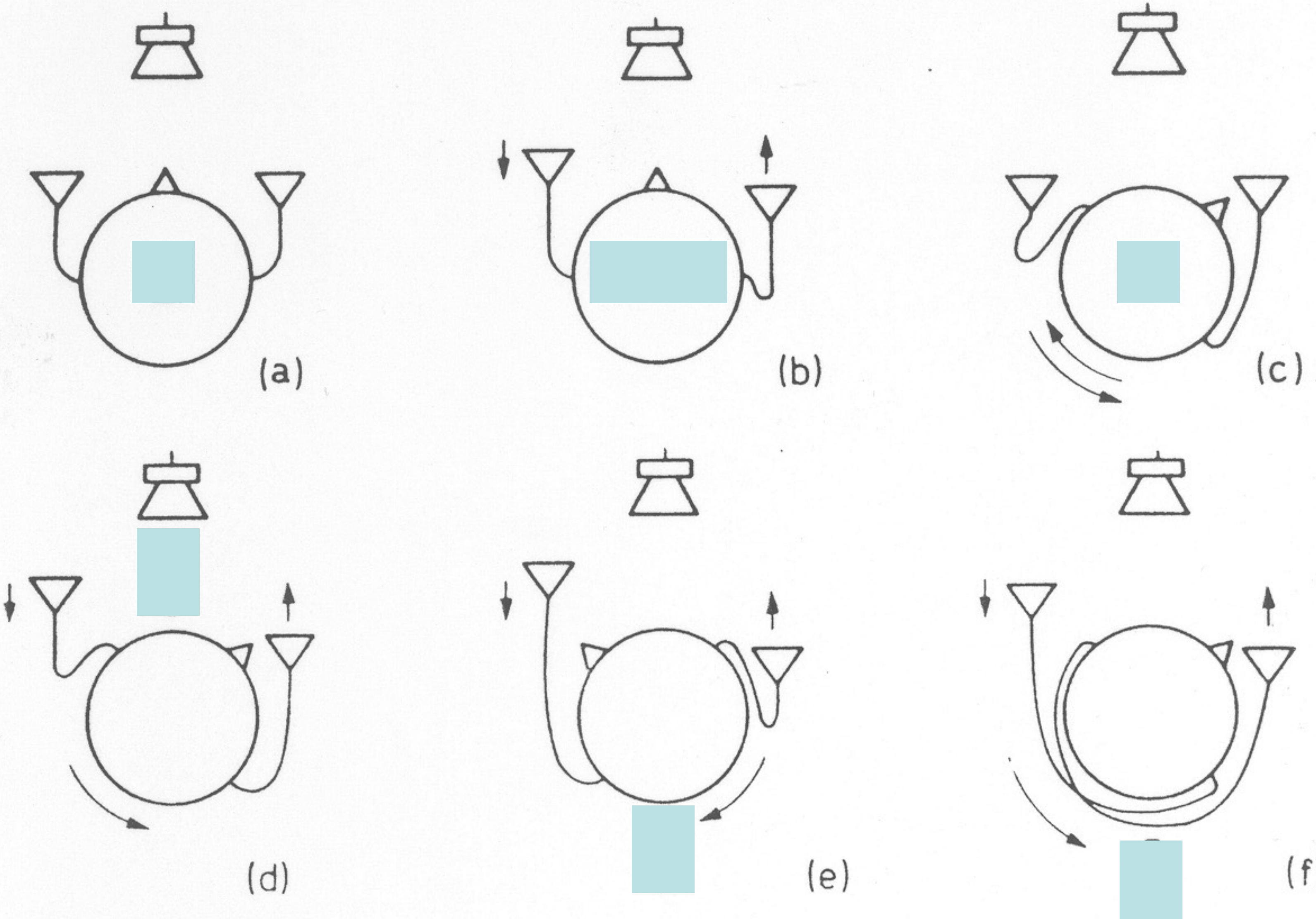
The Problem of Sound Localization

Binaural Hearing and Sound Source Localization

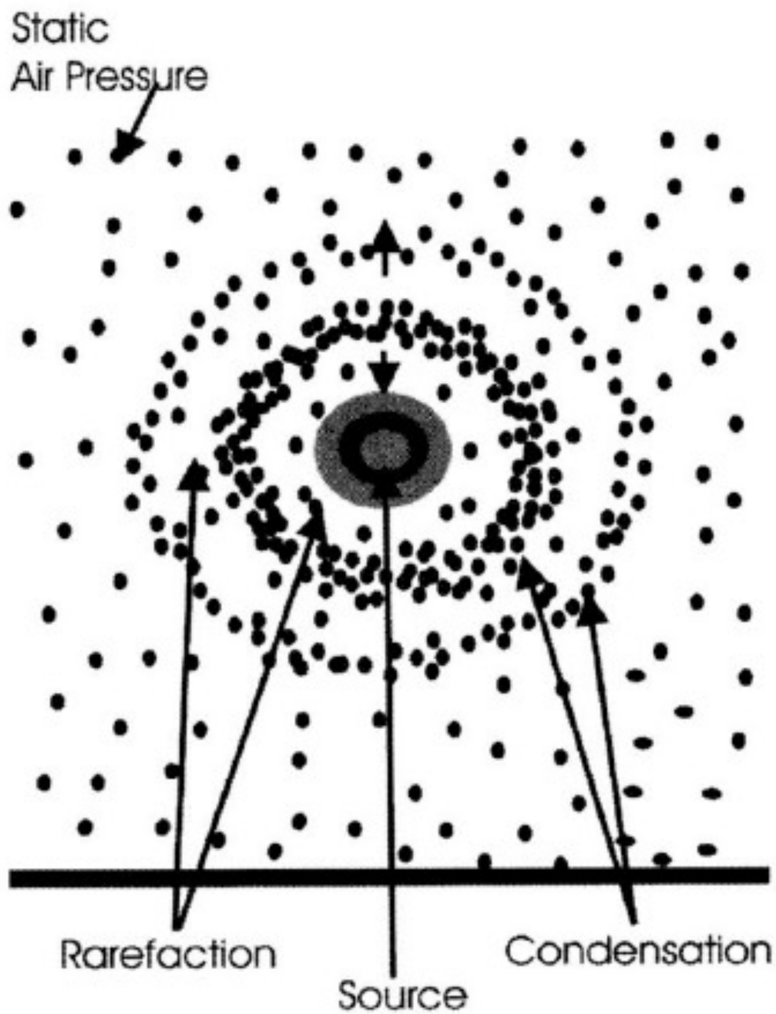
Wu Xihong

School of Artificial Intelligence


(a)
(b)
(c)
(d)
(e)
(f)

Basic knowledge of acoustics

Sound propagation

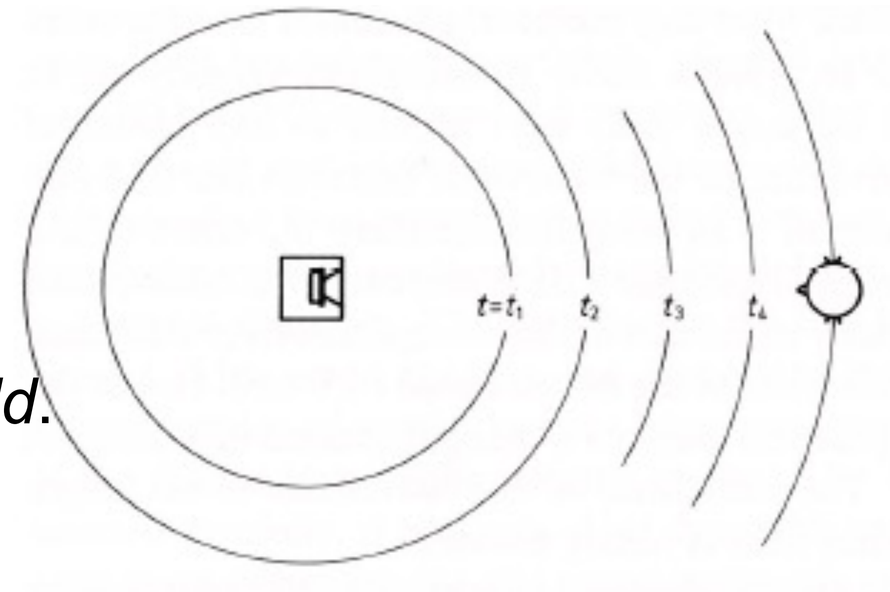


How does sound level attenuate?

intensity = power / unit area

$$I = P / A = P / (4\pi r^2) \Rightarrow 6\, dB\ per\ 2r$$

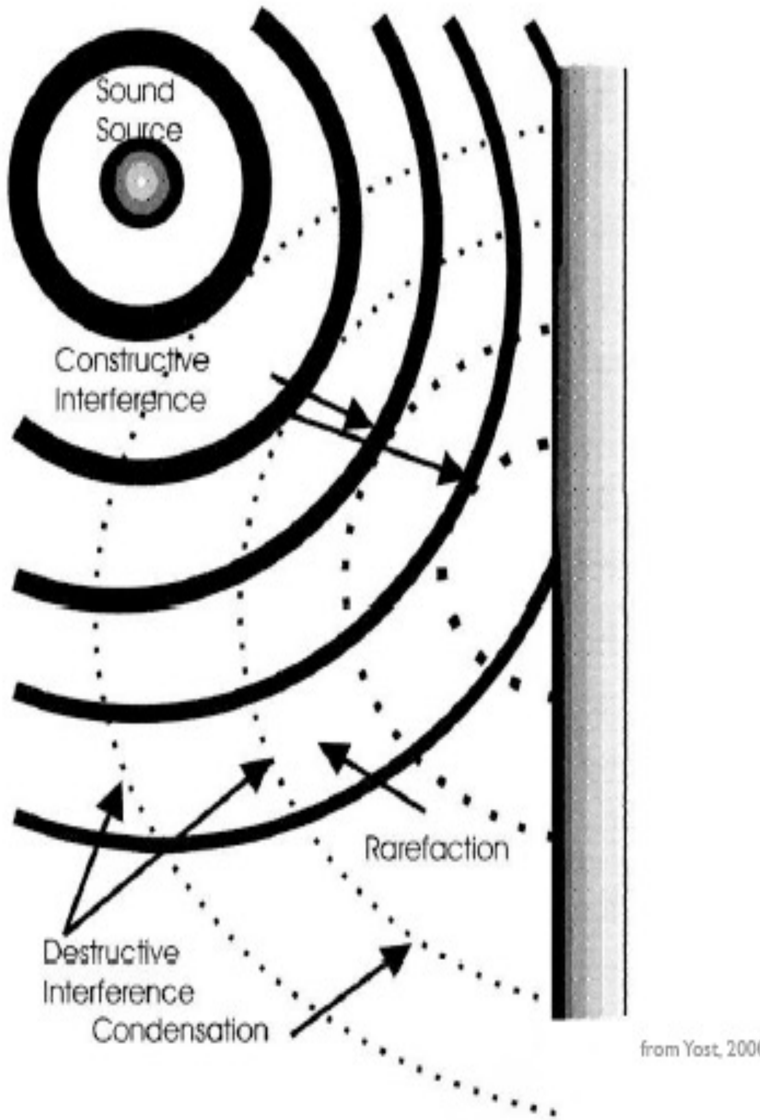
This equation holds for a *free field*.



from Blauert, 1997

How does this fail?

Reflection of sound energy



Sound propagation and objects

- Reflection depends on the relative size of the wavelength and the object:

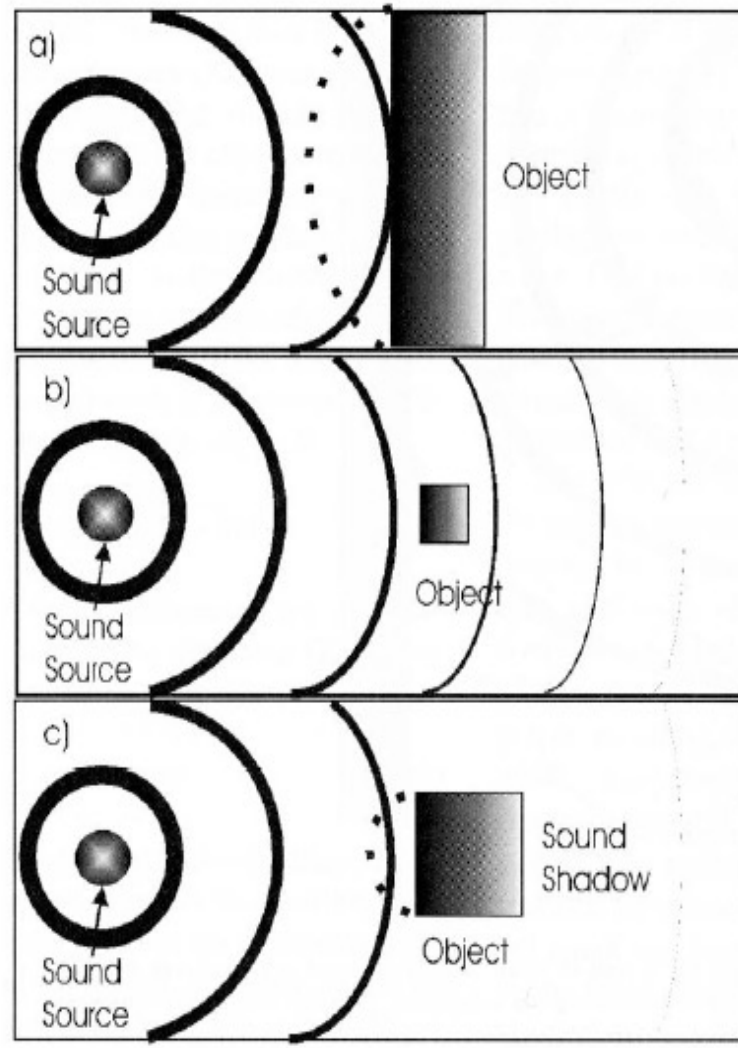
  - “**small**” wavelengths are reflected and create *interference*

  - “**large**” wavelengths pass by an object

  -“**intermediate** wavelengths cast an *acoustic shadow*

- Relevant wavelengths for a typical human head:

  345 m/s / 0.18 m = 1916 Hz



from Yost, 2000

Prominent Features of Binaural Hearing

- Localization

  Formation of Positions of the Auditory Events

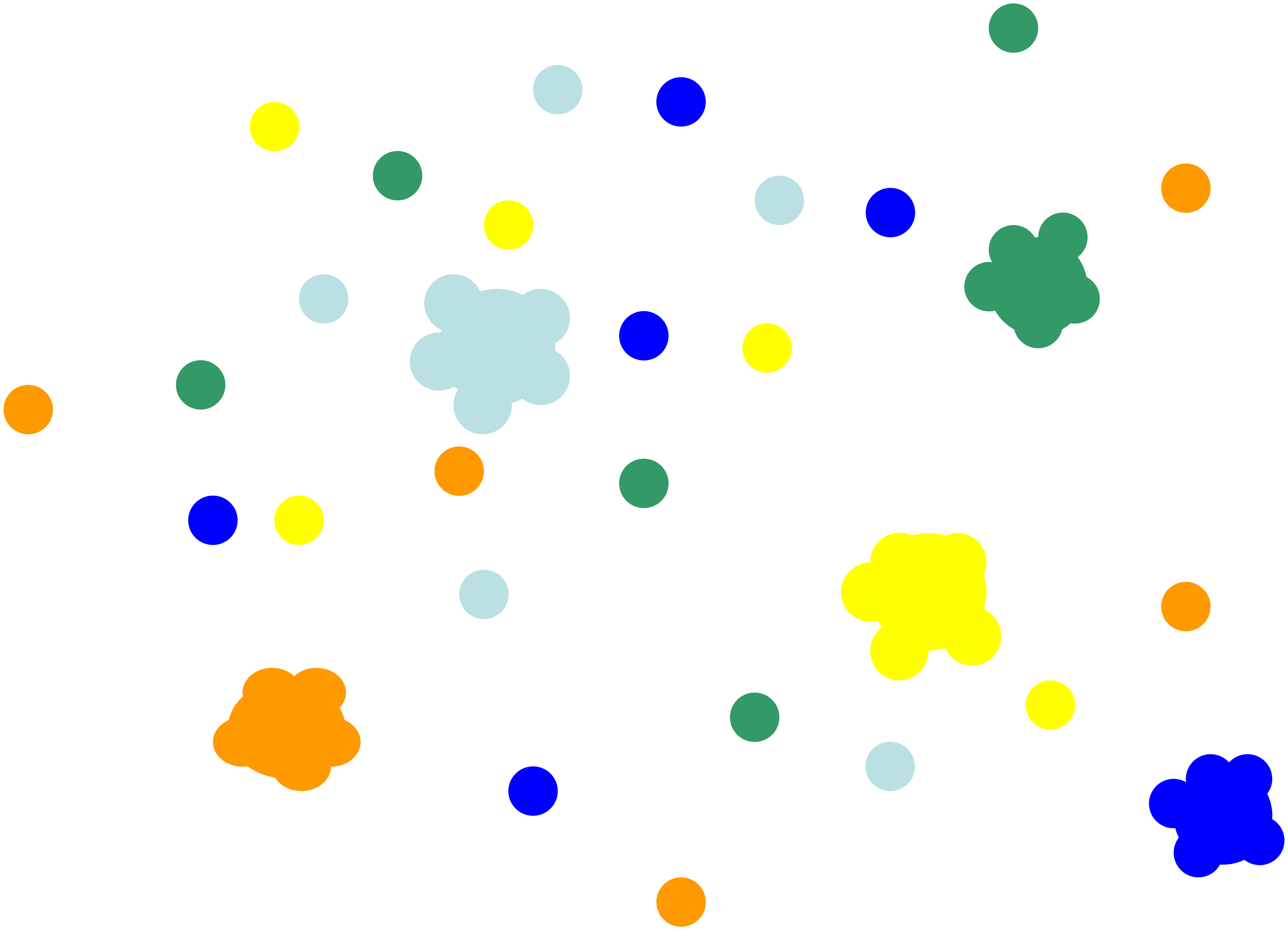
  i.e., Azimuth, Elevation & Distance, Spatial Extent of Auditory Events

- Suppression of

  the Directional Information coming from Reflections

  e.g., Precedence Effect, Localization Dominance, Fusion, Reverberation, Coloration and Noise

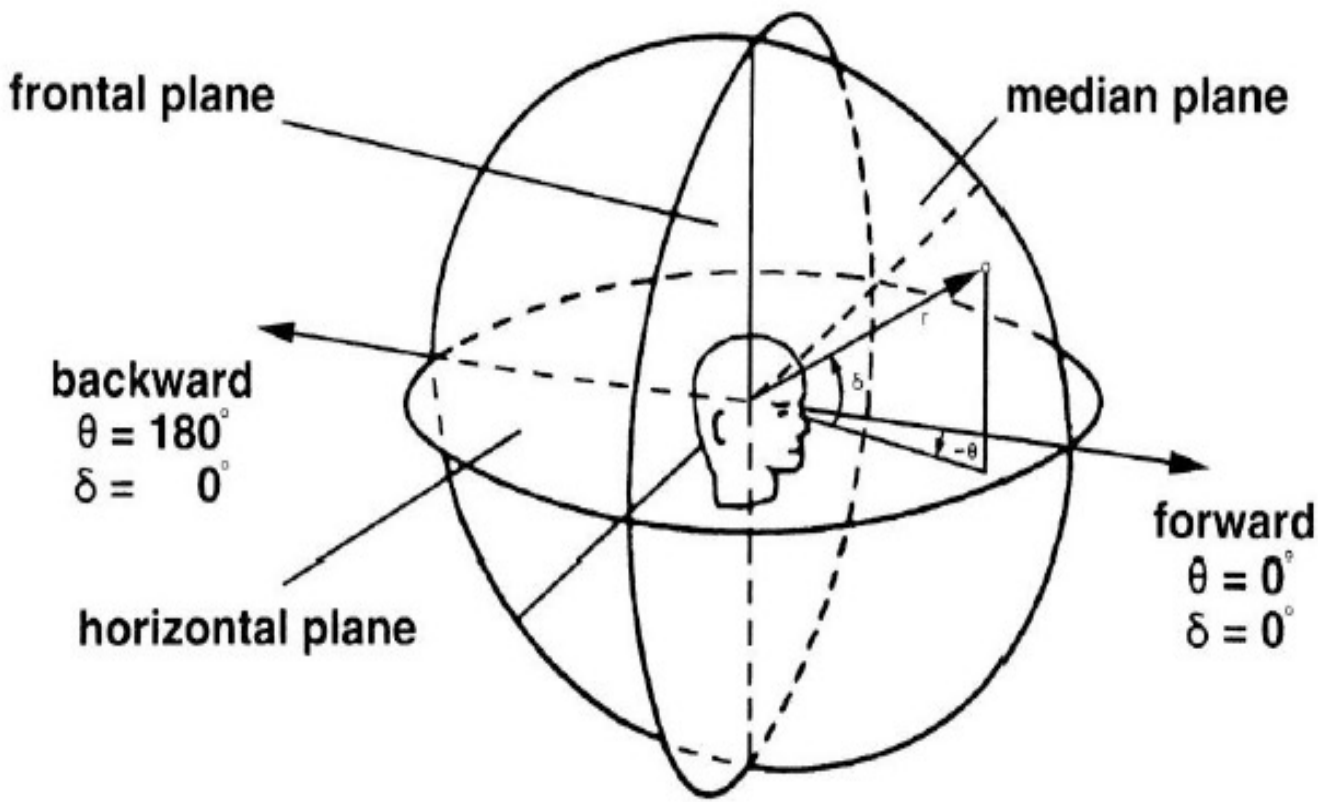
Auditory Scene



Sound source localization

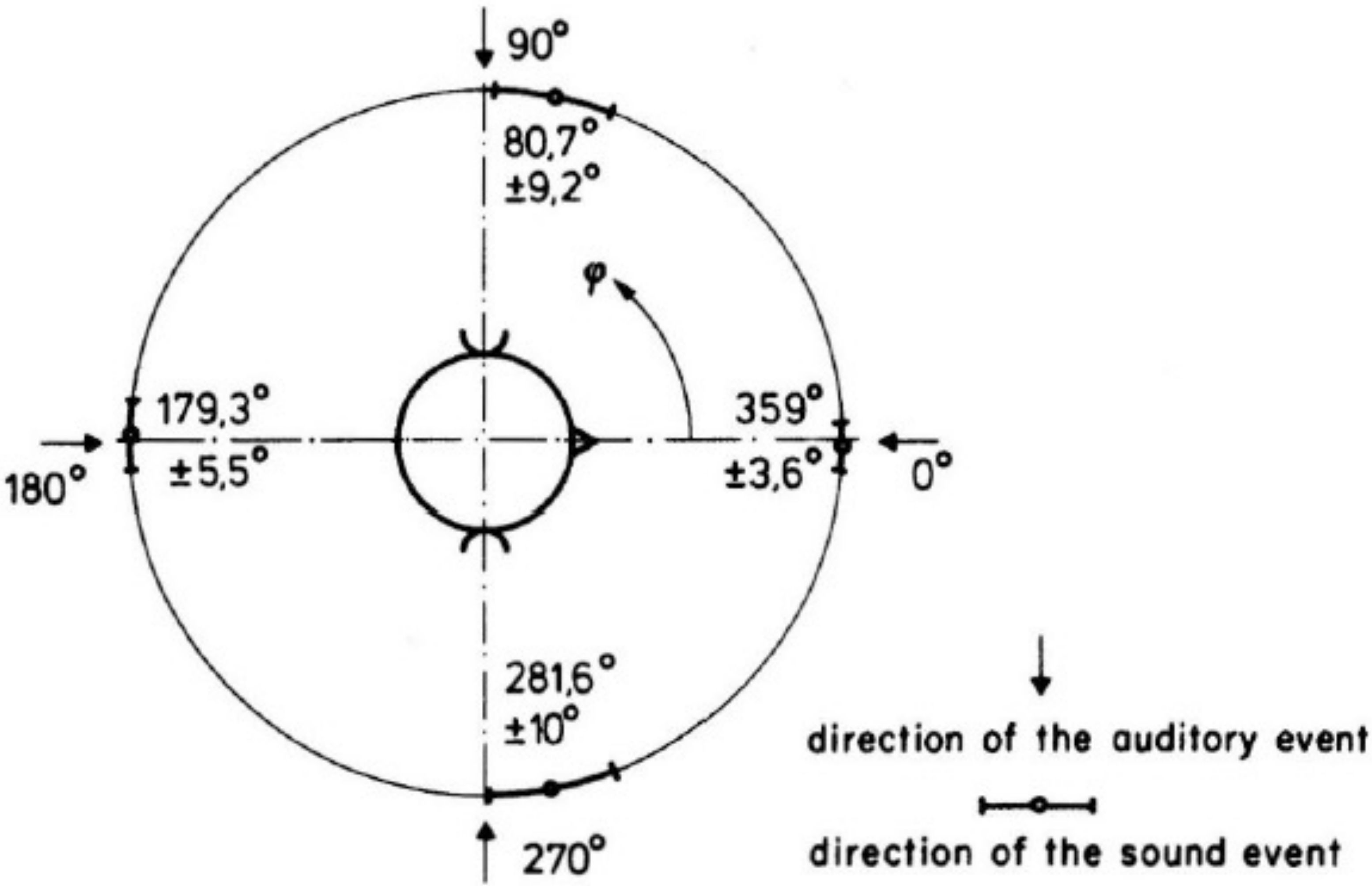
--Formation of Positions of the Auditory Events

Localization in 3D space

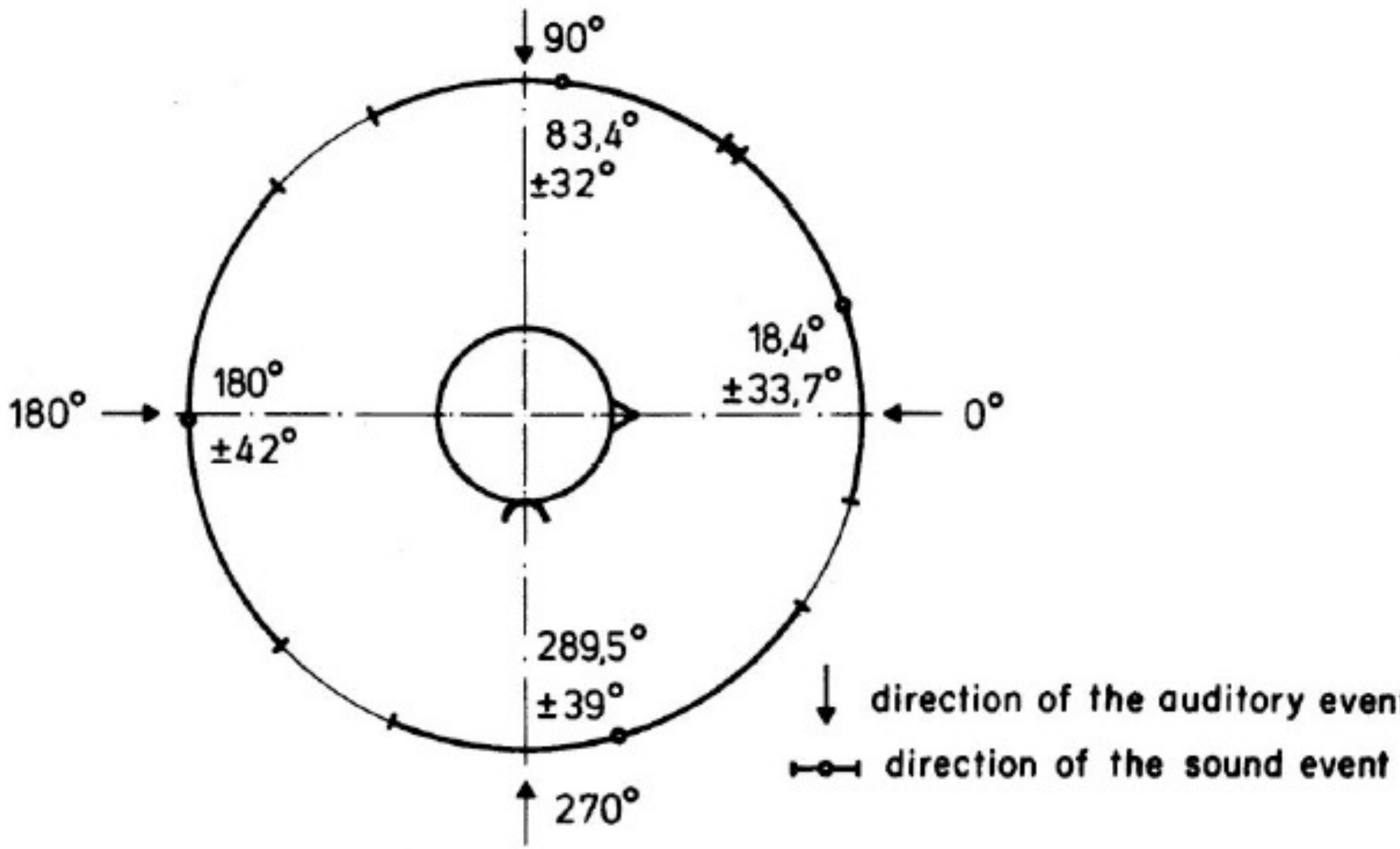


from Moore, 1997

Direction acuity: horizontal plane



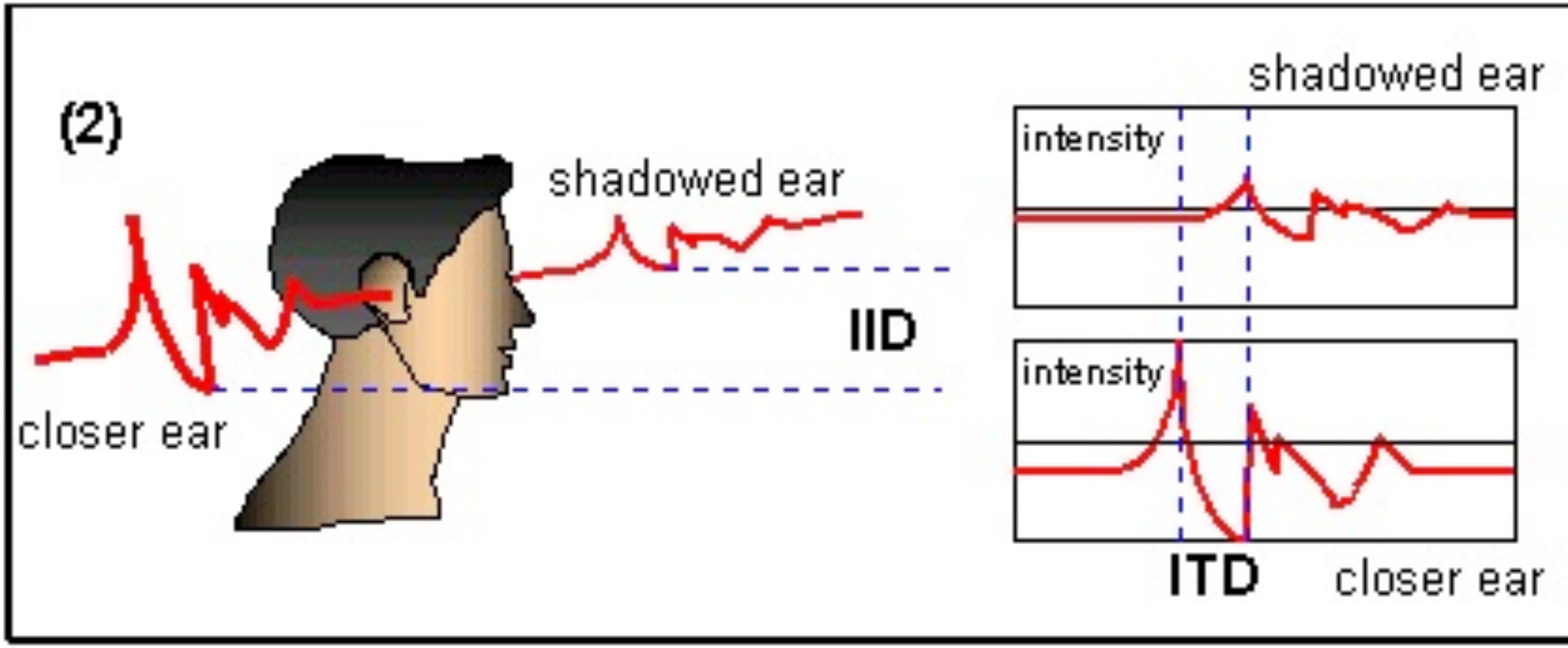
Direction acuity: medial plane



Azimuth for pure tones

- Interaural Level Difference – ILD
- Interaural Time Difference – ITD

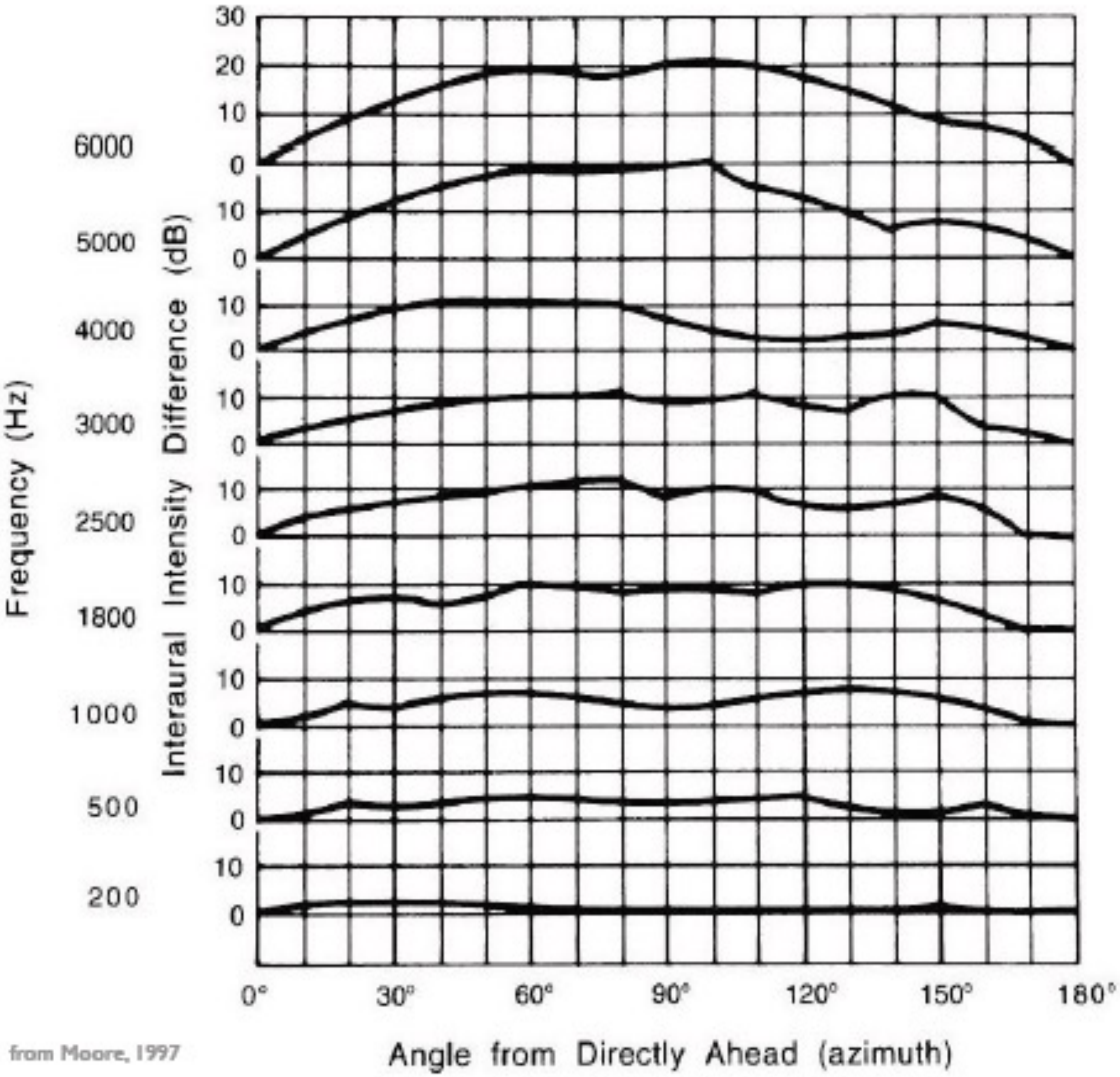
Interaural Level Difference (ILD)



From David McAlpine

Processed in Lateral Superior Olive

Measured ILDs



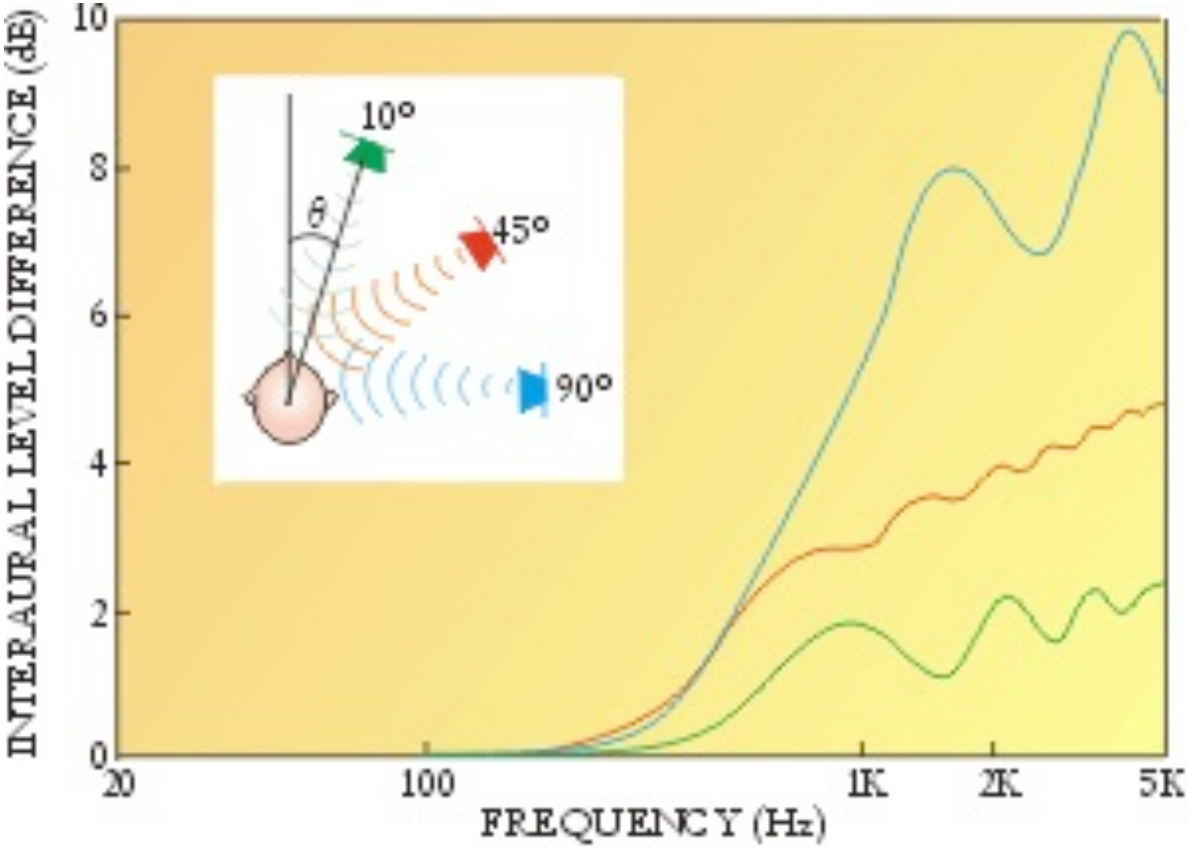
from Moore, 1997

What can you observe?

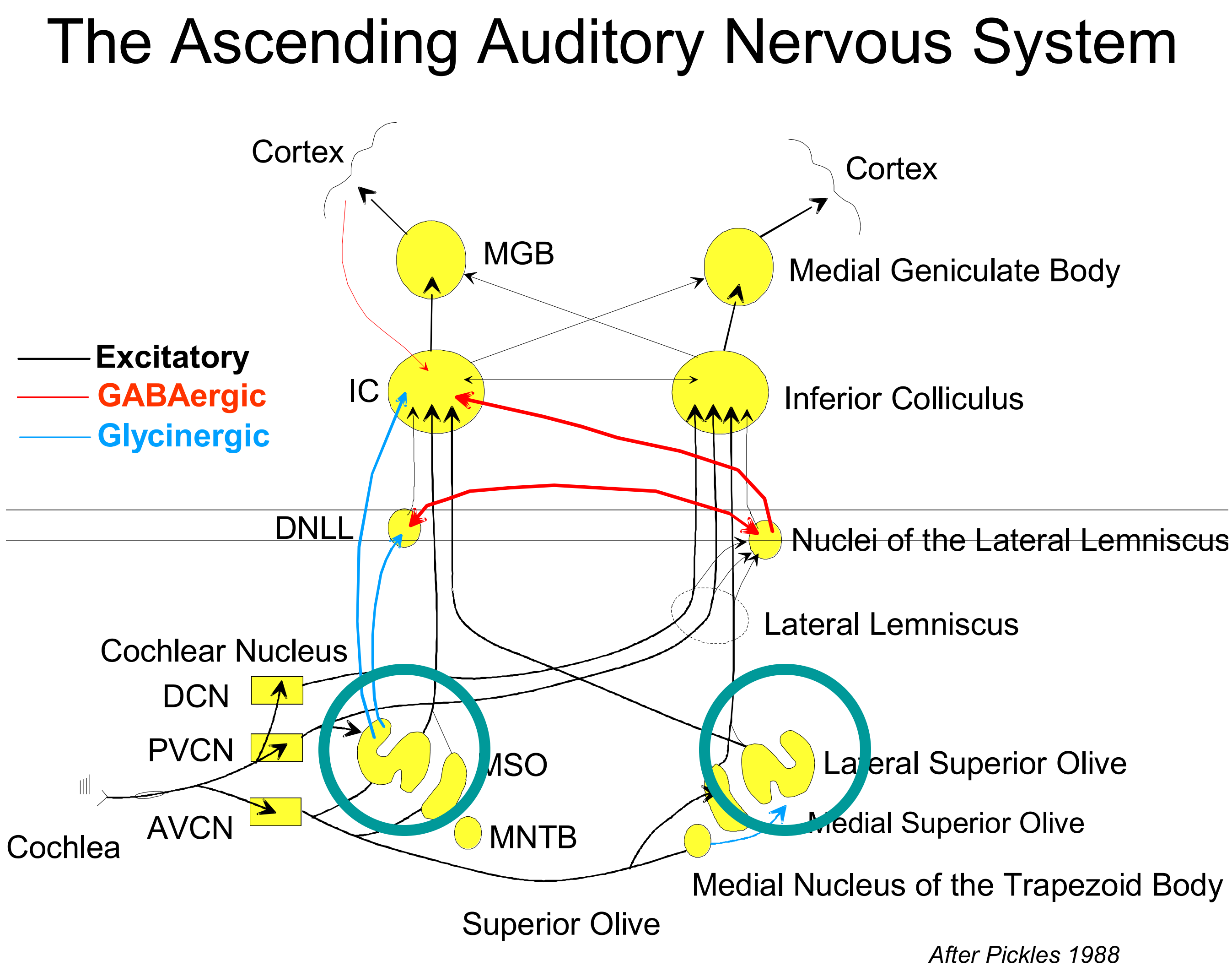
ILD is greater for higher frequencies

Interaural level differences calculated for a source in the horizontal plane. The source is at an azimuth θ of 10° (green curve), 45° (red), or 90° (blue) relative to straight ahead.

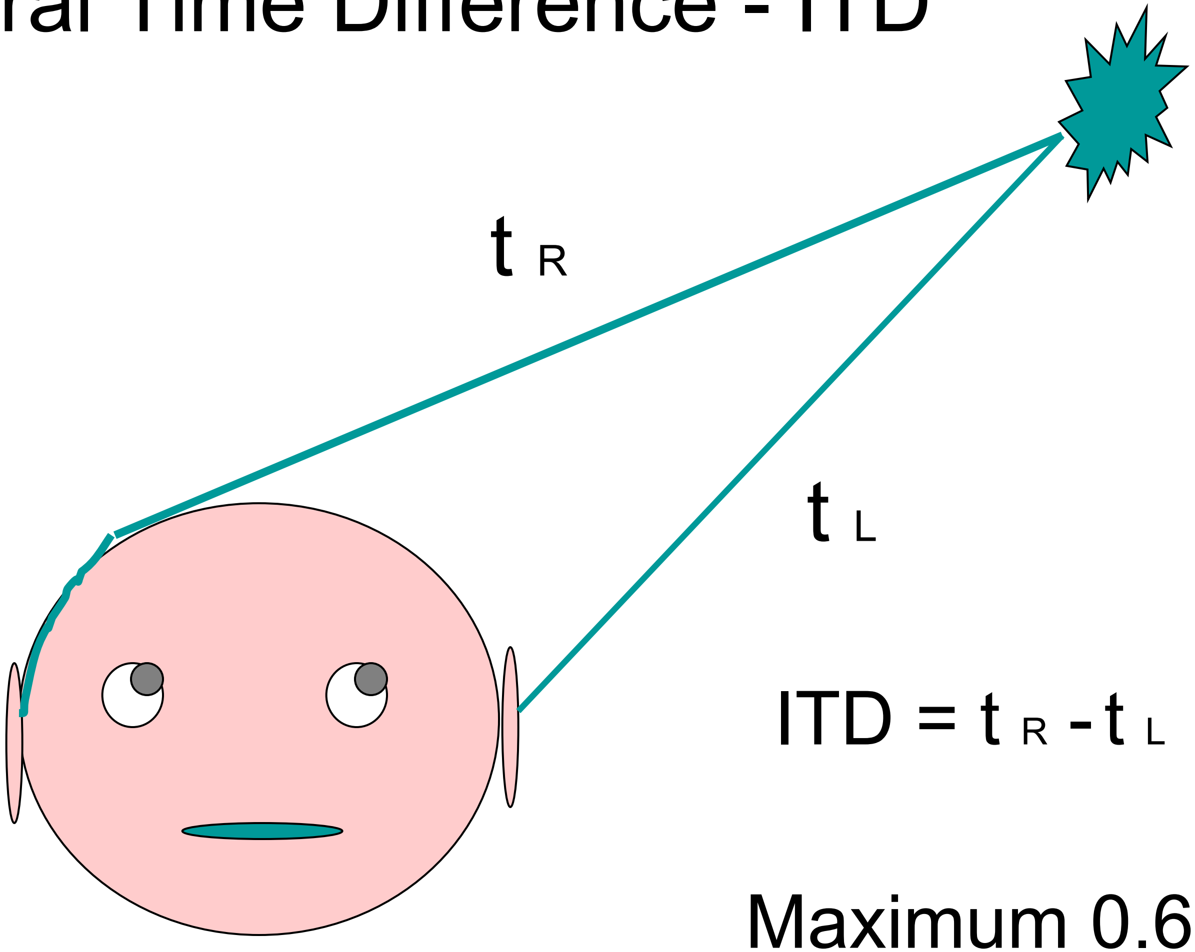
The calculations assume that the ears are at opposite poles of a rigid sphere.



Anatomy of the auditory system



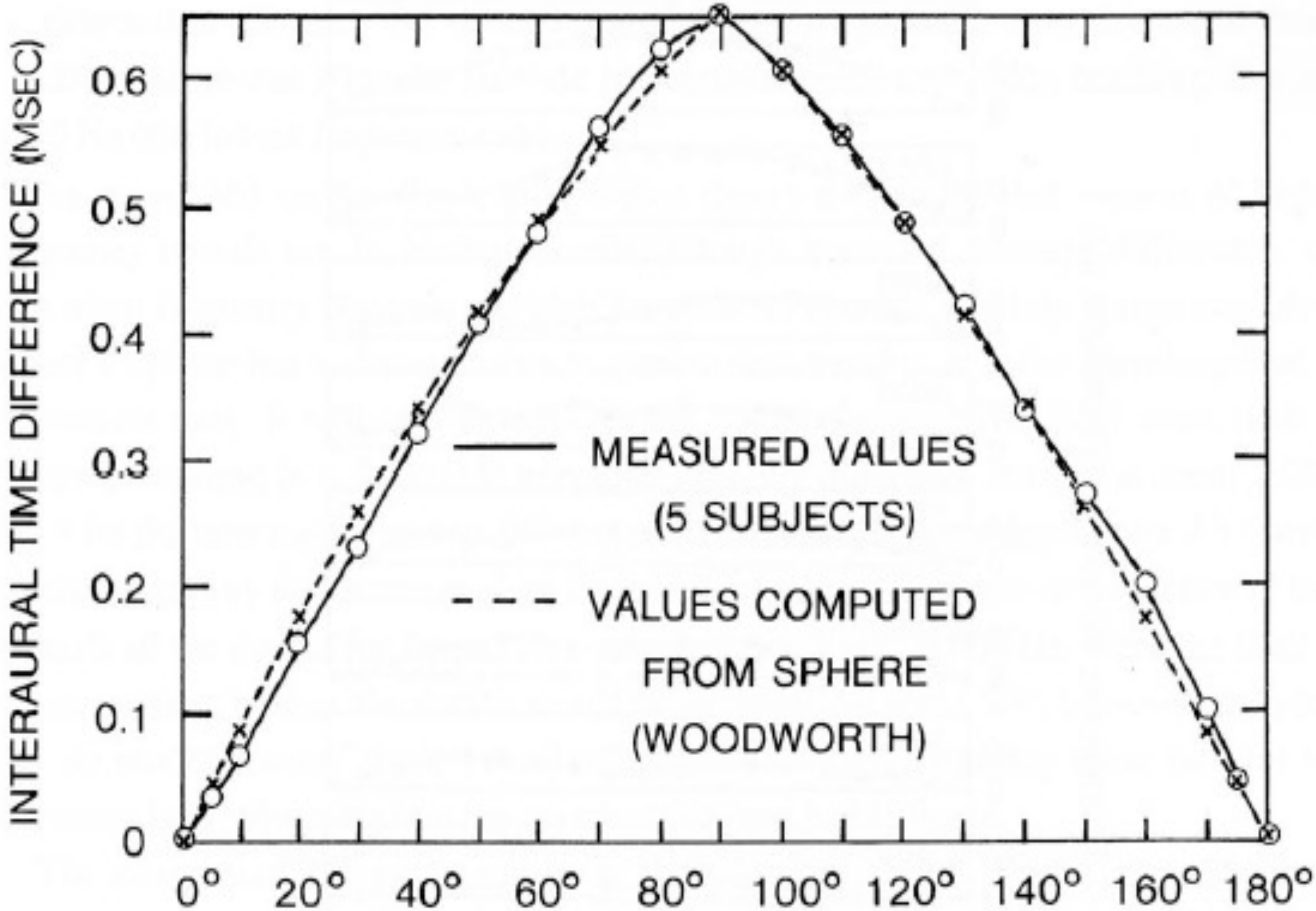
Interaural Time Difference - ITD



$$ITD = t_R - t_L$$

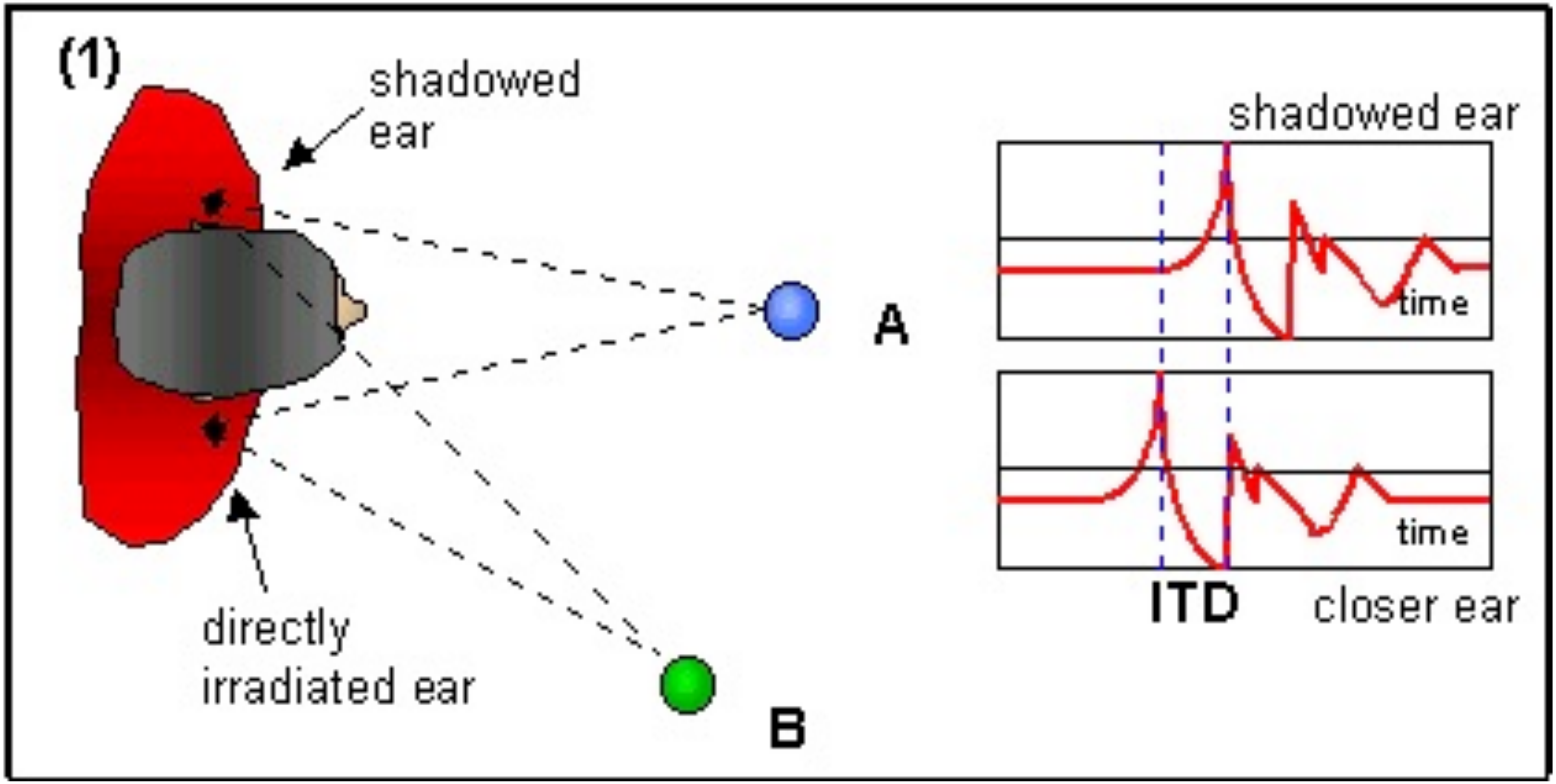
Maximum 0.6 ms

ITD: measured vs predicted



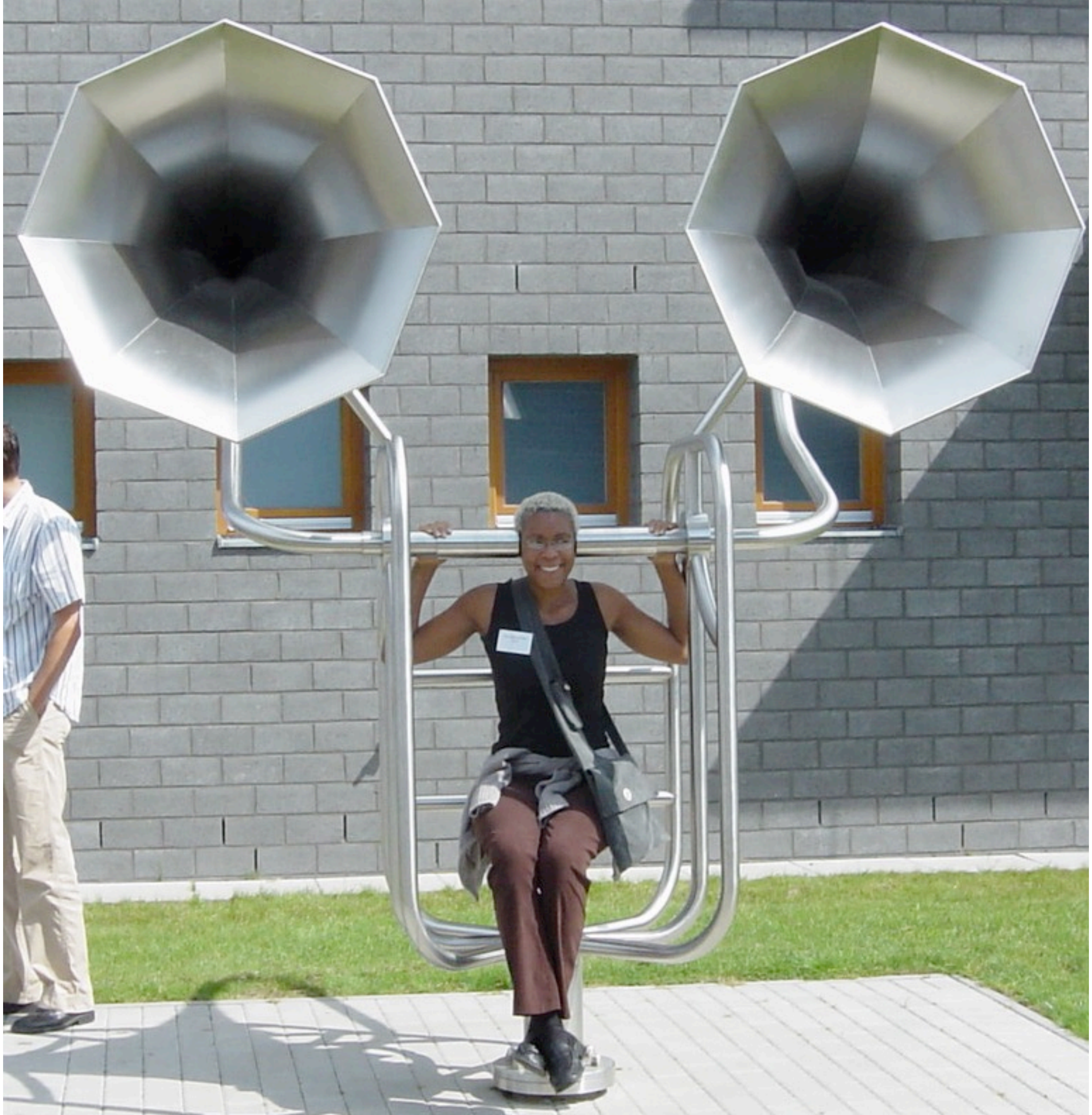
from Warren, 1999

Interaural Time Difference (ITD)

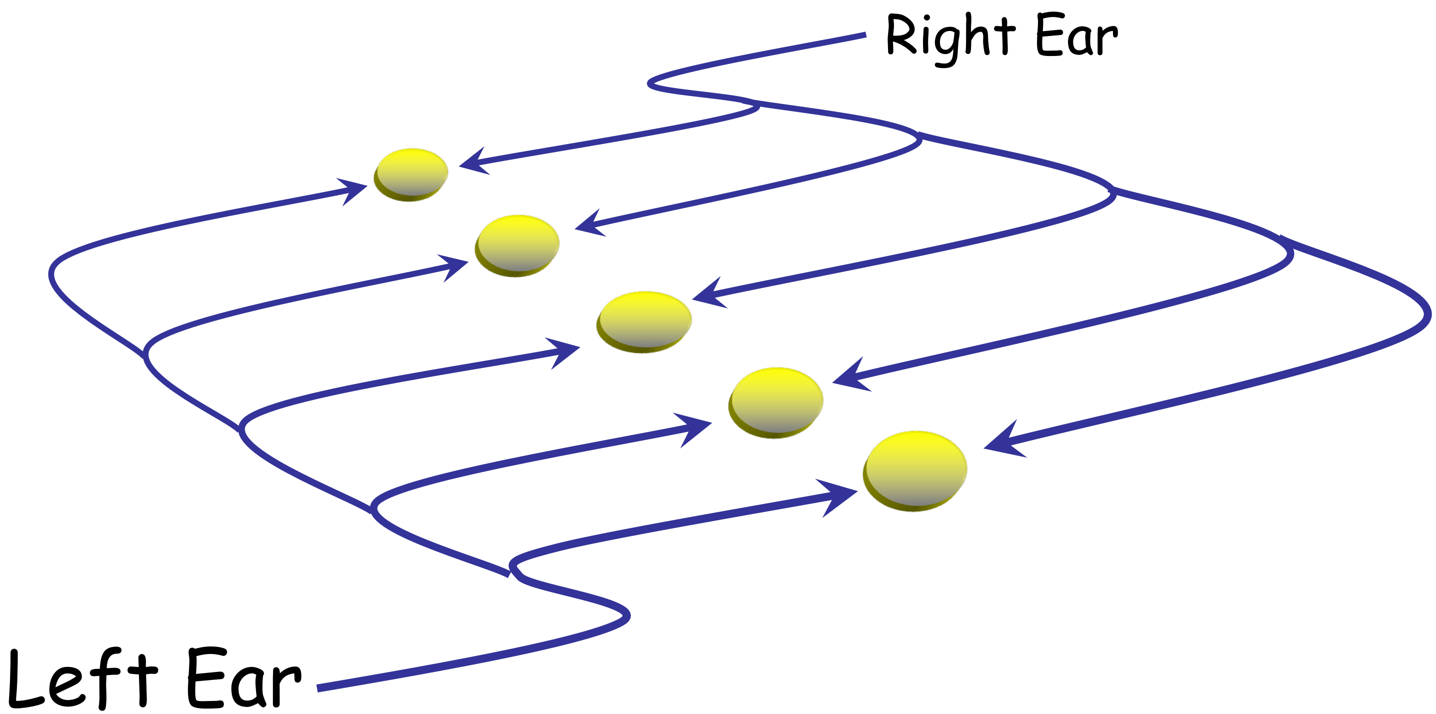


From David McAlpine

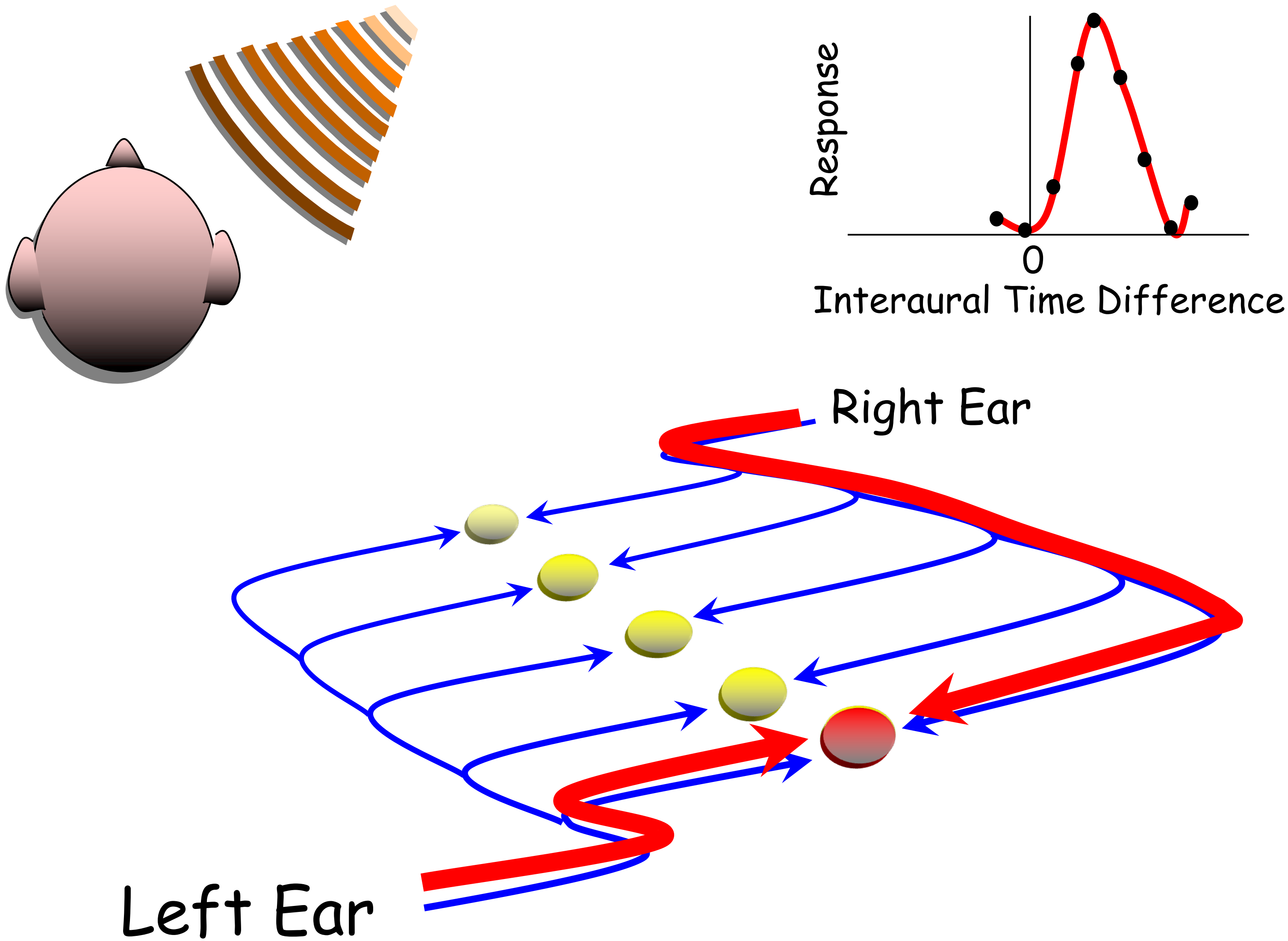
Processed in Medial Superior Olive



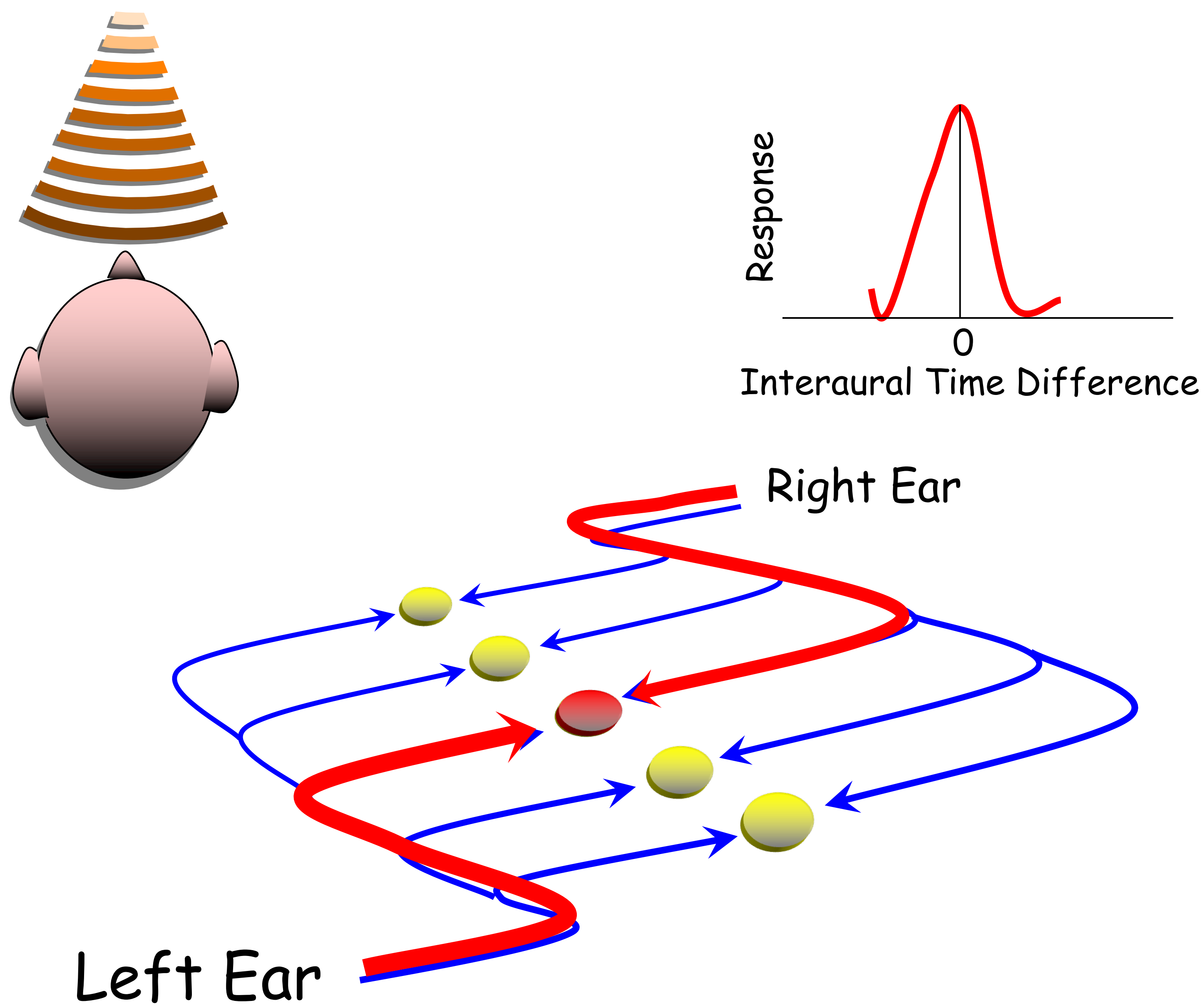
The coincidence detection model of Jeffress (1948) is the widely accepted model for low-frequency sound localisation



From David McAlpine

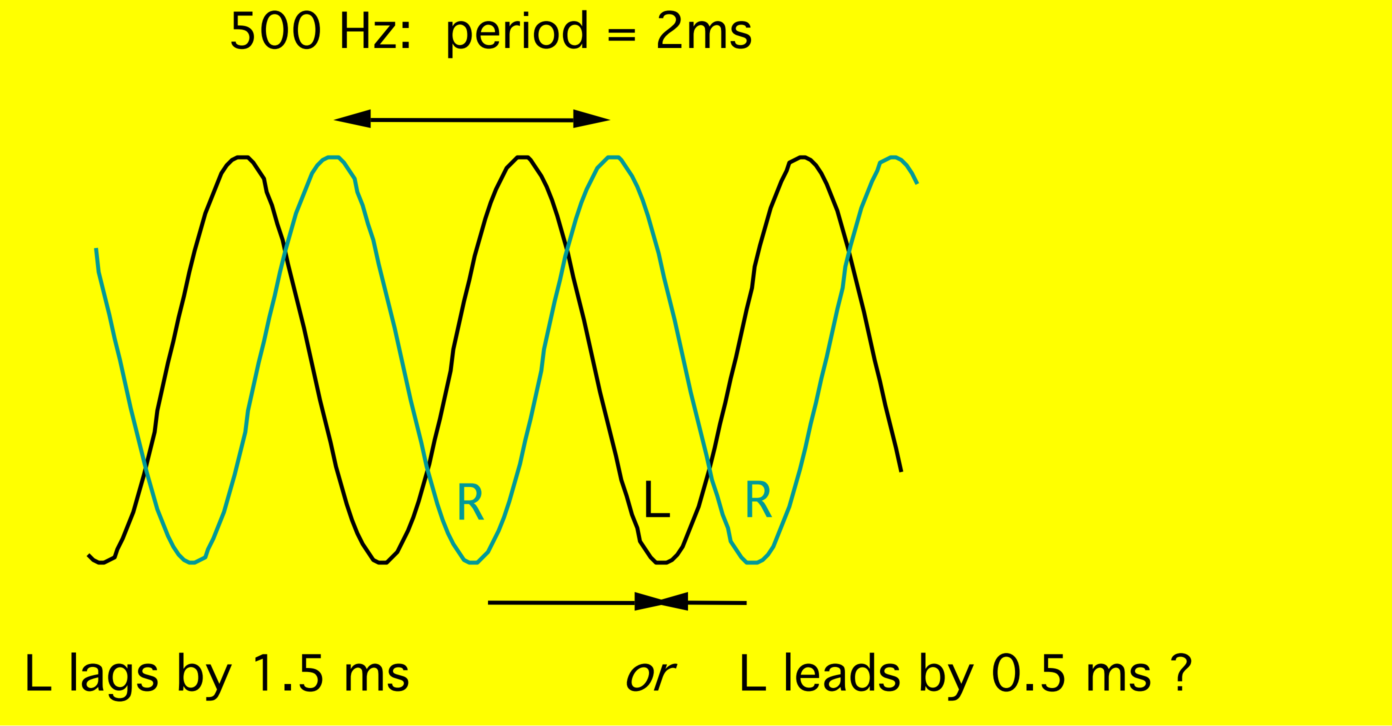


From David McAlpine



From David McAlpine

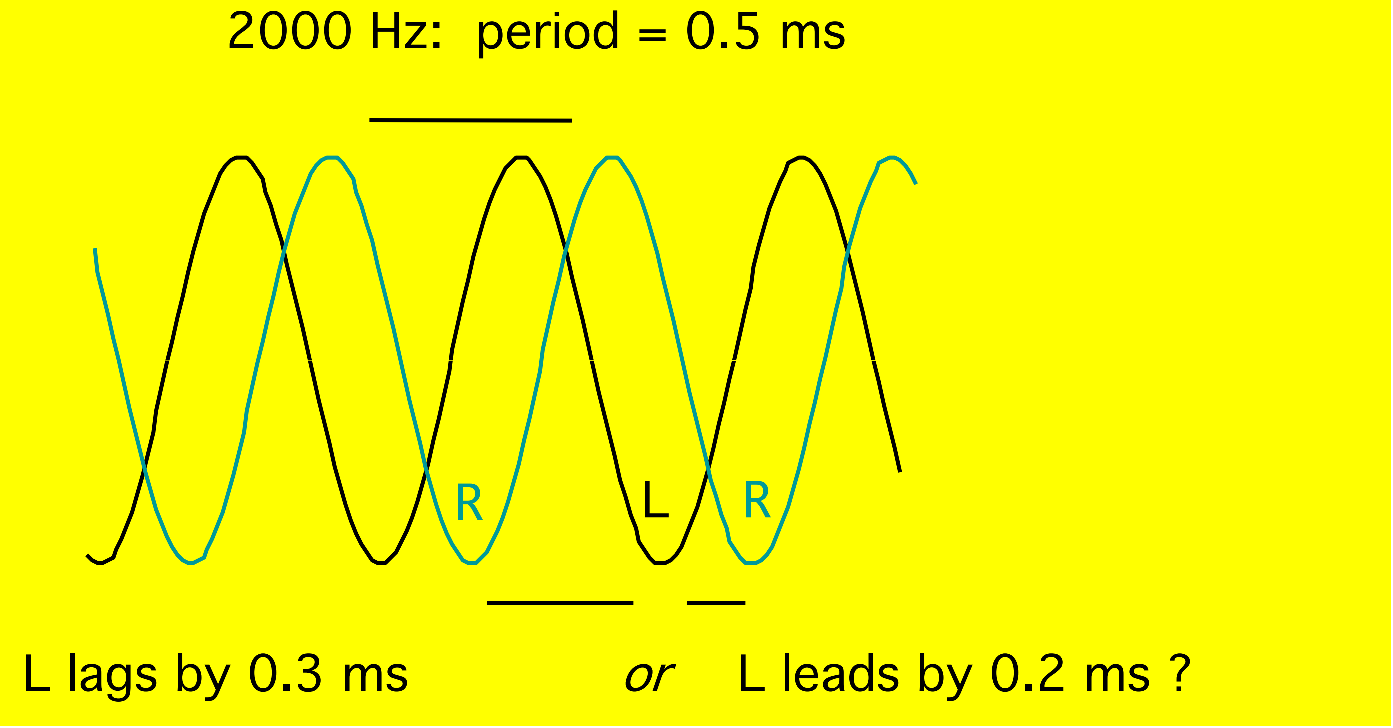
Phase Ambiguity



This particular case is not a problem since max ITD = 0.6 ms

But for frequencies above 1500 Hz it IS a problem

Phase Ambiguity



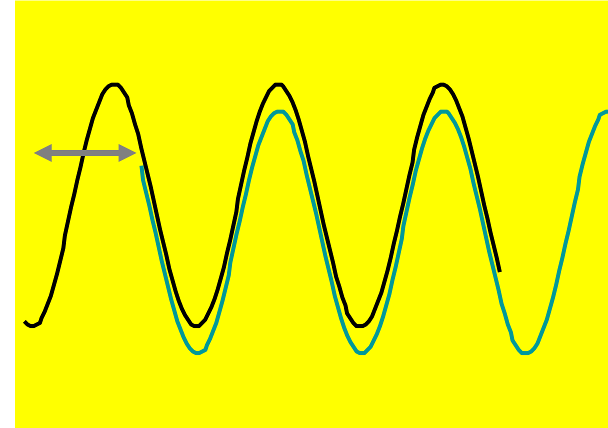
Both possible times are less than the maximum ITD of 0.6 ms

Onset-time

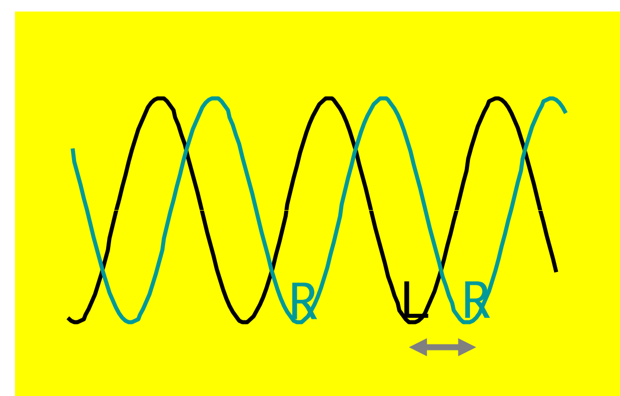
versus

ongoing phase differences

Works for high- and low-frequency sounds

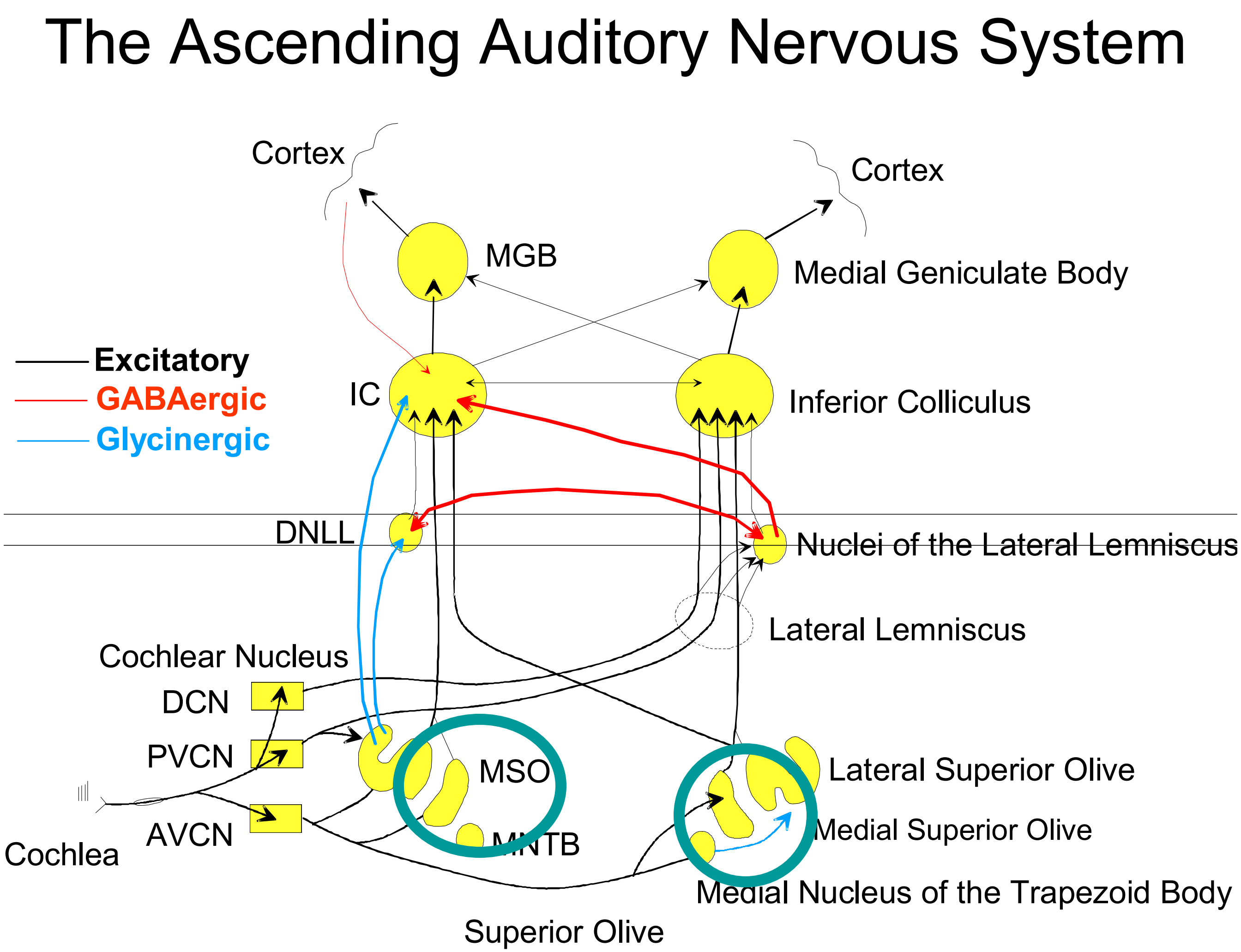


Does not work for high-frequency pure tones:
- no phase locking above 4kHz
- phase ambiguity above 1.5 kHz



Natural sounds have both

Anatomy of the auditory system



After Pickles 1988

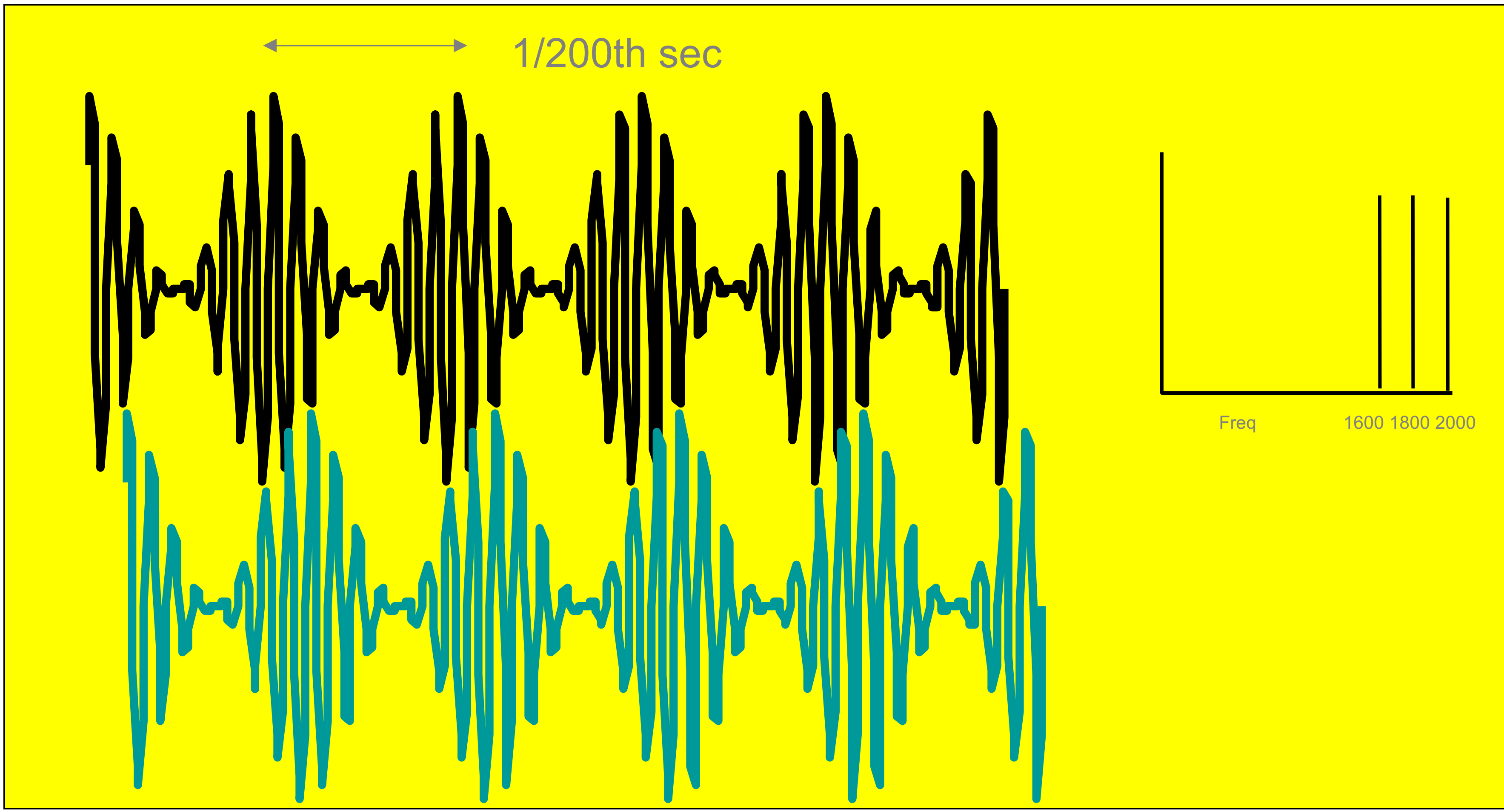
Rayleigh’s Duplex theory for pure tones

- Low frequency pure tones (<1500 Hz) localised by interaural time differences
  - Very small interaural level difference for low-frequency tones.
  - Phase locking present for low frequency tones (<4kHz)
  - Limited by phase ambiguity: Maximum ITD= 670 μs corresponding to a whole cycle at 1500 Hz

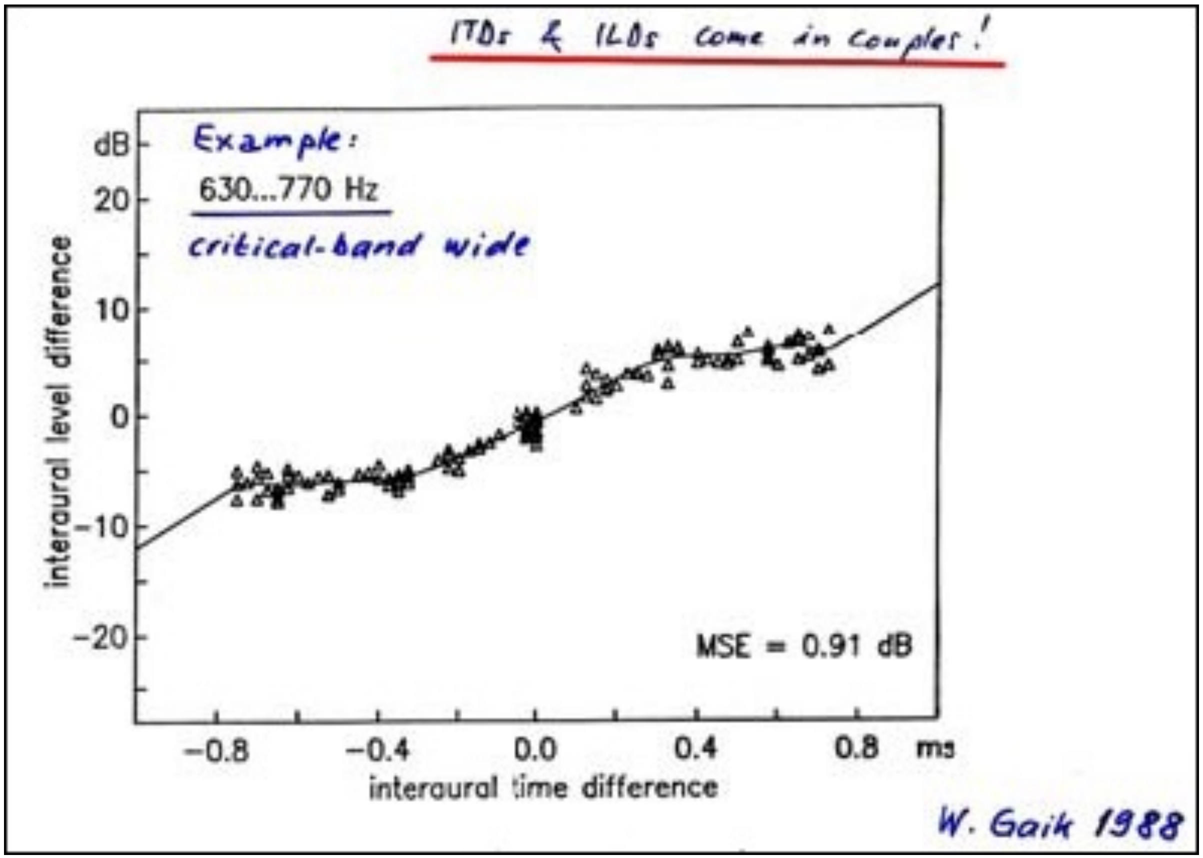
- High frequency pure tones localised by intensity differences
  - Shadow cast by head greater at high (20 dB at 6 kHz) than low frequencies (3 dB at 500 Hz) i.e. head acts as a lowpass filter

Azimuth for complex sounds

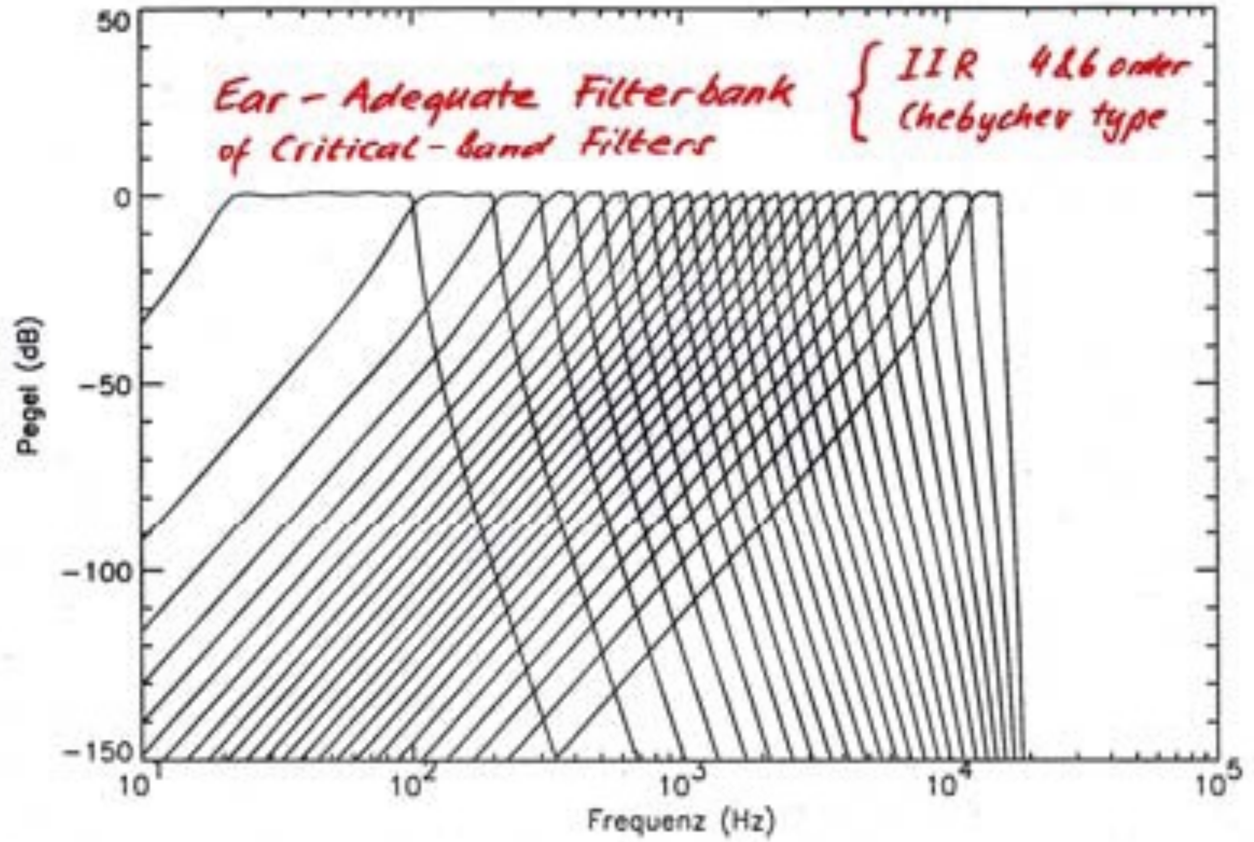
- Complex sounds contain both low and high frequencies
- Phase ambiguity not a problem for complex high-frequency tones



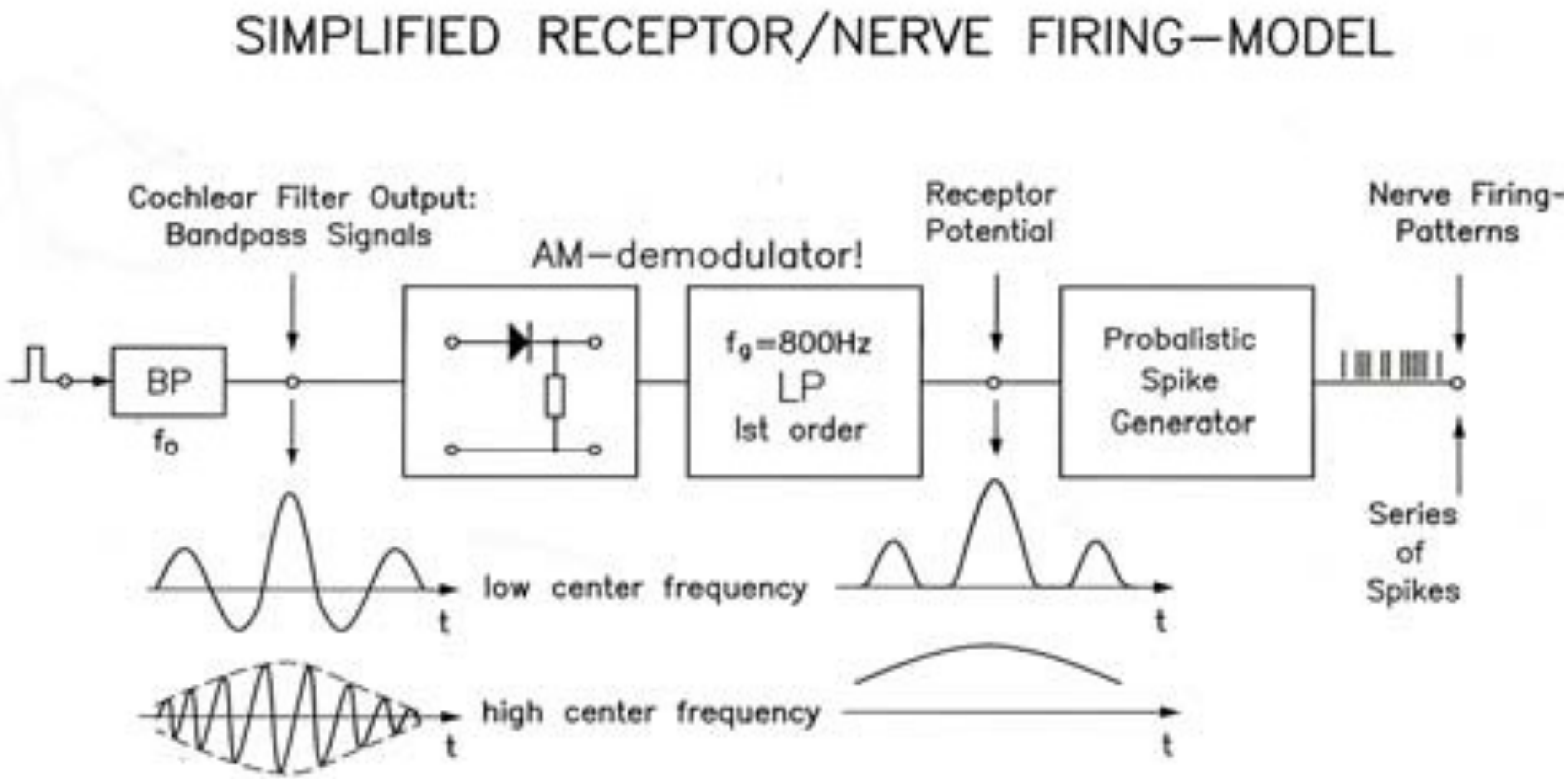
“Natural Combinations” of ITDs and ILDs



Ear-Adequate Band-Pass-Filter Bank



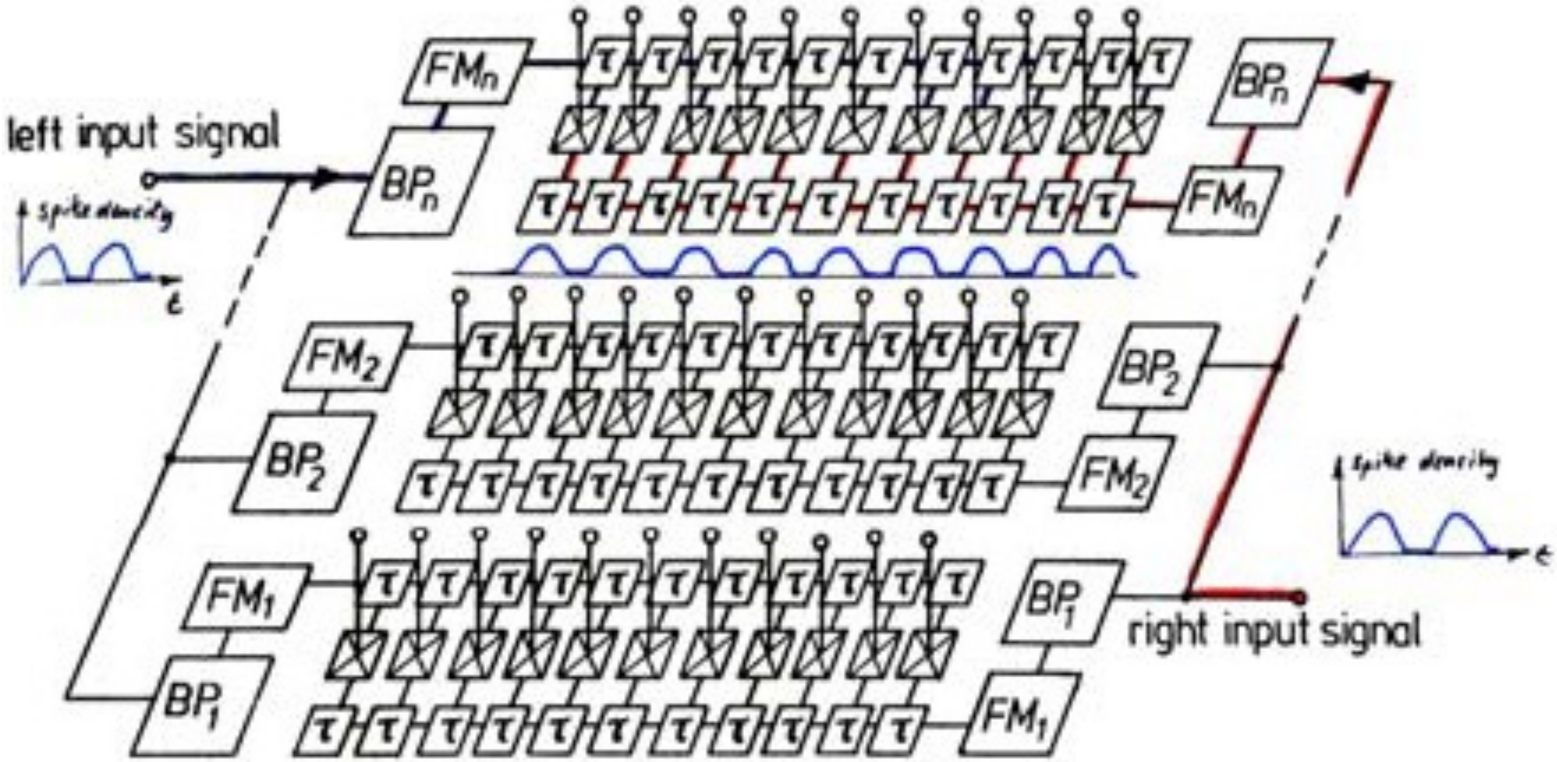
A Simplified Functional Model of the Hair Cells



The Binaural-Coincidence Processor

after *Jeffress 1948*

$$\Psi_Y(\tau) = 1/(t_1 - t_0) \sum_{t=t_0}^{t_1} Y_l(t) Y_r(t+\tau)$$

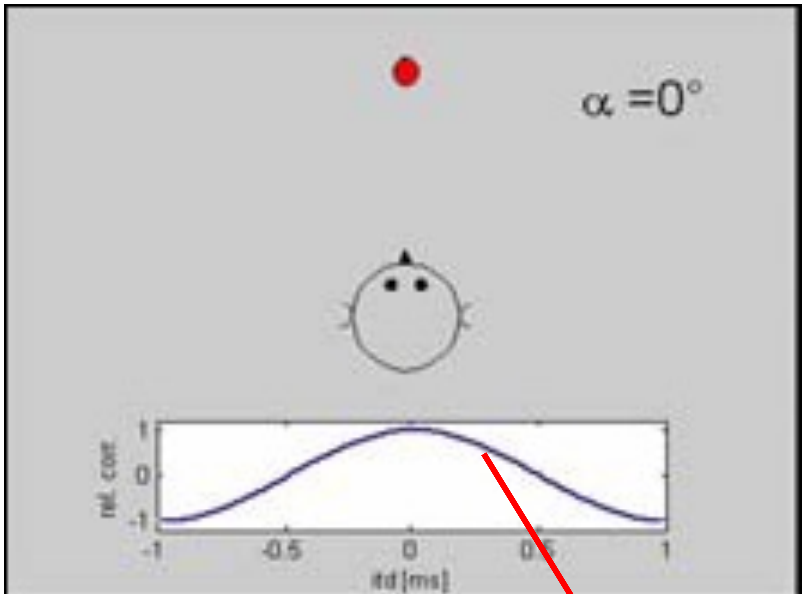


FM … firing model
BP … band-pass filter

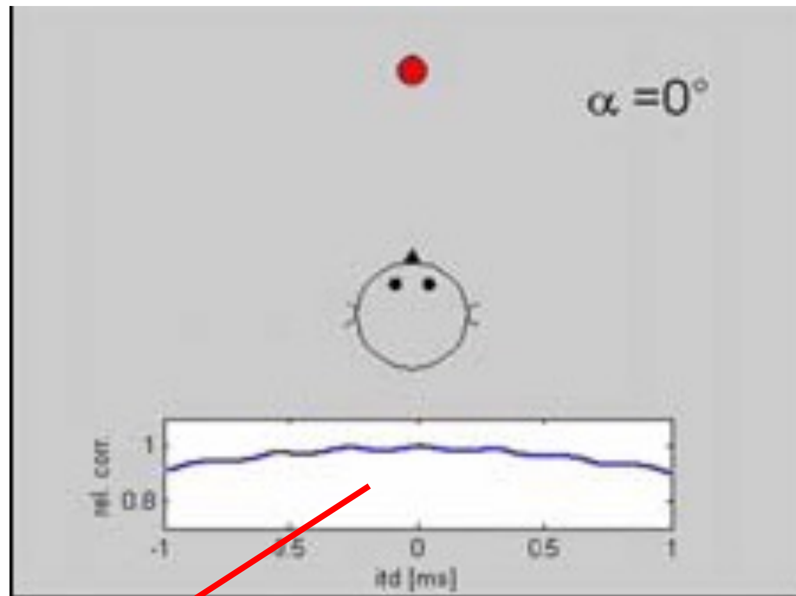
There are many of these channels in parallel.

Output of *Jeffress' Coincidence Processor* for Different Angles of Sound Incidence

about 0.7 kHz



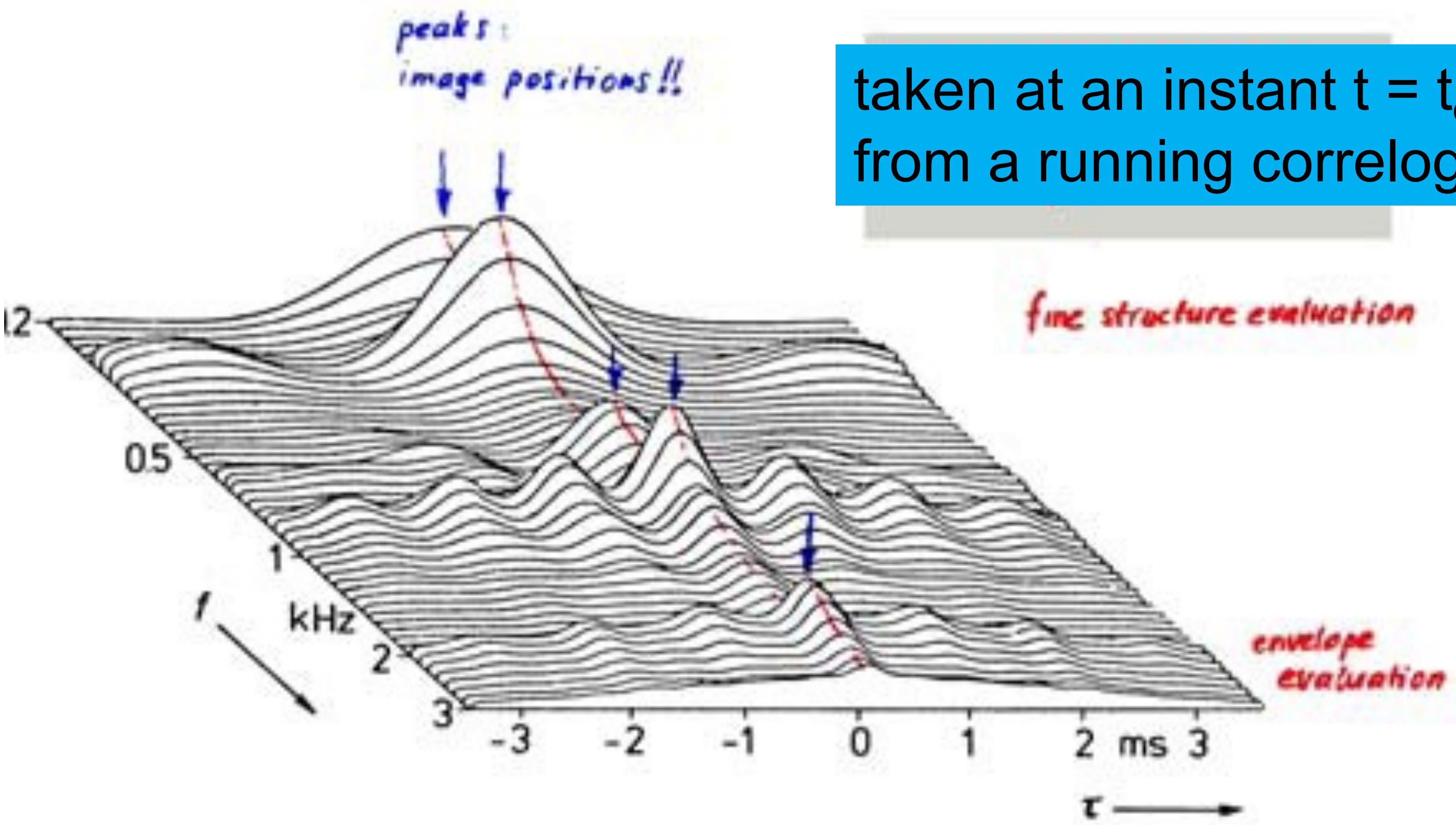
about 6 kHz



estimates of the interaural cross-correlation function

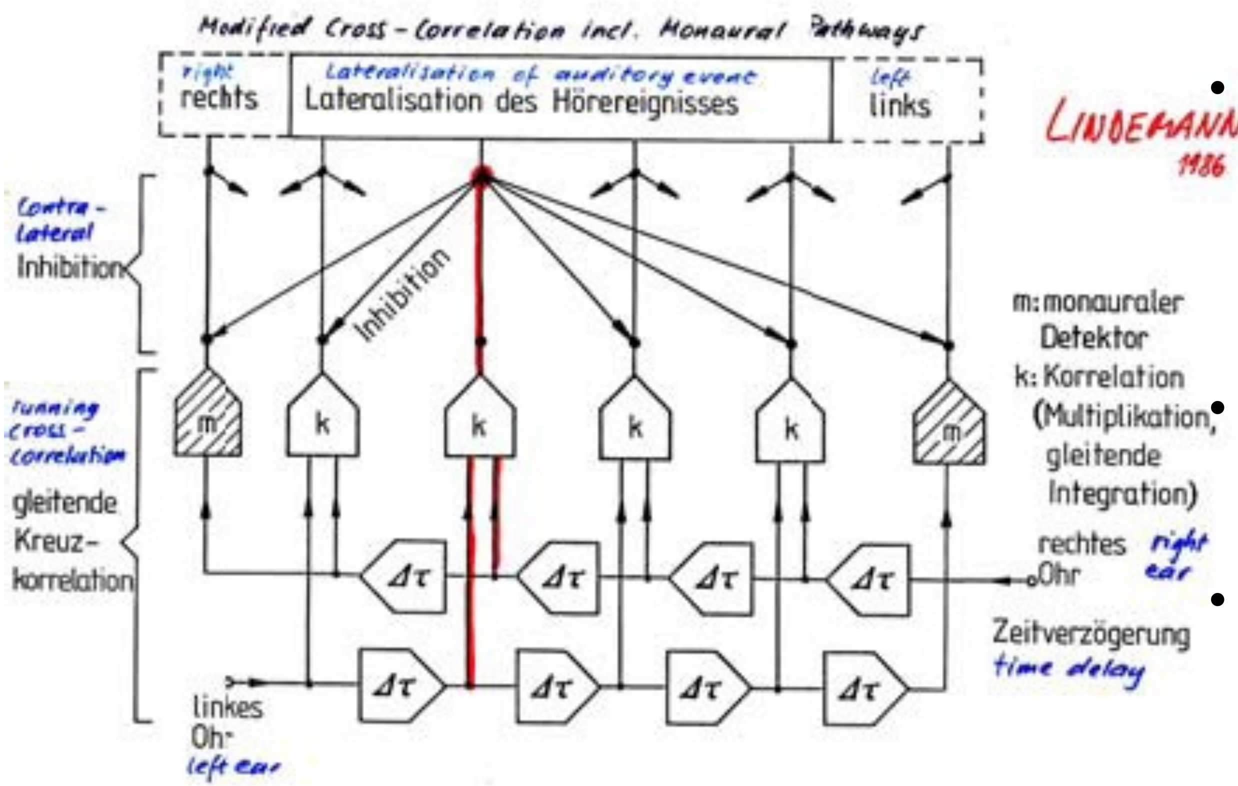
Sample Output of the Binaural Model

taken at an instant t = t_o
from a running correlogram



one frontal sound source, sending out a musical chord

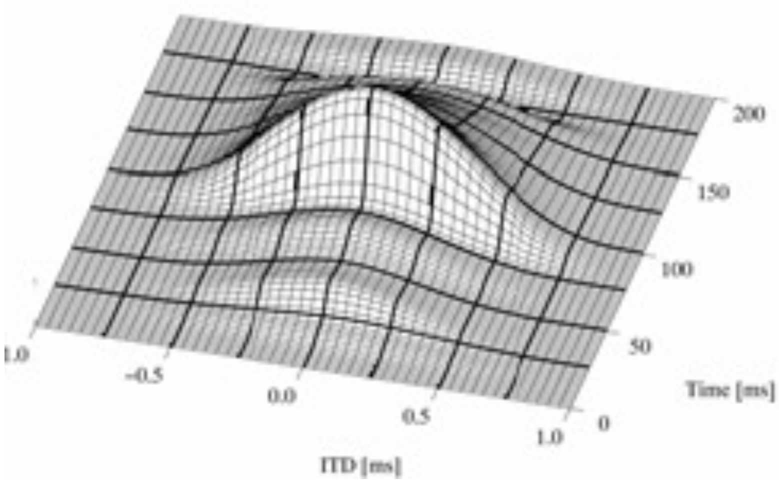
The *Lindemann Processor*



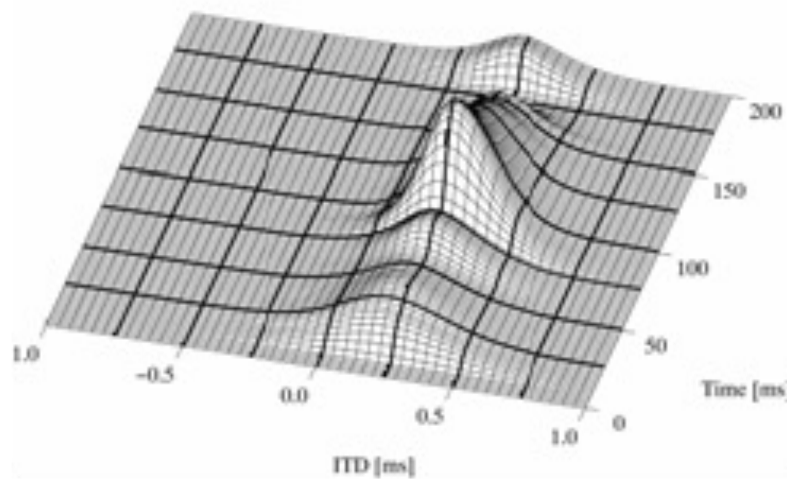
- Cross-correlation at one position (i.e. delay) inhibits activity at other positions
- reduces effects of echos and reverberations.
- does not allow for frequency dependent level differences

Cross Correlation vs. *Lindemann*

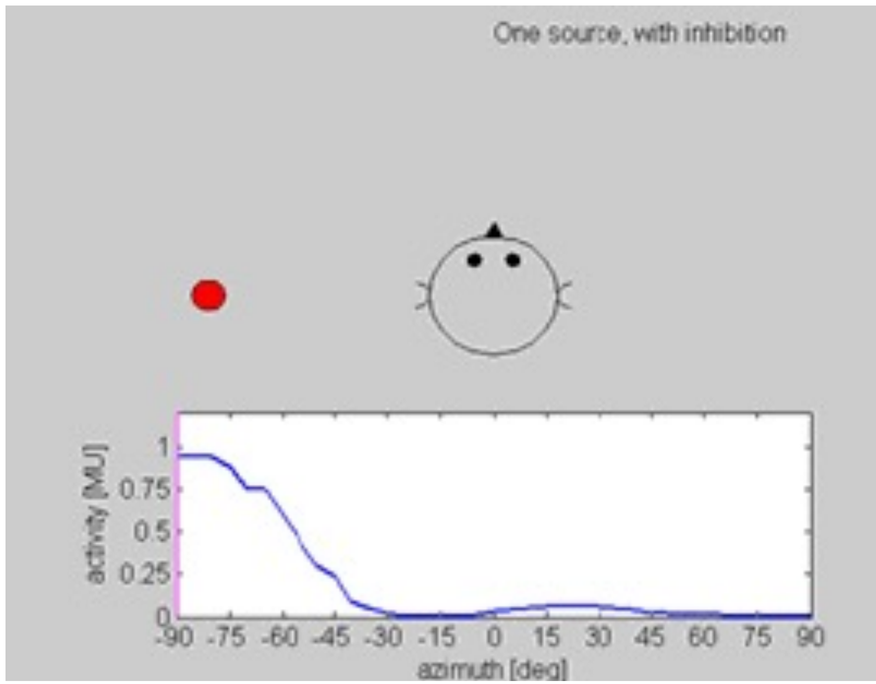
pure cross correlation



Lindemann



Output of *Lindemann's Coincidence Processor* with Lateral Inhibition



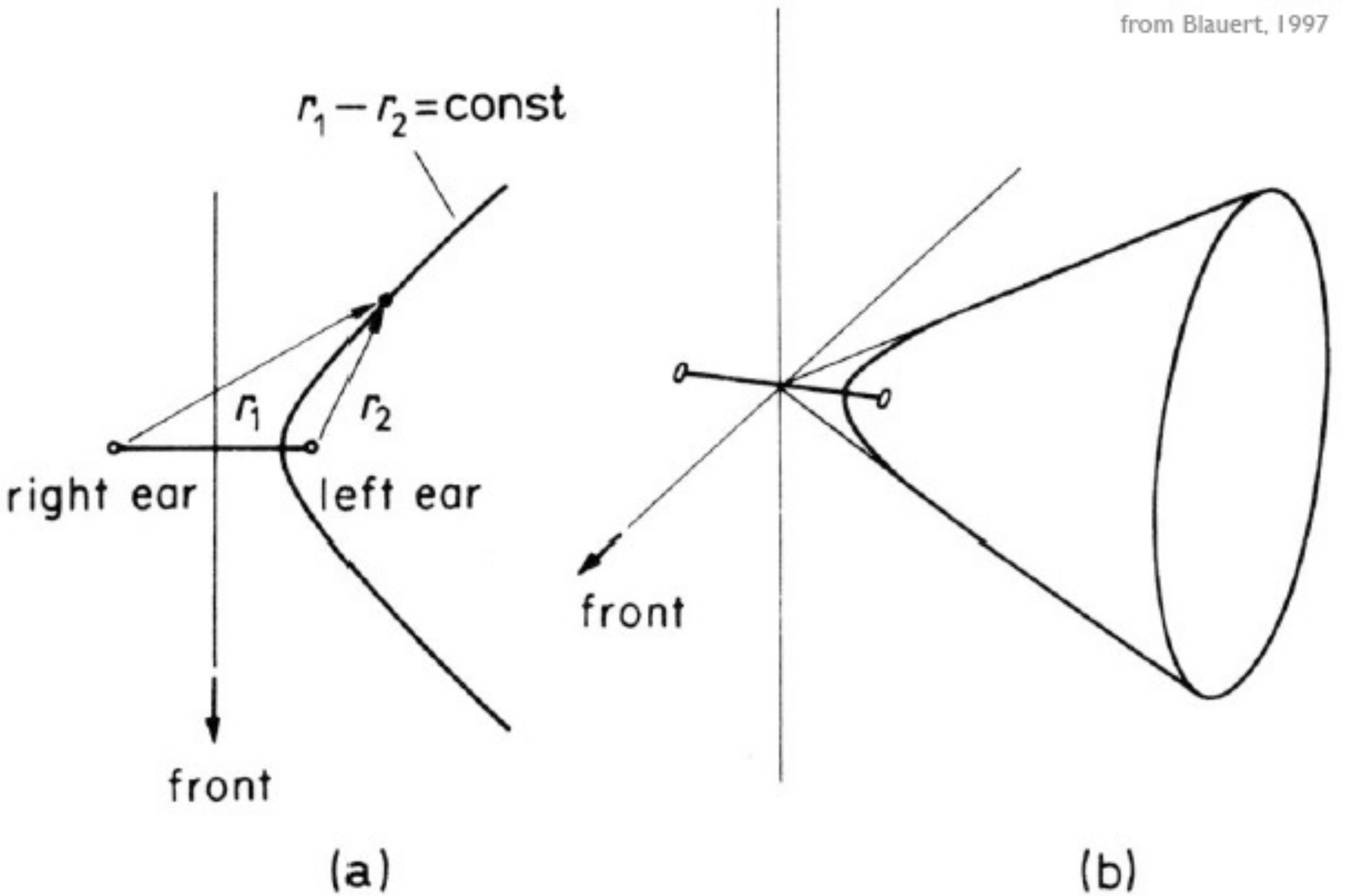
about 7 kHz

estimate of the inhibited interaural cross-correlation function

Limitations of the Duplex Theory

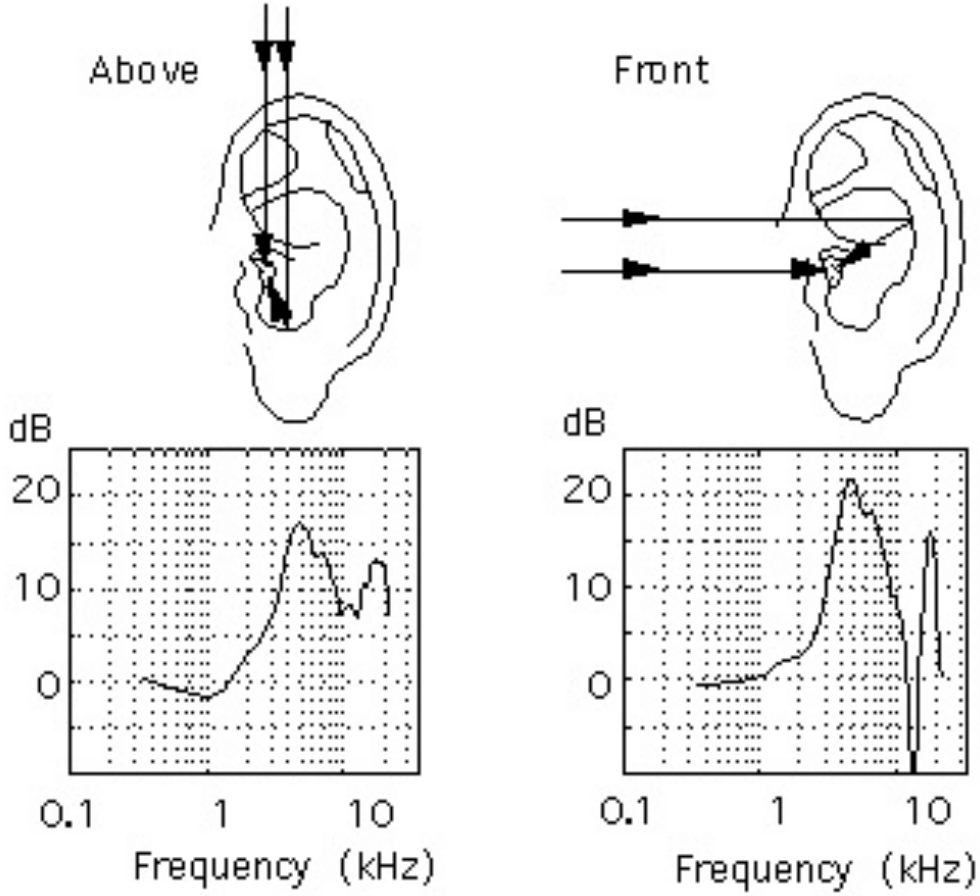
- limited to lateralization, doesn’t explain why are sounds are outside your head
- doesn’t do front-back discrimination
- Cone of confusion

Cone of confusion



All points on cone are same relative distance to each each ear.

Pinna notch



Timmear

Batteau's theory

- Echos produced by pinnae provide lateralization and elevation cues.
  - Listening through pinna casts caused externalization

Head-Related Transfer Function

- Take into account the effect of the pinnae, head and body.

- How to characterize the filtering?

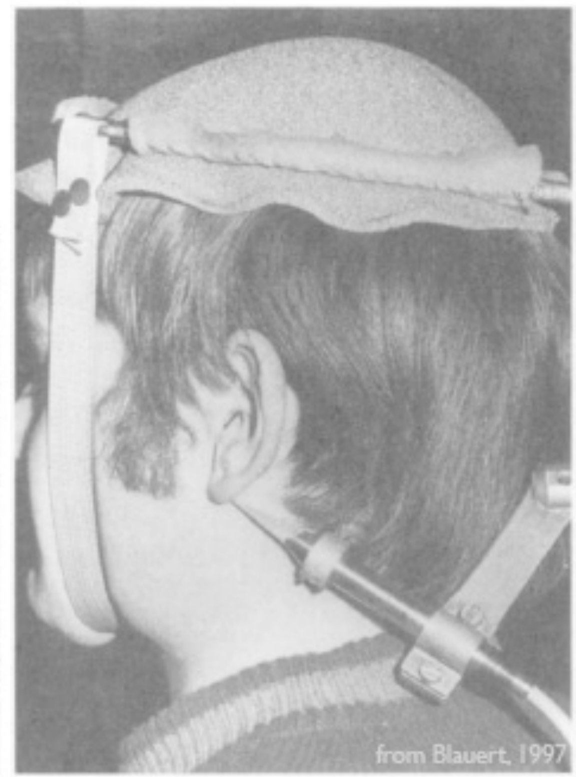
  - Measure the transfer function: the ratio of pressure at sound source to pressure of sound reaching eardrum – the *head-related transfer function* (HRTF)

Measuring HRTFs

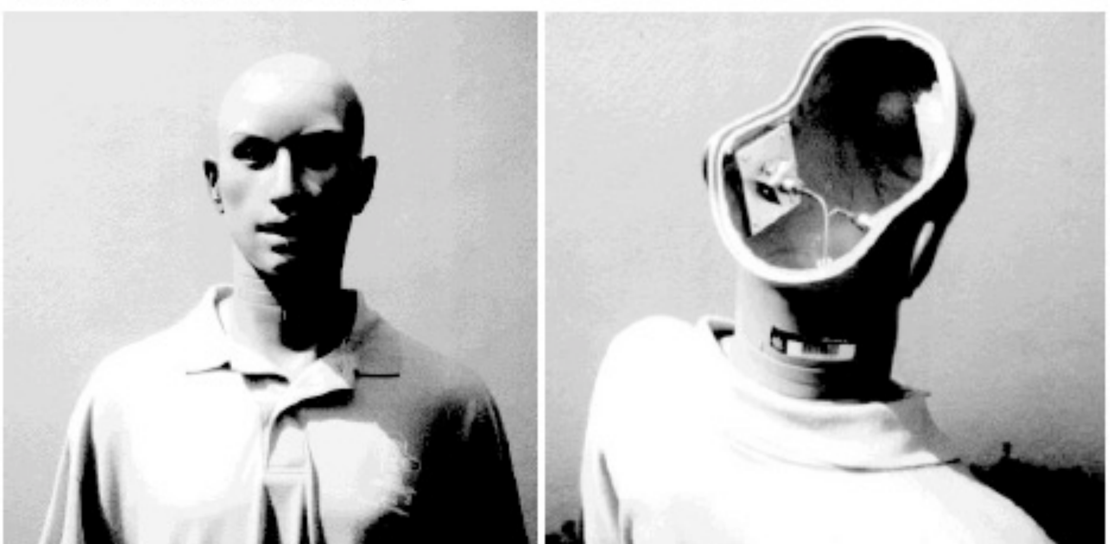
Two types of HRTFs

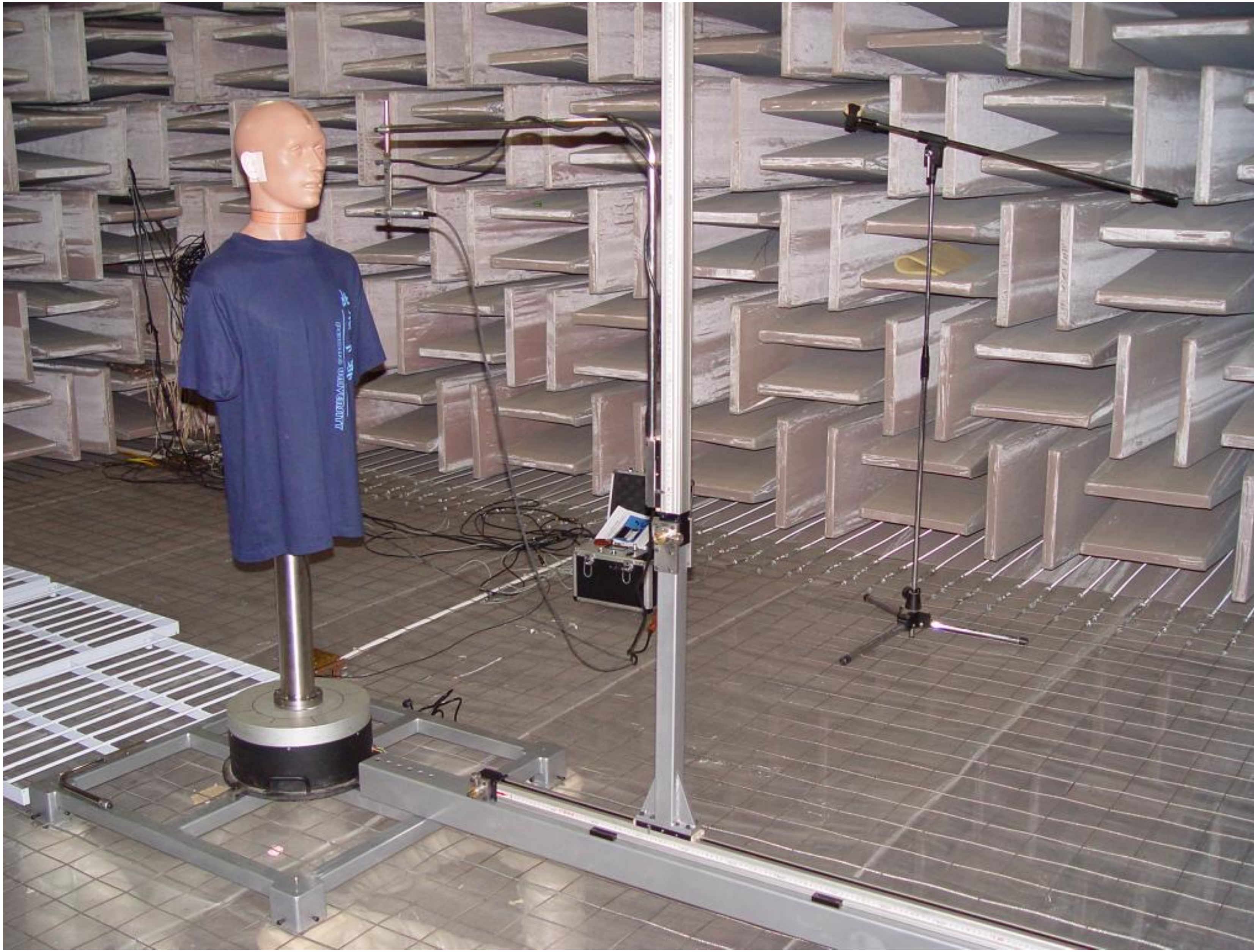
- monaural: pressure at source vs. ear drum
- binaural: pressure difference for two corresponding points in the ear canal

Subject with probe mics

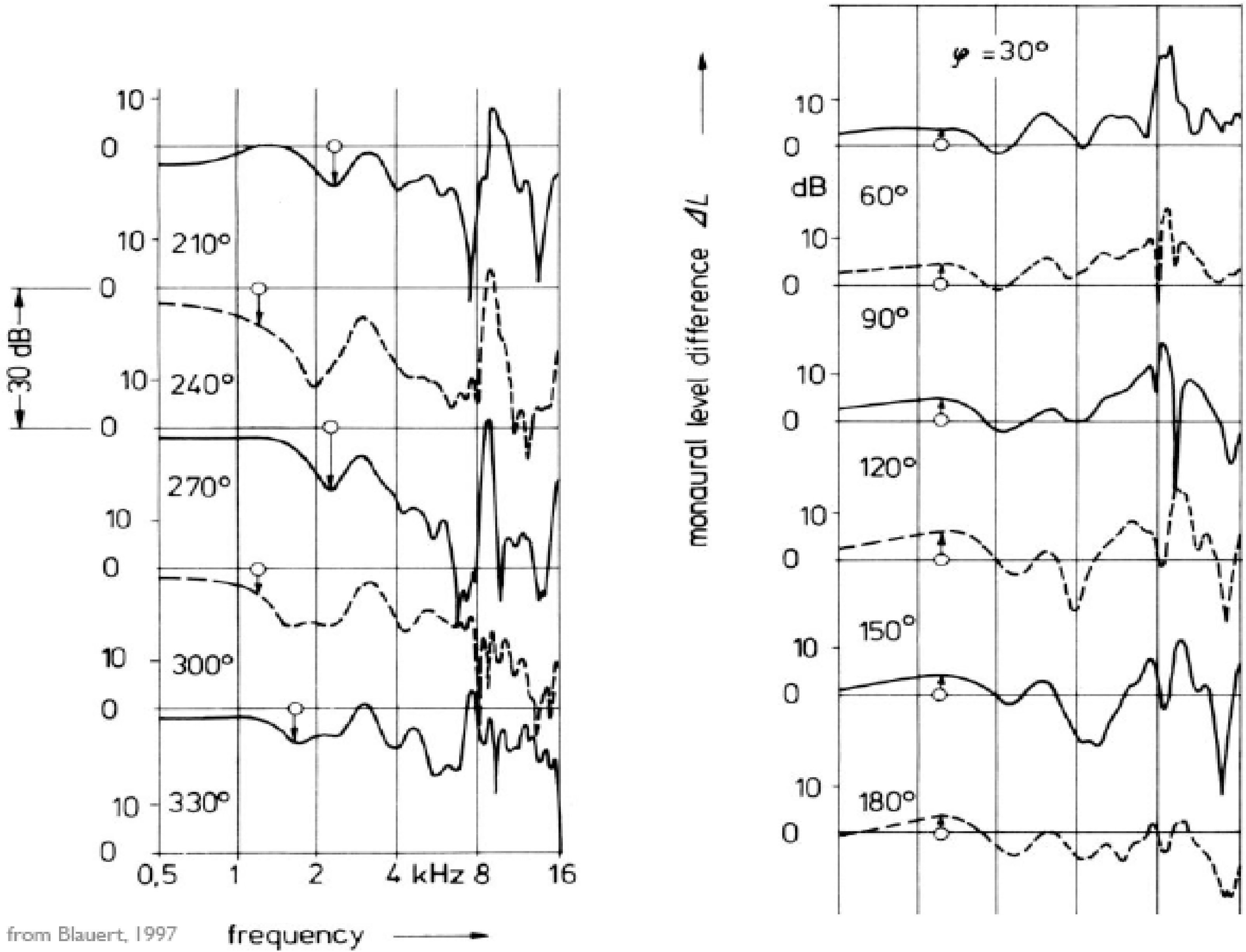


Kemar the sound dummy





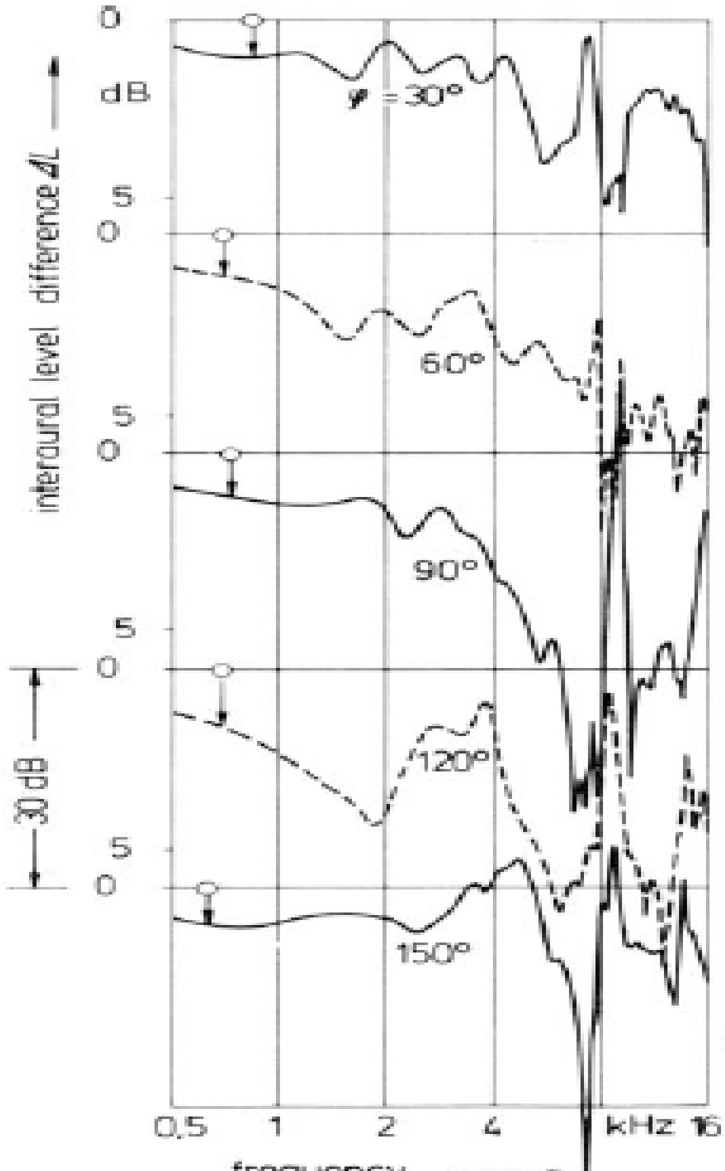
Measured monaural HRTF

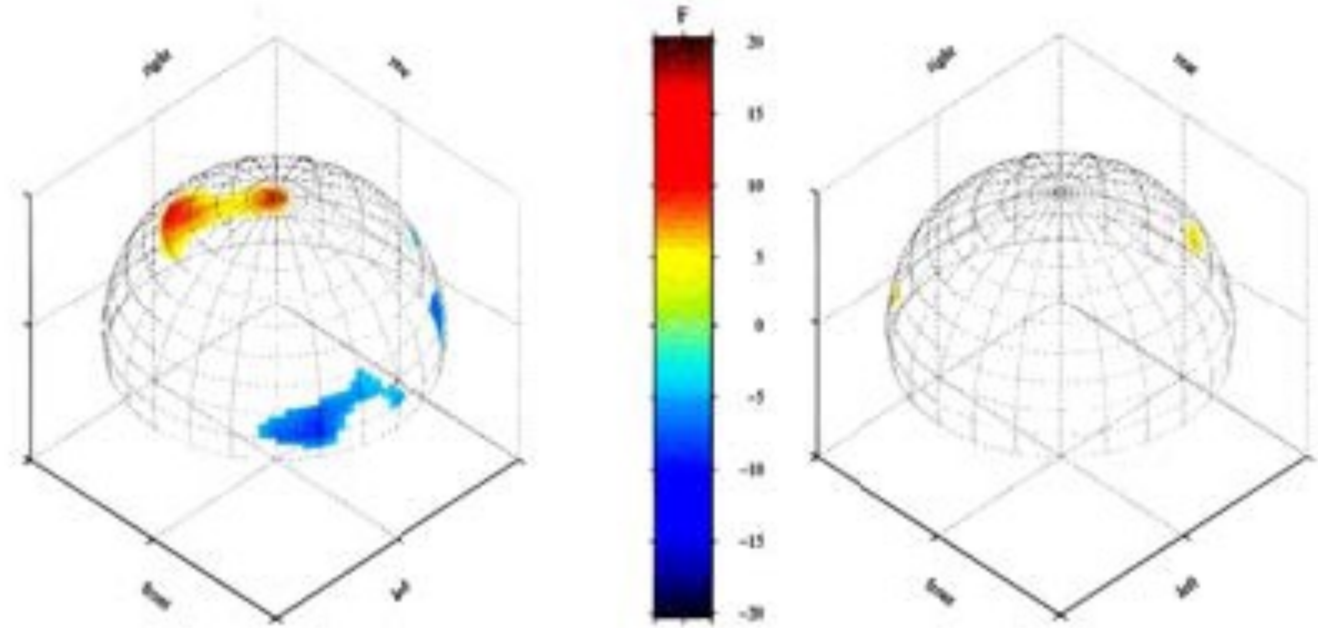


from Blauert, 1997

Measured binaural HRTF

from Blauert, 1997



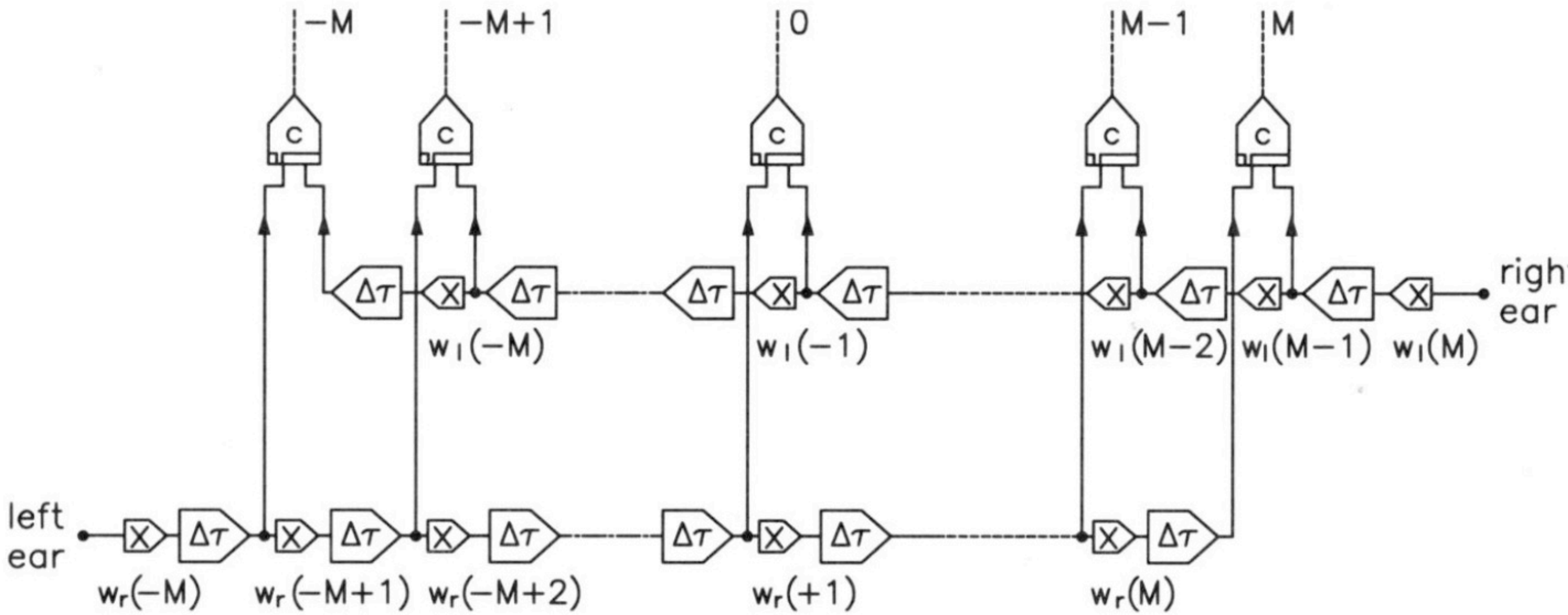


own HRTFs

other animal‘s HRTFs

Single-Cell Responses from the Central Nucleus of the Inferior Colliculus

Extension of Lindemann model: Gaik (1993)

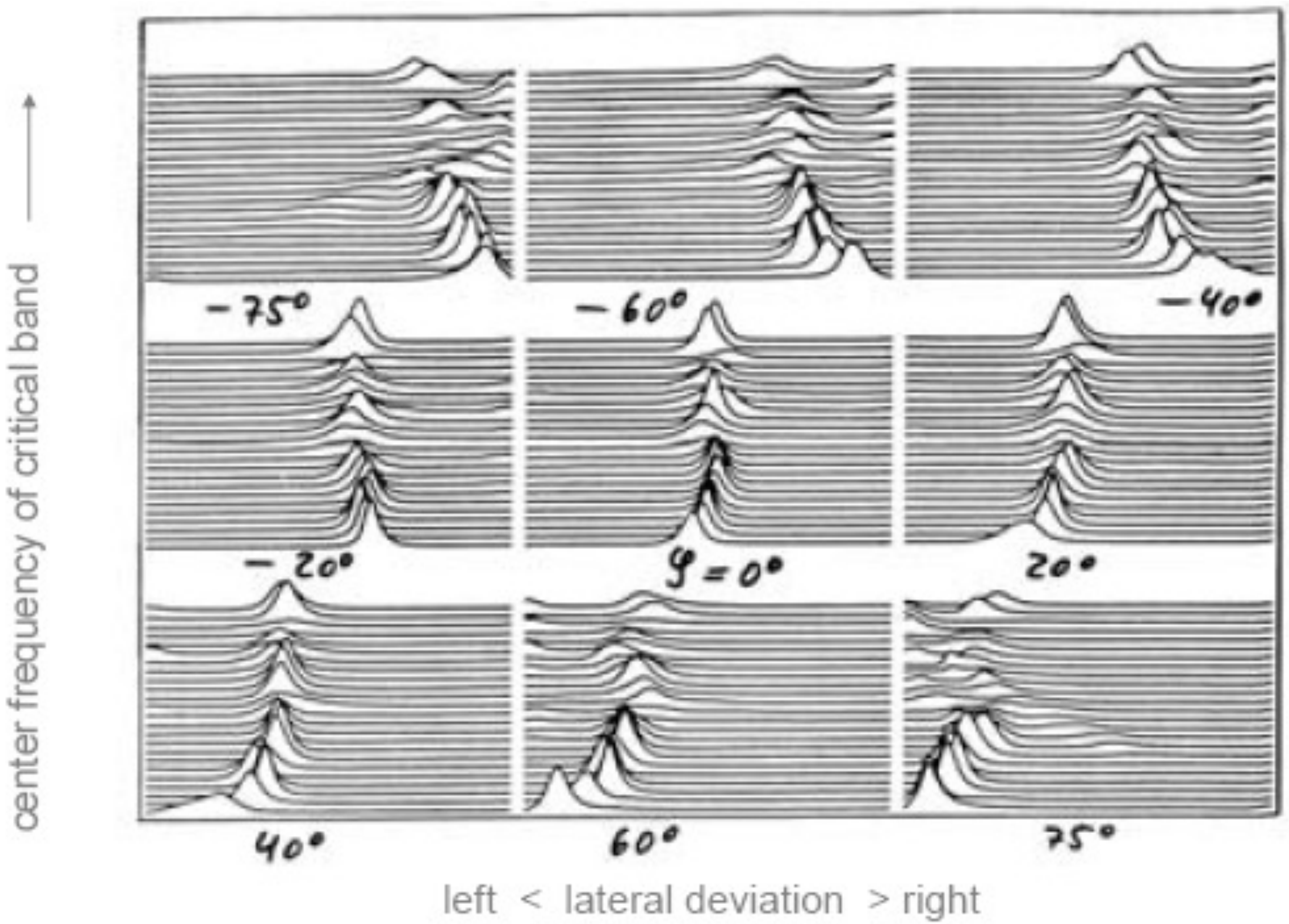


Weightings are adjusted according to interaural transfer function.

Limitations:

- contains a number of parameters that must be set.
- For ongoing signals like speech and music in reverberant environments can have problems distinguishing direct sound from reflections
- These problems compound when multiple sounds are active

Output of the *Jeffress-Lindemann-Gaik Model*



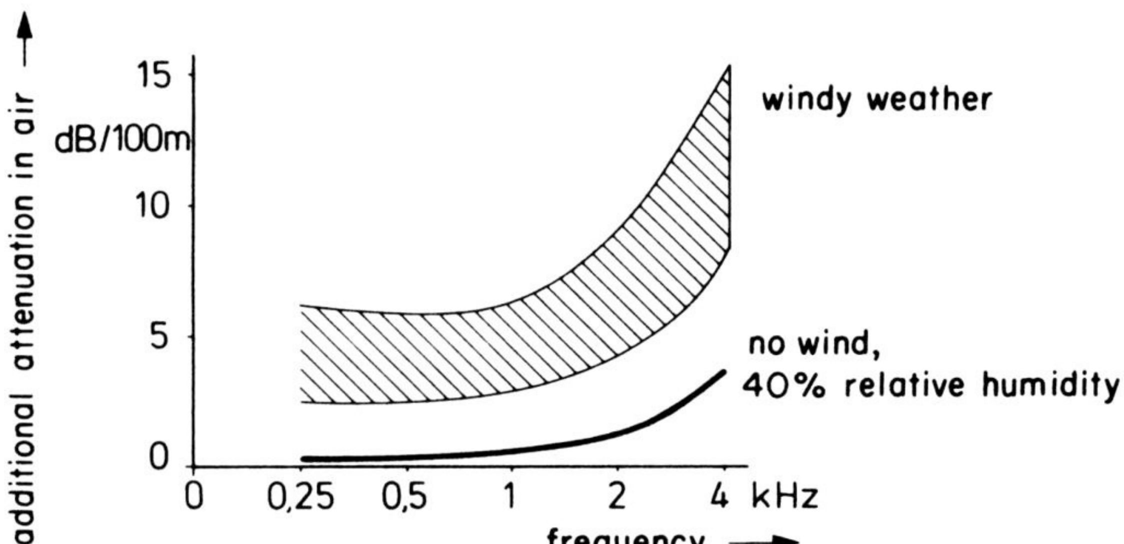
frontal broad-band sound source

More than just direction: cues for sound distance

- HRTF depends on distance (close sounds <1.5m)
- freq. dependent attenuation (long distances)

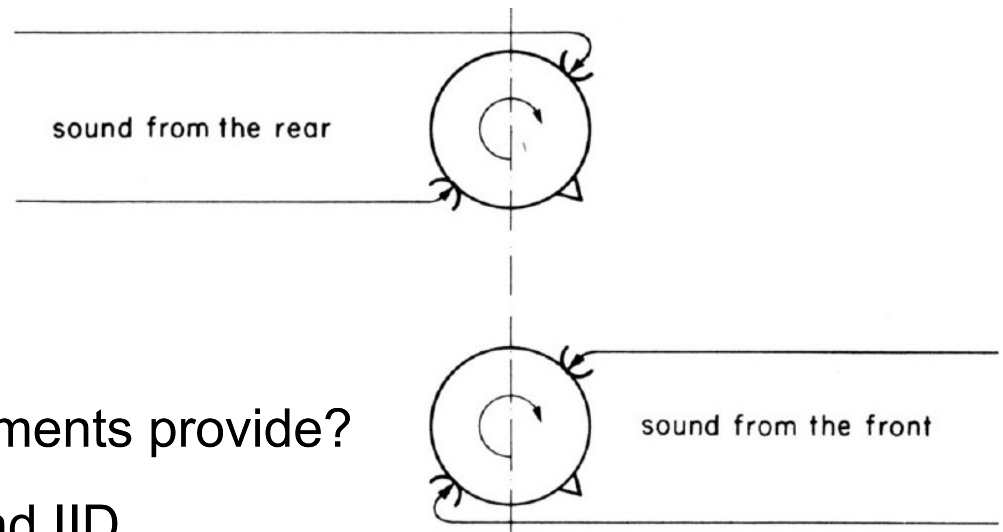
– Beyond 1.5m the difference in level between the ears is less than 1 dB.

- pressure attenuation – properties of sound source



How do we make front-back discriminations?

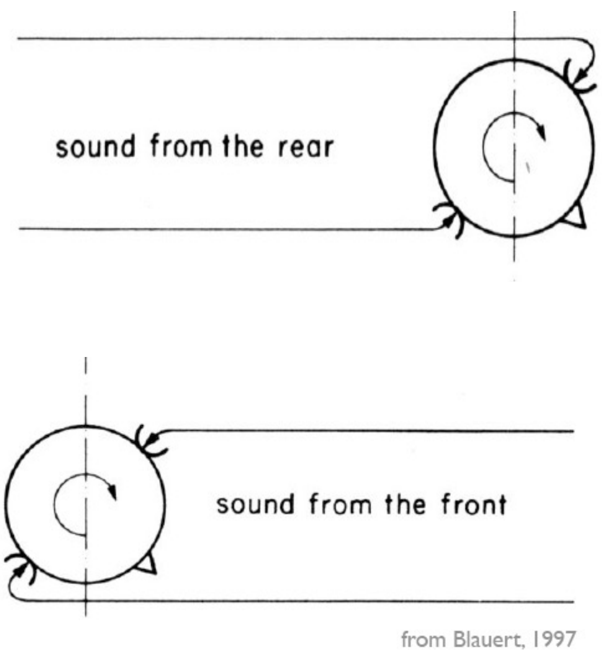
- pinna are asymmetric
- head movements



from Blauert, 1997

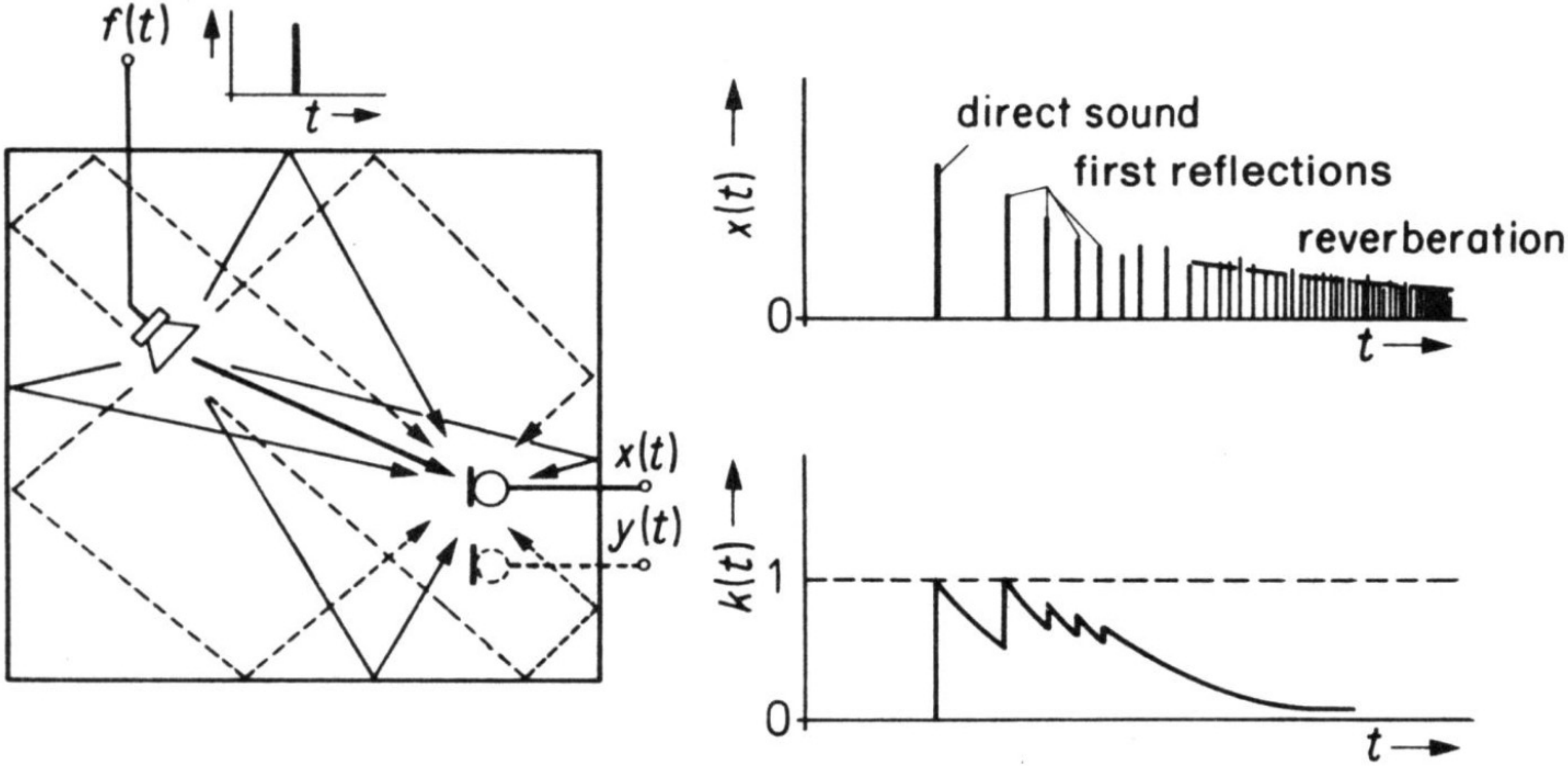
- What cues do head movements provide?
  - changes in ITD and IID
  - changes in sound spectrum

- The change in ITD can be positive or negative depending on whether the sound waves arrives from the front or the near.
- Also same for ILD

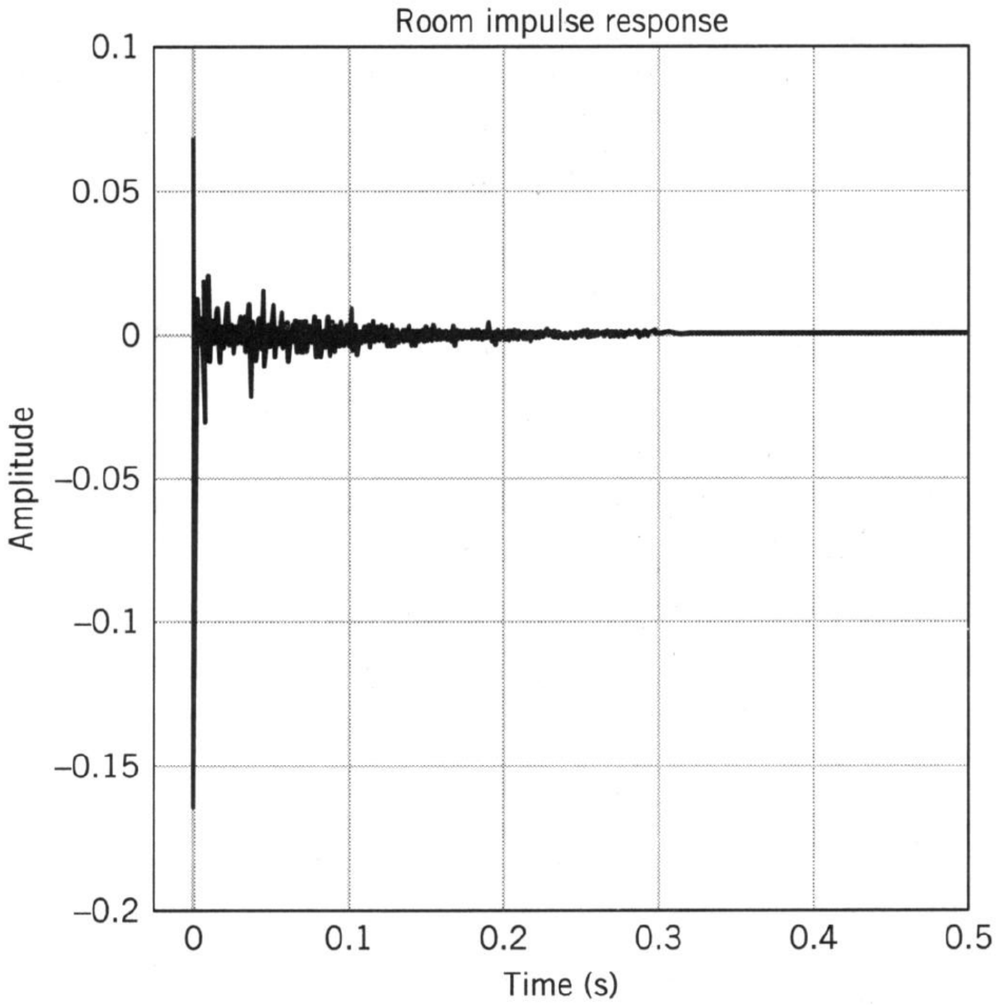


Suppression of
the Directional Information coming from Reflections

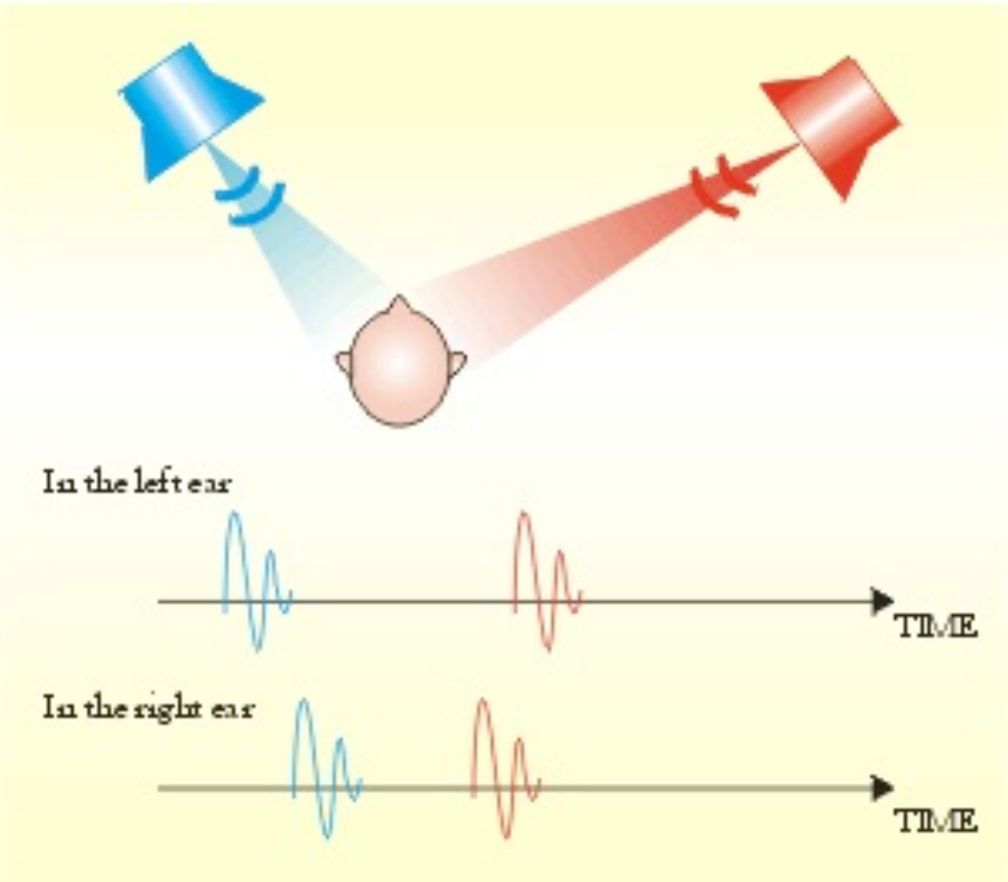
Reverberation in the ideal case



Reverberation in a room



Precedence (or Haas) effect

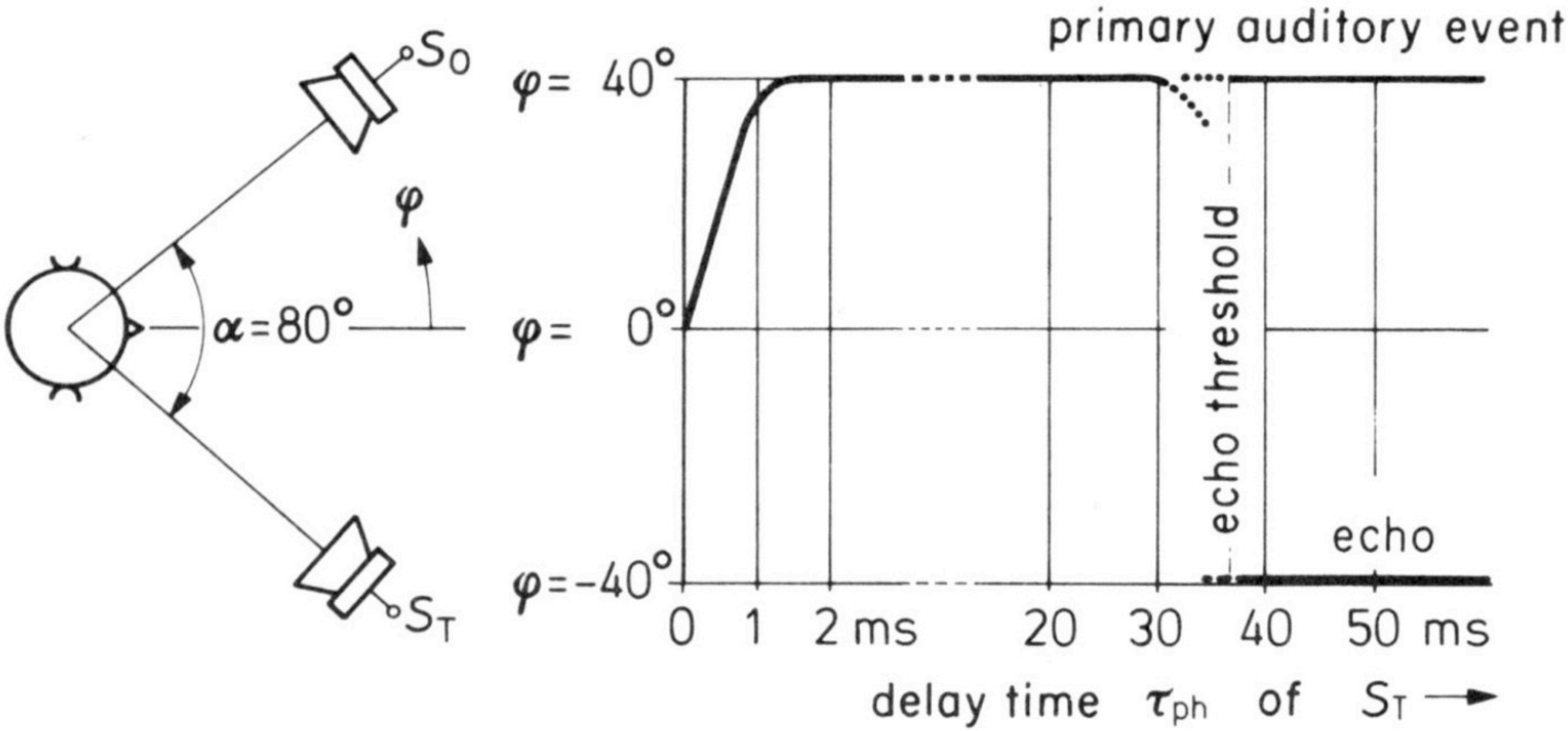


blue ITD vs red ITD to center the single sound

Lots of red ITD needed to offset a little blue

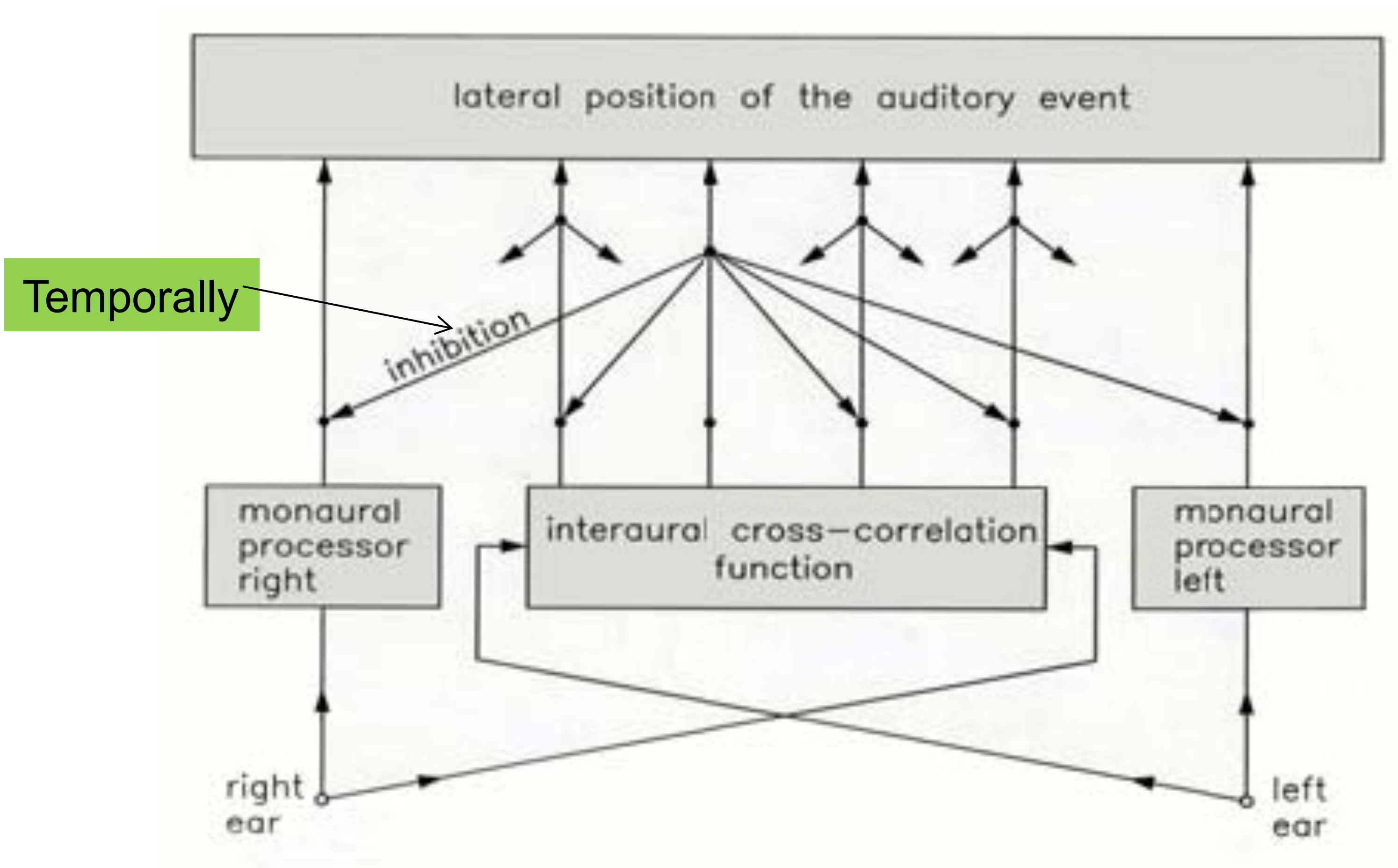
What will happen perceptually as the delay is varied?

The precedence effect

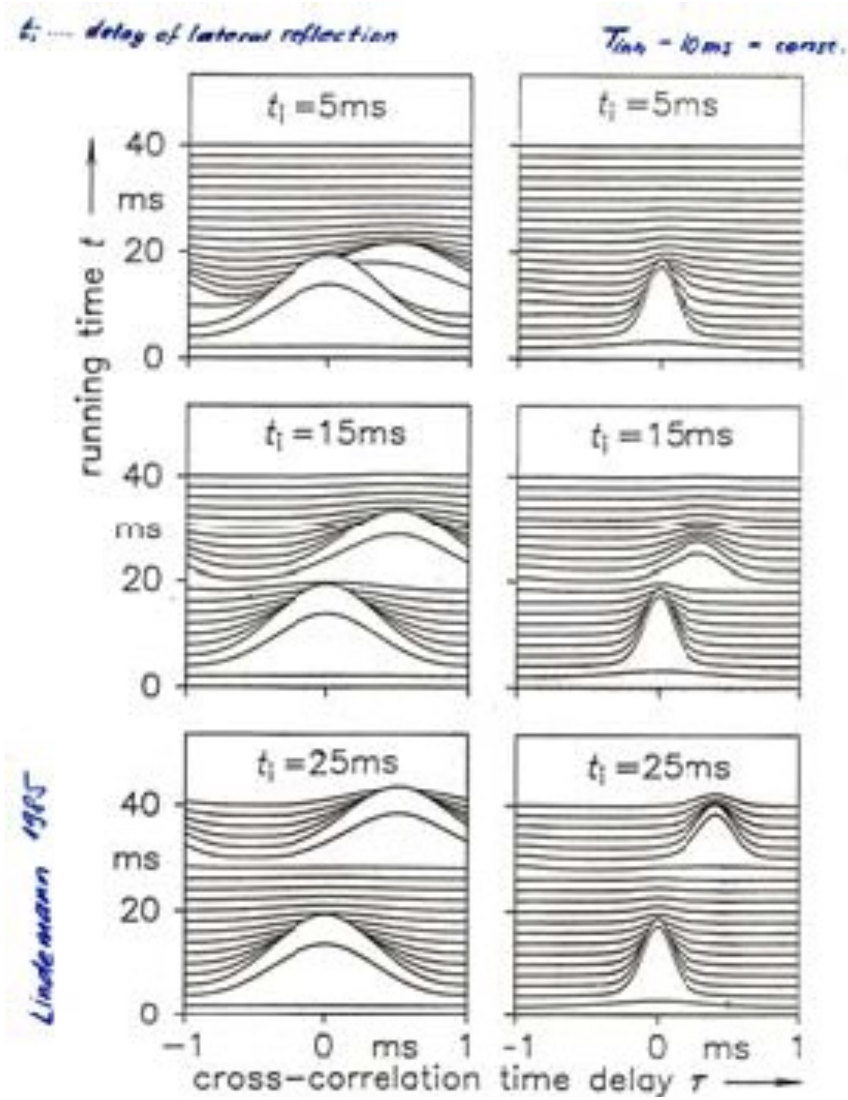


Lateralization shifts from 0 to 1ms.
Only first impulse is perceived from 1ms to 30ms
Beyond that two impulses are perceived separately (echo)

Schematic View of the *Lindemann Model*



Output of the Binaural Model for a Frontal Sound plus One Lateral Reflection



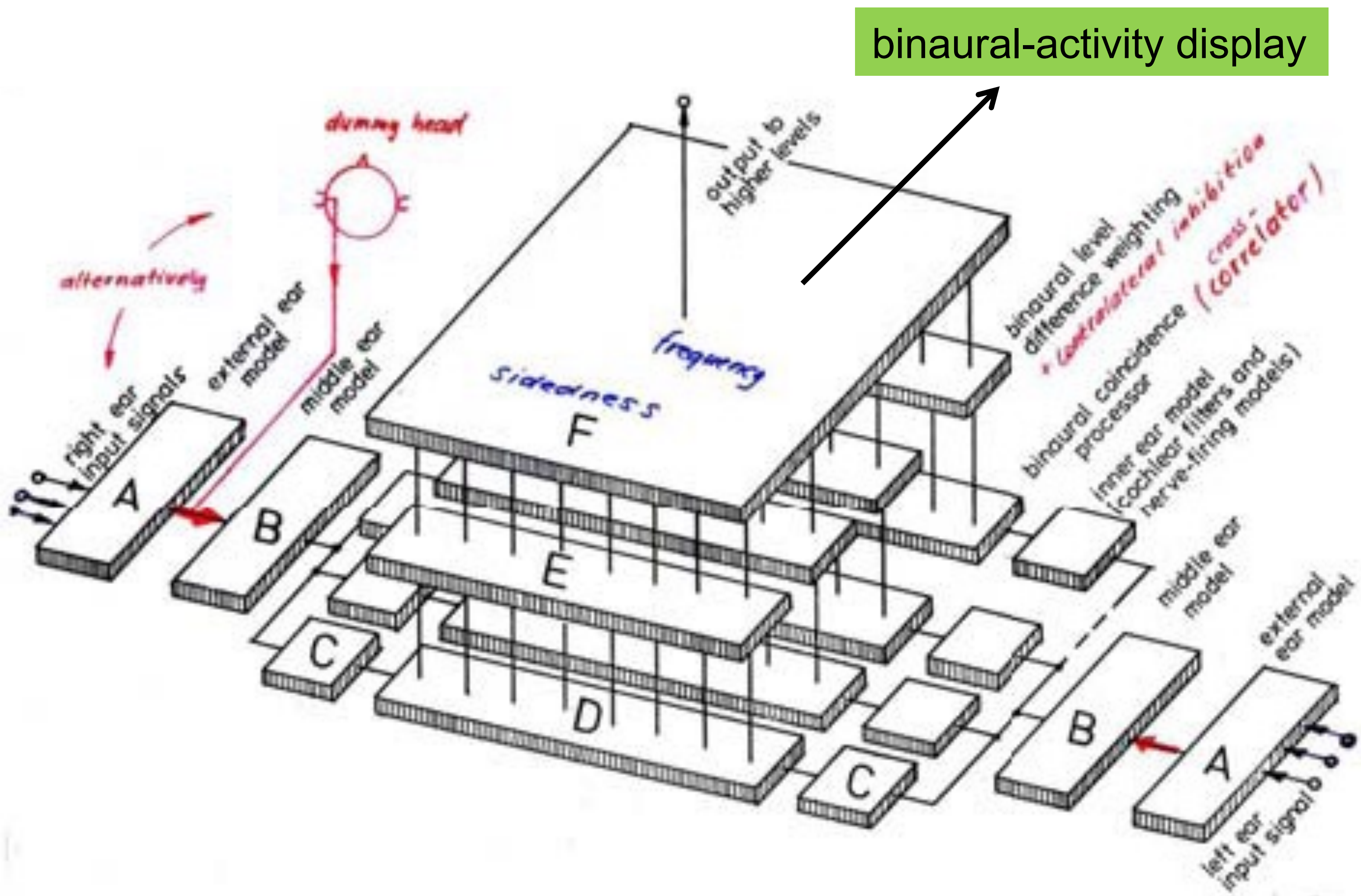
left panel:

cross correlation only

right panel:

cross correlation plus contralateral inhibition

Architecture for a Model of Binaural Hearing

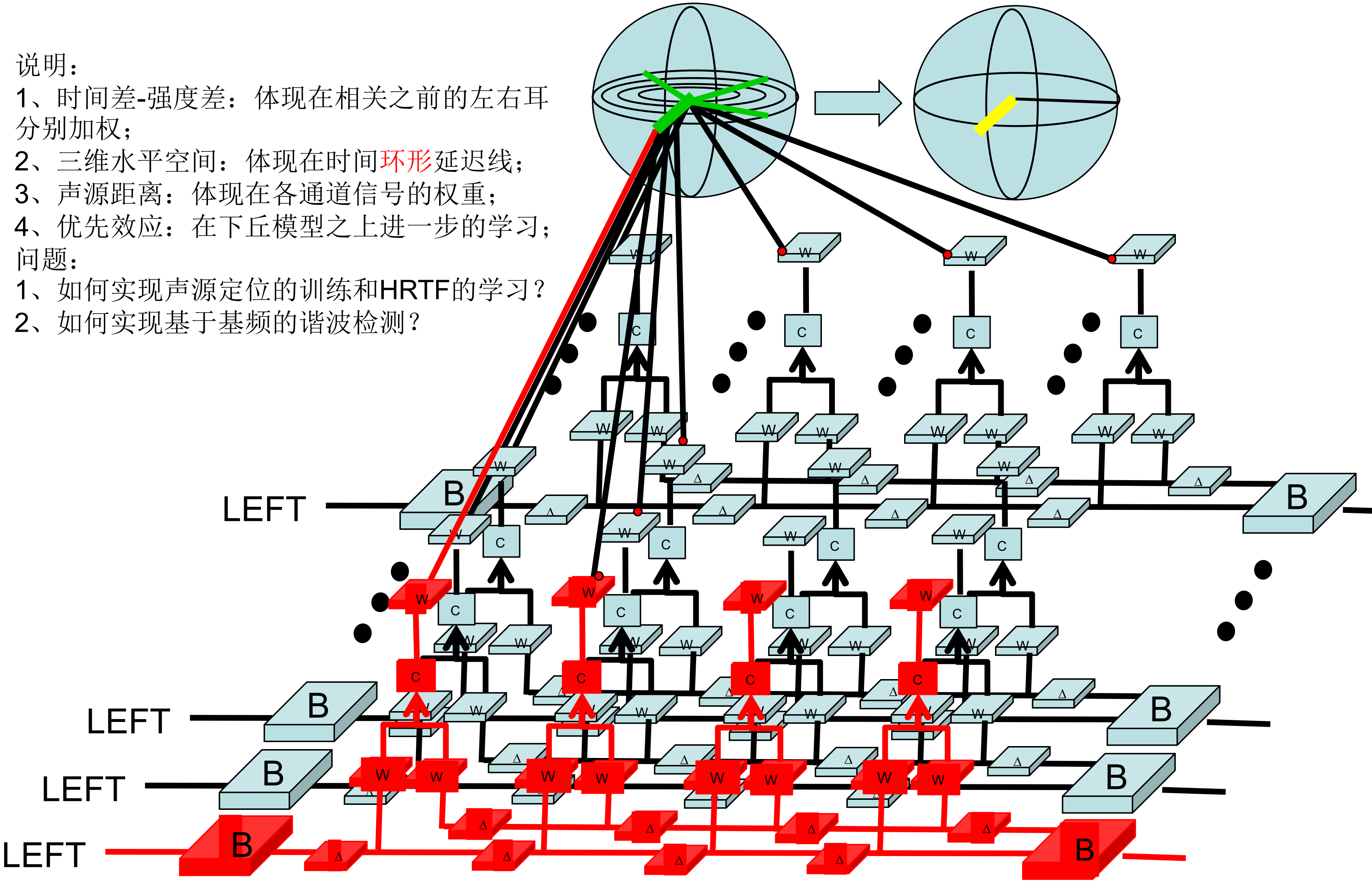


双耳加工及声源定位模型

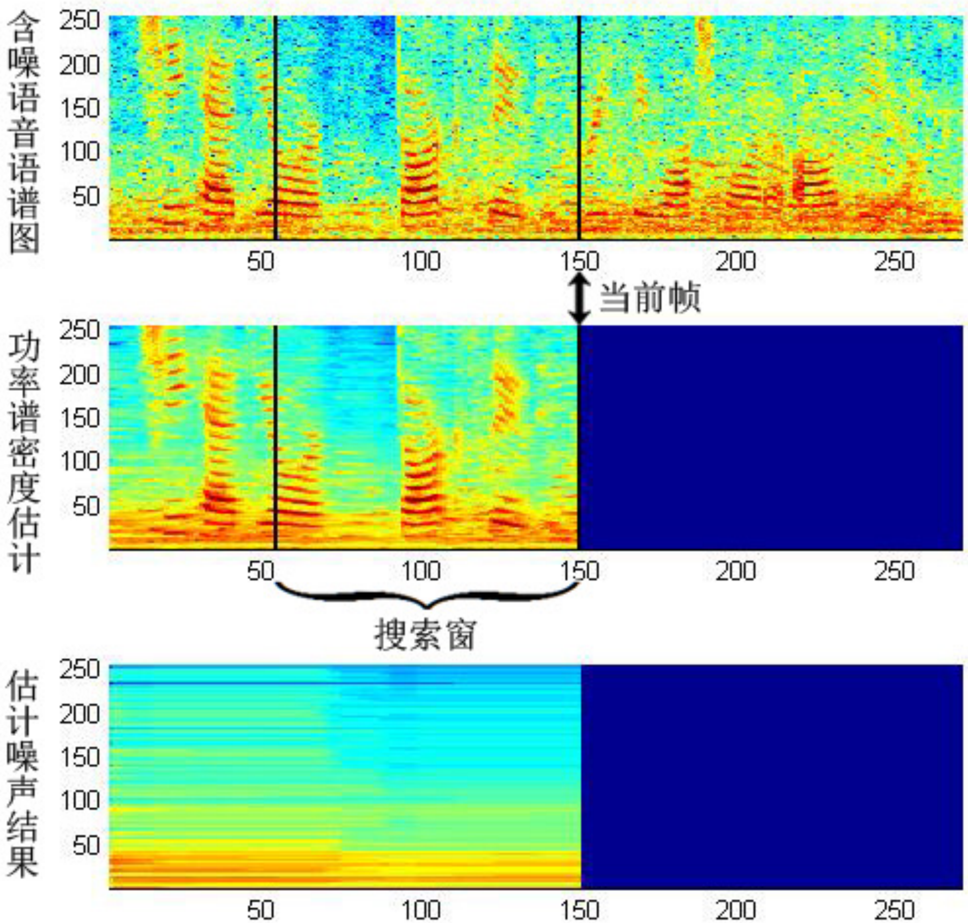
绿色体现了声场的空间感；
粗绿线代表直达声；
黄色代表空间方位及距离；

说明：
1、时间差-强度差：体现在相关之前的左右耳分别加权；
2、三维水平空间：体现在时间环形延迟线；
3、声源距离：体现在各通道信号的权重；
4、优先效应：在下丘模型之上进一步的学习；

问题：
1、如何实现声源定位的训练和HRTF的学习？
2、如何实现基于基频的谐波检测？



语音增强和抗噪声能力



含噪语音功率谱密度：

$$P(\lambda,k) = \alpha(\lambda,k)P(\lambda,k)+(1-\alpha(\lambda,k))\left|Y(\lambda,k)\right|^2$$

固定搜索窗长估计噪声：

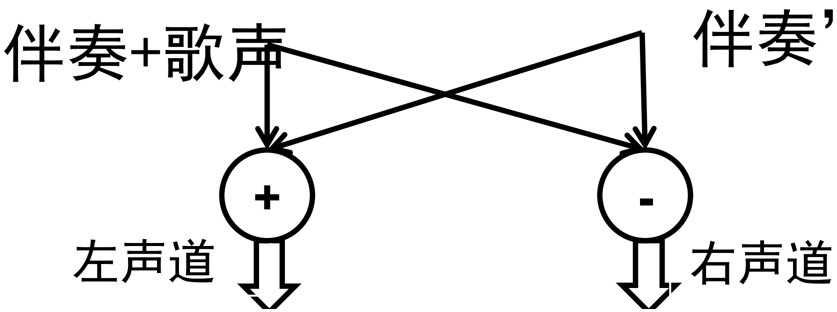
$$\hat{\sigma}_N^2(\lambda,k) = B_{\min}\left(D,\hat{\mathrm{var}}\{P(\lambda,k)\},\hat{\sigma}_N^2(\lambda-1,k)\right)P_{\min}(\lambda,k)$$

动态搜索窗长估计噪声：

$$\hat{\sigma}_N^2(\lambda,k) = B_{\min}^{P(\lambda,k)}\left(D,\hat{\mathrm{var}}\{P(\lambda,k)\},\hat{\sigma}_N^2(\lambda-1,k)\right)P_{\min}(\lambda,k)$$

原始混合语音　　双耳增强　　双耳增强＋单耳增强

歌声提取和抗噪能力



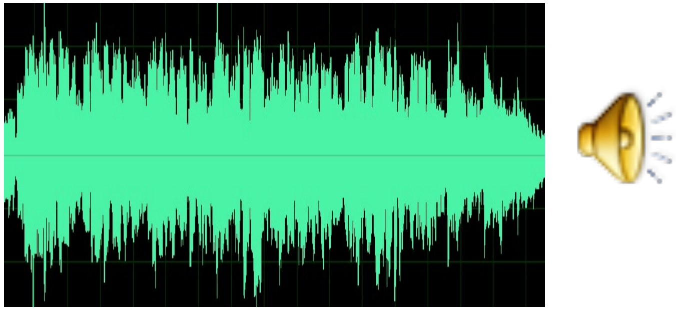
FFT　　　FFT

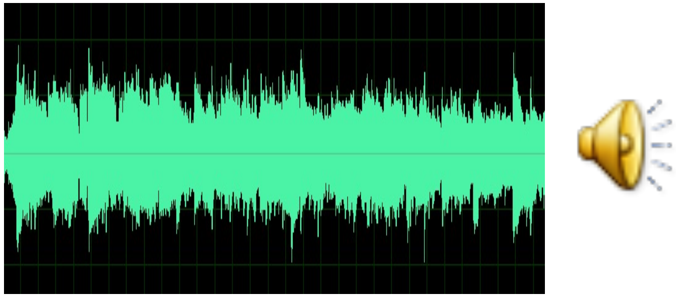
$$G = \max\left(\frac{\cos(\Delta\phi)}{1 + k(|\Delta A|/\mathrm{A})}, 0\right)$$

$$F' = G \cdot (F_L + F_R)$$

IFFT

歌声





Conclusion

- Azimuth (left/right)
  - Binaural cues: ITD and ILD
- Median-plane (front, up, back, down)
  - Pinna-induced spectral cues
  - Head movements
- Distance
  - Absolute level, excess IID (inverse-square law), spectral balance, reverberation
- HRTF
- Jeffress, Lindemann, Gaik models

Q & A?