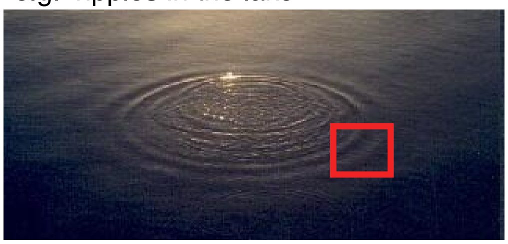
Lecture 8: Spatial sound

Dan Ellis <dpwe@ee.columbia.edu>
http://www.ee.columbia.edu/~dpwe/e6820/

Columbia University Dept. of Electrical Engineering
Spring 2006

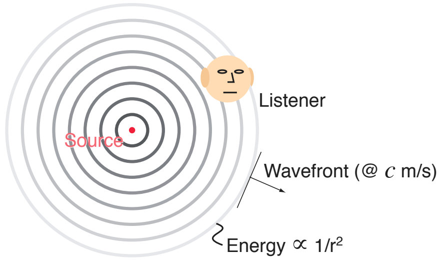
1 Spatial acoustics

- **Received sound = source + channel**
  - so far, only considered ideal source waveform
- **Sound carries information on its spatial origin**
  - e.g. “ripples in the lake”



  - evolutionary significance
- **The basis of scene analysis?**
  - yes and no - try blocking an ear

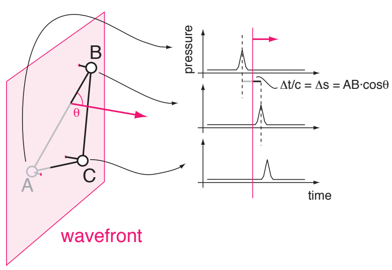
Ripples in the lake



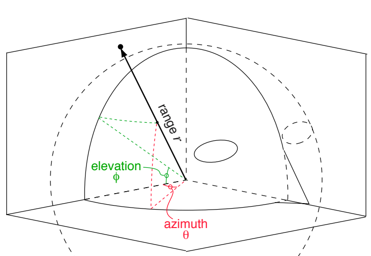
- **Effect of relative position on sound**
  - delay = $\Delta r/c$
  - energy decay ~ $1/r^2$
  - absorption ~ $G(f)^r$
  - direct energy plus reflections
- **Give cues for recovering source position**
- **Describe wavefront by its normal**

Recovering spatial information

- **Source direction as wavefront normal**
  - moving plane found from timing at 3 points



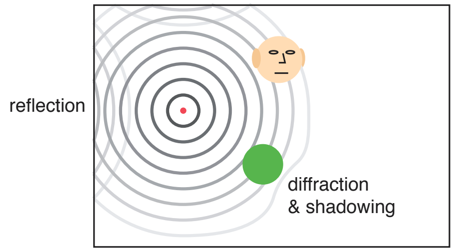
  - need to solve correspondence



- **Space: need 3 parameters**
  - e.g. 2 angles and range

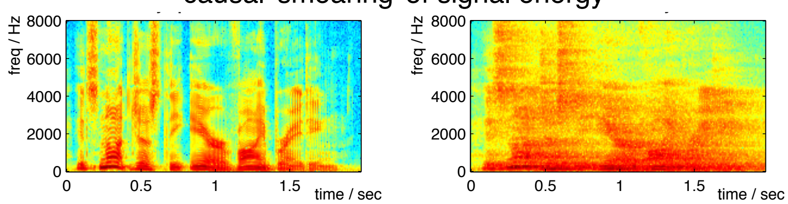
The effect of the environment

- **Reflection causes additional wavefronts**



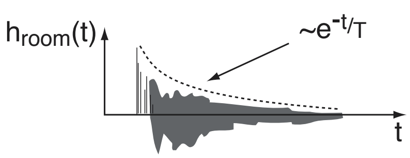
  - + scattering, absorption
  - many paths → many echoes

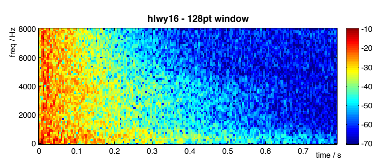
- **Reverberant effect**
  - causal 'smearing' of signal energy



Reverberation impulse response

- **Exponential decay of reflections:**





- **Frequency-dependent**
  - greater absorption at high frequencies → faster decay
- **Size-dependent**
  - larger rooms → longer delays → slower decay
- **Sabine's equation:**

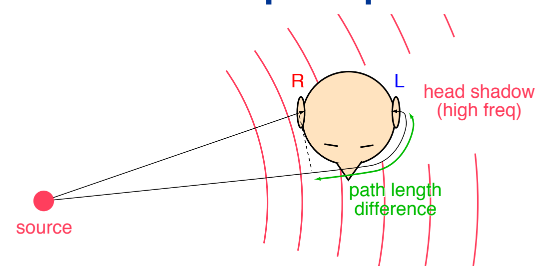
$$RT_{60} = \frac{0.049V}{S\bar{\alpha}}$$

- **Time constant as size, absorption**

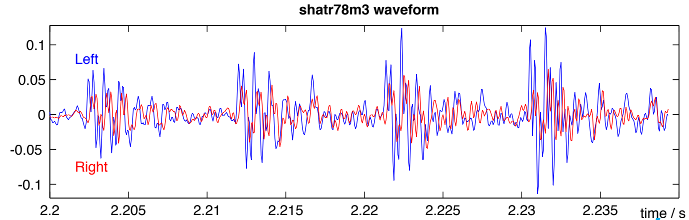
Outline

1. Spatial acoustics
2. **Binaural perception**
   - The sound at the two ears
   - Available cues
   - Perceptual phenomena
3. Synthesizing spatial audio
4. Extracting spatial sounds

2 Binaural perception

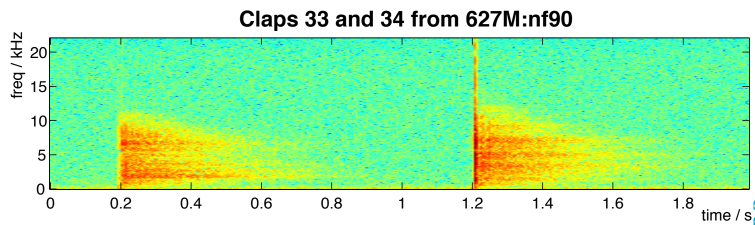


- **What is the information in the 2 ear signals?**
  - the sound of the source(s) (L+R)
  - the position of the source(s) (L-R)
- **Example waveforms (ShATR database)**



Main cues to spatial hearing

- **Interaural time difference (ITD)**
  - from different path lengths around head
  - dominates in low frequency (< 1.5 kHz)
  - max ~ 750 μs → ambiguous for freqs > 600 Hz
- **Interaural intensity difference (IID)**
  - from head shadowing of far ear
  - negligable for LF; increases with frequency
- **Spectral detail (from pinna relfections) useful for elevation & range**
- **Direct-to-reverberant useful for range**

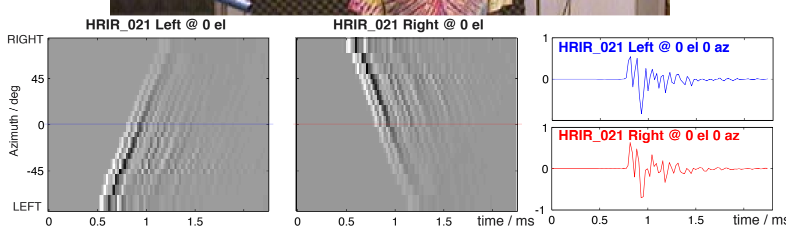


Head-Related Transfer Fns (HRTFs)

- **Capture source coupling as impulse responses**

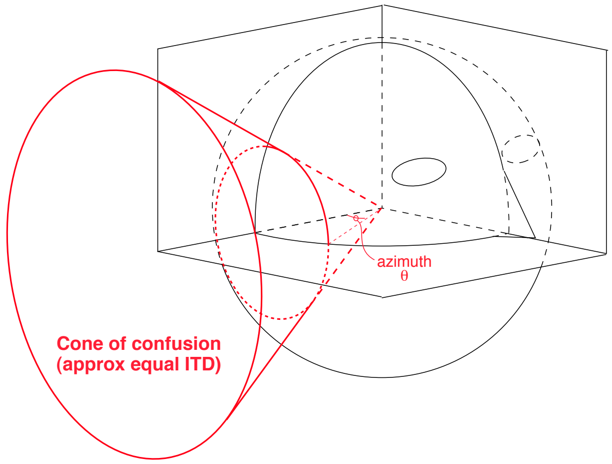
  $\{l_{\theta, \phi, R}(t), r_{\theta, \phi, R}(t)\}$

- **Collection:** (http://interface.cipic.ucdavis.edu/)



- **Highly individual!**

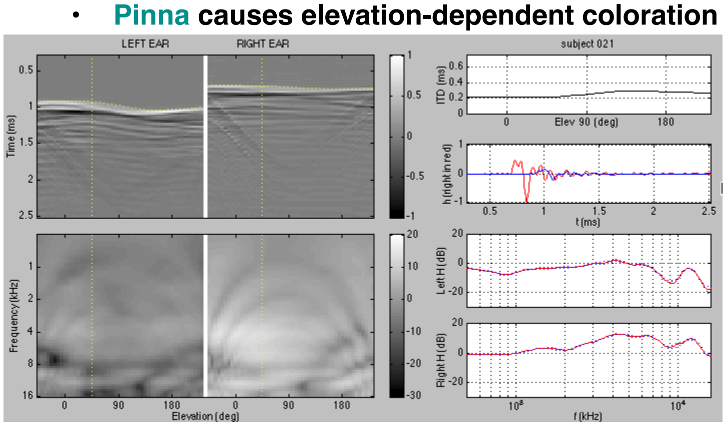
Cone of confusion



- **Interaural timing cue dominates (below 1kHz)**
  - from differing path lengths to two ears
- **But: only resolves to a cone**
  - Up/down? Front/back?

Further cues

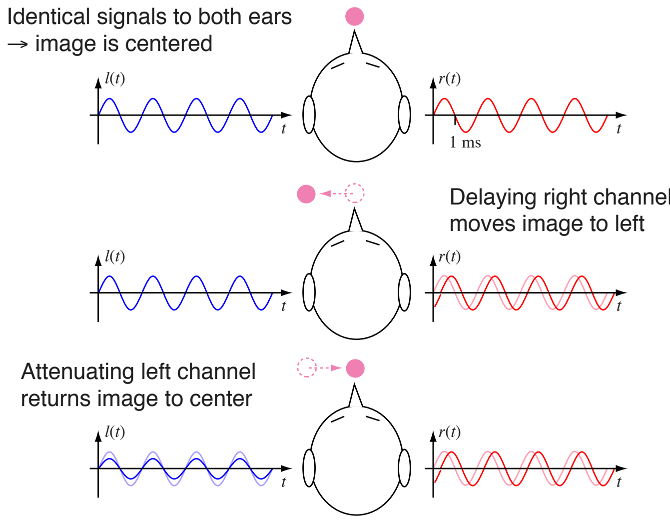
- **Pinna causes elevation-dependent coloration**



- **Monaural perception**
  - separate coloration from source spectrum?
- **Head motion**
  - synchronized spectral changes
  - also for ITD (front/back) etc.

Combining multiple cues

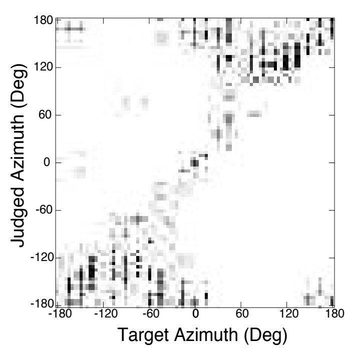
- **Both ITD and ILD influence azimuth; What happens when they disagree?**



  - trading @ around 0.1 ms / dB

Binaural position estimation

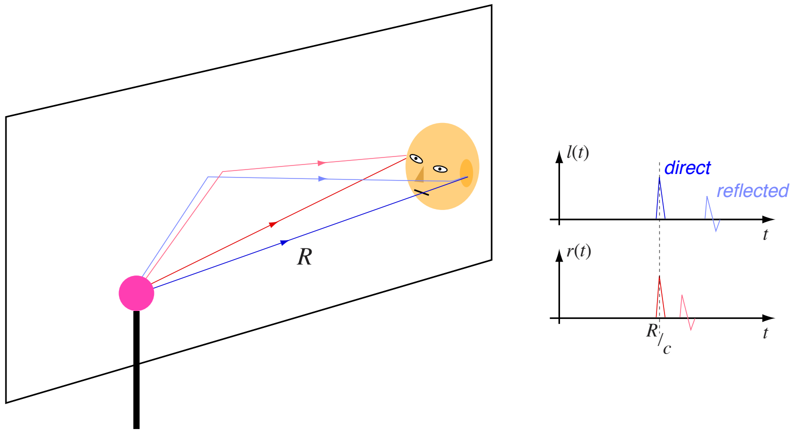
- **Imperfect results:** (Arruda, Kistler & Wightman 1992)



  - listening to 'wrong' hrtfs → errors
  - front/back reversals stay on cone of confusion

The Precedence Effect

- **Reflections give misleading spatial cues**

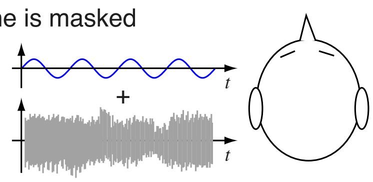


- **But: Spatial impression based on 1st wavefront then ‘switches off’ for ~50 ms**
  - .. even if ‘reflections’ are louder
  - .. leads to impression of room

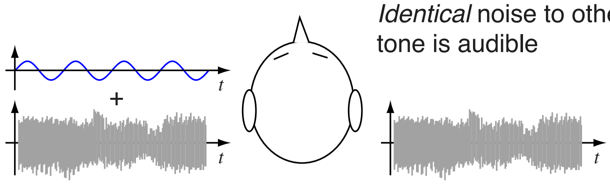
Binaural Masking Release

- **Adding noise to reveal target**

Tone + noise to one ear:
tone is masked



Identical noise to other ear:
tone is audible



  - why does this make sense?

- **Binaural Masking Level Difference up to 12dB**
  - greatest for noise in phase, tone anti-phase

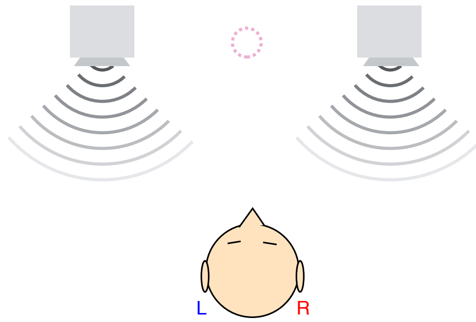
Outline

1. Spatial acoustics
2. Binaural perception
3. **Synthesizing spatial audio**
   - Position
   - Environment
4. Extracting spatial sounds

3 Synthesizing spatial audio

- **Goal: recreate realistic soundfield**
  - hi-fi experience
  - synthetic environments (VR)
- **Constraints**
  - resources
  - information (individual HRTFs)
  - delivery mechanism (headphones)
- **Source material types**
  - live recordings (actual soundfields)
  - synthetic (studio mixing, virtual environments)

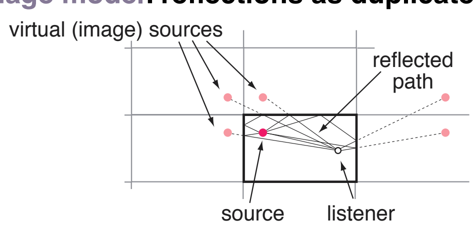
Classic stereo



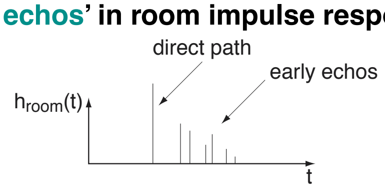
- **'Intensity panning': no timing modifications, just vary level ±20 dB**
  - works as long as listener is equidistant (ILD)
- **Surround sound: extra channels in center, sides, ...**
  - same basic effect - pan between pairs

Simulating reverberation

- **Can characterize reverb by impulse response**
  - spatial cues are important - record in stereo
  - IRs of ~ 1 sec → very long convolution
- **Image model: reflections as duplicate sources**



- **'Early echos' in room impulse response:**

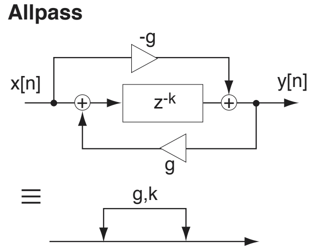


- **Actual reflection may be $h_{reflect}(t)$, not $\delta(t)$**

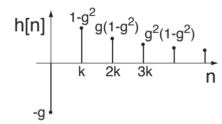
Artificial reverberation

- **Reproduce perceptually salient aspects**
  - early echo pattern (→ room size impression)
  - overall decay tail (→ wall materials...)
  - interaural coherence (→ spaciousness)
- **Nested allpass filters** (Gardner '92)

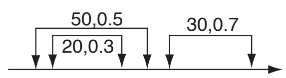
Allpass



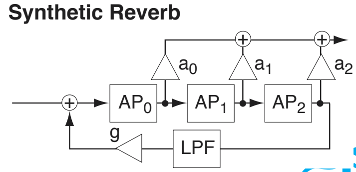
$$H(z) = \frac{z^{-k} - g}{1 - g \cdot z^{-k}}$$



Nested+Cascade Allpass

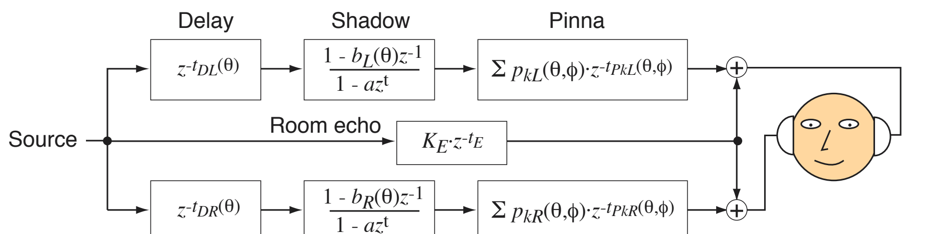


Synthetic Reverb



Synthetic binaural audio

- **Source convolved with {L,R} HRTFs gives precise positioning**
  - ...for headphone presentation
  - can combine multiple sources (by adding)
- **Where to get HRTFs?**
  - measured set, but: specific to individual, discrete
  - interpolate by linear crossfade, PCA basis set
  - or: parametric model - delay, shadow, pinna



(after Brown & Duda '97)

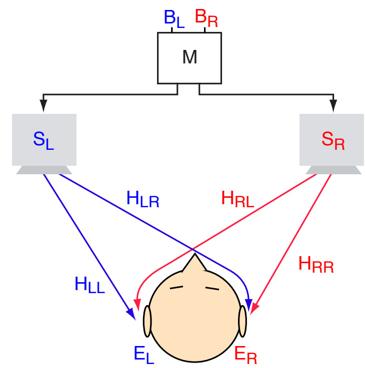
- **Head motion cues?**
  - head tracking + *fast* updates

Transaural sound

- **Binaural signals without headphones?**
- **Can cross-cancel wrap-around signals**
  - speakers $S_{L,R}$, ears $E_{L,R}$, binaural signals $B_{L,R}$.

$$S_L = H_{LL}^{-1}(B_L - H_{RL}S_R)$$

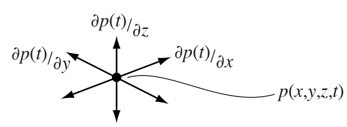
$$S_R = H_{RR}^{-1}(B_R - H_{LR}S_L)$$



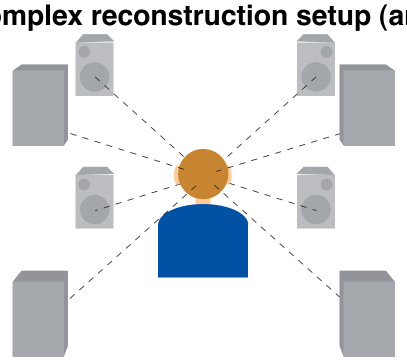
- **Narrow 'sweet spot'**
  - head motion?

Soundfield reconstruction

- **Stop thinking about ears**
  just reconstruct pressure + spatial derivatives



  - ears in reconstructed field receive same sounds
- **Complex reconstruction setup (ambisonics)**



  - able to preserve head motion cues?

Outline

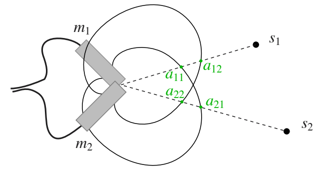
1. Spatial acoustics
2. Binaural perception
3. Synthesizing spatial audio
4. **Extracting spatial sounds**
    - Microphone arrays
    - Modeling binaural processing

4 Extracting spatial sounds

- **Given access to soundfield, can we recover separate components?**
    - degrees of freedom: >N signals from N sensors is hard
    - but: people can do it (somewhat)
- **Information-theoretic approach**
    - use only very general constraints
    - rely on precision measurements
- **Anthropic approach**
    - examine human perception
    - attempt to use same information

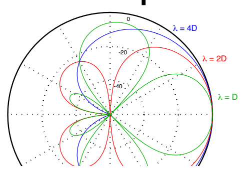
Microphone arrays

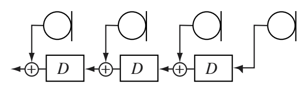
- **Signals from multiple microphones can be combined to enhance/cancel certain sources**
- **'Coincident' mics with diff. directional gains**



$$\begin{bmatrix} m_1 \\ m_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \cdot \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \Rightarrow \begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \end{bmatrix} = \mathbf{A}^{-1} \cdot \mathbf{m}$$

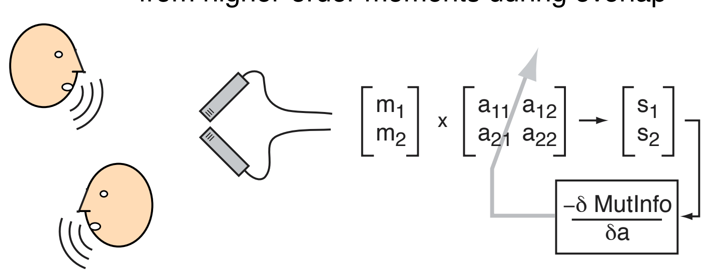
- **Microphone arrays (endfire)**





Adaptive Beamforming & Independent Component Analysis (ICA)

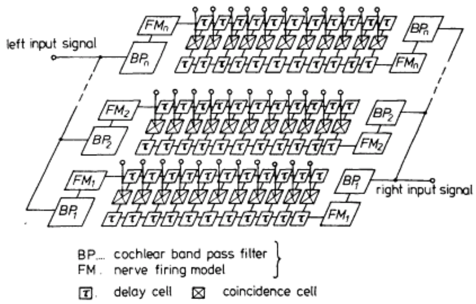
- **Formulate mathematical criteria to optimize**
- **Beamforming: Drive interference to zero**
  - cancel energy during nontarget intervals
- **ICA: maximize mutual independence of outputs**
  - from higher-order moments during overlap

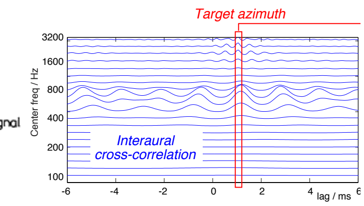


- **Limited by separation model parameter space**
  - only NxN?

Binaural models

- **Human listeners do better?**
  - certainly given only 2 channels
- **Extract ITD and IID cues?**

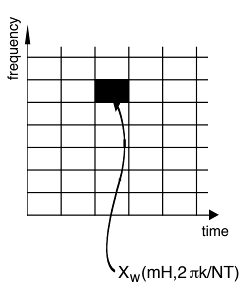


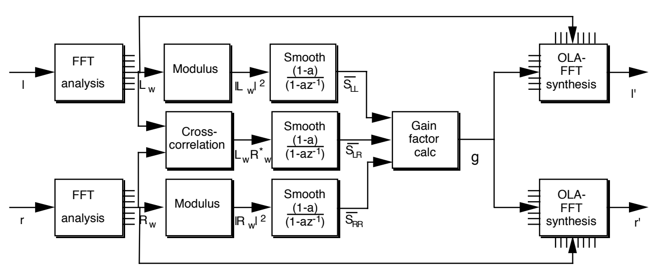


  - cross-correlation finds timing differences
  - 'consume' counter-moving pulses
  - how to achieve IID, trading
  - vertical cues...

Nonlinear filtering

- **How to separate sounds based on direction?**
  - estimate direction locally
  - choose target direction
  - remove energy from other directions
- **E.g. Kollmeier, Peissig & Hohman '93**





  - IID from $|L_w|/|R_w|$; ITD (IPD) from $arg\{L_w R_w^*\}$
  - match to IID/IPD template for desired direction
  - also reverberation?

Summary

- **Spatial sound**
  - sampling at more than one point gives information on origin direction
- **Binaural perception**
  - time & intensity cues used between/within ears
- **Sound rendering**
  - conventional stereo
  - HRTF-based
- **Spatial analysis**
  - optimal linear techniques
  - elusive auditory models

References

B.C.J. Moore, *An introduction to the psychology of hearing* (4th ed.) Academic, 1997.

J. Blauert, *Spatial Hearing* (revised ed.), MIT Press, 1996.

R.O. Duda, *Sound Localization Research*, http://www.engr.sjsu.edu/~duda/Duda.Research.frameset.html