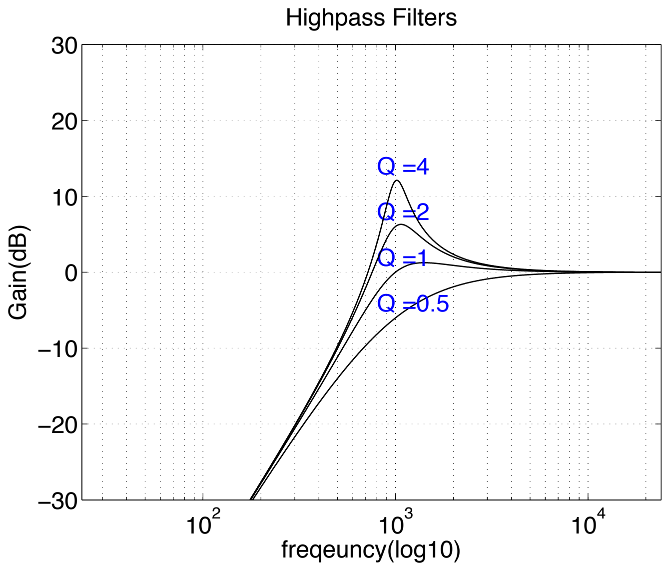


High-pass Filter

- Transfer Function

$$H(z) = \left(\frac{1+\cos\Theta}{2}\right)\frac{1-2z^{-1}+1z^{-2}}{(1+\alpha)-2\cos\Theta z^{-1}+(1-\alpha)z^{-2}} \qquad \alpha = \frac{\sin\Theta}{2Q} \qquad \Theta = 2\pi f_c / f_s$$

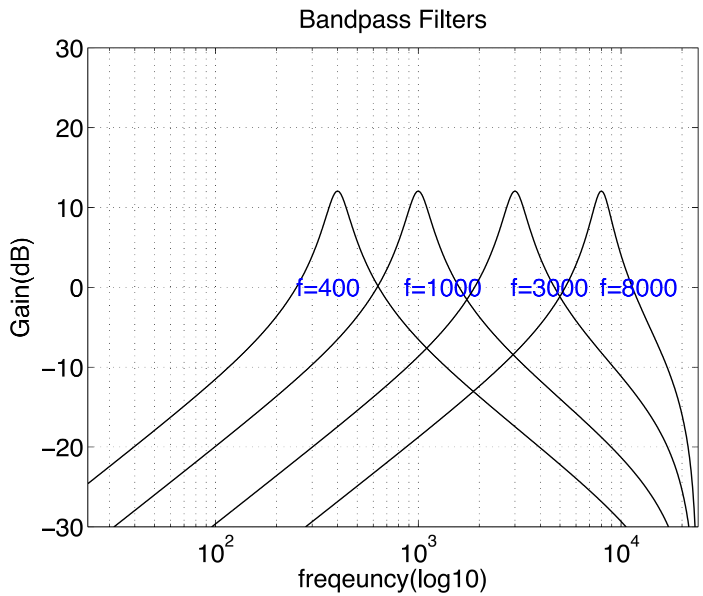


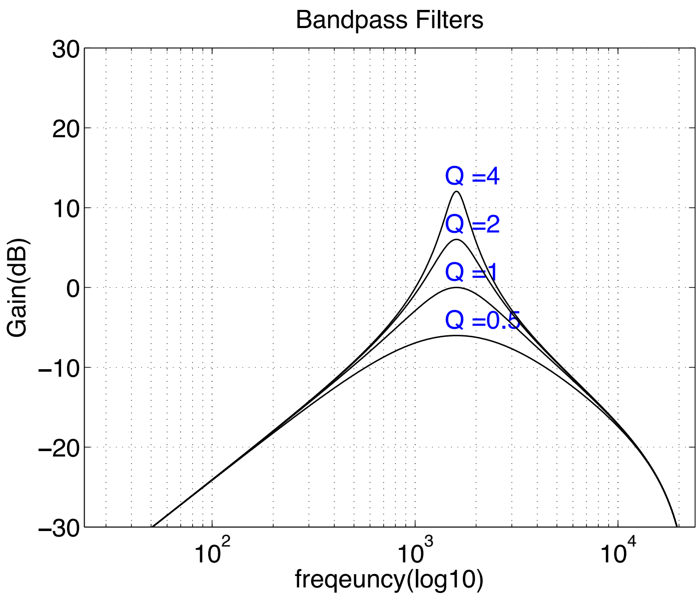


Band-pass filter

- Transfer Function

$$H(z) = \left(\frac{\sin\Theta}{2}\right)\frac{1 - z^{-2}}{(1+\alpha) - 2\cos\Theta z^{-1} + (1-\alpha)z^{-2}} \qquad \alpha = \frac{\sin\Theta}{2Q} \qquad \Theta = 2\pi f_c / f_s$$

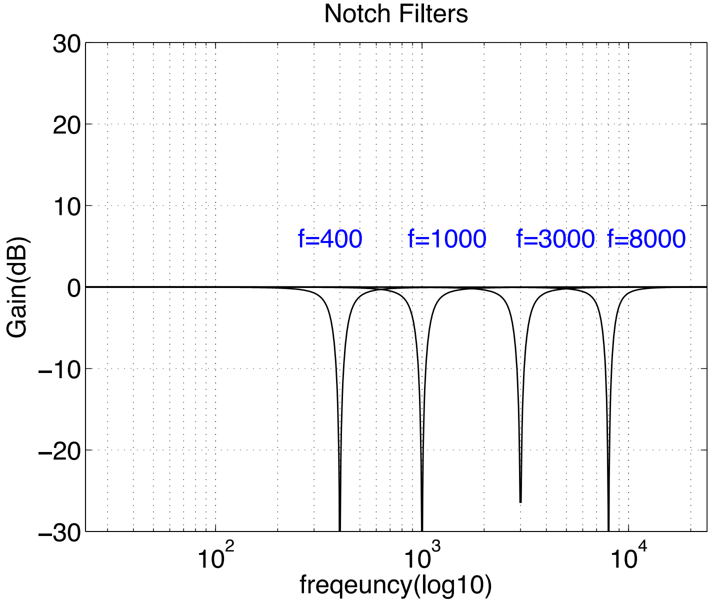


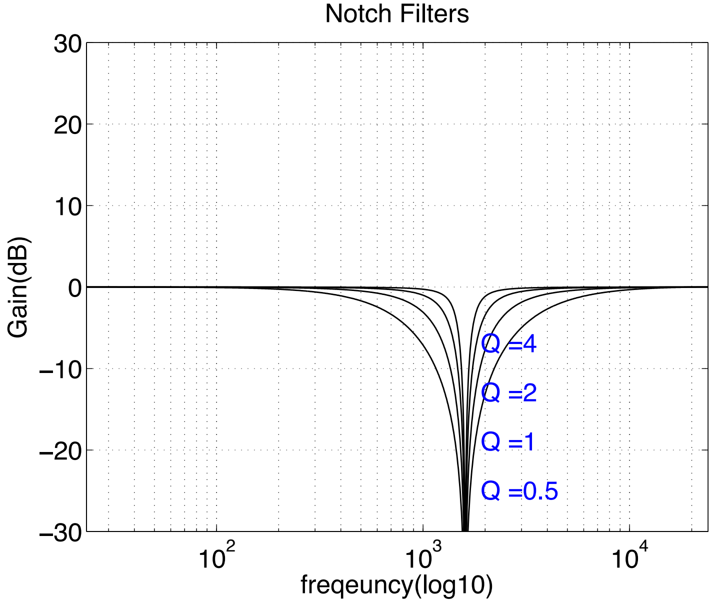


Notch filter

- Transfer Function

$$H(z) = \frac{1 - 2\cos\Theta z^{-1} + z^{-2}}{(1+\alpha) - 2\cos\Theta z^{-1} + (1-\alpha)z^{-2}} \qquad \alpha = \frac{\sin\Theta}{2Q} \qquad \Theta = 2\pi f_c / f_s$$

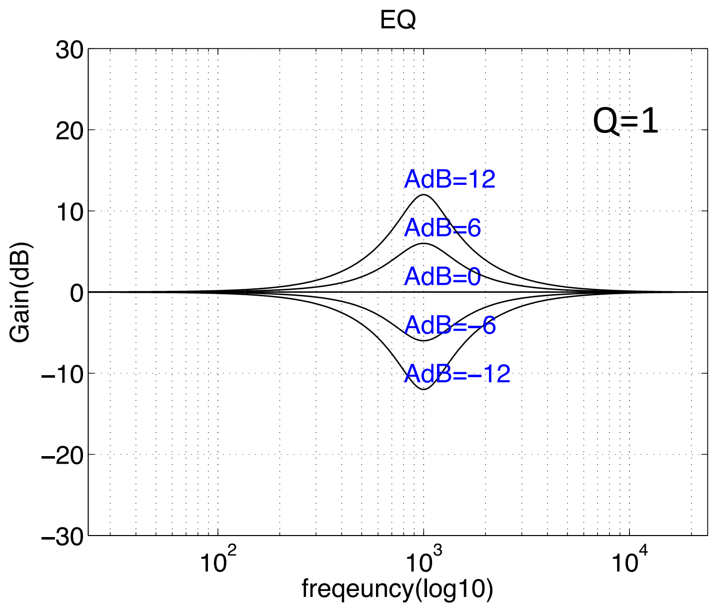


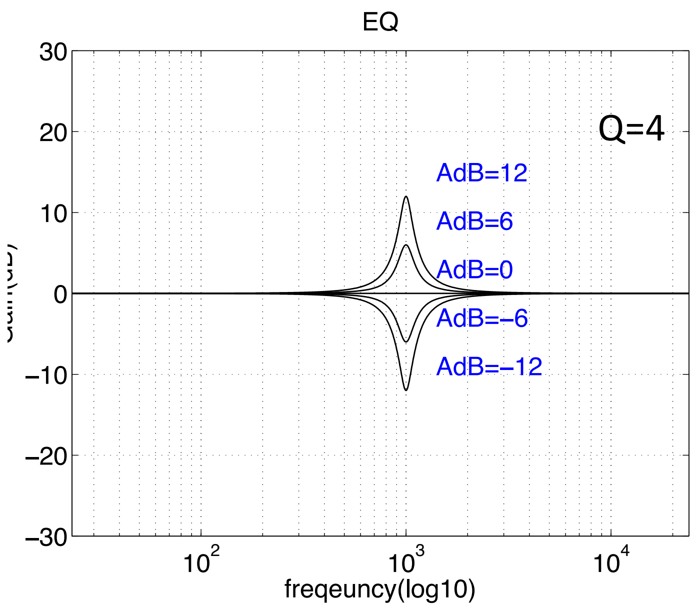


Equalizer

- Transfer Function

$$H(z)=\frac{(1+\alpha\cdot A)-2\cos\Theta z^{-1}+(1+\alpha\cdot A)z^{-2}}{(1+\alpha/A)-2\cos\Theta z^{-1}+(1-\alpha/A)z^{-2}} \qquad \alpha=\frac{\sin\Theta}{2Q} \qquad \Theta=2\pi f_c/f_s$$





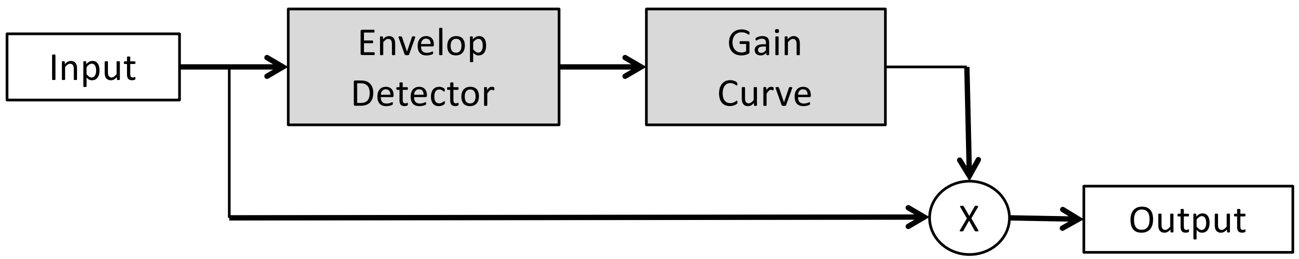
References

- Cookbook formulae for audio EQs based on biquad filter (R. Bristow-Johnson)
  - http://www.musicdsp.org/files/Audio-EQ-Cookbook.txt

Compressor

- Audio effect unit for automatic gain control
  - Boost the level for soft signals and suppress it for loud signals
  - Typically used as a front-end processor in sound recording

- Signal Processing Pipeline



Envelope Detector

- Detecting the level of signal

$$\boxed{\text{Input}} \rightarrow \boxed{\text{Full-wave rectification}} \rightarrow \boxed{\text{Leaky Integrator}} \rightarrow \boxed{\text{envelope}}$$

- Different sensitivity for increasing (attack) and decreasing (release) levels
  - During attack:

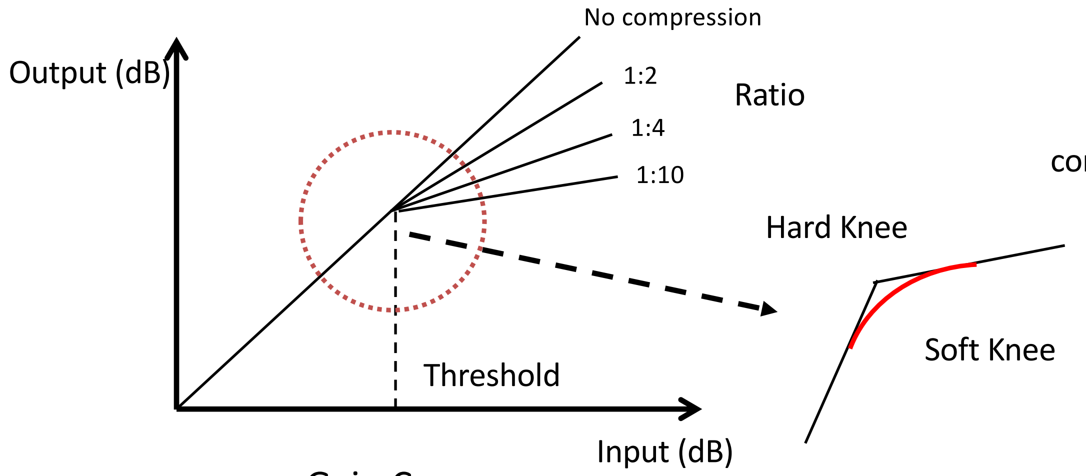
$$y(n) = y(n-1) + (1 - e^{-1/(attack\_time * fs)})(|x(n)| - y(n-1))$$

  - During release:

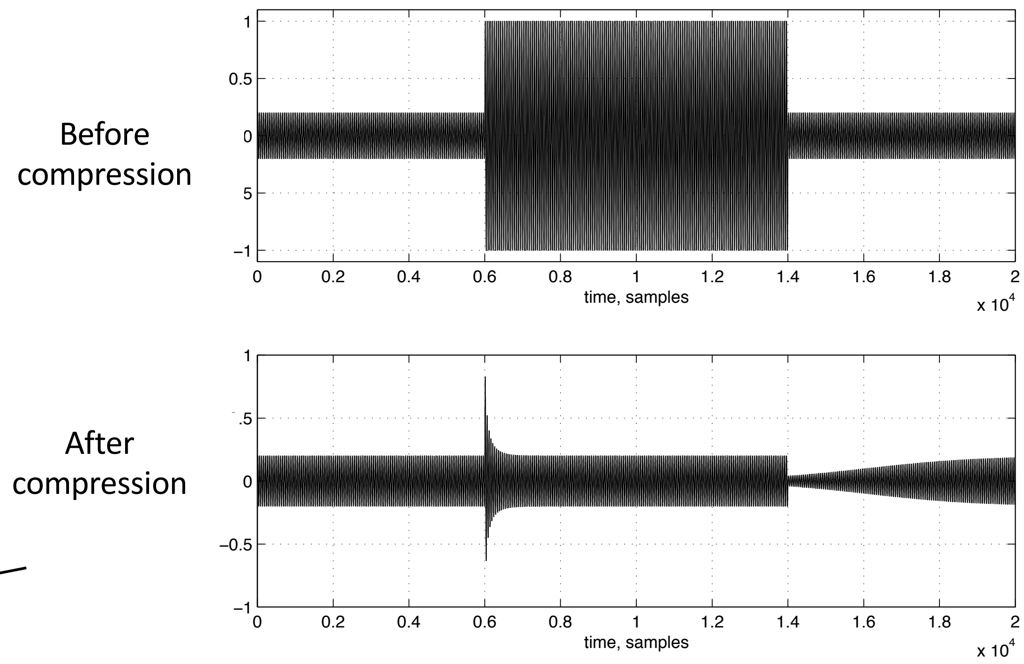
$$y(n) = y(n-1) + (1 - e^{-1/(release\_time * fs)})(|x(n)| - y(n-1))$$

Gain Curve

- Parameters
  - Threshold: level
  - Attack/Release: sensitivity
  - Ratio: amount of compression
  - Knee: smoothing



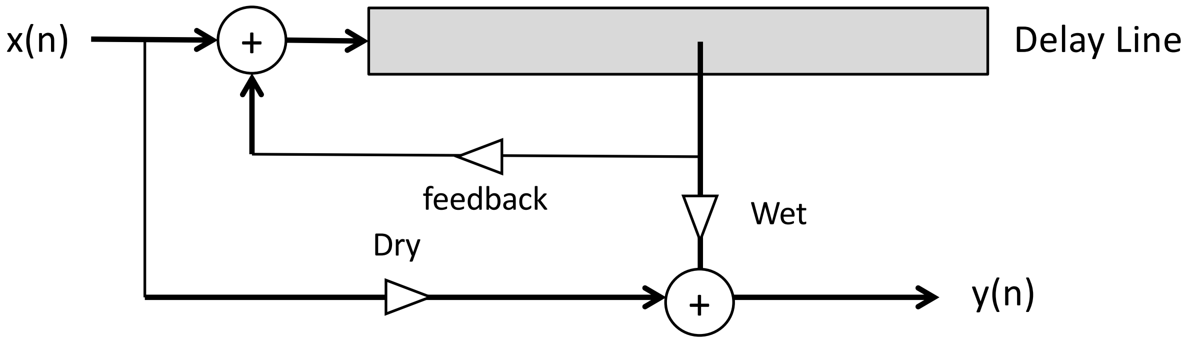
Gain Curve



Delay-based Audio Effects

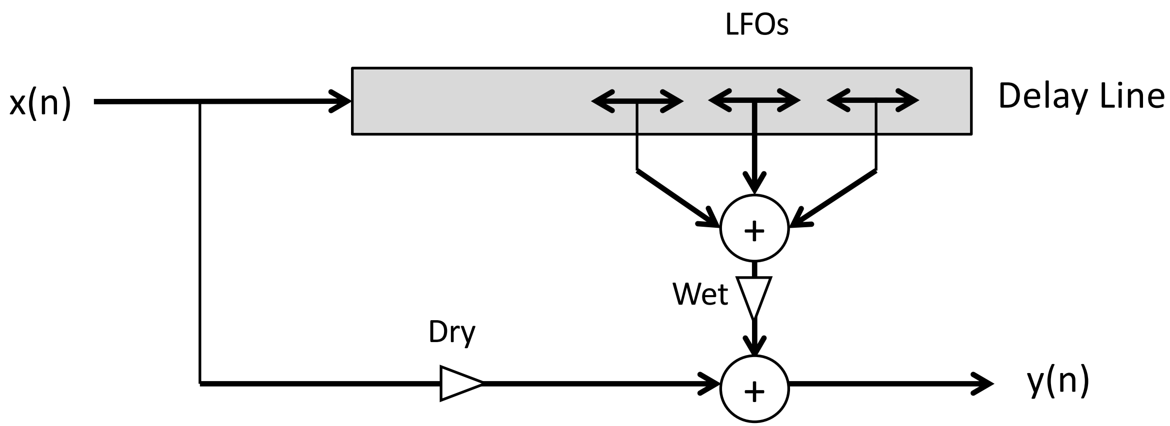
- Types of delay-based audio effect
  - Delay
  - Chorus
  - Flanger
  - Reverberation

Delay



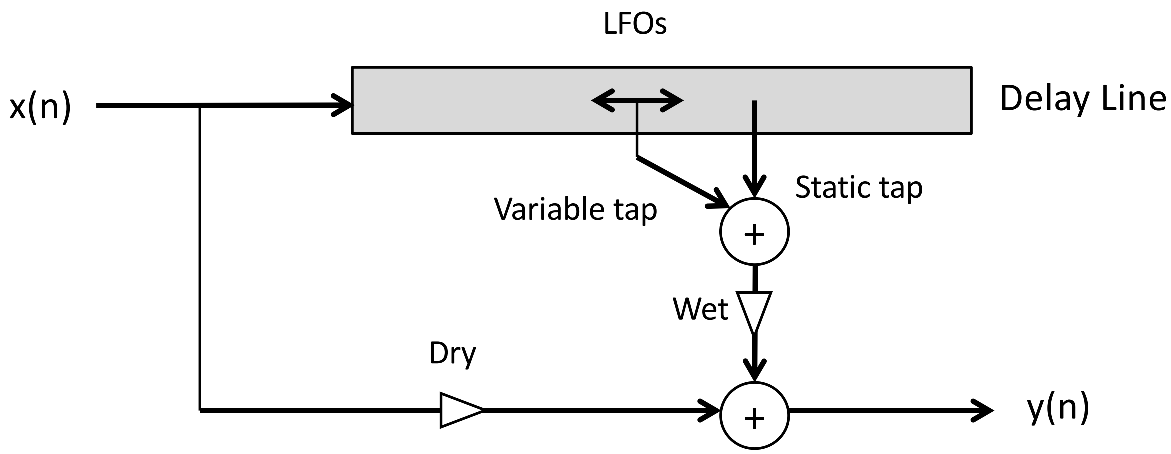
- Delay effect
  - Generate repetitive loop delay
  - Feedback coefficient controls the amount of delayed input
  - Can be extended to stereo signals such that the delay output is "ping-ponged" between the left and right channels
  - The delay length is often synchronized with music tempo
  - The delayline is implemented as a "circular buffer"

Chorus



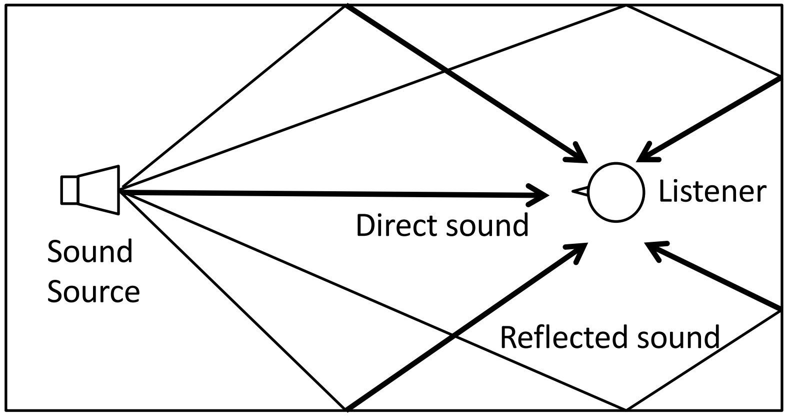
- Chorus effect
  - Gives the illusion of multiple voices playing in unison
  - By summing detuned copies of the input
  - Low frequency oscillators are used to modulate the position of output tops → This causes the pitch of the input (resampling!)

Flanger



- Flanger effect
  - Originally generated by summing the output of two un-locked tape machines while varying their sync (used to be called “reel-flanging”)
  - Emulated by summing one static tap and variable tap in the delay line
    - Feed-forward combine filter where harmonic notches vary over frequency.
  - LFO is often synchronized with music tempo

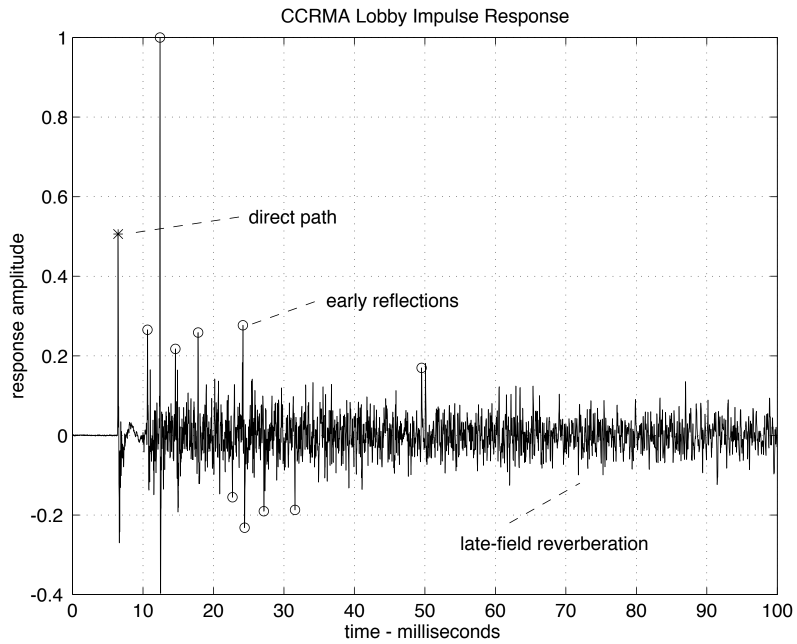
Reverberation



- Natural acoustic phenomenon that occurs when sound sources are played in a room
  - Thousands of echoes are generated as sound sources are reflected against wall, ceiling and floors
  - Reflected sounds are delayed, attenuated and low-pass filtered: high-frequency component decay faster
  - The patterns of myriads of echoes are determined by the volume and geometry of room and materials on the surfaces

Reverberation

- Room reverberation is characterized by its impulse response (IR)
  - E.g. when a balloon pop is used as a sound source
- The room IR is composed of three parts
  - Direct path
  - Early reflections
  - Late-field reverberation: high echo density
- RT60
  - The time that it takes the reverberation to decay by 60 dB from its peak amplitude



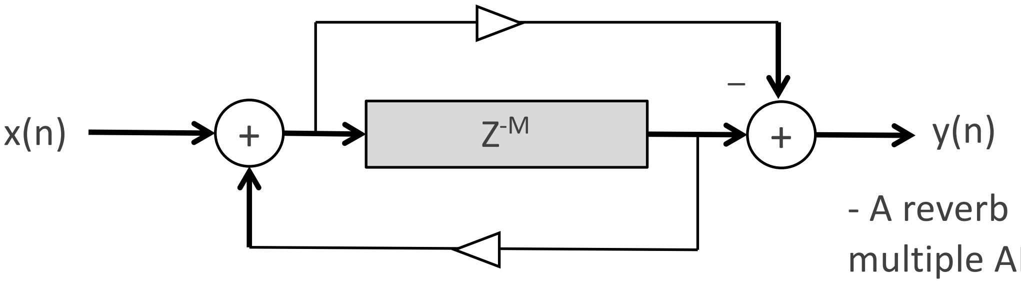
Artificial Reverberation

- Mechanical reverb
  - Use metal plate and spring
  - Plate reverb: https://www.youtube.com/watch?v=XJ5OFpvX5Vs

- Delayline-based reverb
  - Early reflections: feed-forward delayline
  - Late-field reverb: allpass/comb filter, feedback delay networks (FDN)
  - “Programmable” reverberation

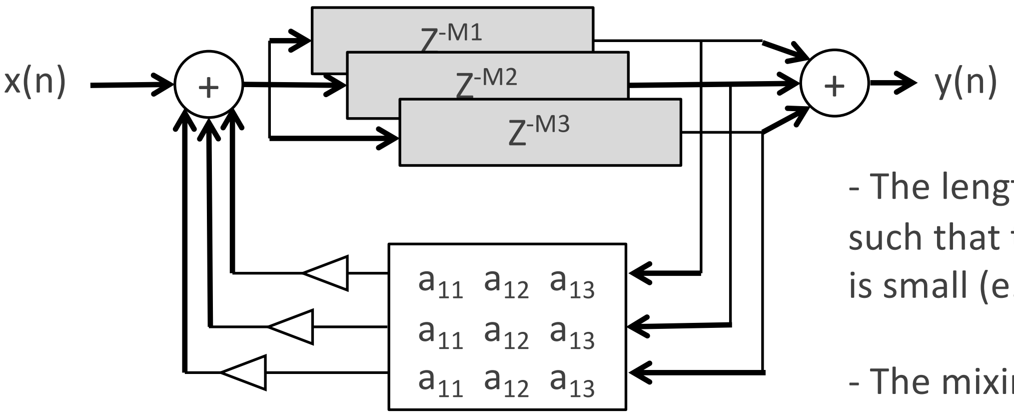
- Convolution reverb
  - Measure the impulse response of a room
  - Do convolution input with the measured IR

Delay-based Reverb



- A reverb is constructed by cascading multiple AP or FFCF units

AllPass filter / Comb filter (when one tap is absent)



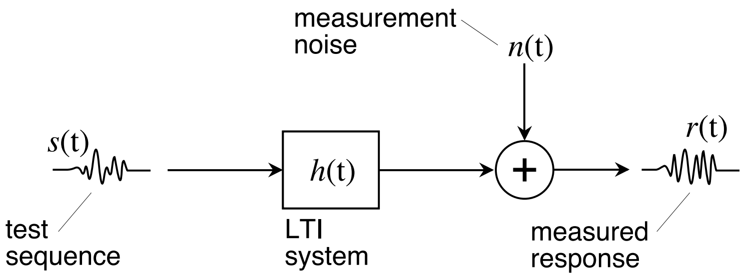
- The lengths of delaylines are chosen such that their greatest common factors is small (e.g. prime numbers)

- The mixing matrix is chosen to be unitary (orthonormal)

Feedback Delay Networks

Convolution Reverb

- Measuring impulse responses



$$r(t) = s(t) * h(t) + n(t)$$

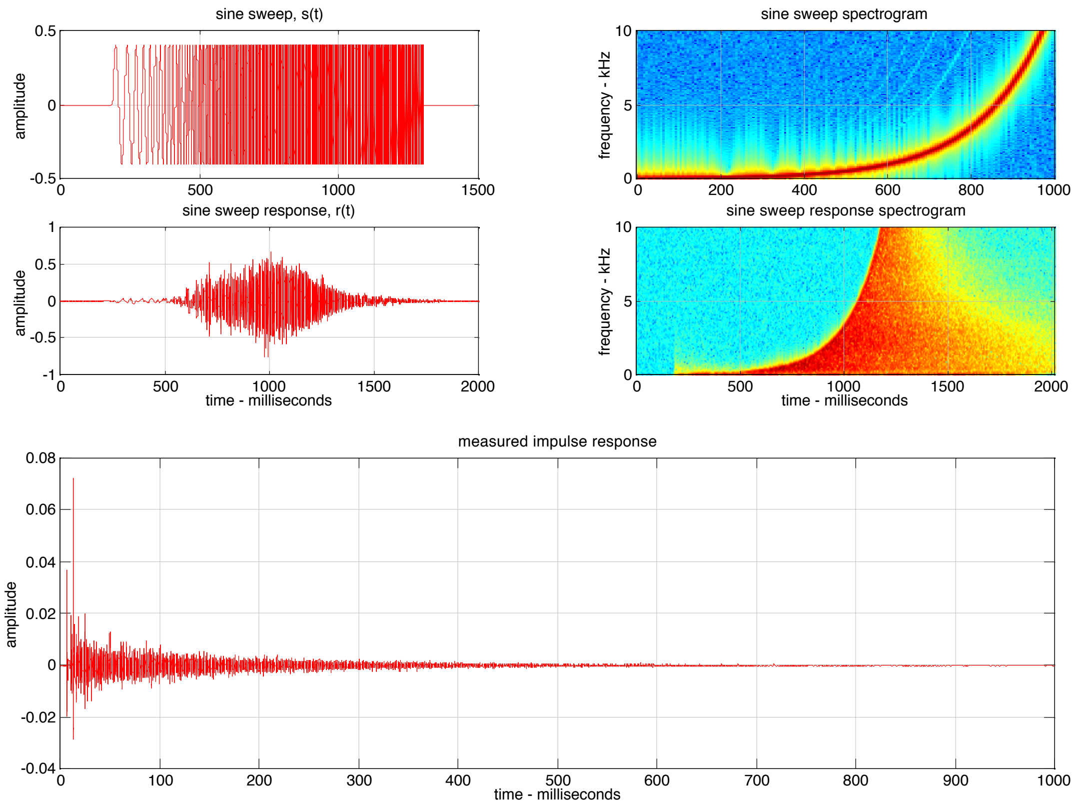
  - If the input is a unit impulse, SNR is low
  - Instead, we use specially designed input signals
    - Golay code, allpass chirp or sine sweep: their magnitude responses are all flat but the signals are spread over time
  - The impulse response is obtained using its inverse signal or inverse discrete Fourier transform

Convolution Reverb

$s(t)$

$r(t)$

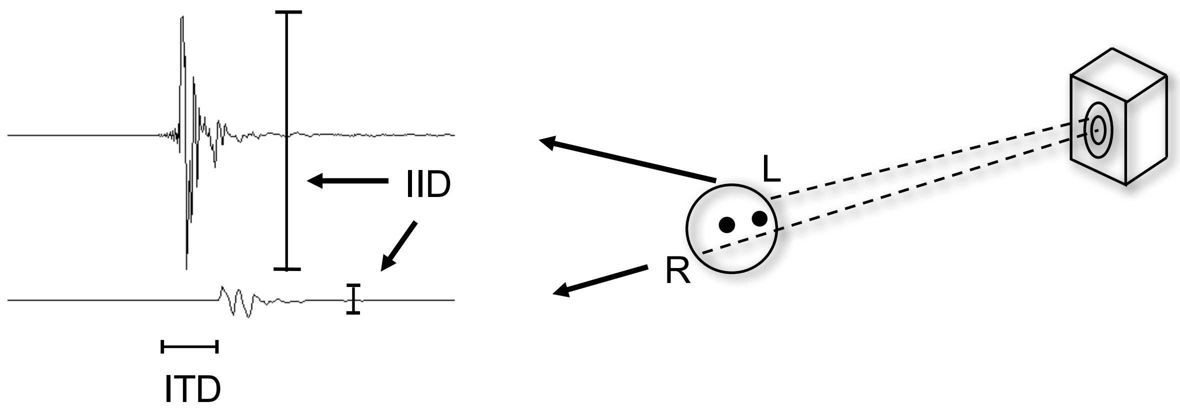
$\hat{h}(t)$



(J. Abel)

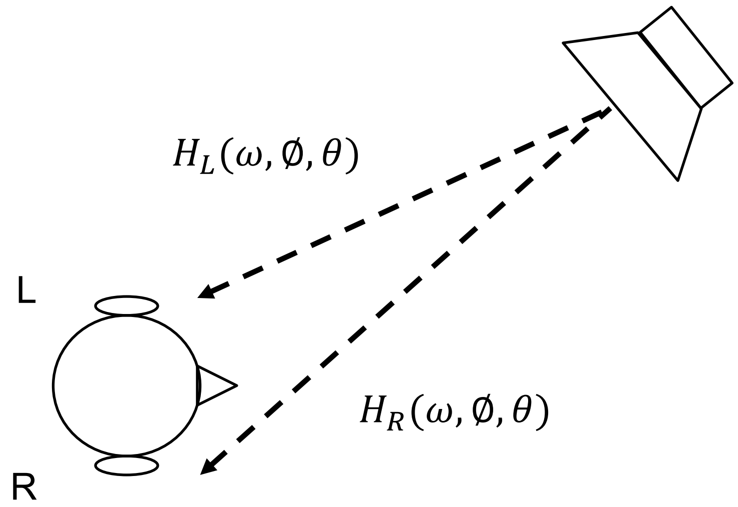
Spatial Hearing

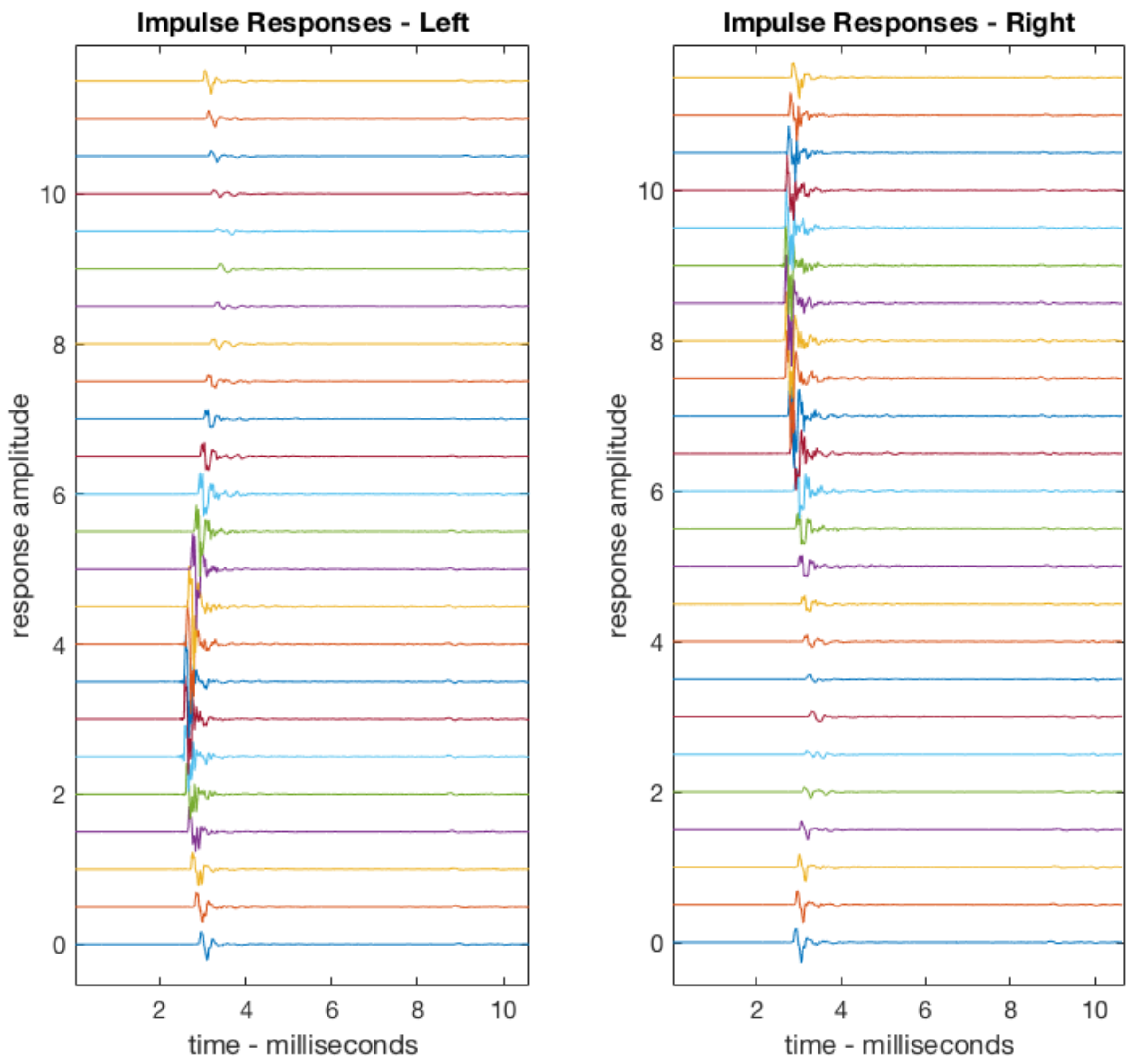
- A sound source arrives in the ears of a listener with differences in time and level
  - The differences are the main cues to identify where the source is.
  - We call them **ITD** (Inter-aural Time Difference) and **IID** (Inter-aural Intensity Difference)
  - ITD and IID are a function of the arrival angle.



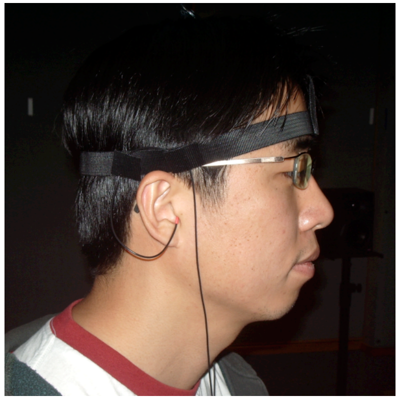
Head-Related Transfer Function (HRTF)

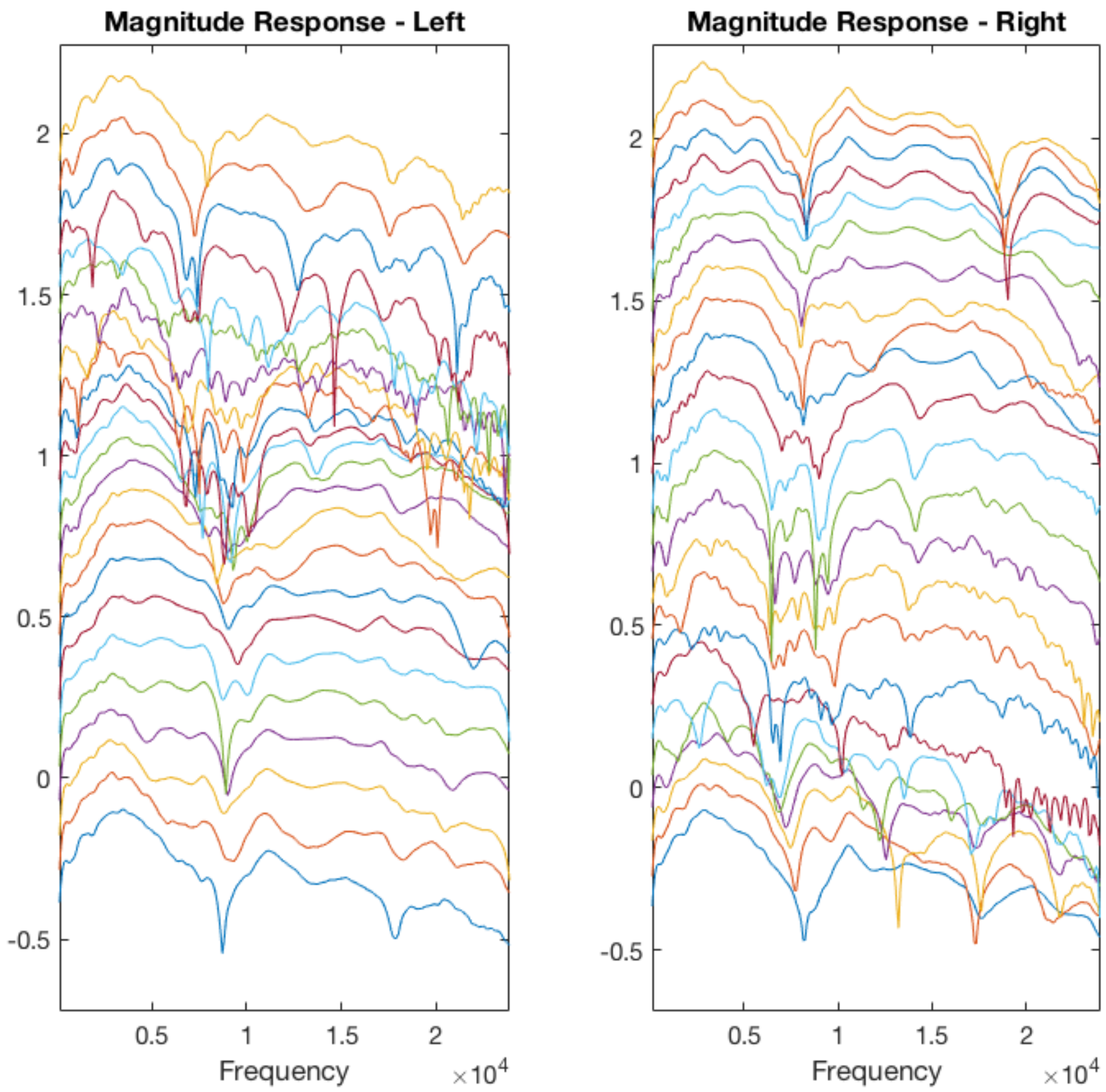
- A filter measured as the frequency response that characterizes how a sound source arrives in the outer end of ear canal
  - Determined by the refection on head, pinnae or other body parts
  - Function of azimuth (horizontal angle) and elevation (vertical angle)





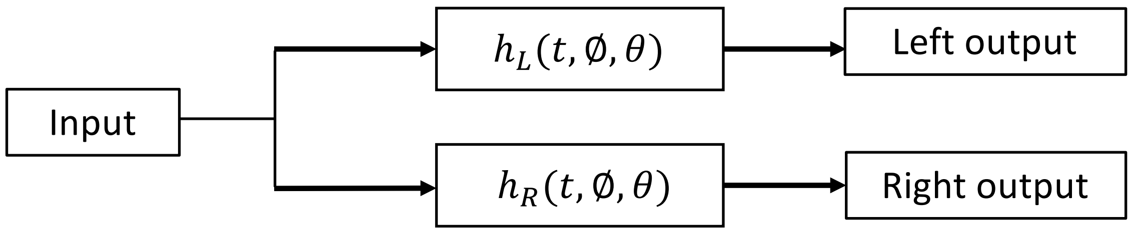
Measured
Head-Related
Impulse Responses





Magnitude response of the HRIRs

Binaural Synthesis



- Rendering the spatial effect using the measured HRIRs as FIR filters
  - HRIRs are typically several hundreds sample long
  - Convolution or modeling by IIR filters

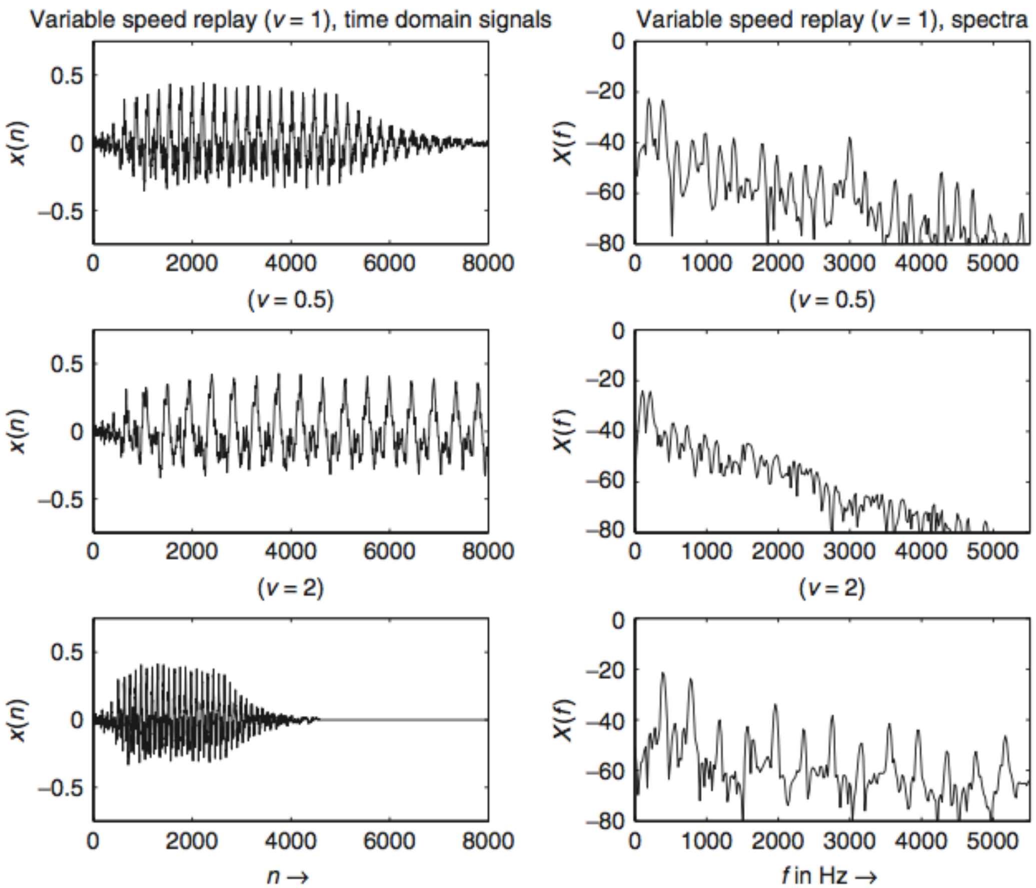
- Individualization of HRTF is a issue

Playback Rate Conversion

- Adjusting playback rate given the sampling rate
  - Analogy to sliding tapes on the magnetic header in a variable speed
  - Speeding down: “monster-like”
  - Speeding up: “chipmunk-like”

Playback Rate Conversion

- Change pitch, length and timbre



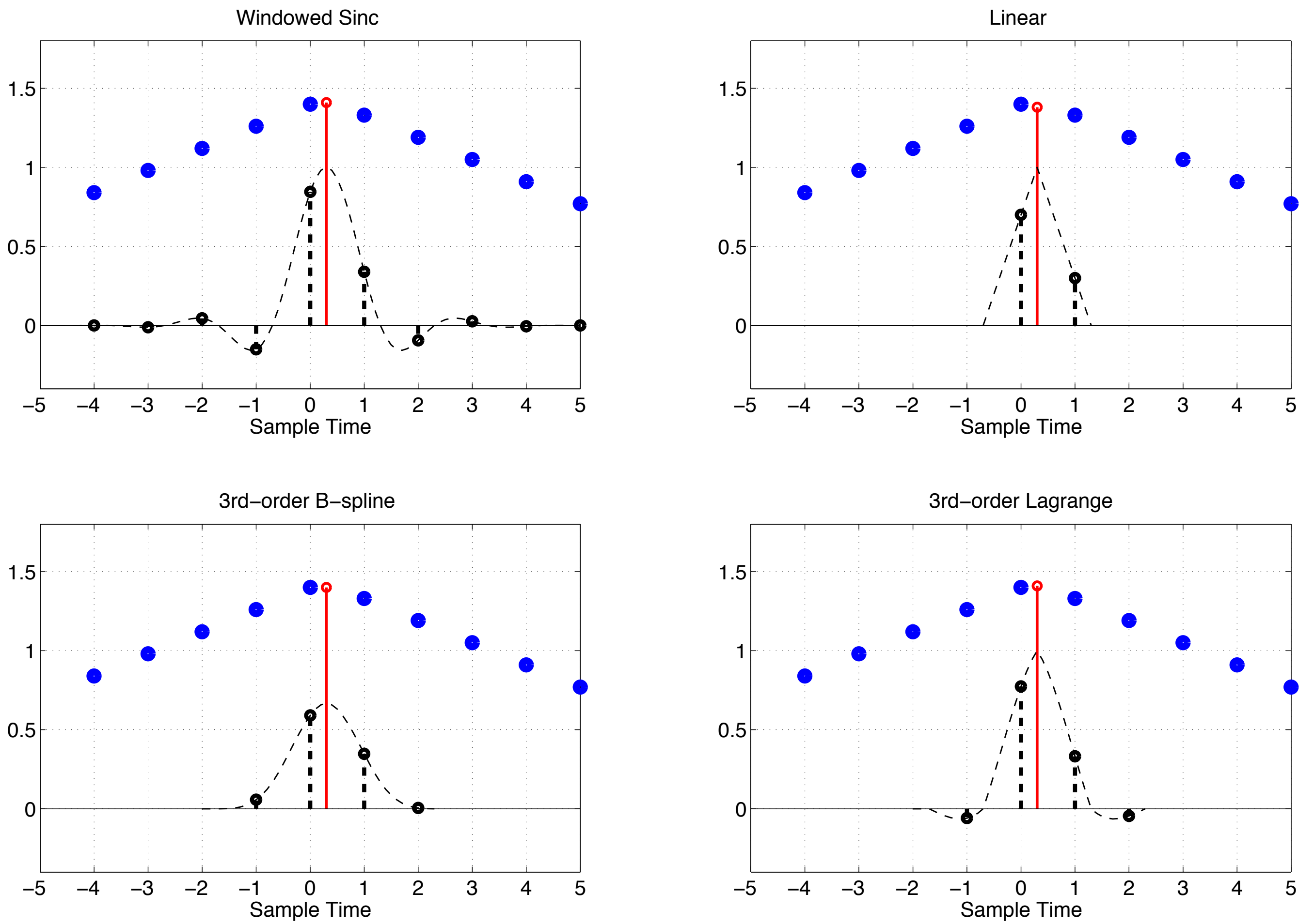
[The DaFX book]

Resampling

- Playback rate conversion is performed by resampling
  - Interpolation on discrete samples
  - Convolution with interpolation filters
  - Need to avoid aliasing for down sampling
    - Narrowing the bandwidth of the lowpass filter

- Two Types
  - Down-sampling: pitch goes up and time shrinks
  - Up-sampling: pitch goes down and time expands

Interpolation Filters



$$h(t) = \mathrm{w}(t)\mathrm{sinc}(t) = \mathrm{w}(t)\frac{\sin(\pi t)}{\pi t}$$

Delayed by d ($0 < d < 1$)

$$x(d) = \sum_{k=-(L-1)}^{k=L} x(k)h(d-k)$$