Notes on Sound Localization.
Physical Cues of Distance, Azimuth, and Elevation. and Behavioral Measures.

Louis D. Braida

Massachusetts Institute of Technology
Department of Electrical Engineering and Computer Science
6.182 - Psychoacoustics Project Laboratory

©L. D. Braida
March 27, 2010
1 Spatial Hearing

Why study spatial hearing (sound localization)?

- Survival value.
  
  Auditory systems must inform their owners of what is threatening and where it is. The location is by far the more important of the two. It dictates the direction of visual contact and it offers the directions for flight. (Tobias, 1972)

- Precise temporal discrimination.

- Spatial diversity in sensory systems.

- Source separation.

- Sense of auditory space.

- Simulated (virtual) acoustic environments.

Why is the study of spatial hearing complicated?

- Multiplicity of acoustic environments.
  - Anechoic – Echoic (reverberant).
  - Quiet/Interfering sounds

- Variety of source characteristics
  - Stationary/Moving sources
  - Single/Multiple sources
  - Spectral/Temporal characteristics

- Non-acoustic constraints
  - A priori information
    - Vision
    - Tactile sensations

- Complicated Geometry
  - Irregular surface of the head and torso
  - Wide range of $\lambda$s relative to head geometry.
  - Wide range of $r$s relative to head geometry.

These notes focus on the auditory localization of single fixed sources, consisting of sinusoids, noises, and clicks.
2 Coordinate systems

![Coordinate system](image)

Figure 1: Coordinate system used to describe the location of a sound source located at a distance \( r \) from the center of the head. The angle \( \theta (\pi/2 \leq \theta + \pi/2) \) is the azimuth of the source and the angle \( \phi (\pi \leq \phi + \pi) \) is the elevation of the source.

Most studies of sound localization focus on the “free field” or anechoic acoustic environment. Assume that a source in anechoic space creates a pressure \( P_0 e^{j\omega t} \) at the position of the center of the head when the head is absent, and that when the head is present the pressure waveforms at the left and right ears are \( P_L e^{j\omega t} \) and \( P_R e^{j\omega t} \). Then the head-related transfer functions (HRTFs) are defined as

\[
H_L(\omega, \theta, \phi, r) = \frac{P_L}{P} \quad (1)
\]
\[
H_R(\omega, \theta, \phi, r) = \frac{P_R}{P} \quad (2)
\]

If the source is a simple point source of strength \( U_0 \),

\[
P = j \frac{2\rho_0 f U_0}{r} e^{-j \frac{2\pi f r}{c}} \quad (3)
\]

When determining the location of the source, the listener has access to pressure waveforms at the two ears whose complex amplitudes are

\[
P_L = j \frac{2\rho_0 f U_0}{r} e^{-j \frac{2\pi f r}{c}} \times H_L(\omega, \theta, \phi, r) \quad (4)
\]
\[
P_R = j \frac{2\rho_0 f U_0}{r} e^{-j \frac{2\pi f r}{c}} \times H_R(\omega, \theta, \phi, r) \quad (5)
\]
and is often assumed to compute the complex ratio
\[
G(\omega, \theta, \phi, r) = \frac{P_L}{P_R} = \frac{P_L/P}{P_R/P} = \frac{H_L(\omega, \theta, \phi, r)}{H_R(\omega, \theta, \phi, r)}
\] (6)

Note that, unlike \(P_L\) and \(P_R\), \(G\) is independent of \(U_0\). Moreover, when \(r\) is large enough that the waves arriving at the head can be approximated as plane waves, \(G\) is independent of \(r\), i.e.,
\[
\lim_{r \to \infty} H_L(\omega, \theta, \phi, r) = H'_L(\omega, \theta, \phi) \quad (7)
\]
\[
\lim_{r \to \infty} H_R(\omega, \theta, \phi, r) = H'_R(\omega, \theta, \phi) \quad (8)
\]
\[
\lim_{r \to \infty} G(\omega, \theta, \phi, r) = G'(\omega, \theta, \phi) \quad (9)
\]

### 2.1 Symmetries

For an arbitrarily shaped head, it is difficult to make accurate statements about \(H_L\), \(H_R\), or \(G\). If the head is symmetric about the vertical median plane, however, then
\[
H_L(\omega, \theta, \phi, r) = H_R(\omega, -\theta, \phi, r) \quad (10)
\]
so that
\[
G(\omega, 0, \phi, r) = 1 \quad (11)
\]

Under some conditions, the sound pressures \(P_L\) and \(P_R\) are nearly the same as at points on the interaural axis of a rigid sphere with the same diameter \(D\) as a head (typically 18 cm). In that case, the transfer functions satisfy stronger symmetry properties.
3 Distance

When the sound source is in the vertical median plane ($\theta = 0$) the physical cues for distance are expected to be the same whether the listener localizes on the basis of the sound at the left or right ears, alone (i.e., using monaural cues). The physical cues for distance $r$ are then restricted to:

- Overall Sound Level (requires familiarity with source properties).
- Direct-to-reverberant ratio (in echoic environments).

When the source is not in the vertical median plane, the sound waveforms at the two ears are not identical, so that there are are interaural as well as monaural cues. As noted above, however, the interaural ratio of pressures $\bar{G}$ is independent of $r$ when the source is far enough away as to create plane waves at the position of the head. Thus the cues for distance in this case are essentially monaural.

When the source is close to the head, however, $\bar{G}$ depends on $r$ because the distance from the source to the left ear can differ substantially from the distance to the right ear. For a head-sized ($2a \approx 18$ cm diameter) rigid sphere, the dependence of $|\bar{G}|$ on $r$ is more dramatic than that of $\bar{G}$ (Brungart and Rabinowitz, 1999). In this case the interaural amplitude ratio provides a binaural cue that can be used to estimate distance.

![Figure 2: Dependence of direct and reverberant sound pressure levels on distance from a sound source.](image-url)
4 Azimuth

The primary physical cues for azimuth are the interaural amplitude ratio \(|G|\) (sometimes referred to as the interaural level difference) and the interaural phase \(\angle G\). It is often easier to deal with the interaural time difference \(\tau = \angle G/\omega\) than with \(\angle G\) itself.

4.1 Theory

![Diagram](image)

Figure 3: Approximations for the acoustics of the head with radius 2\(a\) introduced into a plane-wave sound field at low (left panel) and high (right panel) frequencies.

For distant sources, the physical bases for interaural magnitude and phase relations can be appreciated in the two limiting cases illustrated in Fig. 3:

1. \(ka \ll 1\). In the limit of very low frequencies, the pressures at the ears created by a plane wave source in the presence of a spherical head are the same as those at two points in free space on the interaural axis that are centered on the head and separated by 1.5 times the head diameter. When this limiting condition applies, \(|G| \approx 1\) and

\[
\tau \approx 3a \sin \theta / c. \tag{12}
\]

2. \(ka \gg 1\). In the limit of very high frequencies, the pressures at the ears created by a plane wave source can be computed by taking the diffraction induced by the head into account. Although the \(|G|\) is difficult to compute, Woodworth (1962) has shown that

\[
\tau \approx a(\theta + \sin \theta)/c. \tag{13}
\]

The dependence of \(\tau\) on frequency is a manifestation of the dispersive nature of sound propagation near a rigid body. The dispersion is appreciable both for azimuths near the
vertical median plane: \( \tau \approx 3a\theta/c \) at low frequencies and \( \tau \approx 2a\theta/c \) at high frequencies and for azimuths near \( \theta = \pi/2 \): at low frequencies \( \tau = 3a/c \) (785 \( \mu \)sec for an 18 cm diameter head), whereas at high frequencies \( \tau = 2.58a/c \) (672 \( \mu \)sec for an 18 cm diameter head).

### 4.2 Measurements on a Sphere

Measurements of \( |H'_L(\omega, \theta, \phi)| \) made on a rigid sphere model of the head (Wiener, 1947) are show in Fig. 4 and Fig. 5.

**Figure 4:** Measurements and theoretical predictions of Wiener (1947). The HRTF is plotted as a function of \( ka = \omega a/c = 2\pi a/\lambda \), where \( a = 9.7 \) cm. The angle \( \theta \) is the parameter that distinguishes the different plots. The panel for \( \theta = 0 \) corresponds to the source on the interaural axis at a large distance from the “ear” at which pressures are measured.

### 4.3 Measurements of ITD on a Mannikin

Measurements of the Interaural Time Difference, \( \tau \), made with long duration sine-waves on a isolated manikin’s head (Kuhn, 1977) are shown in Fig. 6 and are compared to the predictions of Eq. 12 and 13 in Fig. 7.
Figure 5: The measurements of Wiener (1947) replotted as a function of the angle between the interaural axis and the line from the center of the head to the sound source. Frequency is the parameter that distinguishes the different plots.
Figure 6: Measurements of the interaural time difference for sine-waves made on the isolated head of a mannikin (no torso) by Kuhn (1977). The different curves correspond to azimuths in steps of $15^\circ$ ranging from $15^\circ$ (bottom) to $90^\circ$ (top).
Figure 7: Dependence of the interaural time difference (ITD, $\tau$) on azimuth ($\theta_{inc}$) from for both low-frequency (upper curves) and high-frequency (lower curves) tones. Data were taken with long duration sine-waves on the isolated head of a mannikin. The symbols used for the data are: ● 360 Hz, △ 500 Hz, × 3000 Hz. The ▼ and ○ symbols denote numerical predictions for a sphere of 18.6 cm diameter based on Eq. 12 and 13. From Kuhn (1977).
4.4 Measurements of IAR on a Real Head

Measurements of the Interaural Amplitude Ratio $|G'(\omega, \theta, 0)|$ can be inferred from measurements reported by Shaw\(^1\) (1974) of the sound pressure in the vicinity of the eardrum produced by distant sources as a function of azimuth in the horizontal plane. Shaw reports these data in three ways:

1. As $T(\omega, \theta)$, the magnitude of the ratio of the pressure at the eardrum to the pressure at the location of the center of the head in free space,

2. As $D(\omega, \theta)$,

\[
D(\omega, \theta) = \frac{T(\omega, \theta)}{T(\omega, 0)},
\]

which is the ratio of the pressure at the eardrum produced by a source at azimuth $\theta$ to the pressure at the eardrum produced by a source in the vertical median plane.

3. As $|G'(\omega, \theta, 0)|$, the magnitude of the interaural amplitude ratio. Note that

\[
|G'(\omega, \theta, 0)| = \frac{D(\omega, \theta)}{D(\omega, -\theta)}.
\]

Although the transformation of sound pressure from the field to the eardrum ($T(\omega, \theta)$), includes both the transformation from the surface to the head to the eardrum as well as that from the field to the surface of the head, the first transformation is largely independent of azimuth\(^2\) below roughly 8 kHz. Thus, $D(\omega, \theta)$ may give a more realistic impression of the azimuthal dependence of the transformation than $T(\omega, \theta)$. Moreover, it is nearly identical for the two ears. Thus, the interaural amplitude ratio at the surface of the head should be nearly identical to the interaural amplitude ratio at the eardrums.

Figures 8 and 9 present “self-consistent” estimates of the functions $T(\omega, \theta)$ (Fig. 8) and $D(\omega, \theta)$ (Fig. 9). Note that the function $T(\omega, 0)$ peaks at $\omega \approx 2\pi 2500$ rad/sec and had a relatively deep notch at $\omega \approx 2\pi 9500$ rad/sec. Much of this dependence ($\approx 10$ dB) can be attributed to the transfer function between the sound pressure at the ear drum and entrance to the ear canal. Fig. 9 illustrates the dependence of $D(\omega, \theta)$ on $\omega$ for sources at various azimuths. Note that there is no curve plotted for $D(\omega, 0)$ as this would be the abscissa of the graph.

Fig. 10 illustrates the dependence of $D(\omega, \theta)$ on $\omega$ for sources at azimuths of $\pm \pi/2$ and the corresponding interaural amplitude ratio, $|G'(\omega, \theta, 0)|$. Fig. 11 shows the dependence of $D(\omega, \theta)$ and the interaural amplitude ratio, $|G'(\omega, \theta, 0)|$, on $\theta$ for several values of $\omega$. Note that the interaural difference is negligible for $\theta \approx 0$ and for $\theta \approx \pi$, reflecting the near symmetry of the head. The interaural difference for a source at $\theta = \pi/2$ is small for low frequencies (3 dB at $\omega = 2\pi 200$ rad/sec, but can become quite large at high frequencies (22 dB at $\omega = 2\pi 12000$ rad/sec. Also note that, $D(\omega, +\pi/2)$ is relatively well predicted by modelling the head as a sphere, whereas $D(\omega, -\pi/2)$ is not.

---

\(^1\)Shaw included 12 sets of data in his estimates of transfer functions.

\(^2\)For sounds whose wavelengths are large relative to the roughly 1 cm diameter of the ear canal.
Figure 8: Dependence of $T(\omega, \theta)$ on frequency and azimuth. From Shaw (1974).
Figure 9: Dependence of $D(\omega; \theta)$ on frequency and azimuth. From Shaw (1974).
Figure 10: Measurements (points) on human heads, summarized by Shaw (1974), of $D(\omega, -\pi/2)$ (bottom panel), $D(\omega, +\pi/2)$ (middle panel), and of the interaural amplitude ratio for tones $|\mathcal{G}(\omega, \pi/2, 0)|$ (top panel), all expressed in dB. Solid curves are “self-consistent” curves fitted to a large body of measurements. The dashed curves are computed for an ideal spherical head of radius 8.75 cm.
Figure 11: Measurements on human heads, summarized Shaw (1974), of $D(\omega, \theta)$ (left panels) and the interaural amplitude ratio $|\mathcal{C}'(\omega, \theta, 0)|$ (right panels) as a function of azimuth $\theta$ for tones of low- mid- and high frequency. Data points are measured values, solid curves are “self-consistent” curves fitted to a large body of data. Note that the interaural amplitude ratio is simply $D(\omega, \theta)/D(\omega, -\theta)$. 


5 Elevation

Determination of elevation ($\hat{\phi}$) for sources in the vertical median plane would be impossible if the head were isolated from the torso and exhibited spherical symmetry. If the head were symmetric about the vertical median plane, only monaural cues would be available for localization in the vertical median plane.

At the moment, it is controversial whether such symmetry exists, or whether there is a vertical median surface for real heads. For most heads, asymmetries in sound transmission are most likely to occur at high frequencies, where they would produce binaural cues for elevation.

Measurements of sound transmission for one subject that exhibit such asymmetries (Searle et al., 1975) are shown in Fig. 12. Sound transmission exhibits a mild low-pass characteristic for sources located above the head. Both sources in front of and behind the head, however, exhibit a transmission characteristic with a deep notch whose center frequency is dependent on the elevation of the source.

These measurements suggest the following:

- Localization of sources in the vertical median plane requires the presence of high frequency (\(\gtrsim 4 \text{ kHz}\)) stimulus components.
- For narrow-band sources, localization of sources in the vertical median plane requires either 1) a priori knowledge about the source intensity or 2) the presence of interaural asymmetries.
- For broadband sources with relatively smooth spectra, the frequency of the transmission notch may provide a cue to elevation that is independent of source characteristics. Interaural asymmetries provide a second possible cue.

The origin of the notches is fairly well understood. The center frequencies of the notches notches are typically in the 6–12 kHz region, corresponding to sounds with wavelengths of 3–6 cm. The pinnae are the most obvious structures with dimensions comparable to these wavelengths. Several investigators have proposed that such notches result from the destructive interference that occurs when there are several transmission paths from the sound source to the eardrum, as shown schematically in Fig. 13.

If two different paths, with a path-length difference of $cT$, are available for an impulse to reach the eardrum, then the impulse response corresponding to the transmission characteristic is of the form

$$h(t) = A\delta(t) + B\delta(t - T)$$

and the complex response to a sinusoidal source is proportional to

$$A + Be^{-j\omega T}$$

The magnitude of the the transfer characteristic is proportional to

$$|H(\omega)| = \sqrt{A^2 + B^2 + 2AB \cos \omega T}$$

15
Thus the acoustic transfer characteristic exhibits a minimum $A - B$ at frequencies $f = 1/(2T), 3/(2T), \ldots$ and a maximum $A + B$ at $f = 0, 1/T, 2/T, \ldots$. The notch depth (in dB) is

$$20 \log_{10} \frac{A + B}{|A - B|}$$

An example of the transfer function associated with two-path transmission is shown in Fig. 14.
Figure 12: Transfer functions for sources in the vertical median plane reported by Searle et al. (1975). Left to right pressure ratios are not shown.
Figure 13: Possible origin of the notches in the transmission characteristic associated with multiple transmission paths associated with pinna structures. The posterior wall of the concha reflects frontal sounds, so that both direct and time-delayed sounds from the source enter the external auditory canal. Because of the concha’s geometry, sounds at higher elevations have shorter reflected component path lengths, hence, less time delay, so that the perceived sound’s spectrum will have a notch at a higher frequency. Illustration from Hebrank and Wright (1974).

Figure 14: Magnitude of the transfer function associated with two-path transmission. In this case \( A = 1, B = 0.9 \) and \( T = 50\mu\text{sec} \).
6 The “Minimum Audible Angle”

Mills (1958) defined the “minimum audible angle” as the smallest detectable difference between the azimuths of two identical sources of sound.

6.1 Measurements

Mills conducted experiments to measure the minimum audible angle for tone pulses (Fig. 15) systematically as a function of azimuth ($\theta$) and frequency. Mills presented 1 s tone pulses with 70 ms rise/fall times at a sound pressure level 50 dB above the threshold of hearing (50 dB SL). The tone was initially presented from an angle $\theta$ and 3 s later at an angle of $\theta + \Delta \theta$ or $\theta - \Delta \theta$.

![Figure 15: Configuration used by Mills (1958) to measure the minimum audible angle.](image)

The observer reported whether the second pulse appeared to the left or right of the first. Typical results are shown in Fig. 16, together with Mills’ estimate of the minimum audible angle $\hat{\Delta} \theta$.

The dependence of $\hat{\Delta} \theta$ on $\theta$ and $\omega$, averaged over three young male undergraduate listeners with clinically normal hearing, are shown in Fig. 17. Note that a minimal audible angle of 1 deg corresponds to a time difference of roughly 15 $\mu$sec.

For source frequencies of 500 and 750 Hz., the minimal audible angle $\hat{\Delta} \theta$ is about 1 deg for sources located straight ahead of the listener. $\hat{\Delta} \theta$ increases rapidly as $\theta \to \pi/2$. Evident in this figure is the high relative difficulty of the task at frequencies in excess of roughly 1200 Hz, particularly when $\theta > \pi/4$. 
Figure 16: Typical psychometric function and data analysis used by Mills.
Figure 17: Measurements of the minimum audible angle reported by Mills (1958).
6.2 Analysis

Mills measured the changes in interaural level differences and interaural phase difference corresponding to differences in azimuth equal to the measured minimum audible angle $\Delta \theta$ as a function of frequency for $\theta = 0$. He compared these physical measurements with psychoacoustical measurements of the minimal noticeable differences in interaural phase $\Delta \phi$ and interaural level $\Delta I$ made for dichotic tonal stimuli presented via earphones. The comparison (Fig. 18) suggests that the physical differences are slightly smaller than the minimal noticeable differences. These results indicate that the ability to discriminate changes in interaural phase determine the minimal audible angle for frequencies under 1400 Hz, and the ability to discriminate changes in interaural amplitude determine the minimal audible angle for frequencies in the range 1400–5000 Hz.

Figure 18: Comparison of changes in interaural level and phase that occur when a source is moved from the median plane through a minimum audible angle with the minimal noticeable changes in interaural level and phase measured with earphones.
7 Ambiguity of Interaural Phase

Figure 19: Sinusoidal waveforms (from the bottom) of frequency 350, 700, and 1400 Hz, and versions of the waveforms delayed by 0.714 ms, the period of the 1400 Hz sinusoid.

The effect of an interaural delay of 714 µs (roughly the maximum interaural delay (Kuhn, 1977)) on sinusoidal waveforms with frequencies of 350, 700, and 1400 Hz is illustrated in Fig. 19.

The effect of various interaural delays on the 700 Hz waveforms is shown in Fig. 20. While the waveform delayed by less than half a period appears to lag the reference, the waveform delayed by more than half a period appears to lead the reference.

Mills (1972) notes that a time difference between two pure tones that produces a half cycle or more of phase lead \((\pi + \phi)\) is indistinguishable from an opposite phase difference that produces a phase lag of \((\pi - \phi)\). The minimum acoustical path between the ears (roughly 23 cm in air) is greater than one half of a wavelength for frequencies in excess of 750 Hz.

For a specified frequency \(\omega\), for which \((\pi + \phi')\) is the maximum possible interaural phase difference for the frequency \(\omega\), phase differences less than \((\pi - \phi')\) correspond to unique azimuths, but phase differences greater than \((\pi - \phi')\) can be produced by two different sound sources – one on each side of the head.

For frequencies above 1500 Hz, the acoustical path between the ears exceeds one wavelength, and there are possible source azimuths on both sides of the head that can produce any give phase difference.
Figure 20: Sinusoidal waveforms of frequency 700 Hz, for which half a period is 0.714 ms. Versions of the waveforms delayed by (from the bottom of the panel) 0.0, 0.714-0.200, 0.714, and 714+0.200 ms. Based on ongoing phase, the longest delayed waveform (714+0.200 ms) appears to lead the reference sinusoid (0.717 ms).
8 Earphone Listening

8.1 Discrimination of Interaural Level, Time, and Phase

The use of earphones permits the study of binaural hearing with stimuli whose interaural relations cannot be created by simple sources in anechoic space. In a typical binaural discrimination experiment with tonal stimuli, the listener would attempt to distinguish between the two pairs of sounds presented to the left and right ears ($l(t), r(t)$)

$$l(t) = A_L \cos (2\pi f(t + \tau /2))$$
$$r(t) = A_R \cos (2\pi f(t - \tau /2))$$

and ($l'(t), r'(t)$).

$$l'(t) = A_L(1 + \Delta_A/2) \cos (2\pi f(t + (\tau + \Delta\tau)/2))$$
$$r'(t) = A_R(1 - \Delta_A/2) \cos (2\pi f(t - (\tau + \Delta\tau)/2))$$

The “base” interaural amplitude ratio is $\alpha = A_L/A_R$, the base interaural time difference is $\tau$. $\Delta_A$ the fractional increment in the interaural amplitude ratio, and $\Delta\tau$ the increment in interaural time difference.

Summary measurements of the ability to discriminate changes in interaural level, phase, and time for diotic stimuli are shown in Fig. 21. Summary measurements of the ability to discriminate changes in interaural level for dichotic stimuli are shown in Fig. 23, and of interaural time for dichotic stimuli are shown in Fig. 22.

8.2 Lateralization of Images

Sounds presented via earphones typically produce internal auditory images, even when the interaural relations in level and phase are the same as those that would be produced by a simple source in free space. When the earphone signals are coherent, the position of the image is dependent on interaural parameters. An increase in level or a temporal advance of the sound at one ear relative to the other cause the image to shift towards that ear. The path of motion is usually perceived as along the top of the cranium. Internal sound images can vary in width, with the coherence of the earphone signals largely determining the width.

The degree to which a sound is perceived as arising from an external source, as opposed to a source within the head, is a subjective quantity, usually labile, and difficult to quantify. Reverberant sounds and the accurate reproduction of head/pinna related acoustic transformations typically increase the perceived degree of externalization. Sensory conflicts with kinesthetic, visual, and tactile cues, and with the listener’s a priori knowledge of the acoustic environment typically reduce the perceived externalization.

The apparent degree of lateralization has been measured in several studies. Data on the effect of interaural level and interaural phase on perceived lateralization are presented in Fig. 24 and Fig. 25.
Figure 21: Just noticeable differences for interaural level, phase, and time for tone bursts, as summarized by Durlach and Colburn (1972). These data are for tones of roughly 1 sec duration and for 50 dB SL, and apply to a base interaural amplitude ratio $\alpha = A_L/A_R = 1$ and a base interaural time difference $\tau = 0$. 
Figure 22: Just noticeable differences for interaural time for dichotic tone bursts, as summarized by Durlach and Colburn (1972). The tone bursts had a duration of 300 msec and rise-fall times of 50 msec. Each panel corresponds to a different value of base interaural amplitude ratio ($\alpha = A_L/A_R$). Within each panel, the abscissa is the value of the base interaural time difference $\tau$, and the ordinate is the minimum value of $\Delta \tau$ that can be discriminated. Collateral combinations correspond to $\tau > 0$ and $\alpha > 1$; antilateral configurations correspond to $\tau < 0$ and $\alpha > 1$. The tones were presented at 55 dB SL to the right ear and nonunity values of $\alpha$ were achieved by attenuating the tone to the left ear.
Figure 23: Just noticable differences for interaural time (top row of panels) and amplitude (bottom row of panels) for dichotic tone bursts, as summarized by Durlach and Colburn (1972). The tone bursts had a duration of 300 msec and rise-fall times of 50 msec. The abscissa for the left pair of panels is overall amplitude ($A_L = A_R$); for the middle pair of panels, the abscissa is the base time delay, $\tau$; for the right pair of panels, the abscissa is the base interaural amplitude ratio $\alpha = A_L/A_R$. In panels (a) and (b), $\tau = 0$ and $\alpha = 1$. In panels (c) and (d), $\alpha = 1$ and the tones were presented at 50 dB SL. In panels (e) and (f), the level of the tone in the left ear was 50 dB SL and different values of $\alpha$ were achieved by reducing the level of the tone in the right ear.
Figure 24: Relative lateralization of pure tones produced by interaural level differences, as summarized by Yost and Hafter (1997).

Figure 25: Relative lateralization of pure tones produced by interaural phase differences, as summarized by Yost and Hafter (1997).
It is also possible to displace the image from a centered position using an interaural time difference $\tau$ and recenter it by applying an interaural level difference $\alpha$. The time-intensity trading relation is heavily dependent on the spectral content of the stimulus, with smaller time differences required to match a given level difference when the stimulus contains only frequency components below 1400 Hz (see Fig. 26. The effectiveness of interaural time in determining the trading relation increases substantially as the sound level increases (Fig. 27 taken from Mills, 1972).

Figure 26: Time-level trading relation for filtered clicks as summarized by Durlach and Colburn (1972).

For sounds arising from simple sources in anechoic space, essentially the same interaural time difference applies to the envelope and fine time structure of the sound pulse. When the sounds are produced by earphones, the time difference applied to the envelope can be made to oppose that applied to the waveform fine structure so that there is no net tendency to displace the image of the sound from the centered position. As shown in Fig. 28, for sounds consisting of broadband noise, the envelope disparity is effective at counteracting the fine structure disparity only for relatively brief (30 ms or less) sound bursts. Even for short bursts, however, the envelope disparity must be 3–4 times as large as the fine structure disparity to compensate for it.
Figure 27: Level dependence of the time-level trading relation, from Mills (1972).

Figure 28: Envelope-fine structure trading relation for broadband noise as summarized by Durlach and Colburn (1972).
9 Localization

In the real world we rarely have to perform localization functions similar to the tasks studied by Mills (1958), i.e., making fine grained distinctions between different azimuths. Often a better model of the real-world situation is the pointing or identification task.

Localization performance has been evaluated in both subjective and objective experiments. On each trial of a subjective localization experiment the experimenter presents sound from a given position and the listener reports where the sound appeared to originate from. On each trial of an objective experiment, the experimenter presents sound from one of $N$ possible positions and the listener reports which of the $N$ possible positions produced what was heard.

The key difference between the two types of experiments is that the experimenter can determine whether the listener’s response is correct in an objective experiment, but not in a subjective experiment.

Neither of these two types of experiment is sufficient to describe localization completely. In a subjective experiment, the listener may exhibit large biases, even though responding in a highly consistent fashion, making it difficult to obtain measures of accuracy. In an objective experiment the listener can only make the errors that the experimenter anticipates.
10 Localization - Elevation

The problem of explaining localization in the vertical median plane results from the near-symmetry of the head about the vertical median plane. If the head were spherical, and detached from the body, a cone of confusion (Fig. 29) would exist: the locus of all source locations that give rise to the same interaural level and time differences.

![Cone of Confusion Diagram](image)

Figure 29: Illustrating the “cone of confusion.”

The role of the pinna(e) in the localization of sounds in the vertical median plane has been suspected for more than a century. However the precise contribution of the pinna has been controversial. Unlike the case of localization in the horizontal plane, where interaural level and interaural time provide much more precise bases than other acoustic variables, in the vertical plane there is not so great variation in the precision that several acoustic transformations provide.

Gardner (1973) summarized the resulting uncertainty in these terms:

Less clearly understood, however, is the nature of the clues that are actually utilized during a given period of observation and whether they are primarily monaural or binaural in origin or whether both influences are present and are most effectively used as a complementary combination of both.

Based on measurements of localization accuracy under conditions in which the cavities of the pinnae were subject to various degrees of occlusion, Gardner concluded that
Figure 30: Relative localization accuracy with occluded pinnae (Gardner, 1973).
1. Localization ability decreases with increasing occlusion.

2. Localization ability is better in the anterior than the posterior sector of the median plane.

3. Localization ability increases with the center frequency of the roughly 1/2 octave wide band of noise (over the range 2–10 kHz).

4. Localization ability increases with noise bandwidth for all degrees of occlusion.

To further elucidate vertical plane localization, Gardner studied the ability of listeners to localize wideband noise sources when interaural disparities were introduced by occluding nearly all of one or both pinnae (Fig. 30). Apparent from these results is

1. The existence of localization clues when both pinnae are occluded.

2. The existence of monaural localization clues.

![Figure 31: Effect of filtering on localization accuracy in the vertical median plane (Gardner, 1973).](image)

Gardner performed additional experiments in which the noise source was low- and high-pass filtered (Fig. 31). The results indicated that when the noise spectrum only contained components with frequencies less than 3.0 kHz, localization accuracy was the same whether or not the pinnae were occluded. Physical measurements of sound pressure at the eardrum of a manikin (Fig. 32) indicate an elevation dependent change in the transmission characteristic in the 0.7–3.5 kHz range in the presence of a torso.
Figure 32: Effect of torso on sound transmission to the eardrum of a manikin from sources in the vertical median plane (Gardner, 1973).
10.1 Localization Theory

Mills (1958) analyzed the responses he obtained in his experiments on the minimum audible angle in a straightforward way to obtain an estimate of the physical quantity $\Delta \theta$. Since he compared these estimates with quantities that could have been derived in a similar manner from data obtained in experiments in which headphone listening was used, no more sophisticated analysis was required.

![Diagram of a decision model for localization experiments.](image)

Figure 33: A decision model for localization experiments.

In localization studies, on the other hand, both Mills-like discrimination experiments and identification experiments with more than two possible source locations may be used. In order to relate results, it is necessary to employ some sort of a decision model, such as that shown in Fig. 33. With the aid of such a model, it is often possible to relate results reported in various ways (e.g., average error magnitude, rms error magnitude, etc.) to a common metric of accuracy (e.g., $\sigma$) and to derive a common metric, the standard deviation of localization, to describe accuracy in a wide variety of experiments, as shown in Table II.

Searle et al. (1976) assumed that there were a number of statistically independent cues potentially available to the listener.

1. Interaural phase/time dependence on $\theta$
2. Interaural spectral dependence on $\theta$
3. Monaural spectral dependence on $\theta$
4. Interaural pinna spectral dependence on $\phi$
5. Monaural pinna spectral dependence on $\phi$
6. Shoulder reflection dependence on $\phi$
Each cue was assumed to be described by an independent Gaussian random variable, whose mean was the actual coordinate of the source. Different cues were assumed to have potentially different variances. Some localization data (e.g., lines 1 and 12 in Table 2), as well as the findings of other identification experiments, suggested that this variance was not constant for a given cue, but rather dependent on the angular span $S$ of sources in the experiment:

$$
\sigma_i^2 = \sigma_{i0}^2 + (G_i S)^2
$$

Figure 34: Dependence of localization accuracy on span in four types of experiments.

Searle et al. (1976) were able to fit more than 40 experimental measurements of localization accuracy using this model (with 8 free parameters) by assuming that all the available cues were combined optimally by the listener (Fig. 34). The relative effectiveness of the cues in an experiment with a span of $\pi/2$ was estimated to be

- Interaural time & Interaural/Monaural spectra (unseparable) $5.6^\circ$
- Interaural pinna spectral - $18^\circ$.
- Monaural pinna spectral (per pinna) - $23^\circ$.
- Shoulder reflection (per shoulder) - $67^\circ$


