3-D Sound and Spatial Audio

What do these terms mean?

- Both terms are very general.
- "3-D sound" usually implies the perception of point sources in 3-D space (could also be 2-D plane) whether the audio reproduction is accomplished with loudspeakers or headphones.
- "Spatial audio" is broader, more inclusive in scope and includes the possibility of environmental sound, multi-loudspeaker systems, etc.

Other important terms,

Mono/Monophonic: an audio system or device with one channel of audio information; usually reproduced from a single loudspeaker.

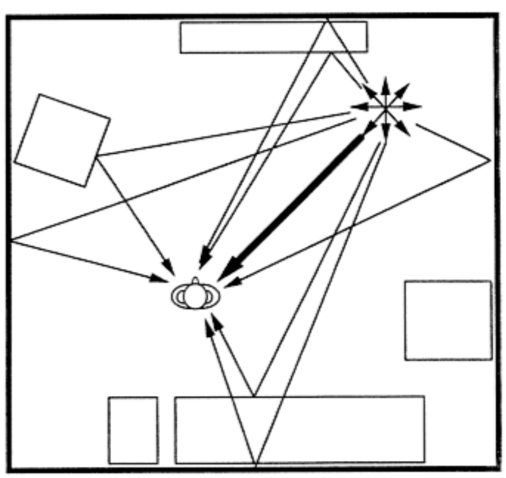
Stereo/Stereophonic: an audio system or device with two channels of information; usually reproduced from two loudspeakers or headphones. The term usually implies that the listener can perceive sound images along a line between the loudspeakers or headphone transducers.

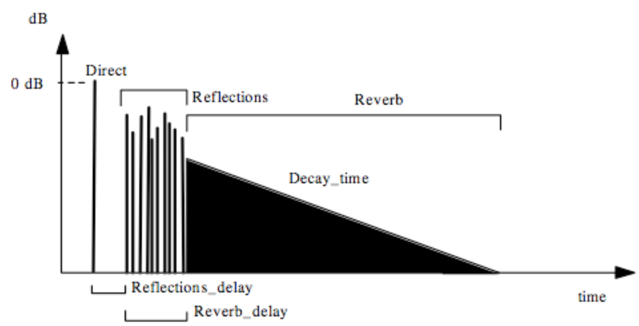
Surround Sound: the term encompasses a variety of audio systems or devices that utilize four or more channels of audio information typically reproduced with four or more loudspeakers. Like stereo, it is usually implies that the listener can perceive sound images between the loudspeakers.

Binaural: two-channel audio system or device that utilizes the spatial hearing cues of everyday life and provides the listener with the perception of three dimensions. The term often implies that the listener uses headphones.

More terms,

Direct and Indirect sound

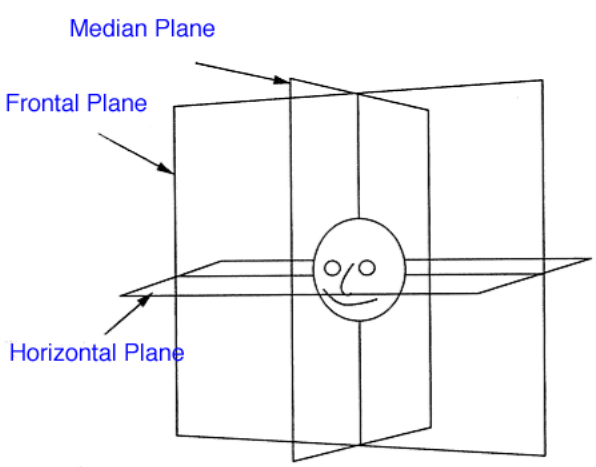


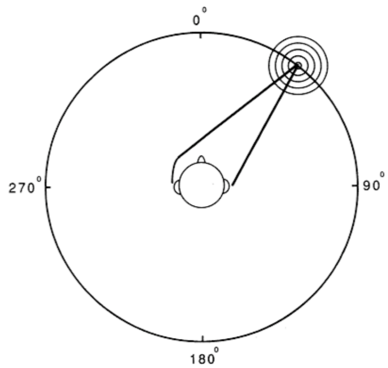


Early Reflections: part of the indirect sound that reaches the listener first and can be decomposed into discrete sound reflections.

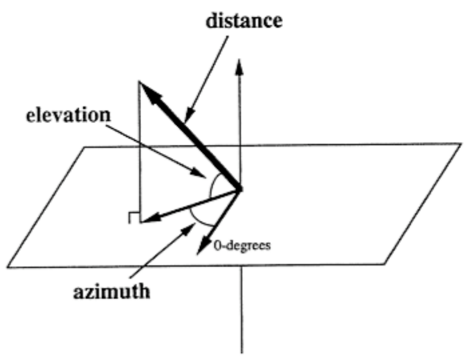
Reverberation: part of the indirect sound that follows early reflections and is made up of a continuous, dense blend of reflections.

2D planes intersecting the head





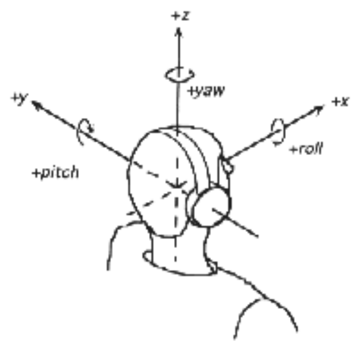
Horizontal Plane Azimuth Angle



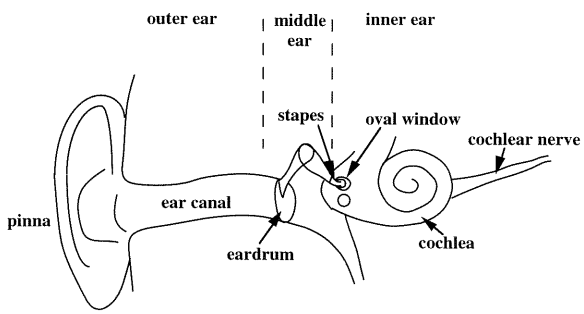
3D Coordinates

Head orientation:

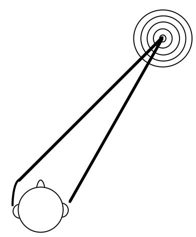
Yaw, pitch and roll.



Parts of the ear.



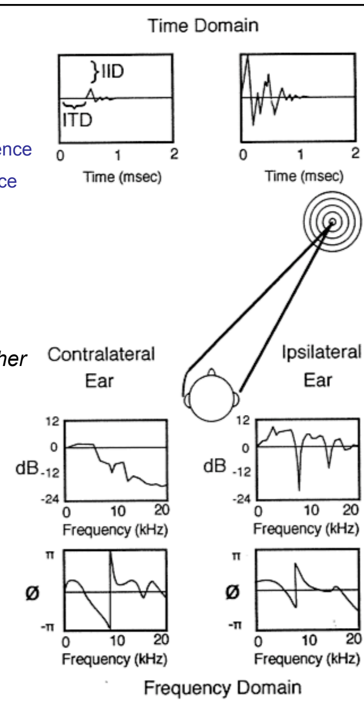
Physical Acoustics and Directional Hearing



Time Domain:
Impulse Response

IID = Interaural Intensity Difference

ITD = Interaural Time Difference



impulse

Contralateral ear *is further from the sound source.*

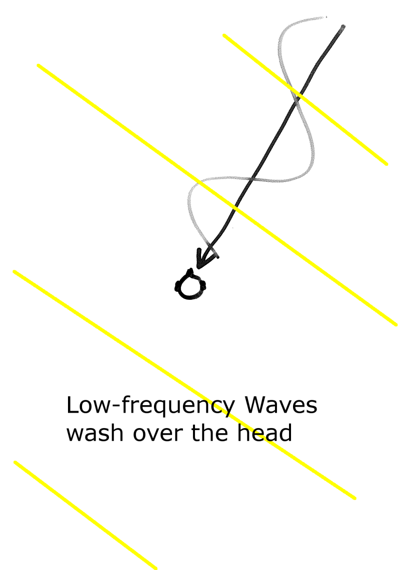
Ipsilateral ear *is closer to the sound source.*

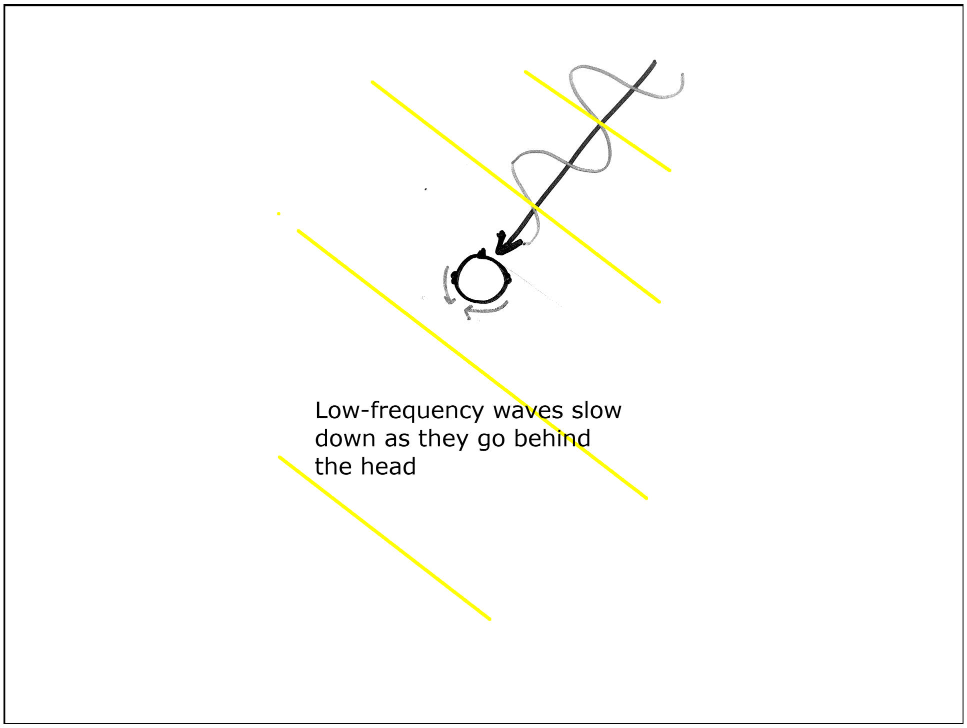
Frequency Domain:
HRTF = Head-Related Transfer Function

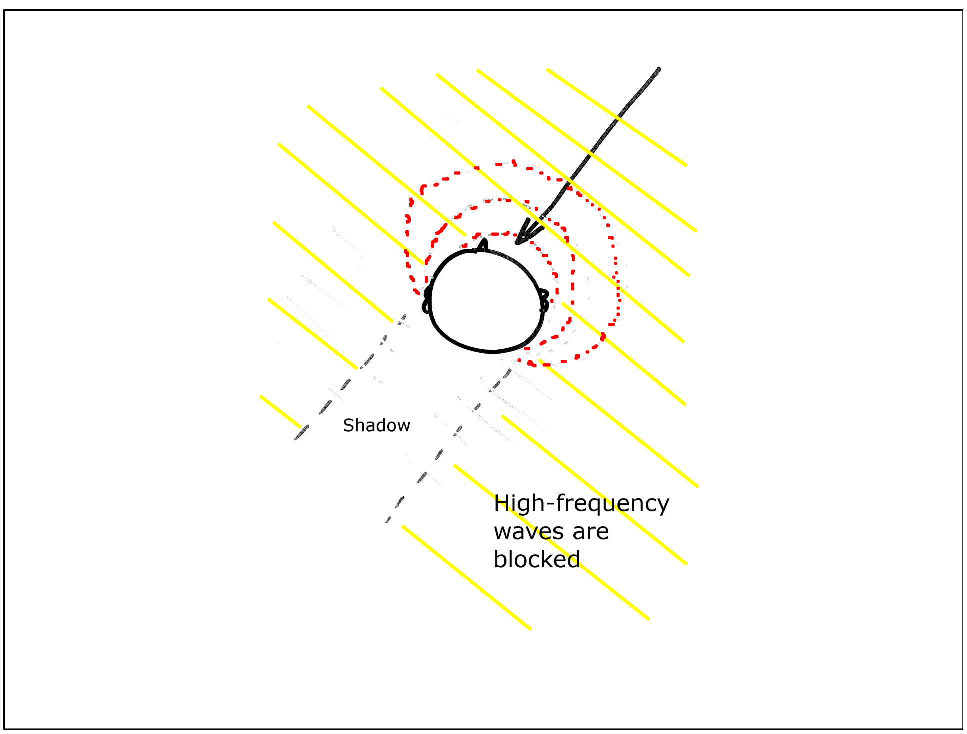
Consider the wavelengths!

Frequency (Hz)	Wavelength
100	11.3 ft
500	2.26 ft
1000	1.13 ft
1500	9 in
2000	6.75 in
4000	3.37 in
8000	1.7 in
16000	.84 in

Wavelength = speed of sound / frequency = 1130 ft/sec / frequency




Low-frequency waves slow down as they go behind the head


Shadow
High-frequency waves are blocked

What are the components of the acoustic system?

Shoulders and Torso (150Hz – 3kHz reflection)

Head (150Hz and up)

Pinna (3.5kHz and up reflection)
Conchae (2-5kHz resonance)

Ear Canal (3kHz and 9kHz resonance)

Each responds differently depending on its size

Pinna Cues

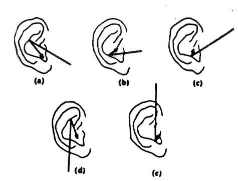


Fig. 7. Drawings of a pinna showing the path of the first major coherent reflection for sound sources at various elevations.

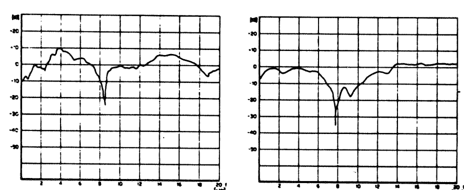


Fig. 12. (a) Pinna response for a sound source at ear level and 45° azimuth. The response is similar to the response shown in Fig. 9. except that the peak at 4 kHz is more pronounced. (b) An example of spectral aberrations that occur when two drivers are misaligned. In this case the acoustic center of one driver was approximately 1 in (25 mm) behind the second driver.

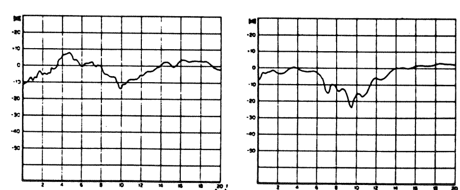
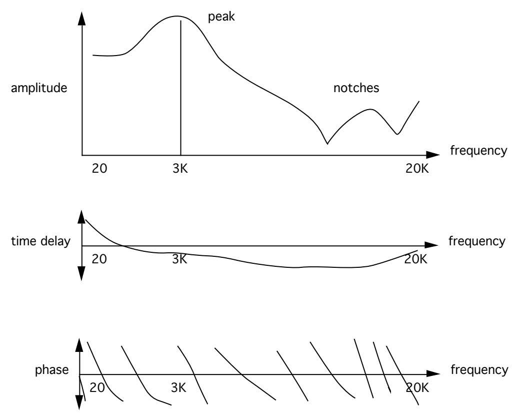
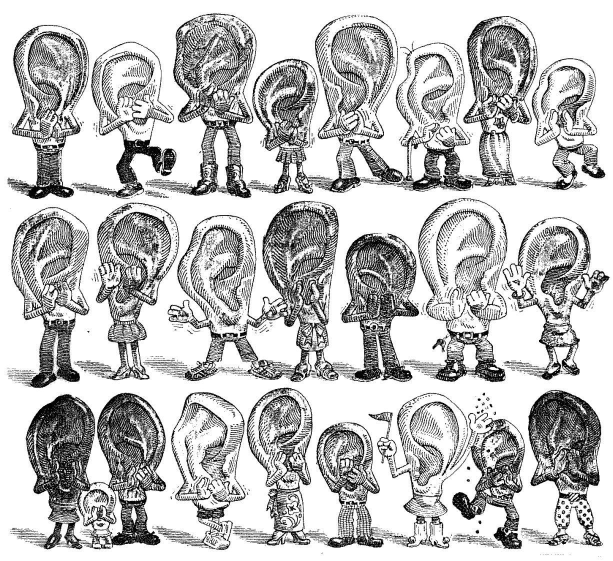


Fig. 13. (a) Pinna response for a sound source above ear level and at 45° azimuth. (b) A spectral minimum created when the acoustic center of one driver was approximately 0.7 in (18 mm) behind a second driver. A delay distance of 0.7 in (18 mm) mimics the pinna response for a sound source above ear level at 45° azimuth.

Composite of the complete acoustic system is captured in a Head-Related Transfer Function (HRTF)



Phase is typically too hard to look at.



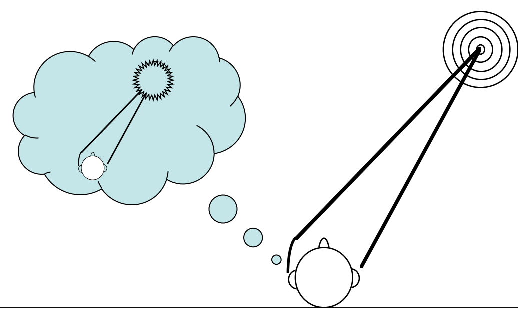
Individual Differences

Psychology of Spatial Hearing

There are *acoustic events* that take place in the environment.

These can give rise to the perception of *auditory events* that are subjective experiences.

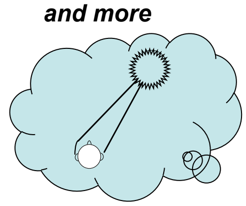
In the same way that sound intensity has a relationship with perceived loudness and frequency has a relationship with perceived pitch, the spatiality of these *acoustic events* has a relationship with the spatiality of the *auditory events* and that relationship is complex.



Psychology of Spatial Hearing

What are some of the spatial aspects of auditory events?

Left/Right Position	**Focus**	**Room dryness/wetness**
Elevation	**Volume**	**Room size**
Close/Far	**Depth**	**Room dryness/wetness**
Front/Back Position	**Direction of motion**	**Speed**
Orientation	**Enclosedness**	**Inside/Outside Head**
Width/Height	***and more***	



Psychology of Spatial Hearing

The scope of terms can depend on context, but for the most part:

What is included in 'spatial hearing'?
Properties of auditory events other than those of the sounding objects themselves---includes the perceived auditory environment and position of events with it. Considers acoustic direct and indirect sound.

What is included in 'localization'?
Spatial properties of auditory events other than those related to the direct perception of the environment---includes the perceived direction and distance of events. Mostly considers direct sound but also considers indirect sound's effect on the perception of direct sound.

What is included in 'directional hearing'?
Localization properties of auditory events---includes the direction and distance of events but only considers direct sound.

The truth is that everything is interrelated in spatial hearing.

Psychology of Direction Hearing

Some important questions

- What are the spatial attributes of auditory events?
- What is the relationship between physical acoustic information and people's perception of these attributes?
- How do you measure the properties of perceived events? How do you ask subjects to respond and what tasks do you give them?

Directional Hearing

Historical Legacy:

Before there was equipment to measure HRTFs, hearing scientists could experiment with ITD and IID

There is a lot to learn with simple means.

Directional Hearing

Duplex Theory of Localization
(Rayleigh 1907)

Good First Approximation

1) Interaural Time Difference (ITD)
   ITD ranges up to 800 microseconds
   ITD works best below 1500 Hz

2) Interaural Intensity Difference (IID)
   IID ranges up to 14 dB at high frequency
   IID works best above 1500 Hz

Duplex Theory of Localization

Let's do some simple experiments

Establish test conditions:

1. **Blocking one ear with ear plug**
2. **Inserting tubes of unequal length over the ears**
3. **Changing the shape of the pinna**
4. **Comparing localization of complex vs impoverished sound sources (square plate with many sine waves vs. circular plate with one sine wave)**

Duplex Theory of Localization

What did/should we learn in our simple experiments?

#1 Blocking one ear with ear plug
The listener's judgment of sound location was biased toward the ear with the greater intensity. (IID)

#2 Inserting tubes of unequal length around the ears
The listener's judgment of sound location was biased toward the ear with the shorter tube. (ITD)

#3 Changing the shape of the pinna (or using two ear plugs)
The listener's judgment of sound elevation was greatly diminished. (Pinna Cues)

#4 Comparing localization of complex vs impoverished sound sources (square plate with many sine waves vs circular plate with one sine wave)
The listener could easily localize the rich sound and hardly localize the sine wave.

Other:
Despite some very unnatural acoustics, listeners perceived a single source location for the sound event.

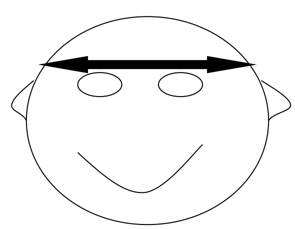
With the invention of electronic equipment,

Testing mostly done with headphones! Because they provide a test situation that seems controlled.

ITD and IID only move sound image left and right inside the head. How do you ask listeners to describe their experience?

"Laterialization"

Left and Right positioning inside the head



In headphone listening, sound sources typically lack 'externalization.'

ITD Alone / IID Constant

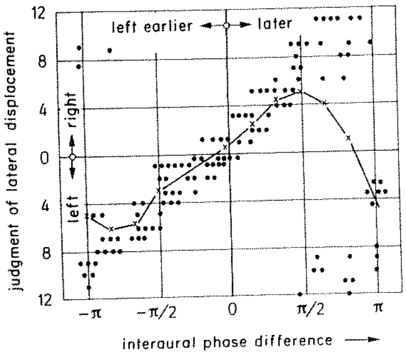
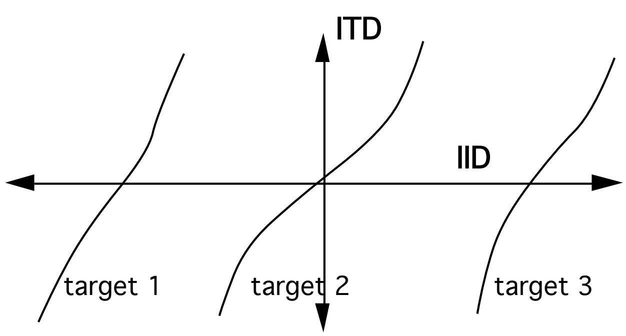


Figure 2.72
Lateralization curve for a 600 Hz sinusoidal signal (after Sayers 1964). The points are individual judgments; the solid line represents the average value.

The pairing of ITD and IID do not have to match 'normal' physical acoustics.

Time and Intensity Trading

ITD and IID can be traded off against one another.



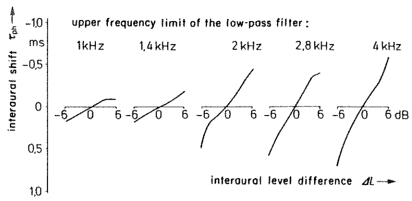


Figure 2.87
Trading curves of a typical subject for low-pass, band-limited clicks at a level of 20 dBSL (after Harris 1960).

High frequency and low frequency content have different trade-offs.

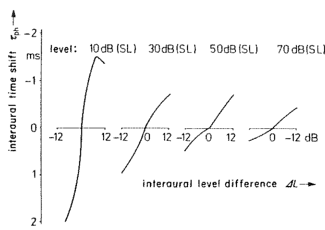


Figure 2.86
Trading curves of a typical subject for broadband clicks at various levels (after David, Guttman, and van Bergeijk 1959).

Absolute level has an impact on the results!

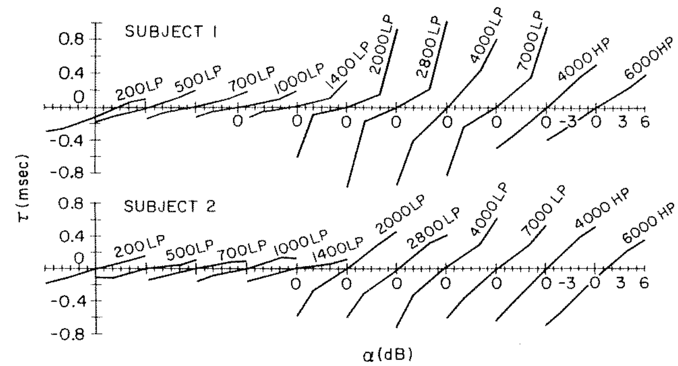
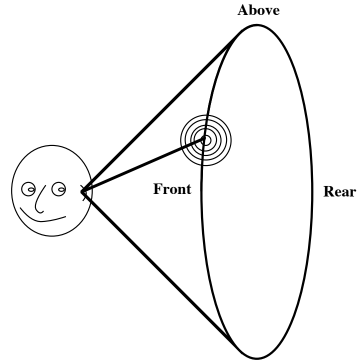


FIG. 9. Time–intensity trading data for clicks as a function of the click cutoff frequency. LP denotes low pass and HP high pass. The overall level of the clicks was 20 dB SL, and the data were obtained using the centering method. [Adapted from Harris (1960).]

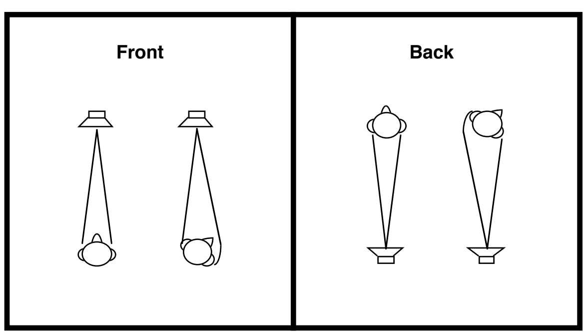
Are IID and ITD sufficient for localization?

No, consider the "Cone of Confusion"



How do listeners resolve the "Cone of Confusion"?

Dominant Cue: Dynamic Head Movement



Wallach, H. 1940. "The Role of Head Movements and Vestibular and Visual Cues in Sound Localization" *Journal of Experimental Psychology* 27(4): 339-68.

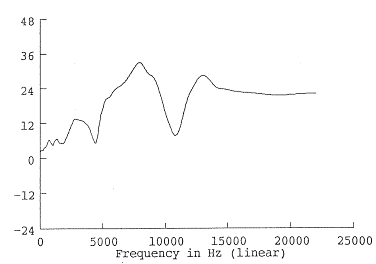
In the absence of head movement?

Frequency-dependent interaural differences in time and intensity (HRTFs)

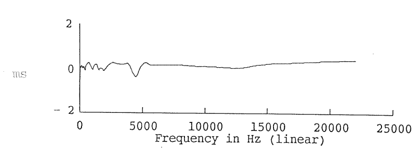
- Frequency-dependent level differences dominate at high frequencies
- Frequency-dependent time differences dominate at low frequencies
- And, . . .

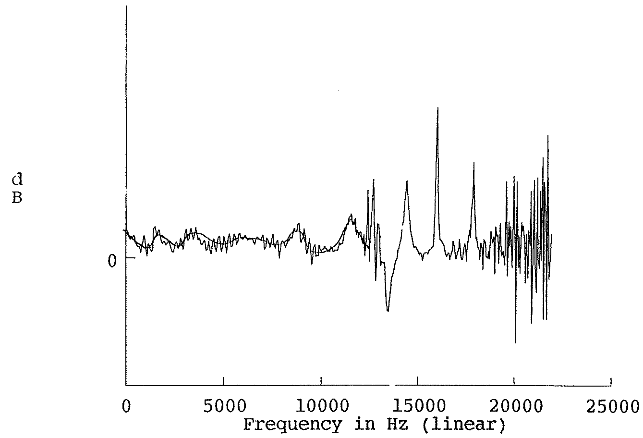
But even with HRTFs, what acoustic information are people using?

Interaural Magnitude Difference?



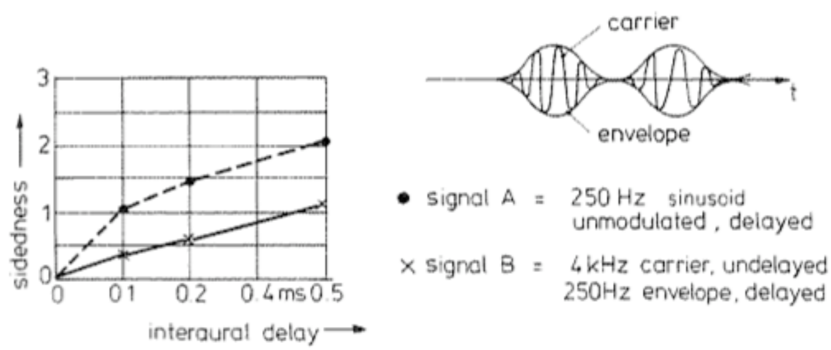
Interaural Time Difference?





group delay diff. (4 to −2 msec) for right, plane=el0, angle=30

On-going time differences in high frequencies affect localization (temporal envelopes)



Blauert, 1982

Low-frequency interaural time differences dominate localization judgments

When different kinds of directional information (IID, ITD, pinna cues) are contradictory:

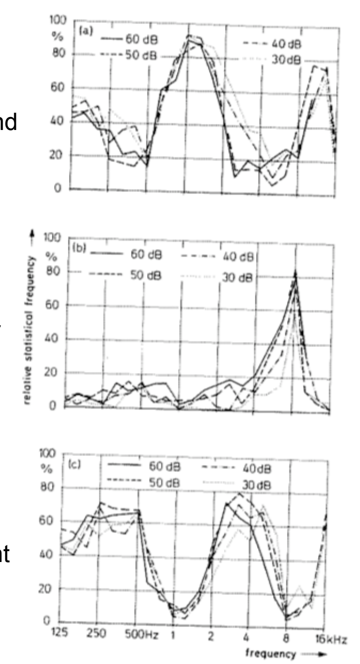
For broadband sounds:
subjects' judgments tend to follow ITD.

For high-pass sounds (without significant low-frequency information):
subjects' judgments tend to follow IID and pinna cues

Wightman and Kistler 1992 "The dominant role of low-frequency interaural time differences in sound localization"

Listeners appear to give preference to particular frequency bands

behind



over

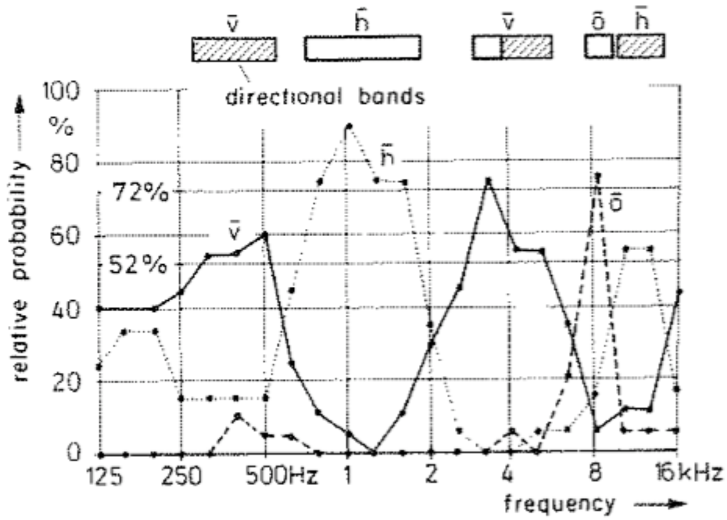
front

1/3-octave noise in median plane:

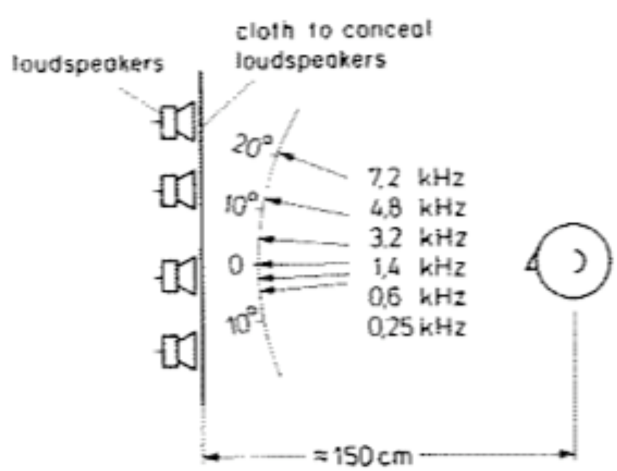
Blauert 1974

Summary of directional bands

v = front, h = behind, o = above



Directional bands may be mapped in a more continuous and systematic way

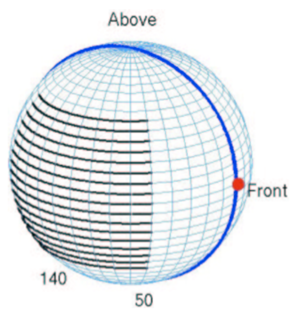


Roffler and Butler 1968

For sounds that are to the side, the contralateral ear appears to be the dominant one

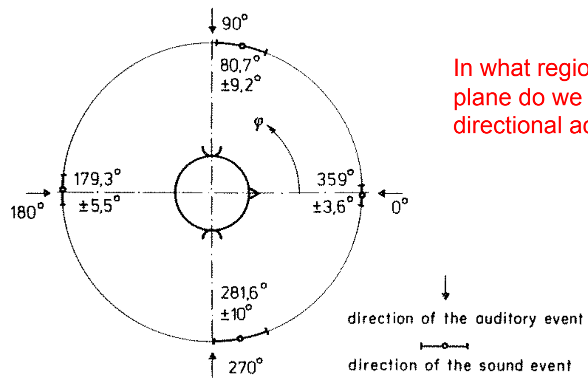
Humanski and Butler 1988

Nakamura, Mano, and Martens 2002



Other Questions:

- How accurately do we localize sound? (“Spatial accuity” / “Spatial Blur”)
- How well do we localize with other people’s ears? (“Individual Differences”)



In what regions of the horizontal plane do we have the greatest directional acuity?

Figure 2.2
Localization blur $\Delta\varphi_{min}$ and localization in the horizontal plane (after Preibisch-Effenberger 1966a and Haustein and Schirmer 1970; 600–900 subjects, white-noise pulses of 100 ms duration, approximately 70 phon, head immobilized).

In what regions of the median plane do we have the greatest directional acuity?

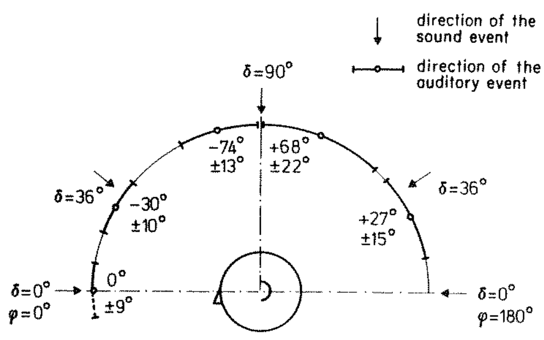


Figure 2.5
Localization and localization blur in the median plane for continuous speech by a familiar person (after Damaske and Wagener 1969; 7 subjects, 65 phon, head immobilized). Note that the view is different from that in figure 2.2.

Why would there be such a difference in comparison to the horizontal plane?

In what range of distance judgments do we have the greatest acuity?

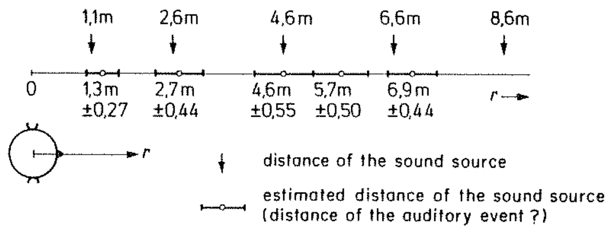


Figure 2.8
Localization blur Δr_{min} and localization between the distances of the sound source and the auditory event in the forward direction (after Haustein 1969; impulse sound, approximately 70 phon, from a distance of 4 m, 20 subjects, head immobilized).

"Individual Differences"

- *Individuals generally localize better with their own HRTFs than with those of others*
- *Some individuals have HRTFs that are superior and these HRTFs can sometimes improve the localization of others*
- *In order for one individual's HRTFs to work for another, the head sizes must be approximately the same, and*
- *Localization can be achieved with synthetic directional transfer functions whose details differ from measured HRTFs*

And, is it one ear or two?

People with asymmetric hearing loss can still localize!

Are people using ITD and IID when available, but HRTF cues are processed monaurally?

How well are we able to localize with only one ear?

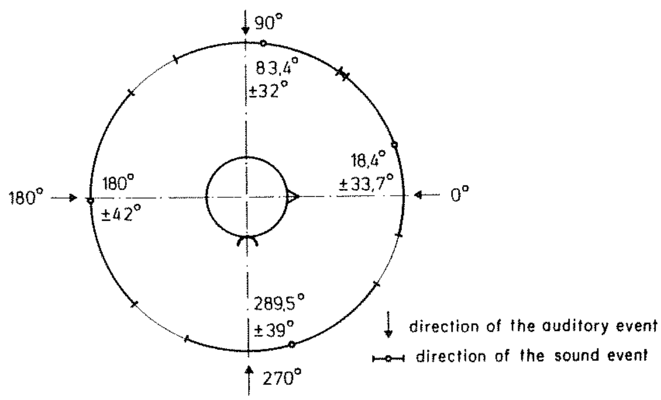


Figure 2.10
Localization blur $\Delta\varphi_{min}$ and localization in the horizontal plane with total deafness in the left ear and normal hearing in the right ear (Preibisch-Effenberger 1966a; white-noise pulses of 100 ms duration, approximately 70 phon, 32 subjects, head immobilized).

What does this tell us?

Monaural Listening

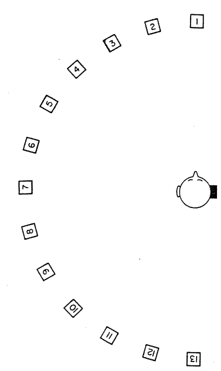


FIG. 1. A diagramatic sketch of the loudspeaker arrangement.

Humanski & Butler 1985 "The influence of monaural spectral cues on binaural localization" JASA 77:1.

Subjects make judgements on the basis of the frequency band only!

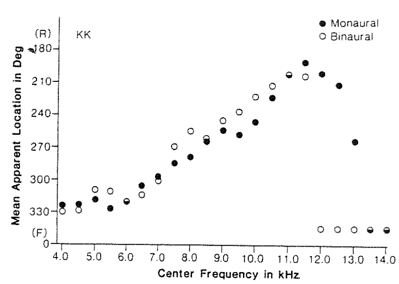


FIG. 2. The mean apparent location of narrow noise bands as a function of center frequency for subject K.K. Actual location was 270 deg azimuth.

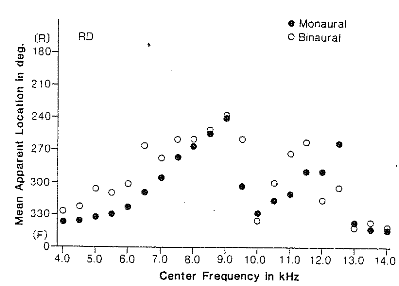


FIG. 3. The mean apparent location of narrow noise bands as a function of center frequency for subject R.D. Actual location was 270 deg azimuth.

Wightman & Kistler (1989)

Headphone simulation of free-field listening

I. Stimulus synthesis

II. Psychophysical validation

I. Stimulus synthesis

Goal is to be able to capture free-field listening acoustics with headphones.

- 200-14,000 Hz
- Greater than 20 dB S/N (*only 20 dB?)*
- 8 loudspeakers on movable arch creating 144 directions
- With & without bite bar

Measure loudspeaker-delivered HRTFs and compare to headphone-delivered HRTFs

HRTF measurement system

Figure 4.22. A subject in the anechoic chamber at the psychoacoustic laboratory at the Waisma Center of the University of Wisconsin, Madison. *Courtesy of Fred Wightman and Doris Kistler*.

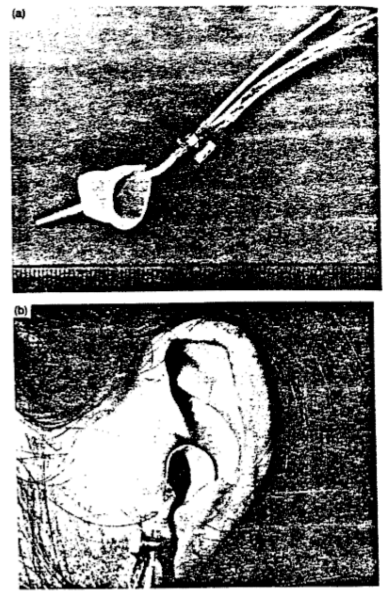


FIG. 1. Photographs of the custom earmold shell microphone holder used for acoustical transfer function measurements. (a) The shell with microphone probe tube inserted. The major divisions on the scale are 1 cm and the minor divisions are 1 mm. (b) The assembly in place in a subject's ear.