GCT535: Sound Technology for Multimedia

Spatial Audio

KAIST Graduate School of Culture Technology

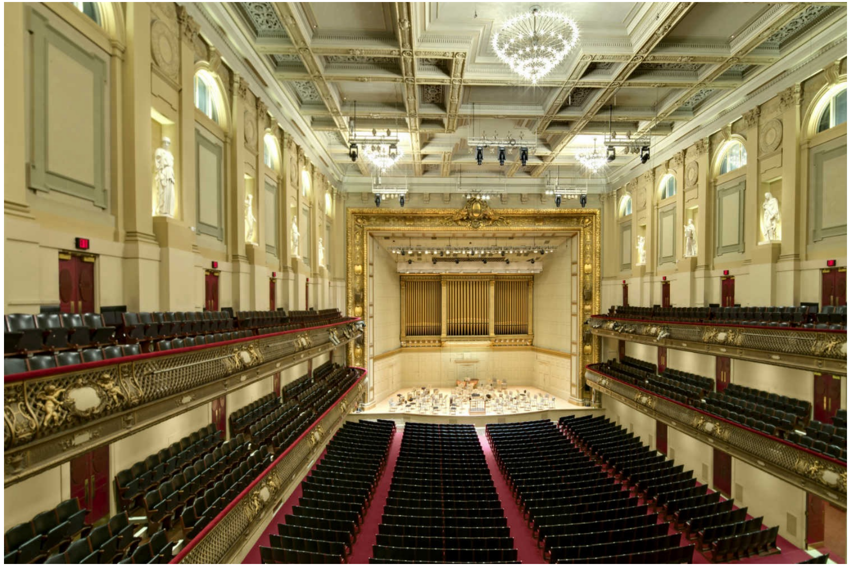
Juhan Nam

Outlines

- Localization using HRTF
- Reverberation

Hearing Sound in Space

- We have different auditory experiences depending on where we listen to sound



Boston Symphony Hall

Lotte Concert Hall

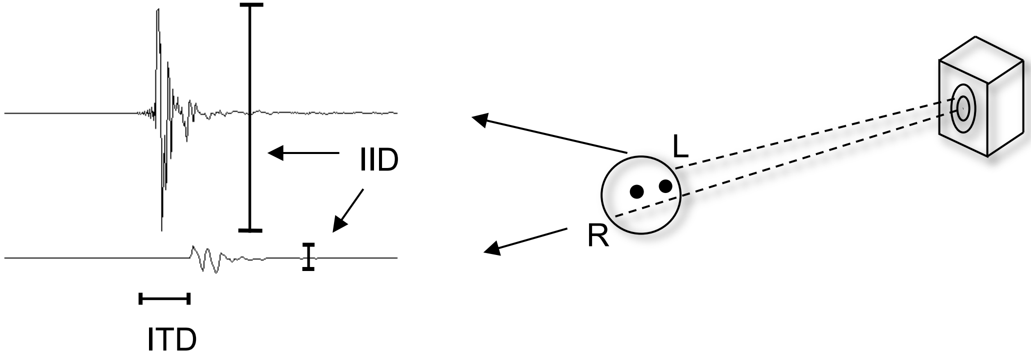
Hearing Sound in Space

- Sound Localization
  - Hear sound as a point source in 3-D space
  - Possible with two ears and human body structure
  - Panning (stereo), 3-D Sound

- Room Effect
  - Hear sound and its myriads of reflections
  - Determined by the room type: size, structure and materials
  - Reverberation effect

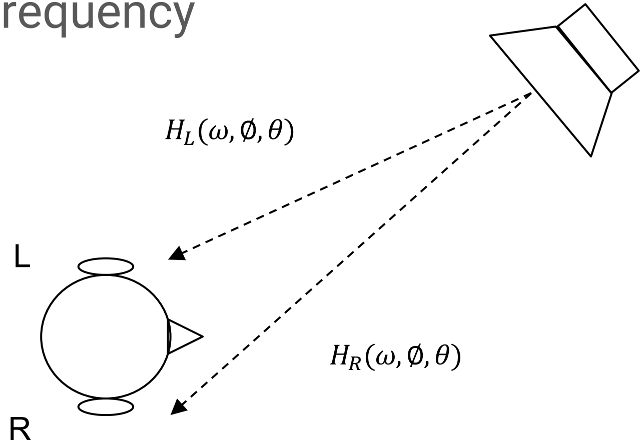
Sound Localization

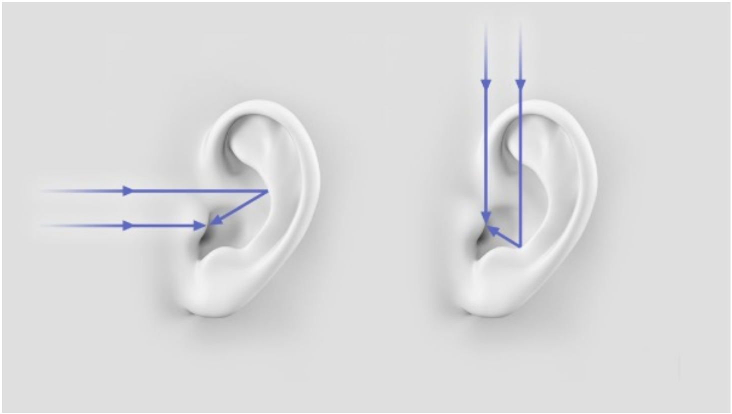
- Perception of Directions
  - A sound source arrives in each of two ears with differences in time and level
  - We call them **ITD** (inter-aural Time Difference) and **IID** (Inter-aural Intensity Difference)
  - IID is used as a main cue of direction above about 1.5 kHz
- Only the level and time differences are enough?



Head-Related Transfer Function (HRTF)

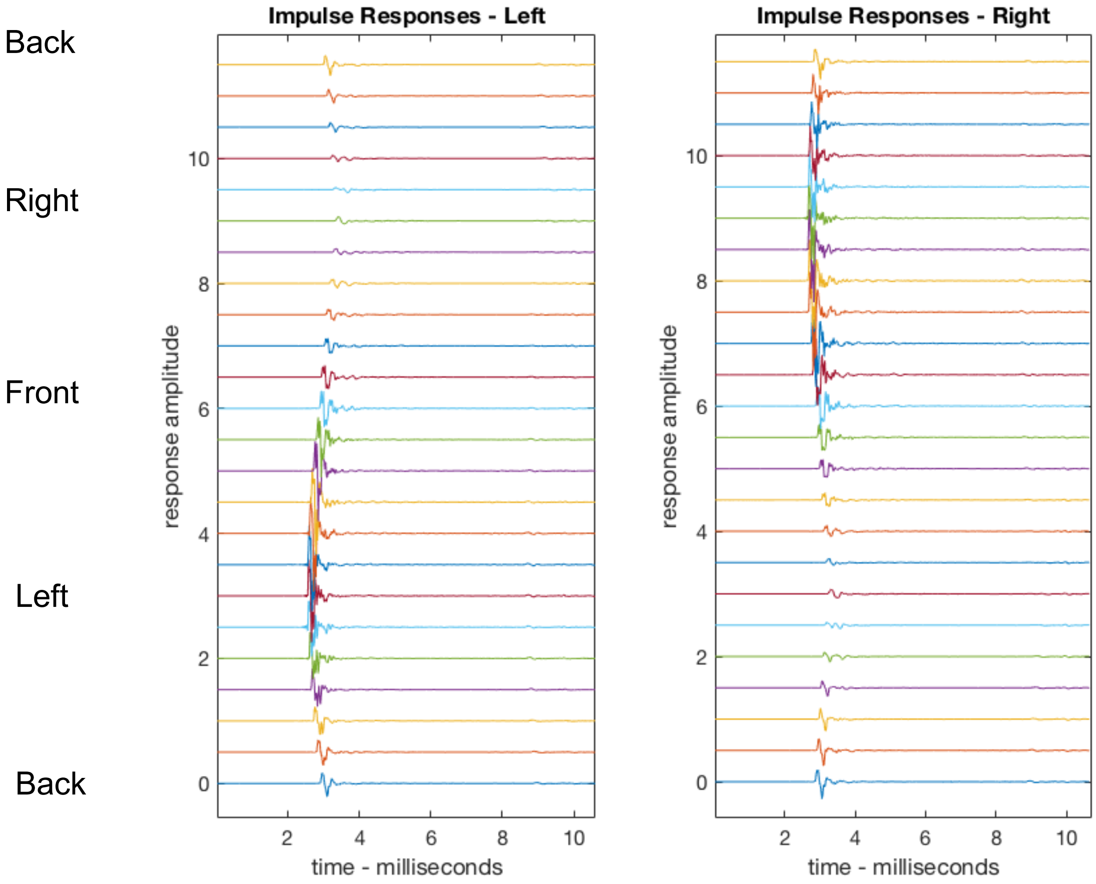
- A filter that characterizes how a sound arrives in the outer end of ear canal from the source
  - Measured from human body or dummy head
  - Determined by the refection on head, pinnae, torso or other body parts
  - Function of azimuth (horizontal angle), elevation (vertical angle) and frequency





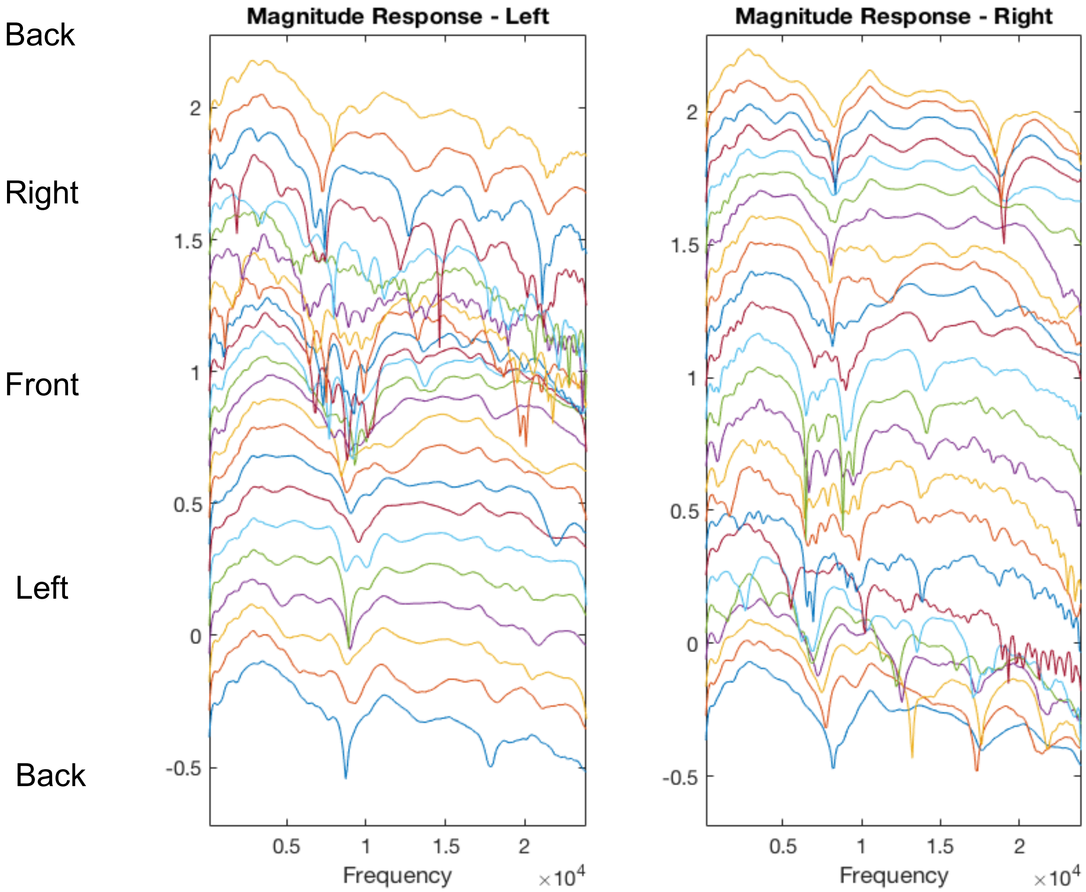
Source: https://developers.google.com/vr/concepts/spatial-audio

HRTF Measurement



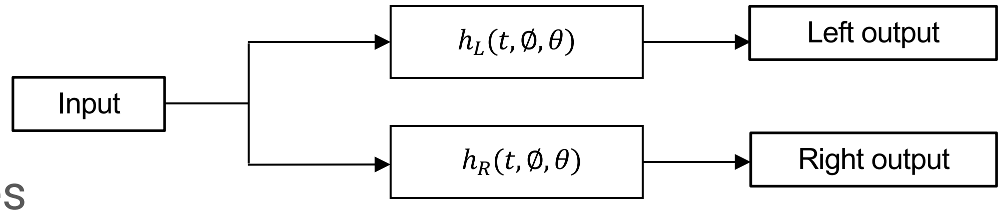
Measured
Head-Related
Impulse Responses

HRTF Measurement



Magnitude response of the HRIRs

Rendering Sound in 3-D Space



- Binaural Synthesis
  - 3D sound effects
  - Typically run with headphones
  - Applications: Game, VR
- Methods
  - Convolution with HRTFs: the IRs are typically several hundreds sample long
  - Modeling HRTFs with biquad filters (e.g. Prony's method)
- Issues
  - Standardization: dummy head
  - Individualization

Head-Related IR Datasets

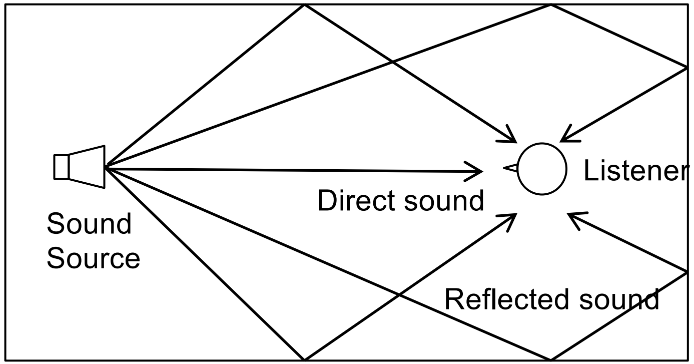
- ARI
  - 85 subjects and 1550 source positions
  - https://www.kfs.oeaw.ac.at/index.php?option=com_content&view=article&id=608&Itemid=857&lang=en
- CIPIC
  - 45 subjects and 1250 source positions
  - http://interface.cipic.ucdavis.edu/sound/hrtf.html
- IRCAM
  - 50 subjects and 187 source positions
  - http://recherche.ircam.fr/equipes/salles/listen/

HRTF Demos

- Virtual Barber Shop Hair Cut
  - https://www.youtube.com/watch?v=IUDTlvagjJA
- OpenAL Example
  - https://www.youtube.com/watch?v=tY9DhuEe1WY
- Google Chrome Omnitone
  - http://googlechrome.github.io/omnitone
- My Research Work
  - https://ccrma.stanford.edu/~juhan/threedee.html

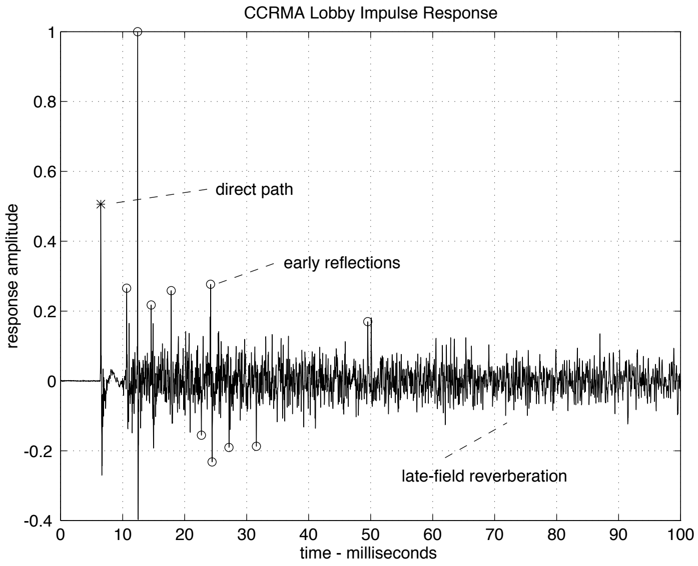
Reverberation

- Acoustic phenomenon when a sound source is played in a room
  - Thousands of echoes are reflected against wall, ceiling and floors
  - The patterns are determined by the volume and geometry of the room and materials on the surfaces
  - We can recognize the geometry and composition of the room from the sound
  - They provide different (often better) feelings of the sound



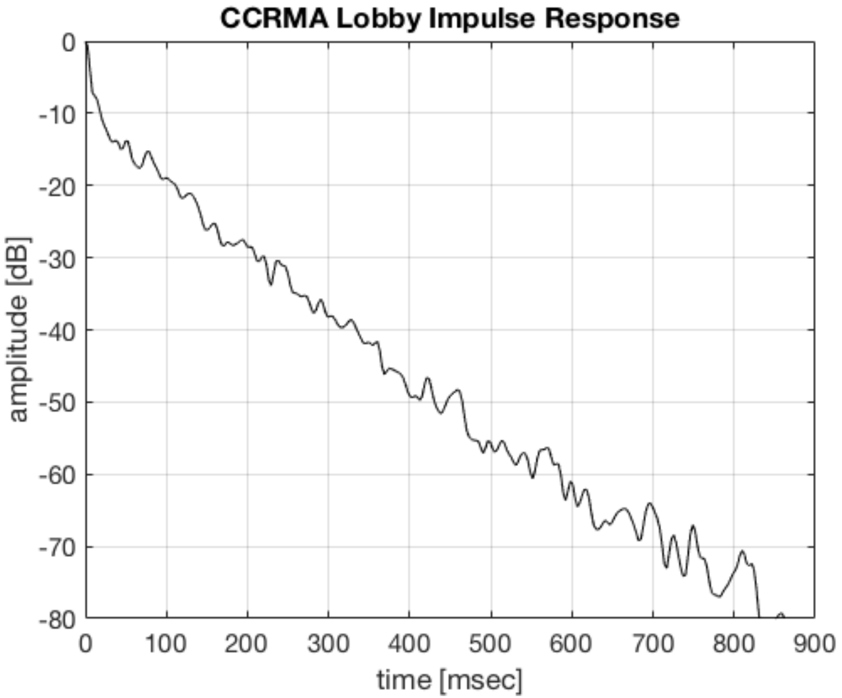
Room Impulse Response

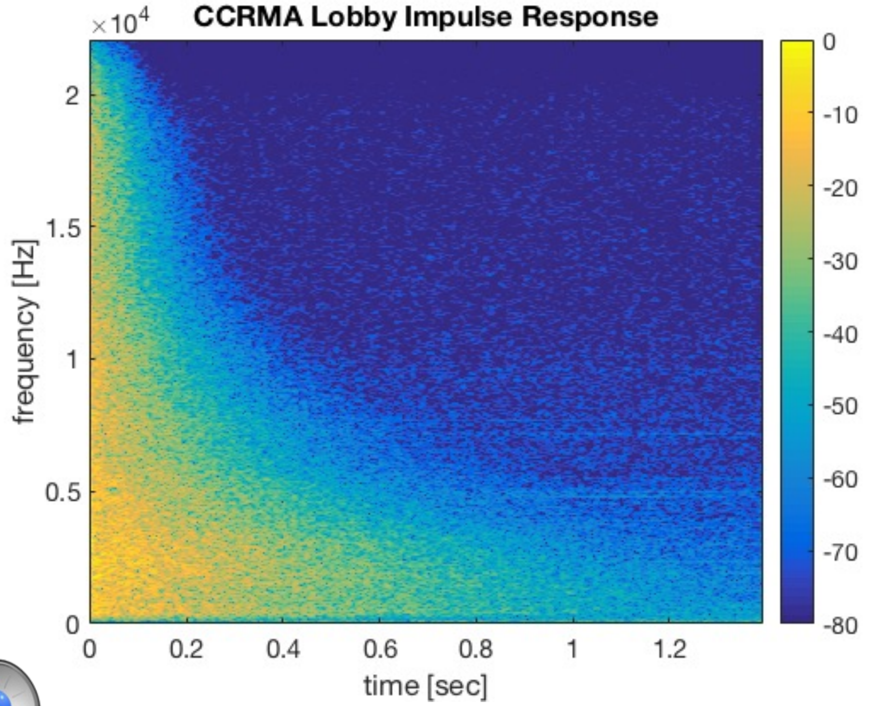
- Room reverberation is characterized by its impulse response (IR)
- The room IR is composed of three parts
  - Direct path
  - Early reflections: convey a sense of the room geometry and size
  - Late-field reverberation: high echo density like noise, determined by room size and materials
- RT60



Room Impulse Response

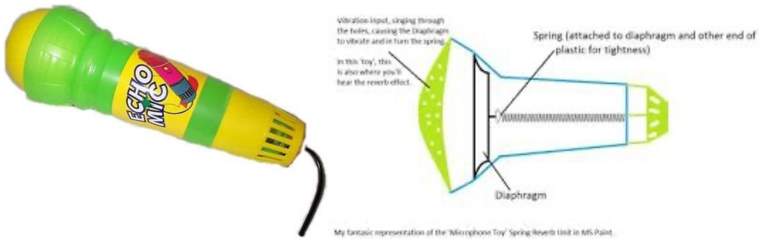
- Energy Envelope and Spectrogram





Artificial Reverberation

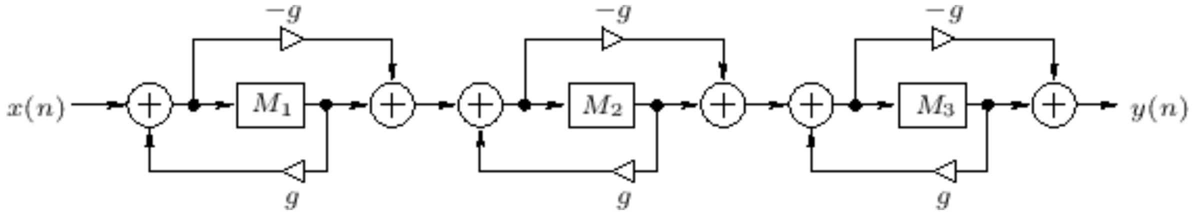
- Convolution reverb
    - Measure the impulse response of a room
    - Convolve input with the measured IR



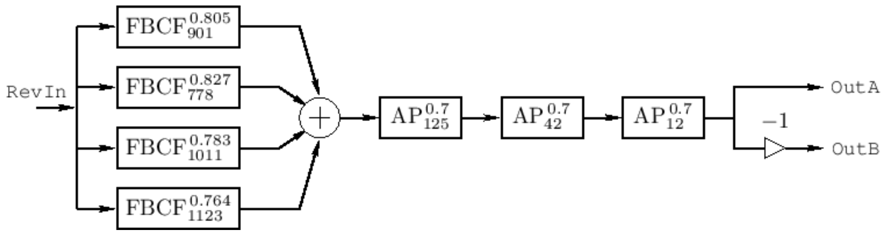
- Mechanical reverb
    - Use metal plate and spring
    - EMT140 Plate Reverb: https://www.youtube.com/watch?v=HEmJpxCvp9M
- Delay-based reverb
    - Early reflections: feed-forward delayline
    - Late-field reverb: allpass/comb filter, feedback delay networks (FDN)
    - “Programmable” reverberation

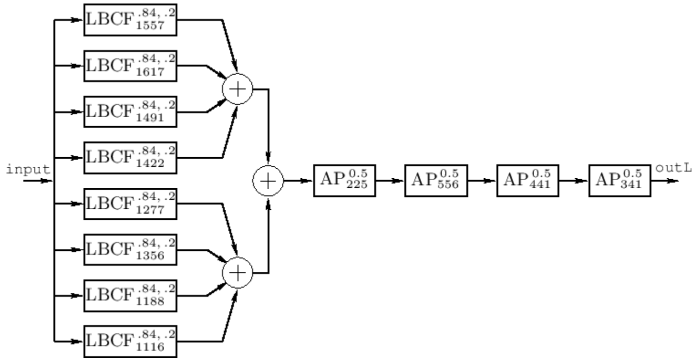
Delay-based Reverb

- Schroeder model
  - Cascade of allpass-comb filters
  - Mutually prime number for delay lengths



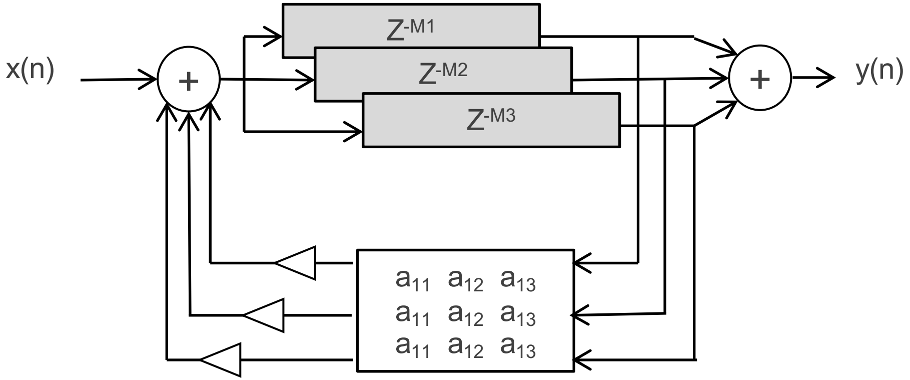
- Variations





Delay-based Reverb

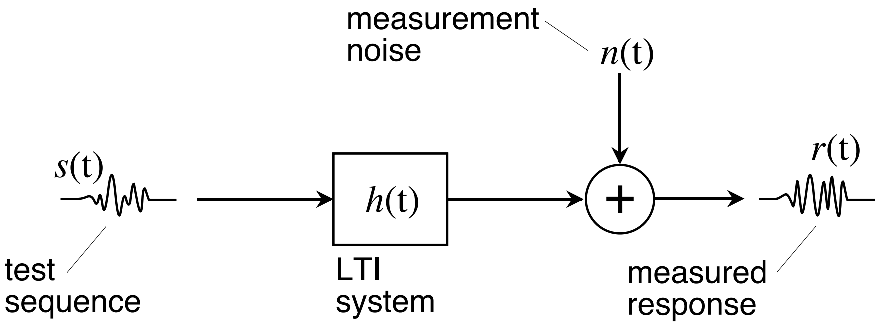
- Feedback Delay Networks
  - Mixing matrix creates “good spreading” of delayed outputs
    - Chosen to be orthonormal (unitary matrix)
  - The lengths of delaylines are chosen to be mutually prime number
  - Should generate a white noise in lossless mode
  - T60 is controlled by the loop gains



Feedback Delay Networks

Measuring Impulse Response

- Measurement Model
  - Assume the system as linear time-invariant
  - Use a test signal and the output to derive the impulse response



$$r(t) = \delta(t) * h(t) + n(t)$$

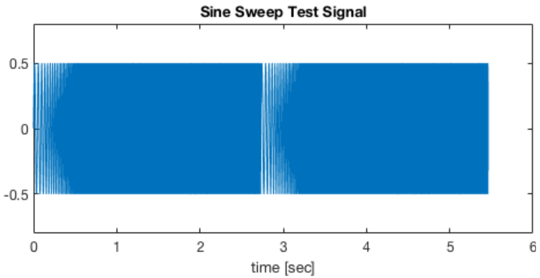
$$r(t) \rightarrow \hat{h}(t)$$

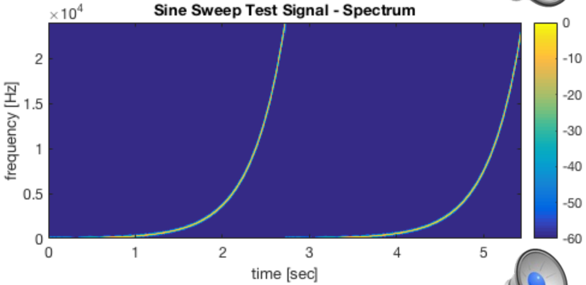
  - Using a sine sweep: based on the convolution theorem

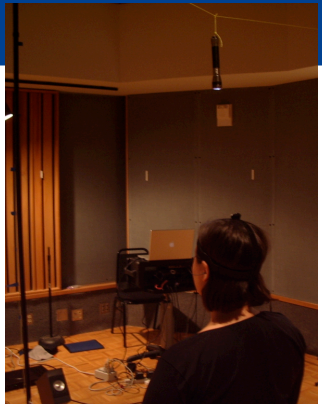
$$\hat{h}(t) = \mathrm{FFT}^{-1}\{\frac{\mathrm{FFT}\,\{r(t)\}}{\mathrm{FFT}\{s(t)\} + \varepsilon(f)}\}$$

Measuring HRTFs

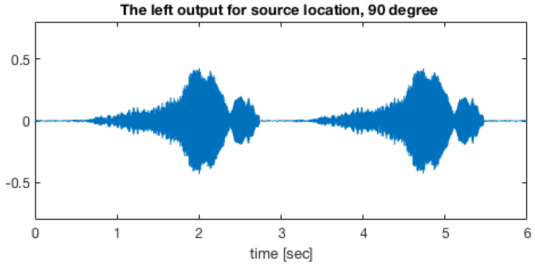
$s(t)$

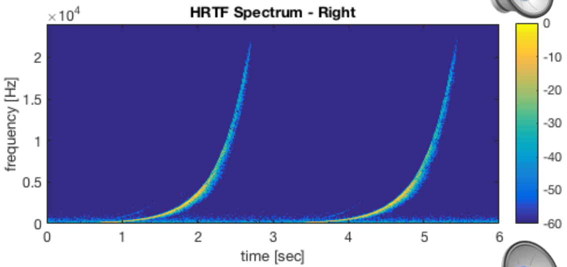




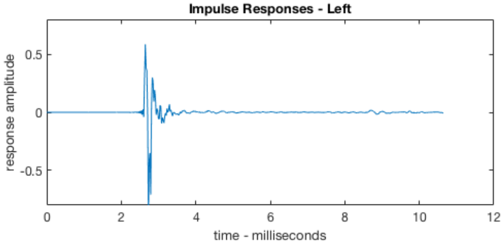


$r_L(t)$

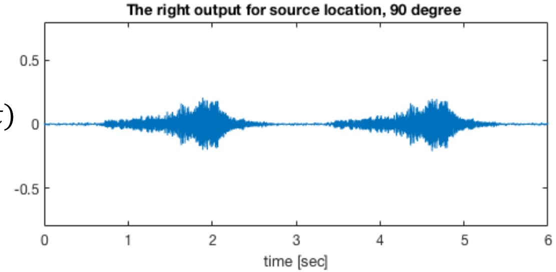


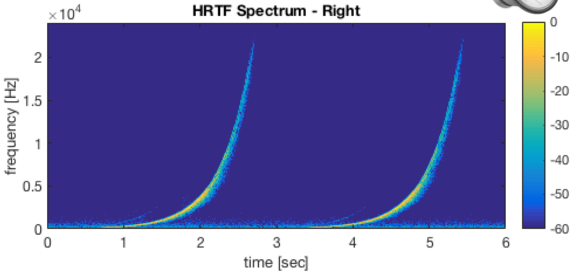


$\hat{h}_L(t)$

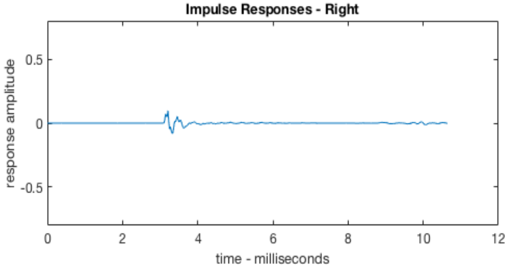


$r_R(t)$





$\hat{h}_R(t)$

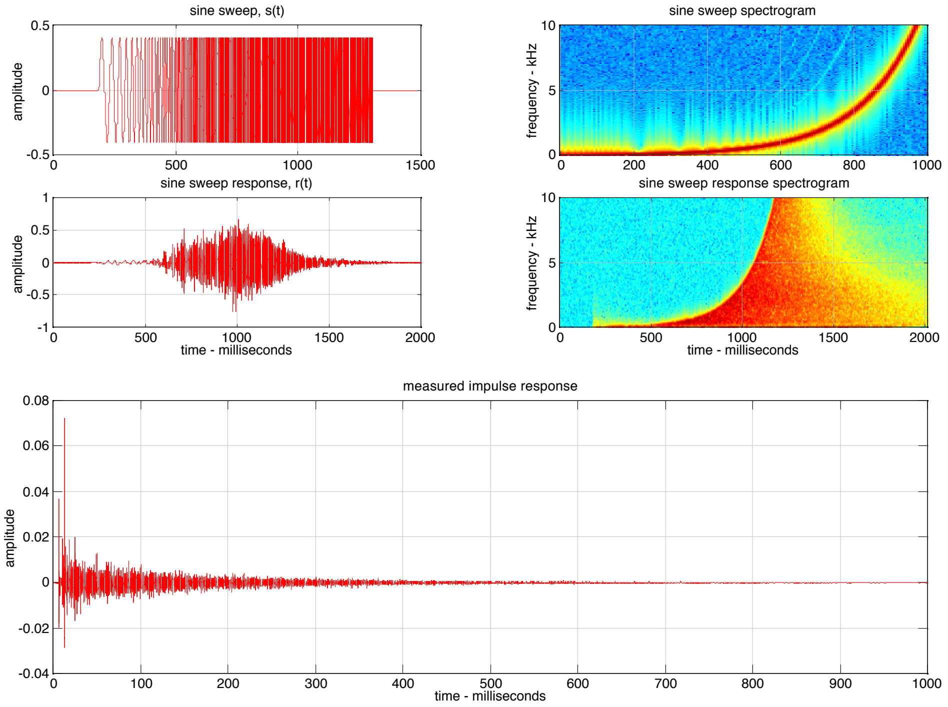


Measuring Room IRs

$s(t)$

$r(t)$

$\hat{h}(t)$



(J. Abel)

Room IR datasets

- Open AIR
  - http://www.openairlib.net/
- Aachen Impulse Response Database
  - http://www.iks.rwth-aachen.de/en/research/tools-downloads/databases/aachen-impulse-response-database/

Convolution by FFT

- Direct convolution in time domain is computationally expensive
  - It has O(n^2) in complexity
  - Especially, when the length of IR is long (e.g. room IR)
- FFT Convolution
  - Using Convolution Theorem: $x(n) * h(n) \leftrightarrow Y(k) = H(k)X(k)$
  - FFT has O($n\log_2 n$) and convolution in frequency O(n) in complexity
- However, the issue is latency when the IR is long. How can we implement it in real-time reverb?
  - Solution: **Partitioned Convolution**
  - https://webaudio.github.io/web-audio-api/convolution.html

References

- Reverberation using Feedback Delay Network
  - https://ccrma.stanford.edu/~jos/pasp/FDN_Reverberation.html
- Impulse Response Measurement
  - http://pcfarina.eng.unipr.it/Public/Papers/226-AES122.pdf