Spatial Hearing: Single Sound Source in Free Field

Piotr Majdak & Bernhard Laback
http://www.kfs.oeaw.ac.at
piotr@majdak.com

ÖAW
Austrian Academy
of Sciences

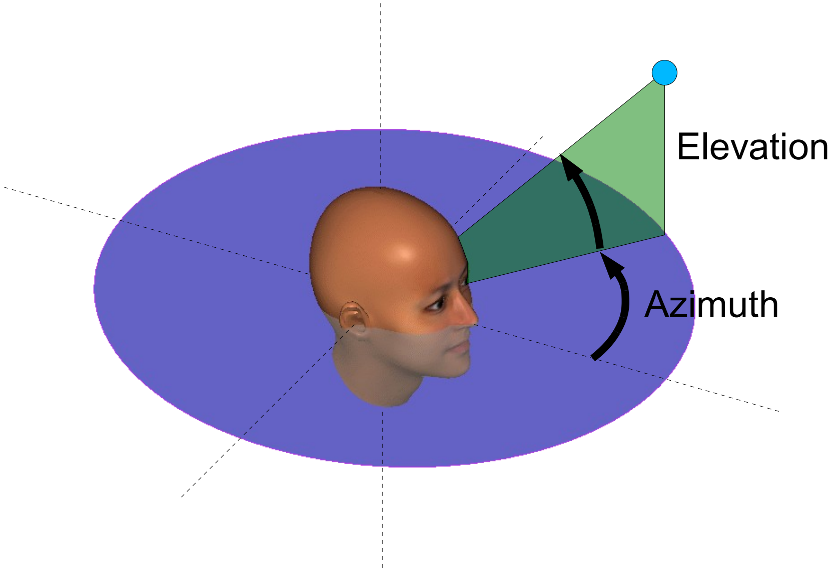
Spatial Hearing

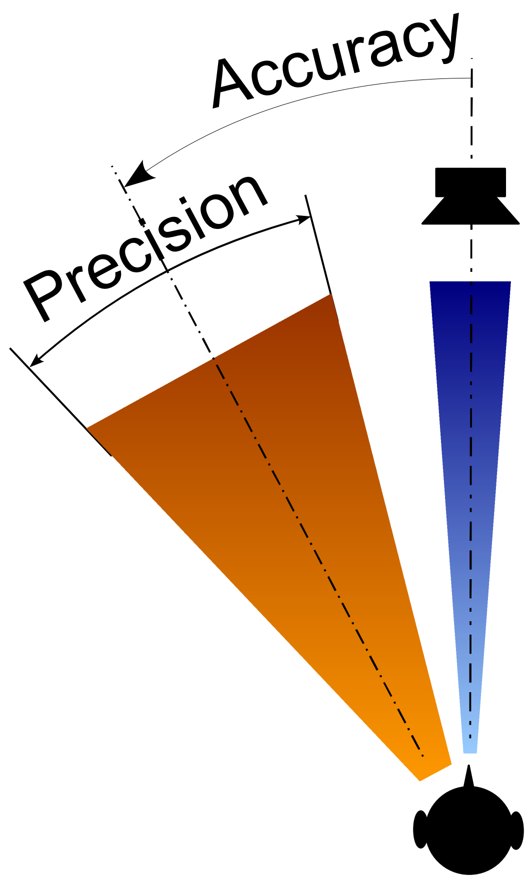
- Object identification (outside the field-of-view)
- Multiple sound sources:
  - Analysis of the auditory scene
  - Focusing on a target (coctail-party effect)
- Enclosed auditory spaces:
  - Multiple echos of the sound (reverberation)
  - Complex perceptual effects:
    - Sound perception (directivity, distance)
    - Room perception
    - Sound-room interaction (Jot, 1999)

ARI

Spatial Hearing: Simple Case

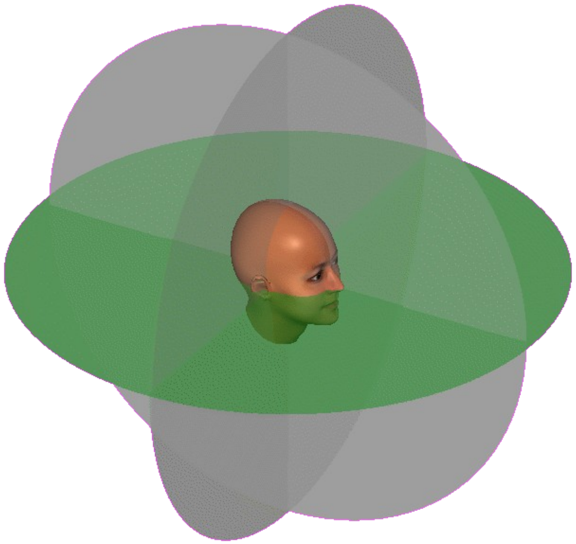
- Single sound source in free field:
  - No room (no reflections)
  - No precedence effect
  - No distance (Direction only)

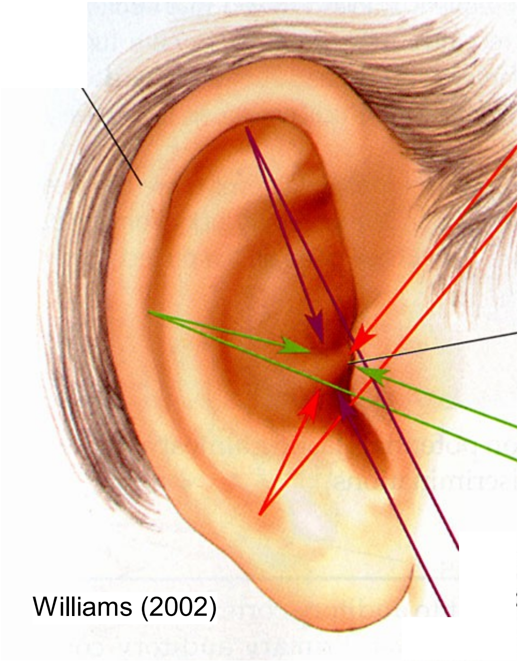




Localization of Sound Sources

- Signals from two receivers available:
  - Aligned in the horizontal plane
  - Asymmetries in the geometry of the individual receivers
    - Direction-dependent spectral change of incoming sound

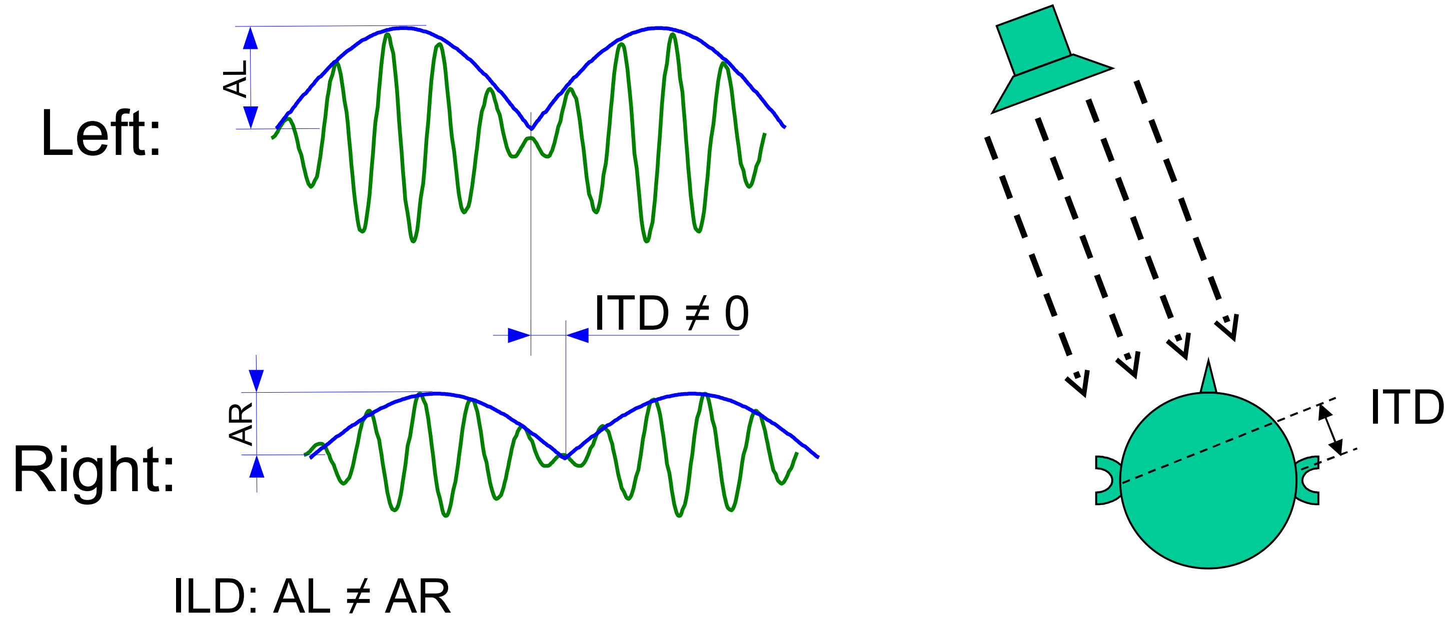




Williams (2002)

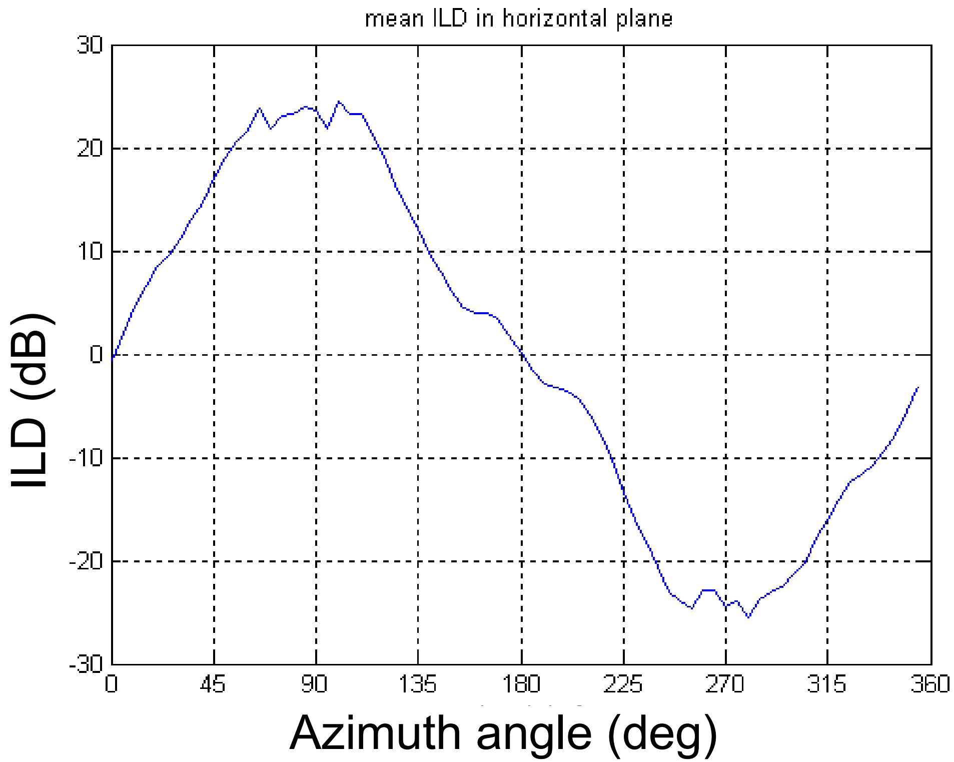
Sound Localization in the Horizontal Plane

- Interaural level differences (ILDs)
- Interaural time differences (ITDs)



Interaural Level Differences

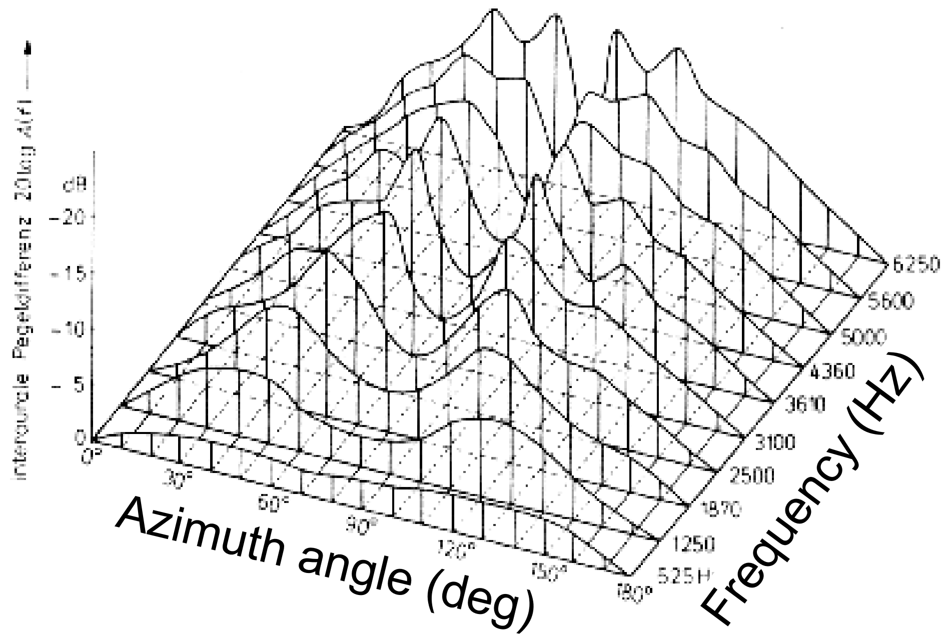
- Broadband ILD of dummy head:



Interaural Level Differences

- Frequency dependent:
  - Wave length larger than head diameter:
    - Little effect
  - Wave length in the range of the head diameter:
    - Diffraction
  - Wave length smaller than head diameter:
    - Large attenuation due to shadowing effects

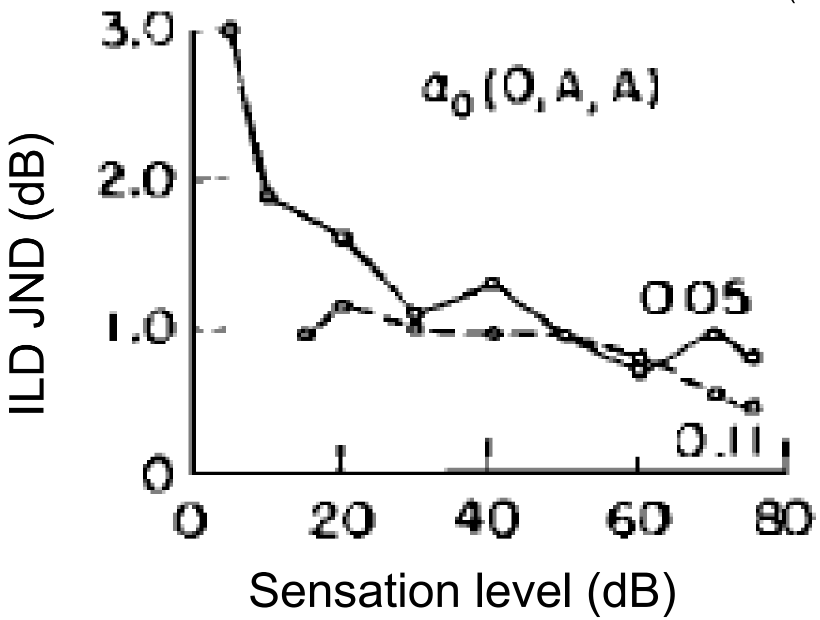
ILD (sphere model)



Blauert (1974)

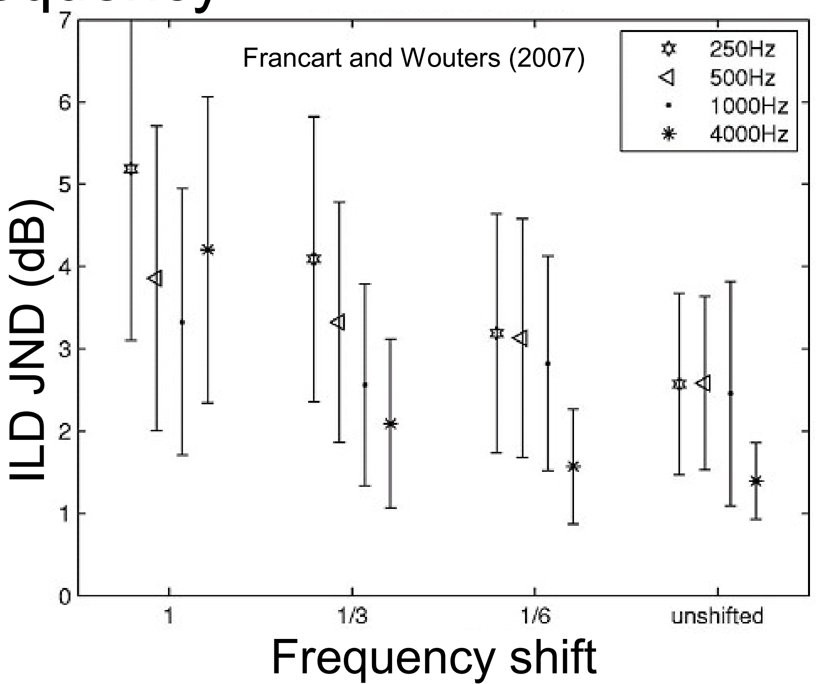
Interaural Level Differences

- Perceptual threshold (just noticable diff., JND):
  - In the order of 1 dB (Hall, 1964 and others)
  - Depends on the sound level (Herschkowitz & Durlach, 1969)



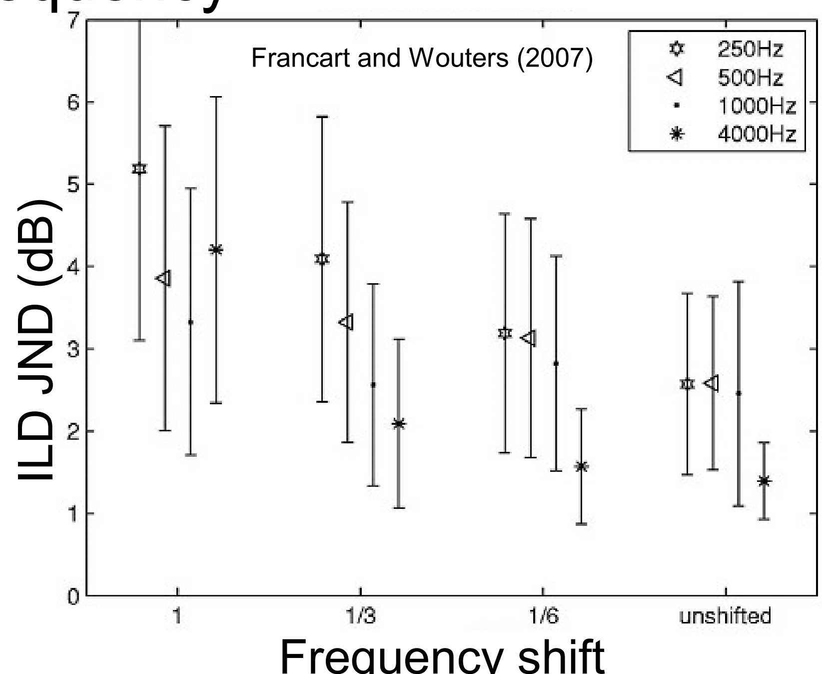
Interaural Level Differences

- Perceptual threshold:
  - In the order of 1 dB (Hall, 1964 and others)
  - Depends on the sound level (Herschkowitz & Durlach, 1969)
  - Small dependency on frequency and frequency shift between the two ears (Francart and Wouters, 2007)



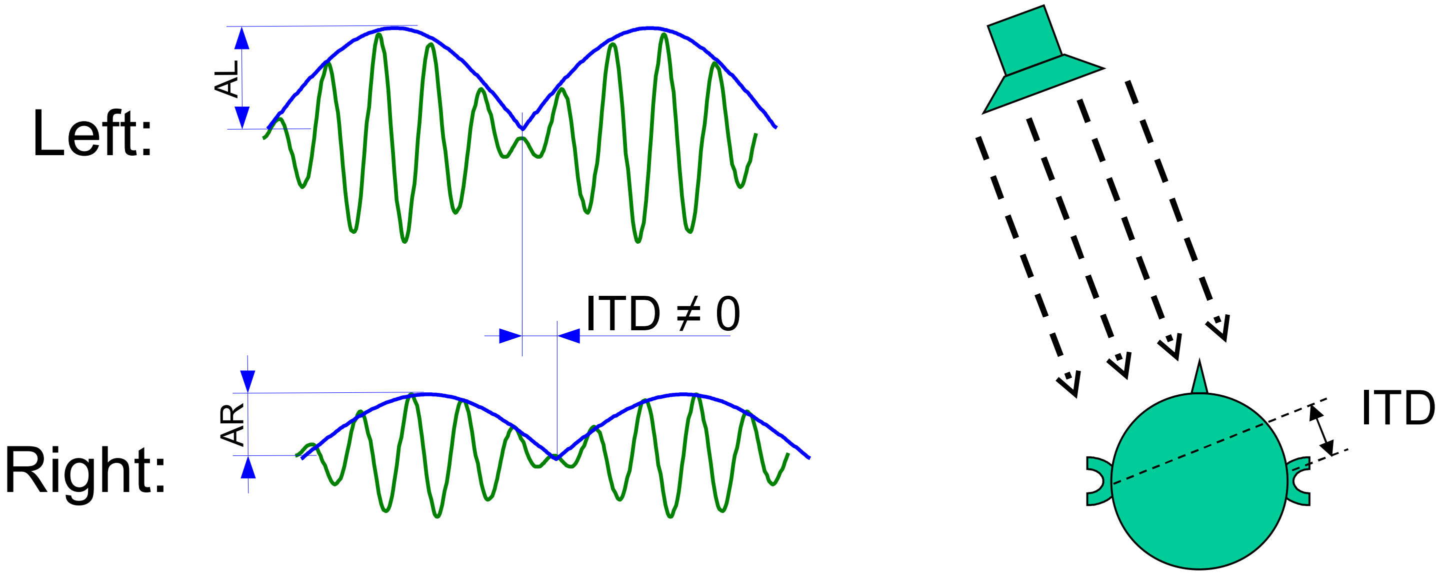
Interaural Level Differences

- Perceptual threshold:
  - In the order of 1 dB (Hall, 1964 and others)
  - Depends on the sound level (Herschkowitz & Durlach, 1969)
  - Small dependency on frequency and frequency shift between the two ears (Francart and Wouters, 2007)
  - Depends on the **lateral sound position** (Bernstein, 2004)



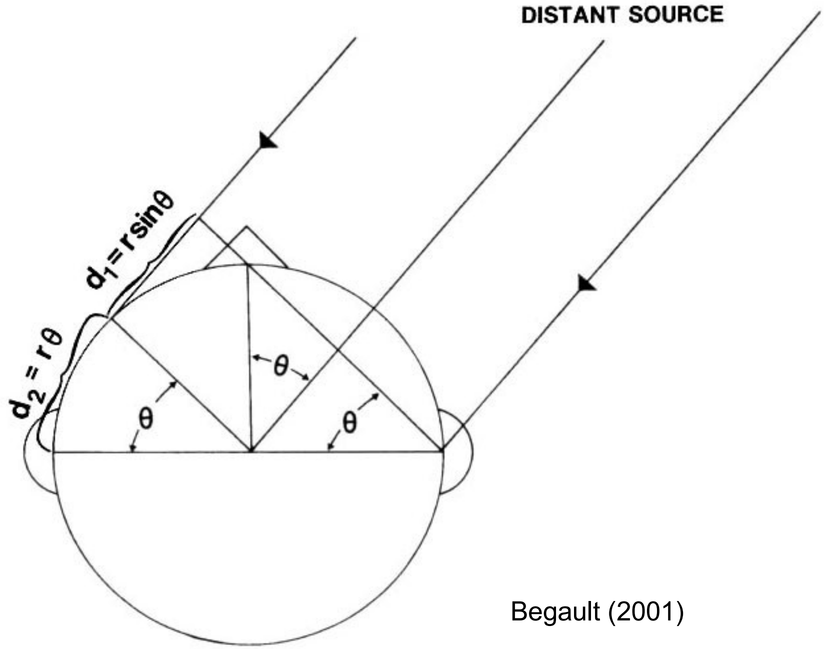
Sound Localization in the Horizontal Plane

- Interaural level differences (ILDs)
- **Interaural time differences (ITDs)**



Interaural Time Differences

- $ITD = \frac{r}{c}\left(\theta + \sin\theta\right)$ (Woodworth & Schlosberg, 1962)

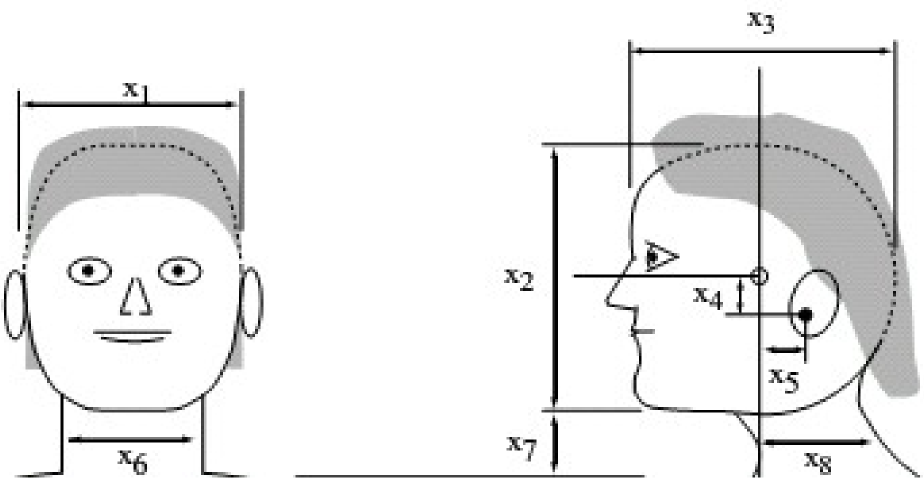


Begault (2001)

Interaural Time Differences

- $ITD = \frac{r}{c}(\theta + \sin\theta)$ (Woodworth & Schlosberg, 1962)
- $r \rightarrow r_e$:

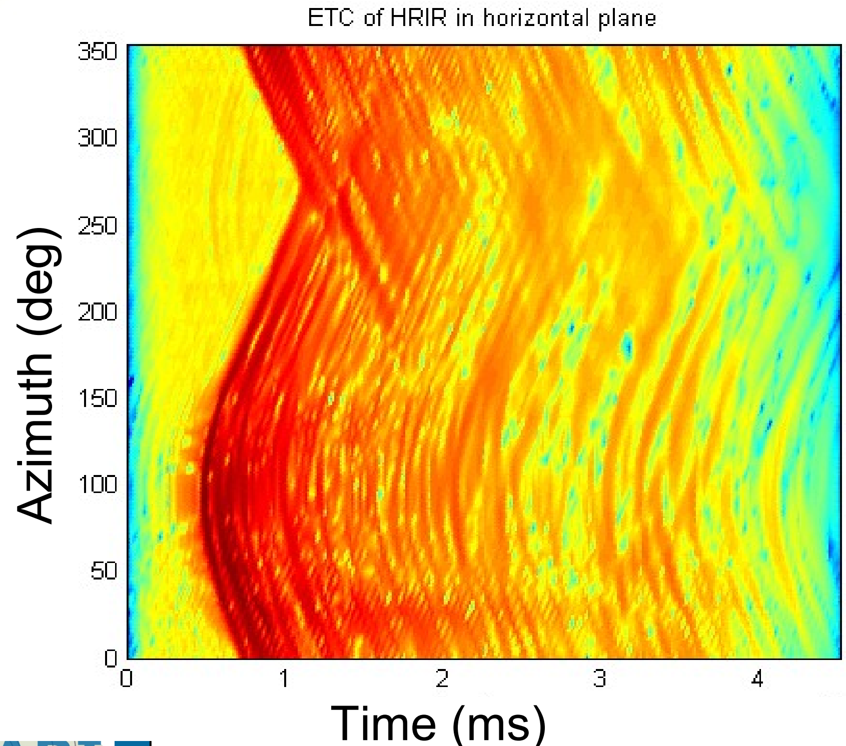
$$r_e = 0.51\,x_1 + 0.18\,x_3 + 0.032$$ (Algazi et al., 2001)

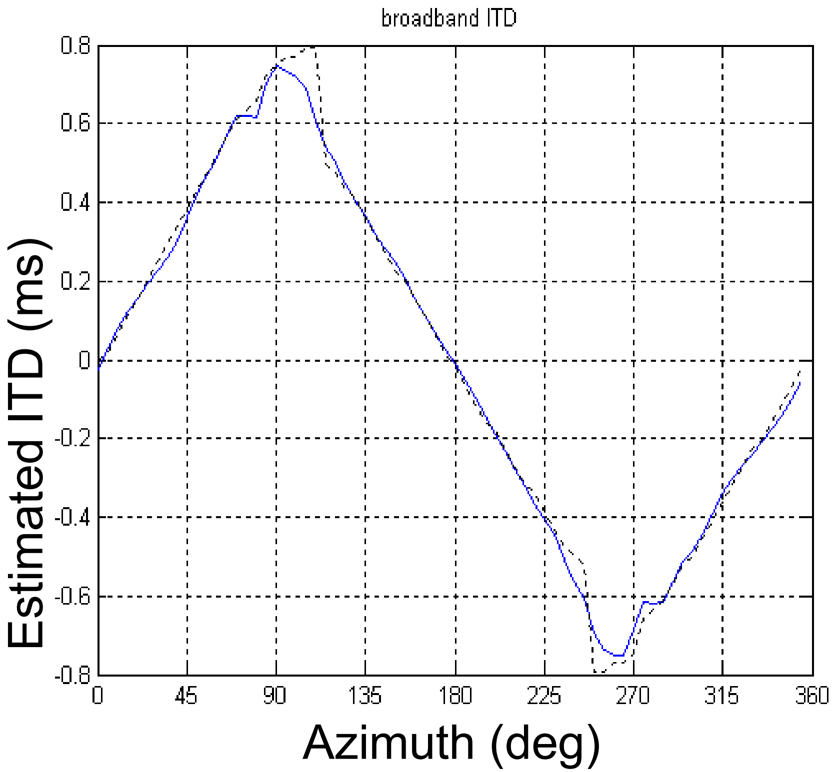


Algazi et al. (2001)

Interaural Time Differences

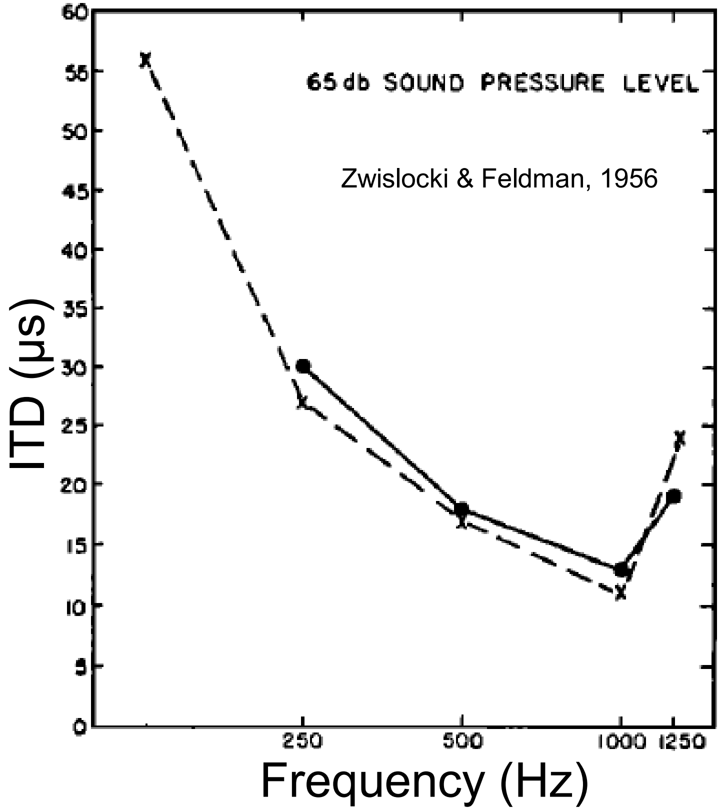
- Physical range: +/- 800 µs





Interaural Time Differences

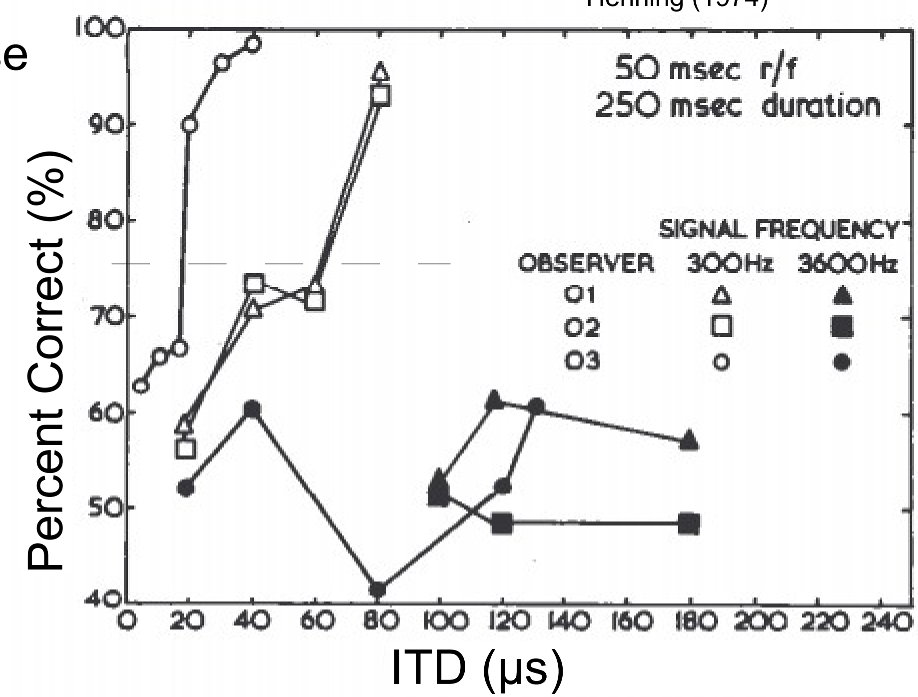
- Perceptual threshold:
  - Best conditions: In the order of 10 µs



Interaural Time Differences

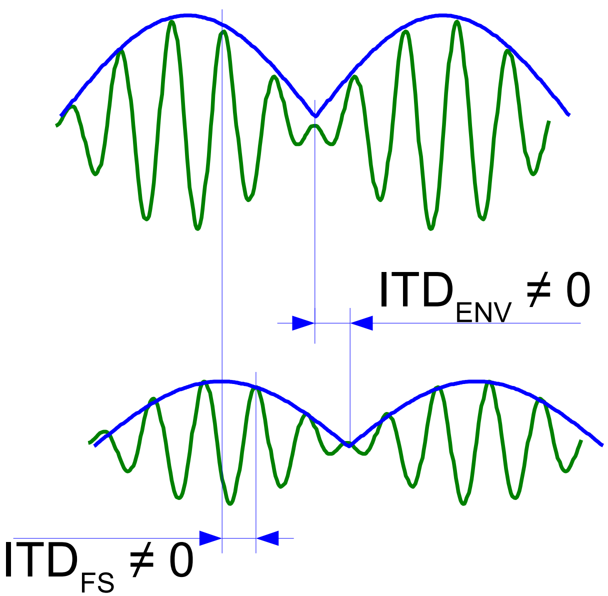
- Perceptual threshold:
  - Best conditions: In the order of 10 µs
  - Pure tones: depends on frequency
    - Ambiguity of the ongoing ITD in the phase
    - Refractory time of the neurons

Henning (1974)

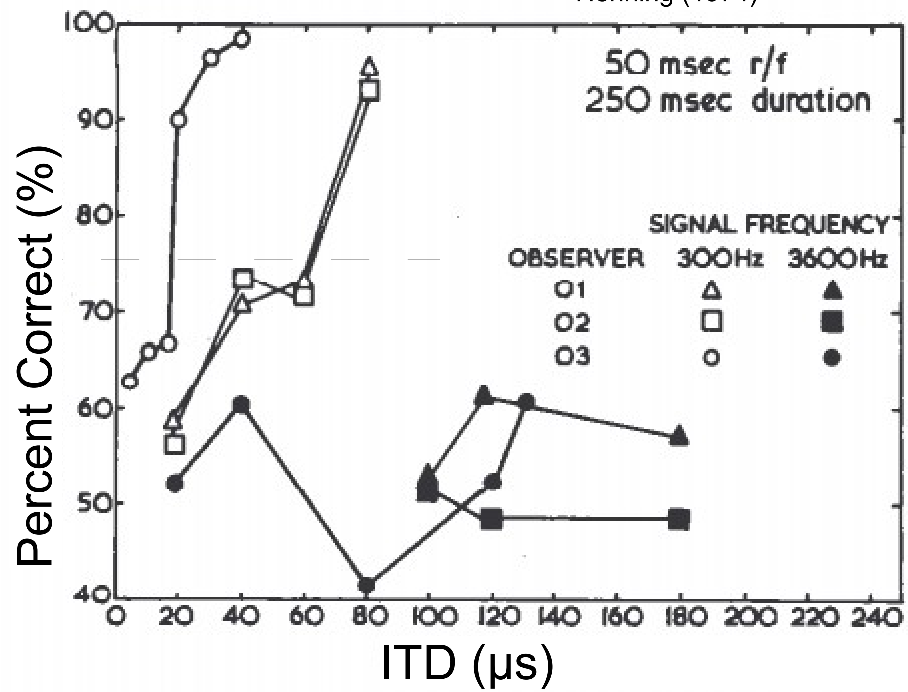


Interaural Time Differences

- Perceptual threshold in complex signals:
  - Pure-tone ITD → ITD in the fine structure

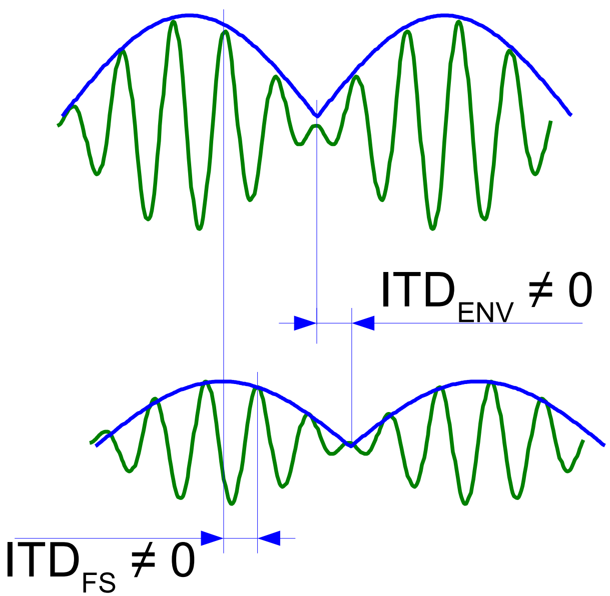


Henning (1974)

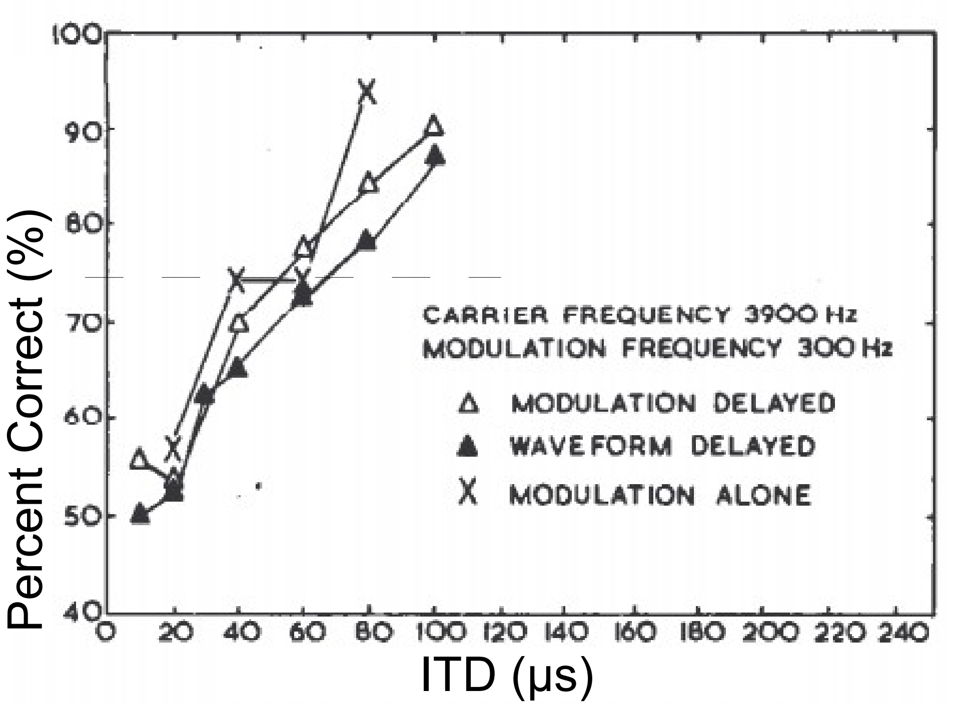


Interaural Time Differences

- Perceptual threshold in complex signals:
  - Pure-tone ITD → ITD in the fine structure
  - Modulations: ITD in the envelope

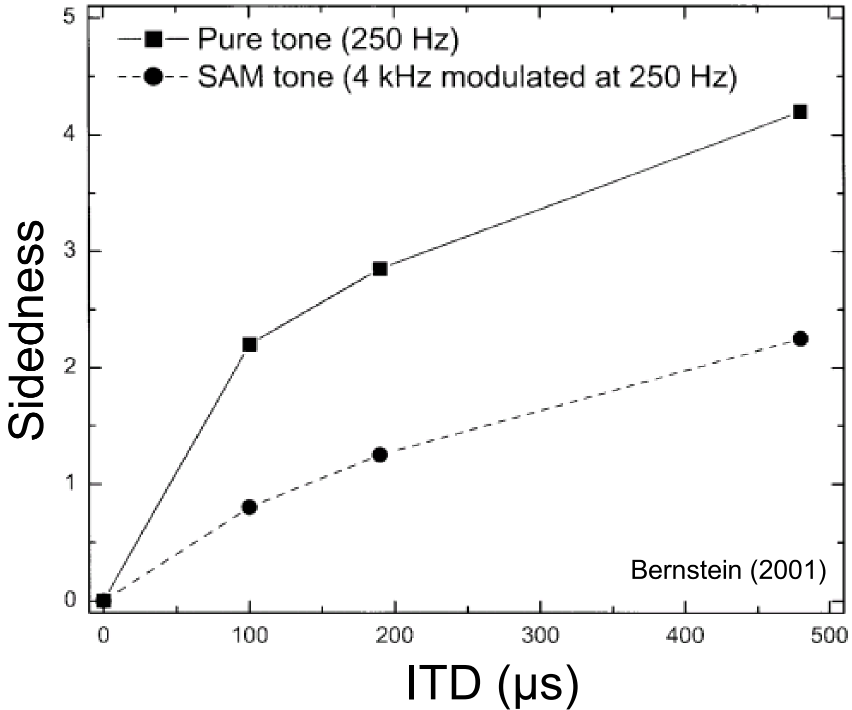


Henning (1974)



Interaural Time Differences

- Lateralization based on ITD:
  - Low-frequency (pure-tone) ITD: strong cue
  - High-frequency (envelope) ITD: weaker cue



Localization Cues for the Horizontal Plane

- ILD (broadband)?
- ITD (broadband)?
- Envelope ITD (high frequencies)?
- Spectral cues?
    - Interaural spectral differences?
    - Monaural cues?
- Duplex theory (Rayleigh 1907 & others)
    - Low frequency range: ITDs
    - High frequency range: ILDs

→ Does the duplex theory still hold?

ÖAW
Austrian Academy
of Sciences

Duplex Theory Revisited

Macpherson & Middlebrooks (2002)

- ILD weight:
  - 0.52 (broadband); 0.24 (low-pass); 0.82 (high-pass)
- ITD weight:
  - 0.82 (broadband); 0.88 (low-pass); 0.24 (high-pass)
  - Envelope ITD weight (broadband):
    dep. on onset and modulation
- Interaural spectral difference weight:
  - Same as broadband ILD
- Monaural (near-ear) spectrum weight:
  - 0.03 (broadband); 0.03 (high-pass)

ARI

Localization Cues for the Horizontal Plane

- High-frequency ILD
- Low-frequency ITD
- Onset-ITD and ITD in the ongoing modulation
- *Spectral information not relevant*

  (neither monaural nor binaural)

- Valid for the lateral dimension in horizontal plane only!

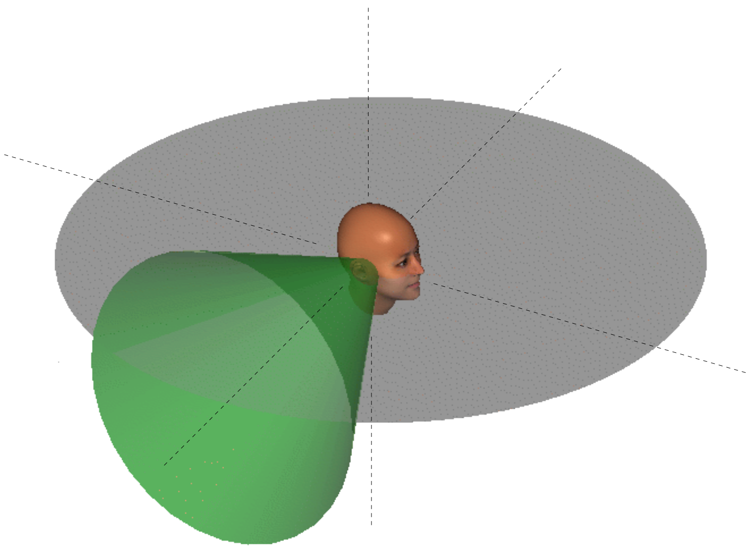
(Rayleigh 1876)

> The possibility of distinguishing a voice in front from a voice behind would thus appear to depend on the compound character of the sound in a way that it is not easy to understand, and for which the second ear would be of no advantage.

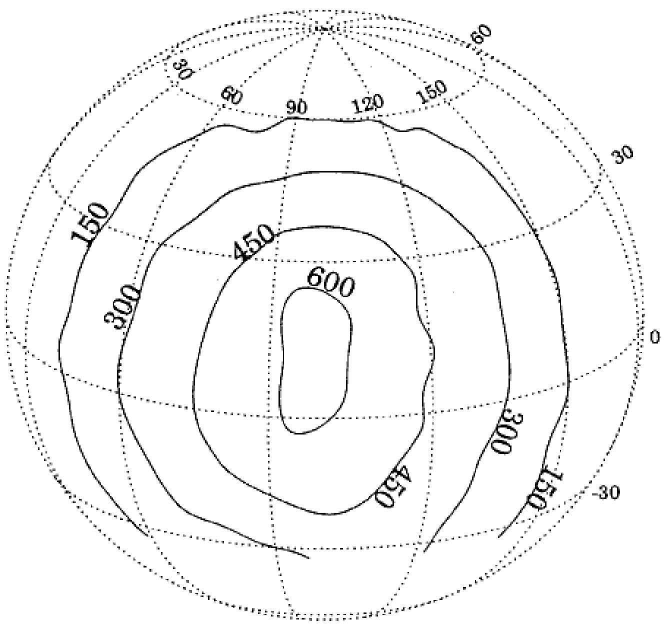
Cone of Confusion

- ITD-based front-back ambiguity:

Modeled



Measured

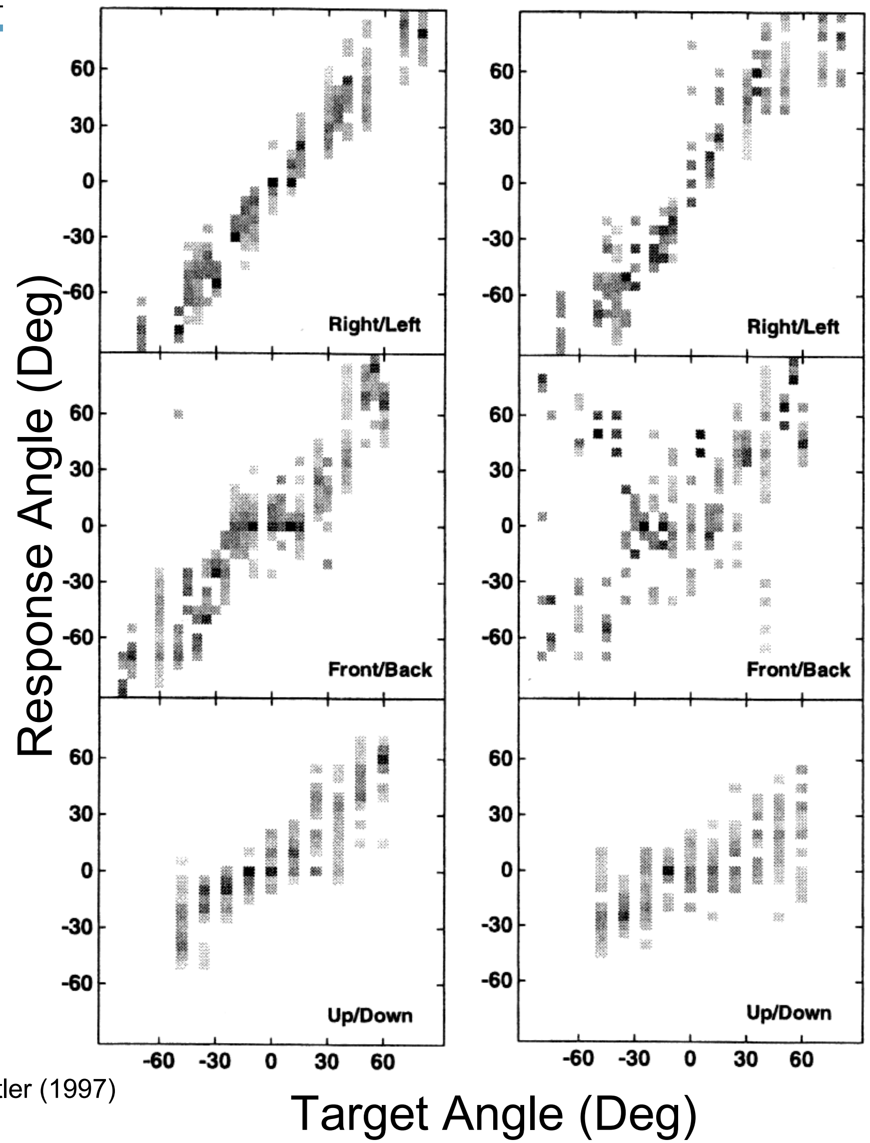


Whigthman & Kistler (1997)

- Can be resolved with the help of spectral cues

Cone of Confusion

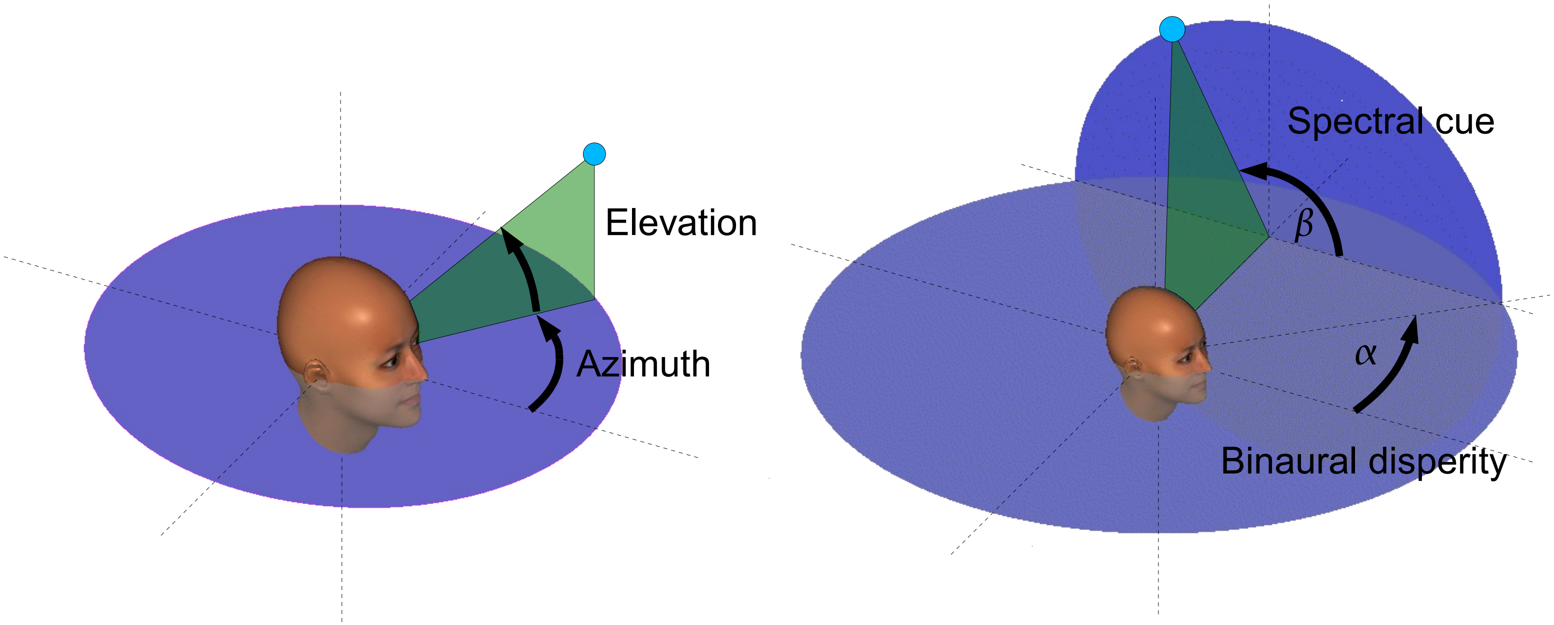
- Left panel:
  - Flat spectrum
- Right panel:
  - Scrambled spectrum
- Results:
  - More front-back confusions
  - Larger elevation error



Whigthman & Kistler (1997)

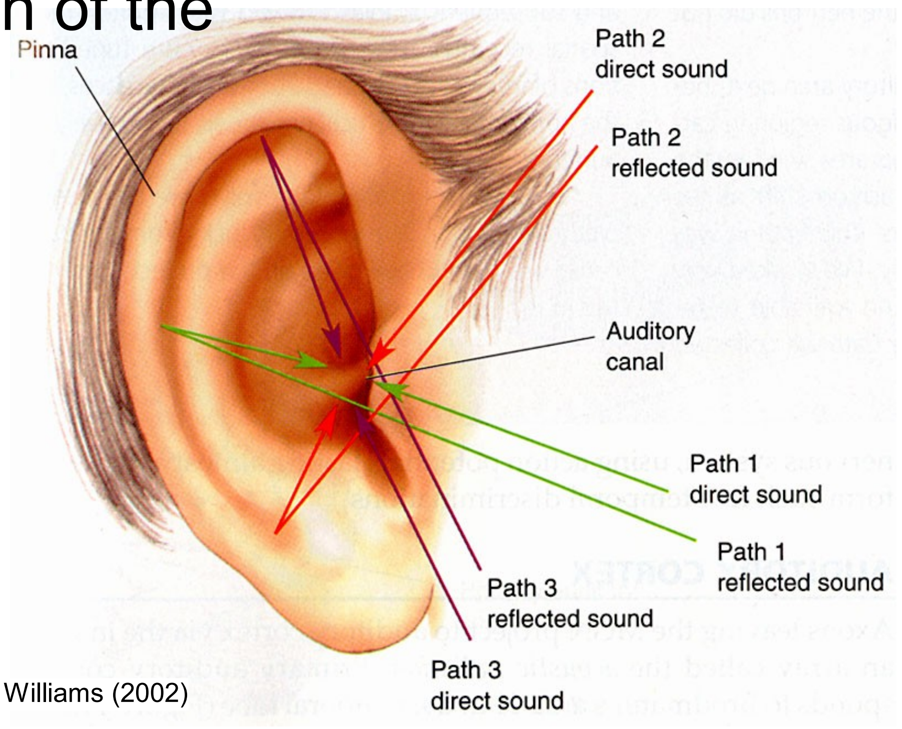
Perceptually-Relevant Coordinate System

- *Geodetic coordinate system: Azimuth & Elevation*
- Horizontal-polar coordinate system:
  - Lateral angle: binaural disperity
  - Polar angle: spectral cues



Spectral Cues

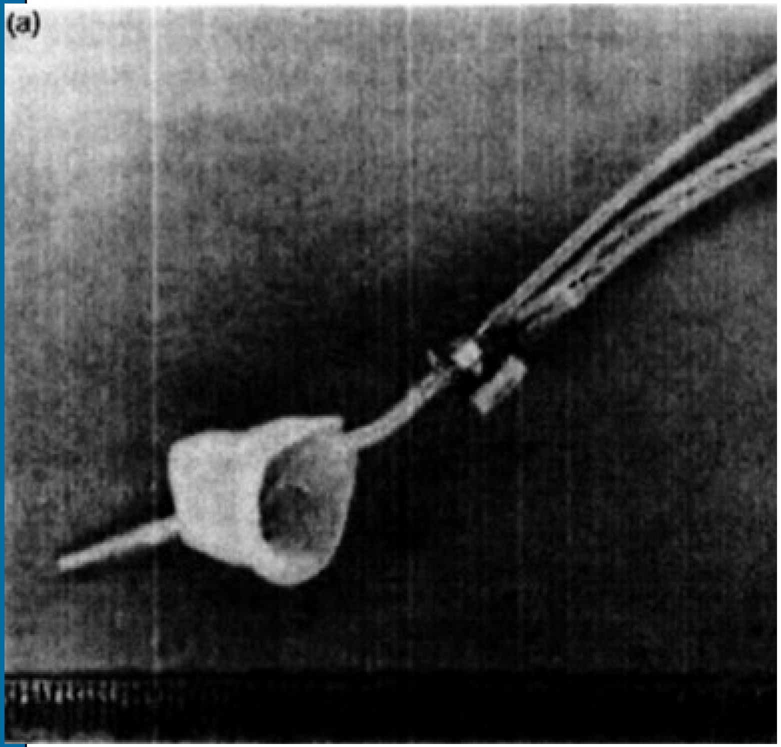
- Head-related transfer functions (HRTFs)
  - Describe the filtering effect of the head, torso, pinna
  - Depend on the position of the sound source
- Time-domain:
  - Head-related impulse responses (HRIRs)



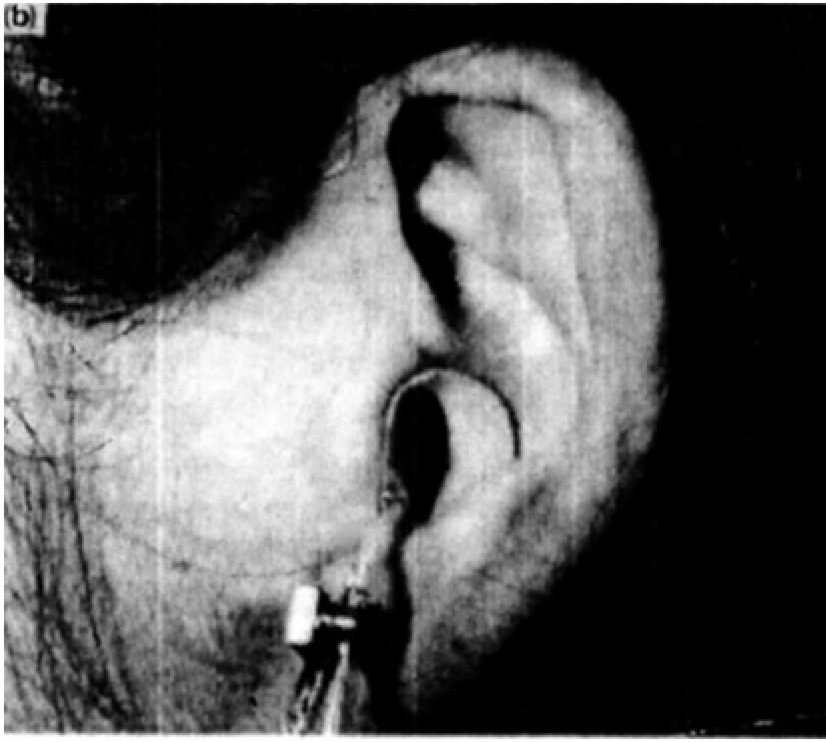
Williams (2002)

HRTF Measurement

- Open ear canal



Wightman & Kistler (1996)



Wightman & Kistler (1996

HRTF Measurement

- Closed ear canal: simpler (less variability)

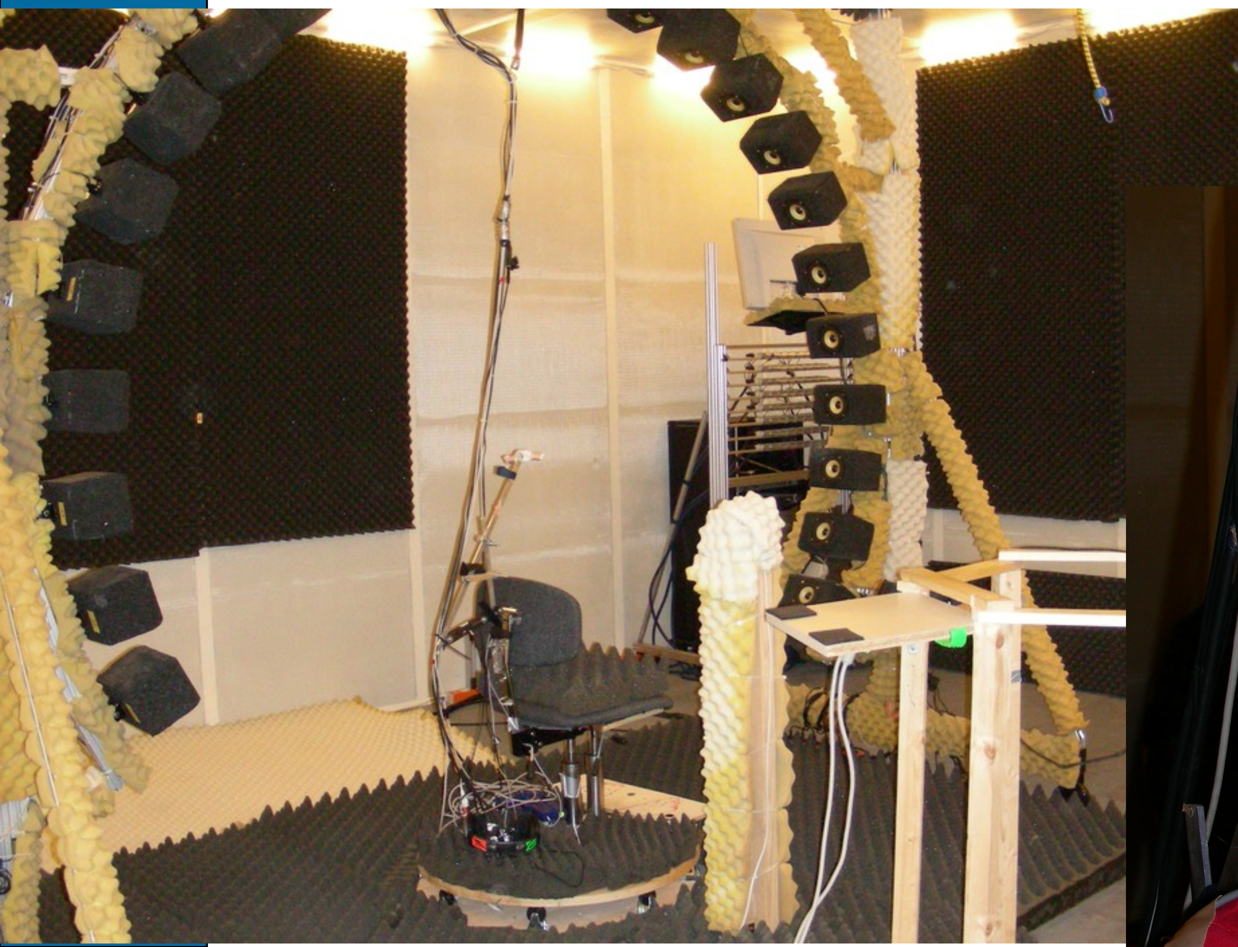
Probe microphone

MØLLER et al (1995)

Electret capsule

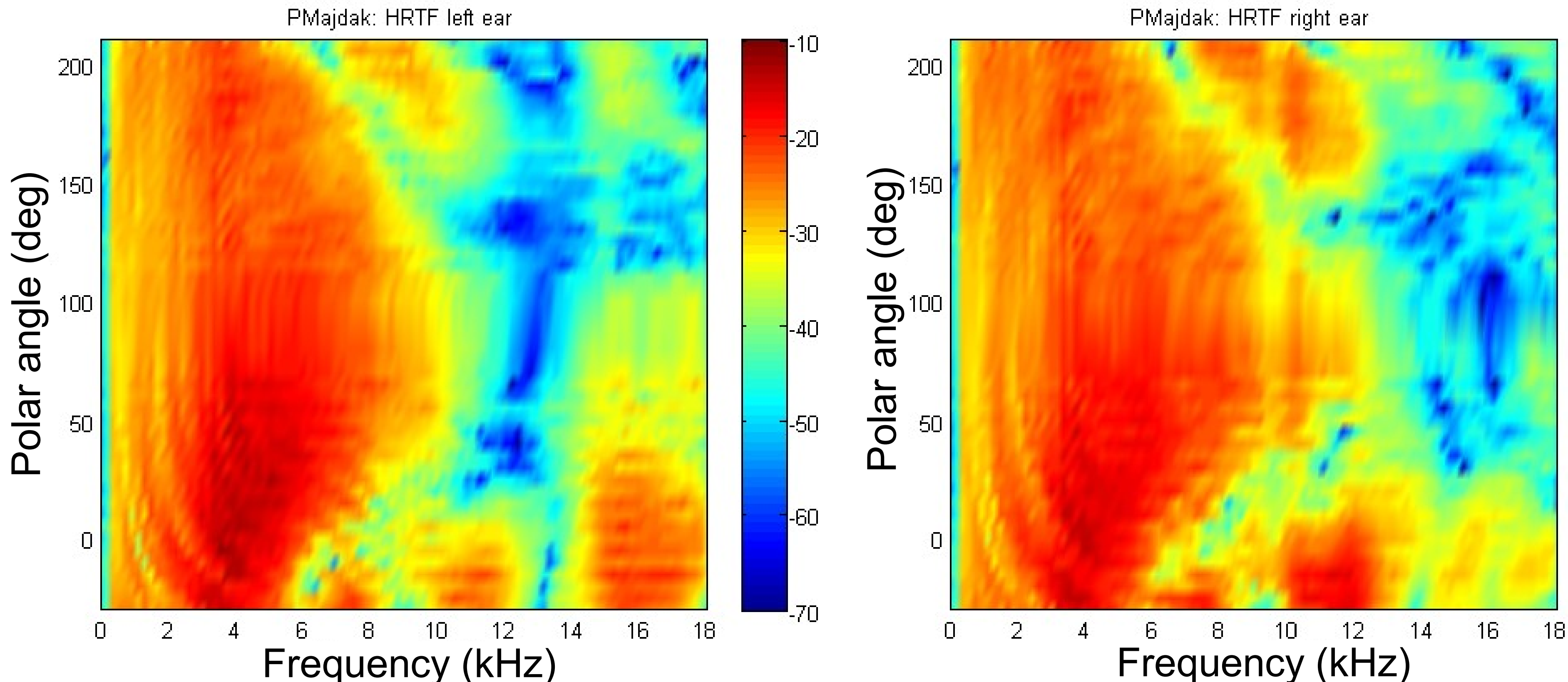
HRTF Measurement

- System identification for many positions



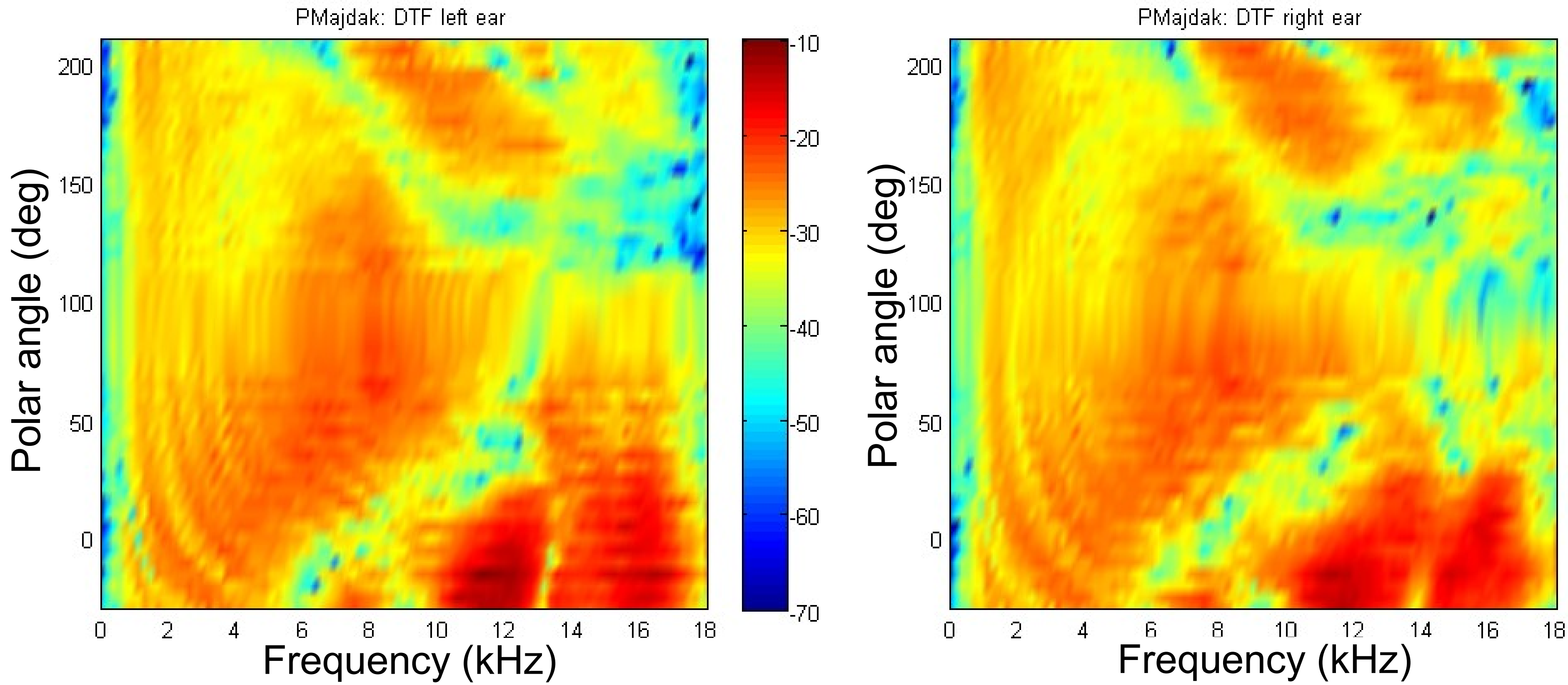
HRTFs

- In the median (mid-sagittal) plane



Directional Transfer Functions (DTFs)

- Model for HRTFs: $H(f) = C(f) \cdot D(f)$
- Consider directional cues only $C(f) = \frac{1}{N}\sum_{i=1}^{N} \log|H_i(f)|$

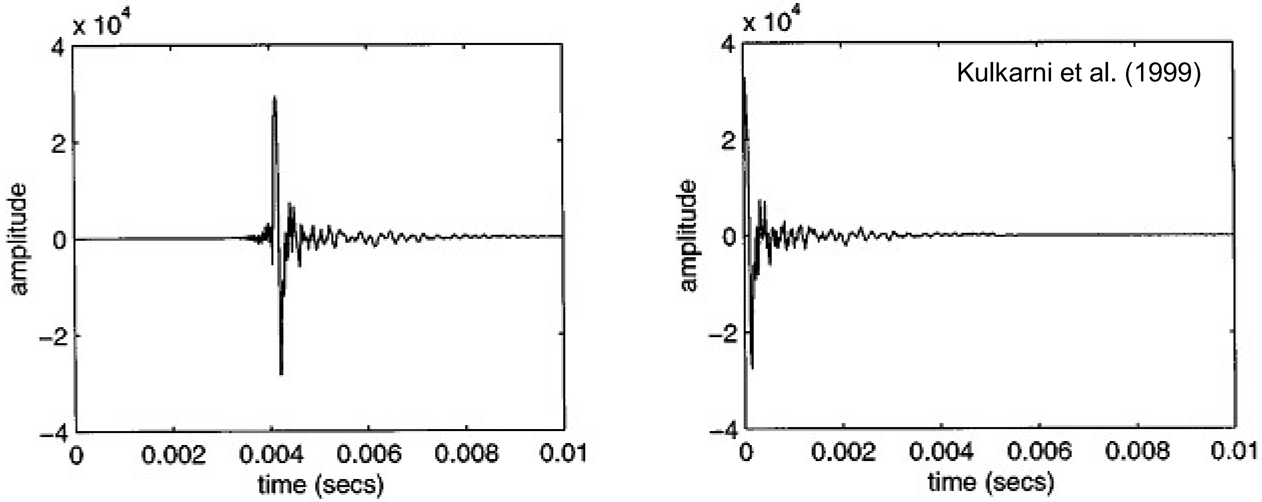


Phase Spectrum?

- Model for the HRTFs:

$$H(f) = H_{ap}(f) \cdot H_{min}(f)$$

  - $H_{ap}(f) = e^{i\varphi_{ap}(f)}$ … All-pass filter (delay → ITD)
  - $H_{min}(f) = |H(f)| e^{i\varphi_{min}(f)}$ … Minimum phase system

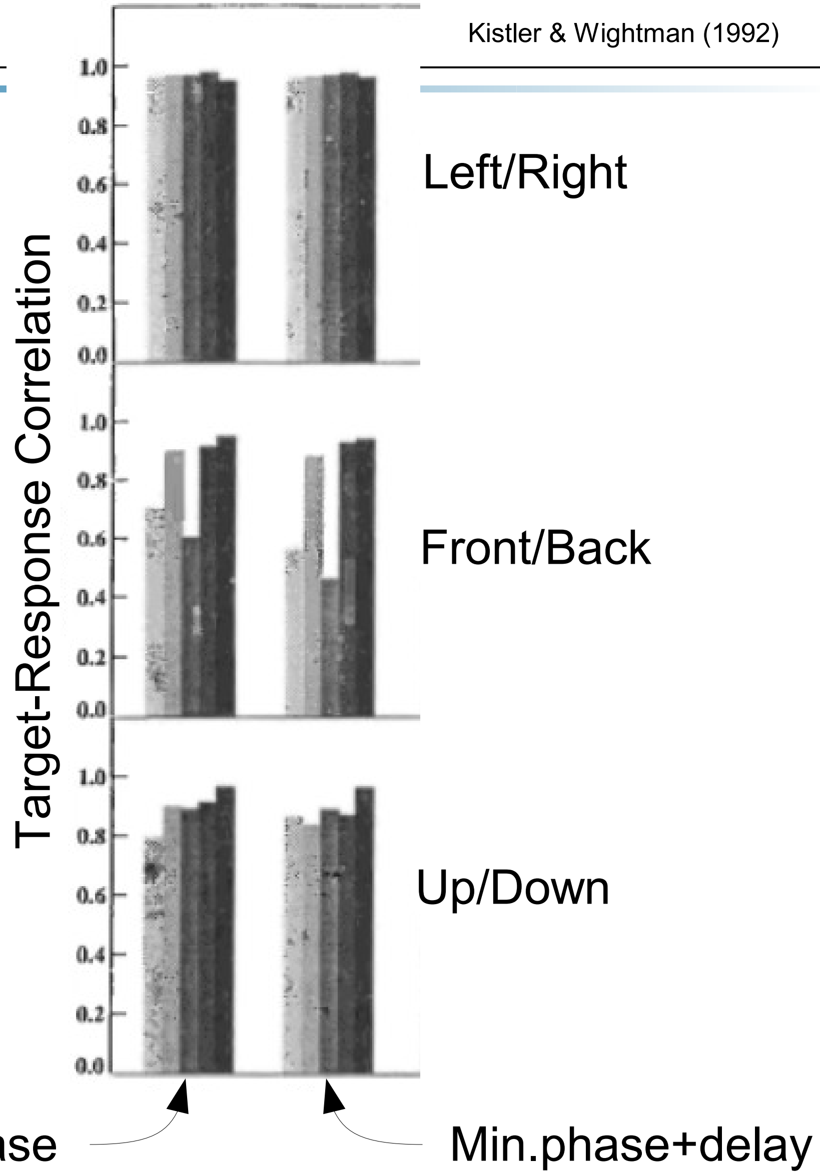


Phase Spectrum

- Perceptually not relevant

Wightman & Kistler (1992)
Kulkarni et al. (1999)
Macpherson & Middlebrooks (2002)
Hartmann et al. (2010)

Kistler & Wightman (1992)

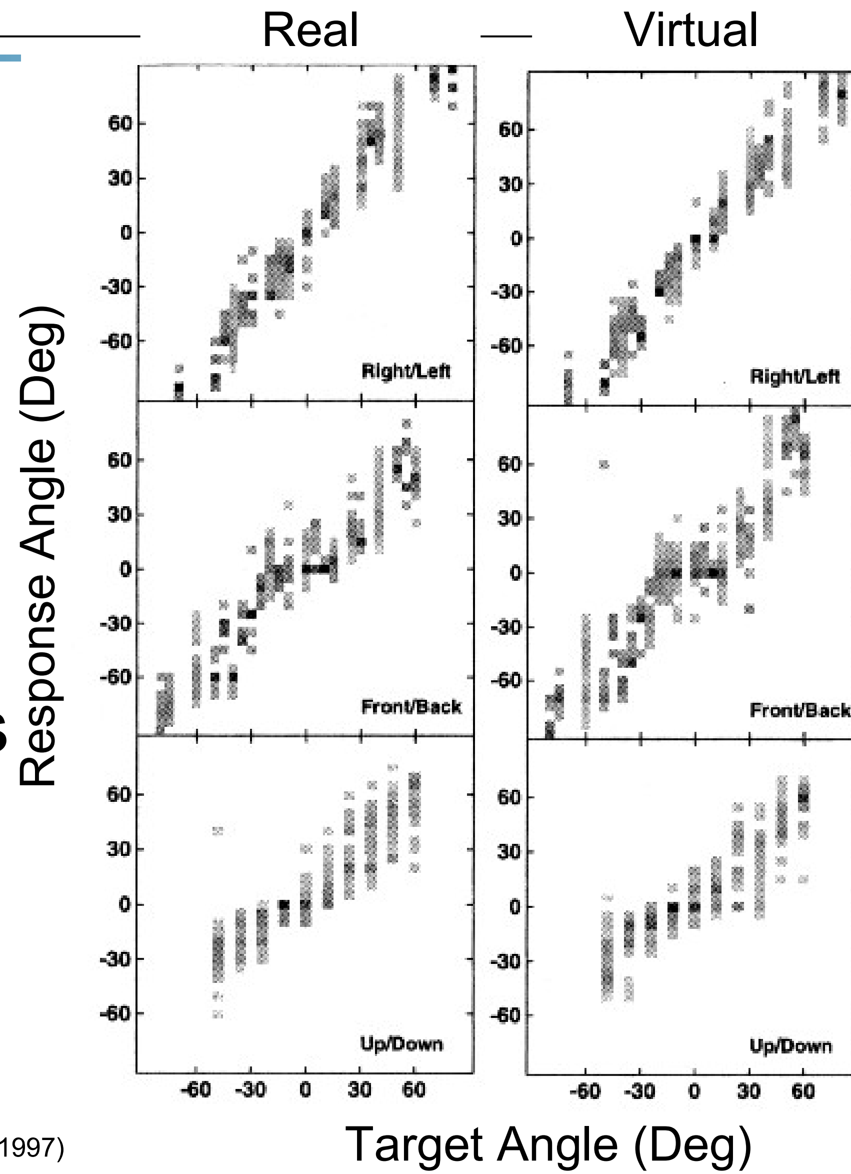


Signal Synthesis for Virtual Acoustics

- Filter signal with corresponding pair of DTFs
- Apply ITD

  (if DTFs modeled by min. phase+delay)

- Present the binaural signals via headphones



Whigthman & Kistler (1997)

Signal Synthesis for Virtual Acoustics

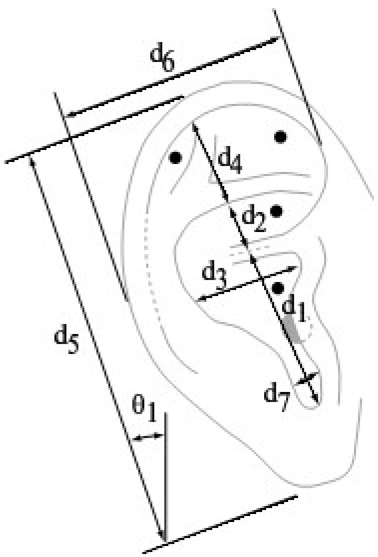
	Free-field	Own-ear virtual
rms lateral error (deg.)	10.6±2.0	14.5±2.2
Magnitude of lateral bias (deg.)	2.9±3.1	3.1±3.9
rms local polar error (deg.)	22.7±5.1	28.7±4.7
Magnitude of elevation bias (deg.)	5.5±4.4	10.2±6.6
Total quadrant errors (% of trials)	4.6±5.9	7.7±8.0
Quadrant error by target quadrant:		
Down-front (% of trials)	0.6±1.1	1.0±2.6
Up-front (% of trials)	1.9±3.3	5.7±6.4
Up-rear (% of trials)	10.6±13.8	21.7±21.1
Down-rear (% of trials)	0±0	1.8±5.0

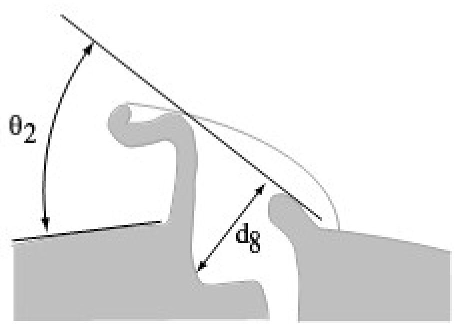
Middlebrooks (1999)

- So: Generic HRTFs for all?

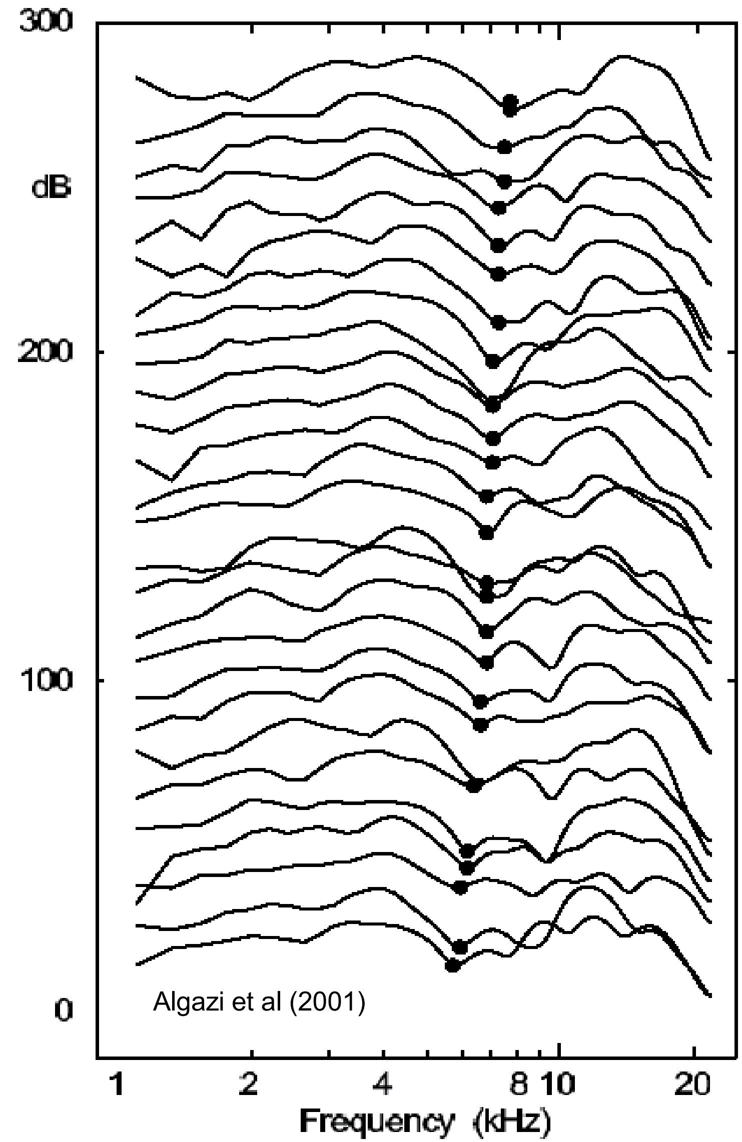
Subject-Dependency of HRTFs

- HRTFs depend on the anthropometry
  - HRTFs of 27 subjects
  - Sorted by the first notch
  - The same position





Algazi et al (2001)

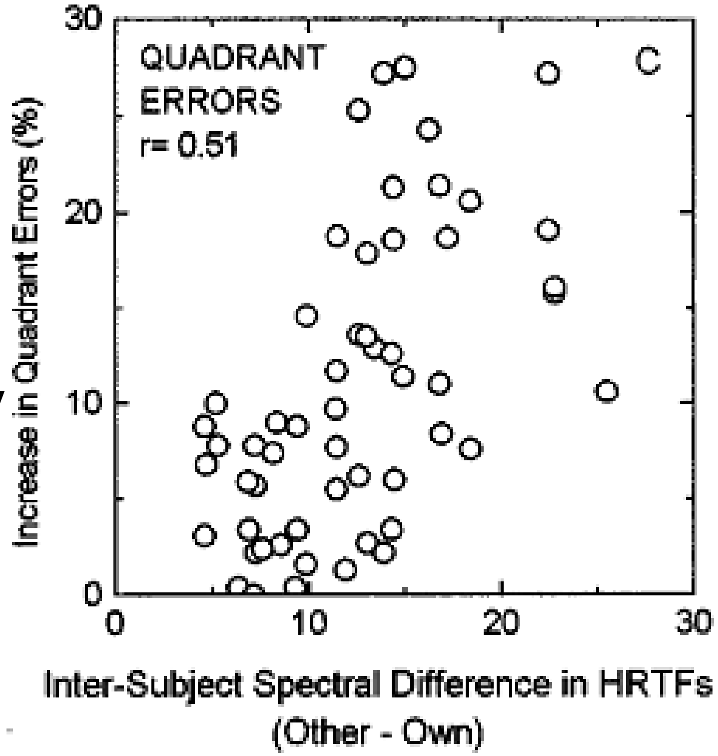


Algazi et al (2001)

Subject-Dependency of HRTFs

- Localization with others' ears?
  - Less externalization
  - More front-back confusions
  - Effect depends on subject compatibility: differences in anthropometry → differences in HRTFs

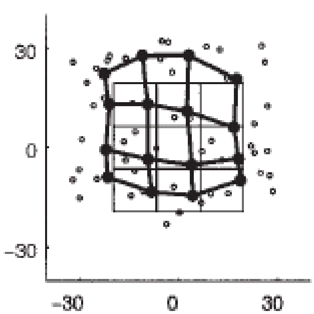
Middlebrooks (1999)

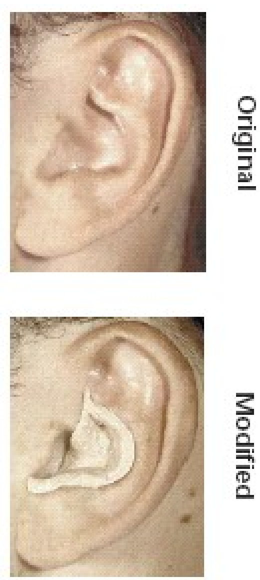


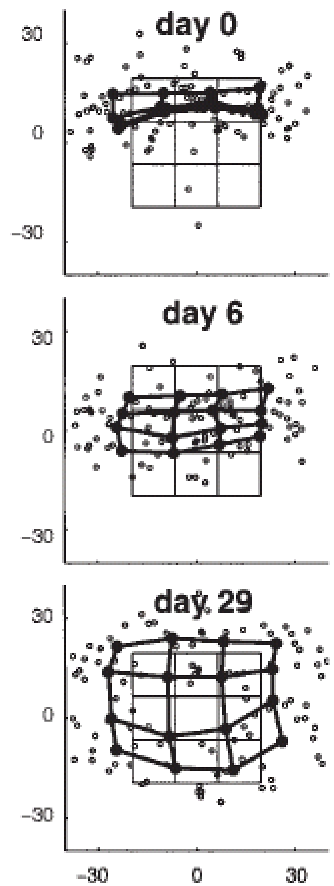
Plasticity in Sound Localization

- Ability to recalibrate the auditory system

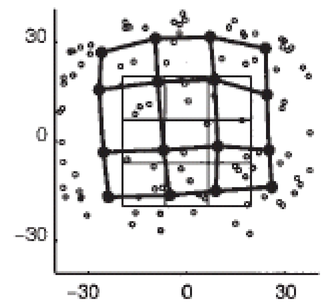
Pre-test







Post-test

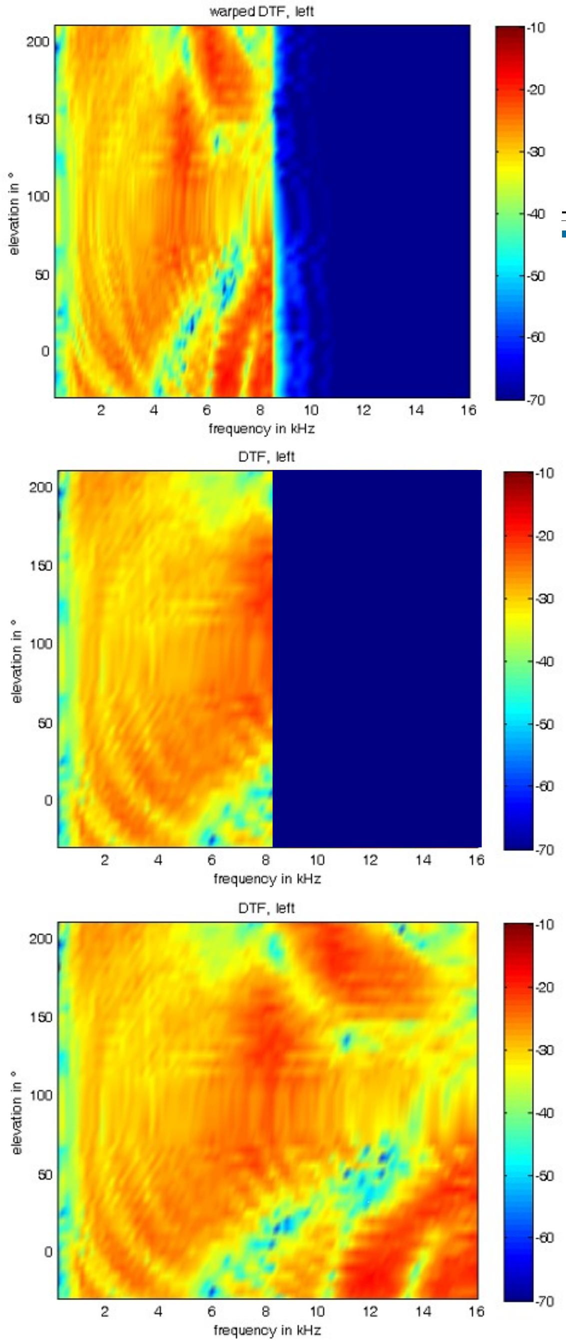


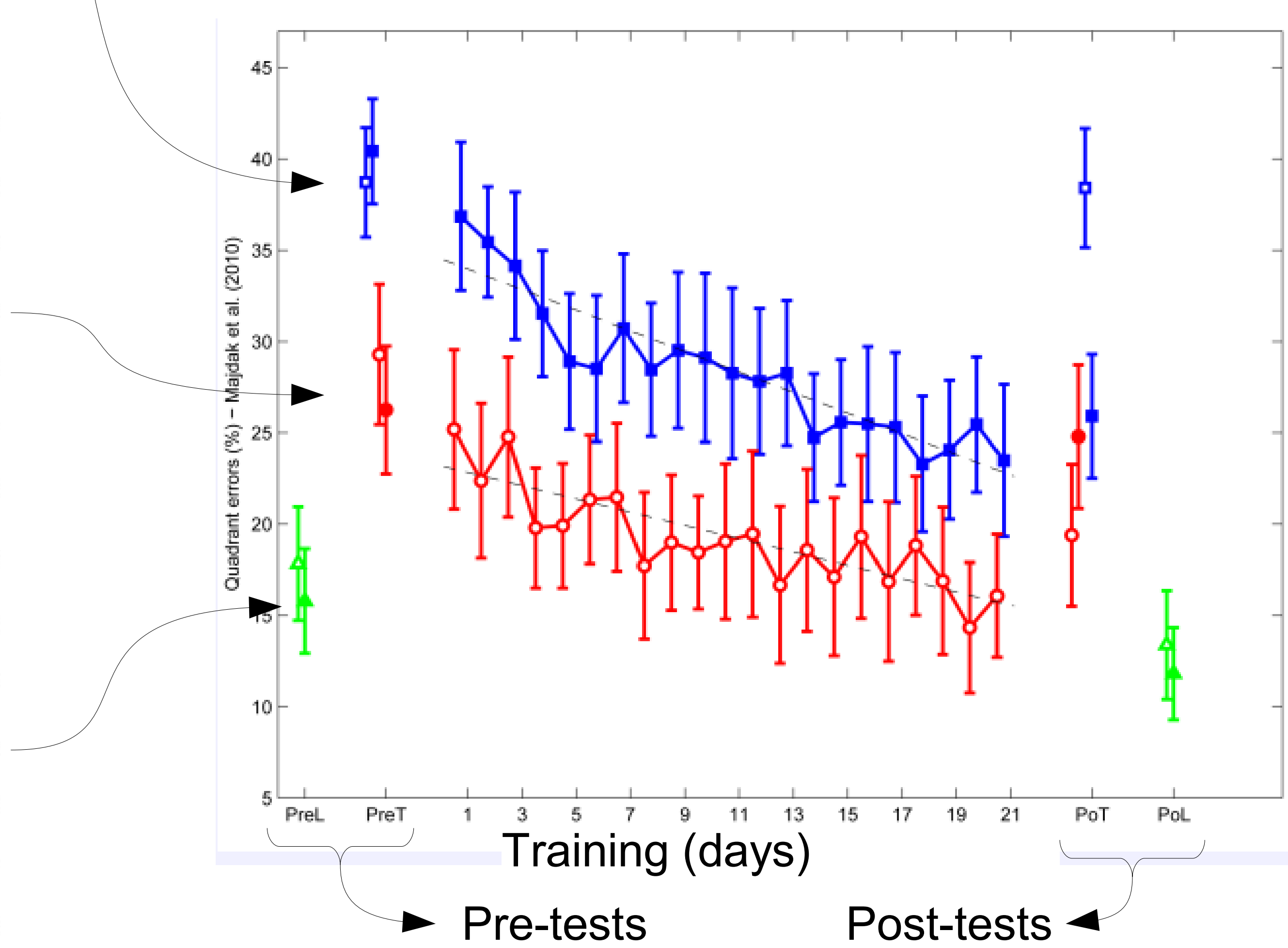
Hofman et al (1998)

Hofman et al (1998)

Plasticity in Sound Localization: Supervised Training

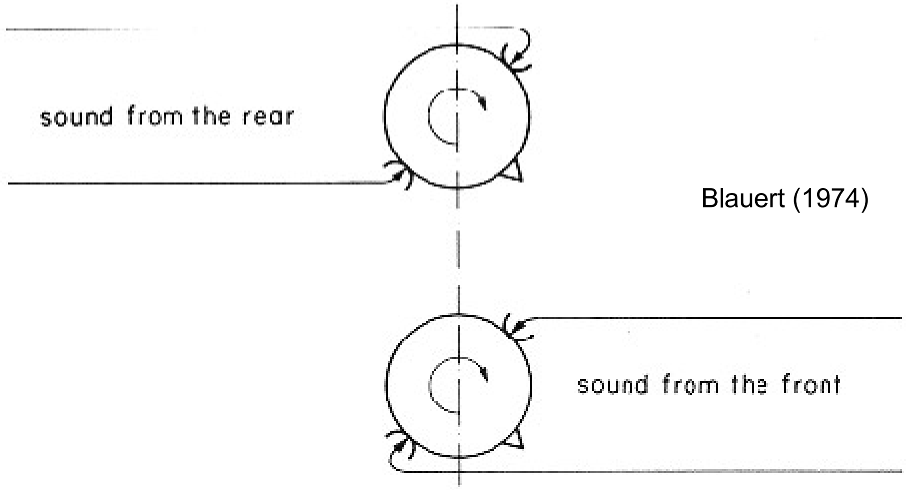
(Walder, Laback, Majdak, 2010)





Further Factors Affecting Sound Localization in Free Field

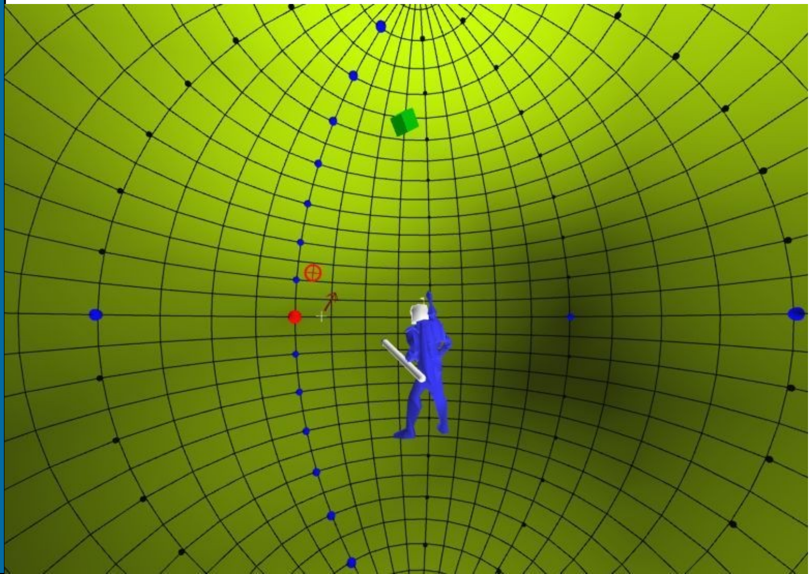
- Head movements:
  – Help to resolve front-back confusions (Perret & Noble , 1997)

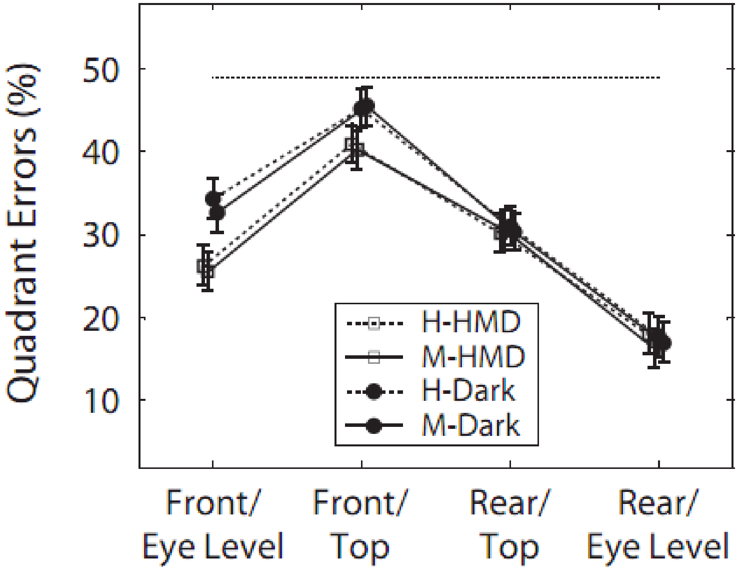


Blauert (1974)

Further Factors Affecting Sound Localization in Free Field

- Head movements:
  - Help to resolve front-back confusions (Perret & Noble , 1997)
- Vision:
  - Visual feedback providing consistent information about the environment helps (Majdak et al. 2010)



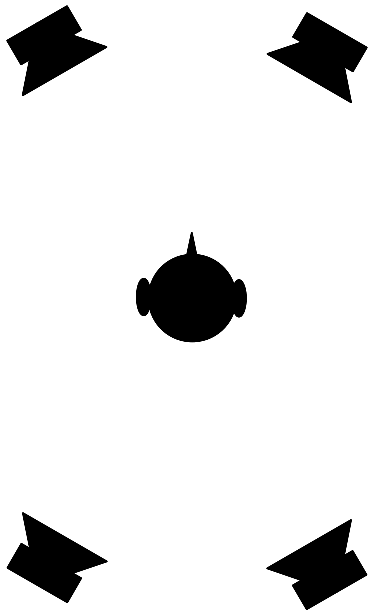


Further Factors Affecting Sound Localization in Free Field

- Head movements:
  - Help to resolve front-back confusions (Perret & Noble , 1997)
- Vision:
  - Visual feedback providing consistent information about the environment helps (Majdak et al. 2010)
- Experience:
  - Training on localization using own HRTFs helps (Majdak et al. 2010)

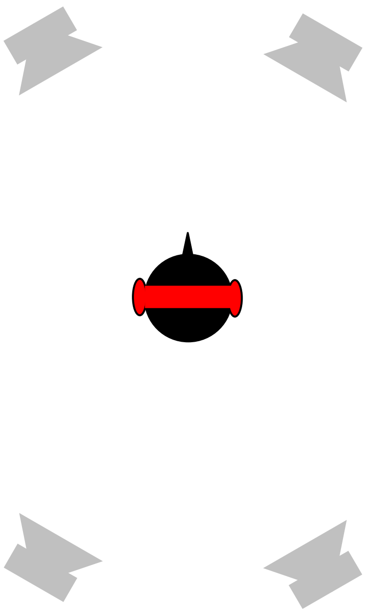
Summary

- Sound localization (single source, no room):
  - Lateral positions:
    - Relevant cues: Binaural and broadband
    - Spectral shape negligible
    - Easily derived from anthropometry
  - Vertical positions (also front vs. back):
    - Relevant cues: Monaural spectral shape
    - Complex relation to the anthropometry
    - Individualized HRTFs required
      - Generic HRTFs: problem not solved yet
    - Long-term recalibration to modified HRTFs possible
      - A way towards an optimized generic HRTF?



Summary

- Sound localization (single source, no room):
  - Lateral positions:
    - Relevant cues: Binaural and broadband
    - Spectral shape negligible
    - Easily derived from anthropometry
  - Vertical positions (also front vs. back):
    - Relevant cues: Monaural spectral shape
    - Complex relation to the anthropometry
    - Individualized HRTFs required
      - Generic HRTFs: problem not solved yet
    - Long-term recalibration to modified HRTFs possible
      - A way towards an optimized generic HRTF?



Summary

- Sound localization (single source, no room):
  - Lateral positions:
    - Relevant cues: Binaural and broadband
    - Spectral shape negligible
    - Easily derived from anthropometry
  - Vertical positions (also front vs. back):
    - Relevant cues: Monaural spectral shape
    - Complex relation to the anthropometry
    - Individualized HRTFs required
      - Generic HRTFs: problem not solved yet
    - Long-term recalibration to modified HRTFs possible
      - A way towards an optimized generic HRTF?