Sound Localization I: Physical Cues

Reading: Yost Ch. 12
Auditory space - surrounds an observer and exists wherever there is sound.

Researchers study how sounds are localized in space by using:
- azimuth coordinates (left to right)
- elevation coordinates (up/down)
- distance coordinates

On average, people localize sounds most accurately when they are in front of them, and least accurately when sounds are to the side or behind.
Sounds have no spatial dimensions: Localization depends on sound interacting with the head and external ears, generating spatial cues useful for localization:

- **Binaural** cues: interaural disparities in time and sound level.
- **Monaural** cues: mainly spectral.

Localization in the *horizontal plane* (azimuth) primarily involves two cues, both generated by distance between the ears on the head:

- **Interaural time differences** (ITD).
- **Interaural level differences** (ILD).

Localization in the *vertical plane* (elevation), as well as front-back discrimination, relies on spectral cues generated by head and pinna filtering.

Localization cues have a head-centered frame of reference, so cues vary with:

- Object location
- Object movement
- Head/body movements relative to sound object.
Binaural Cues for Horizontal Localization

If sound is lateral to midline, then the initial wave front of sound arrives at near ear before far ear. This creates between ear (interaural) differences in arrival time (ITD) and intensity or level (ILD); in particular, relative to the near ear, the sound arriving at the further ear is delayed in time and lower in amplitude.

Onset ITD: depends on path length difference (i.e. distance) between the two ears with respect to the azimuth of source. Onset ITD is independent of stimulus waveform.

ILD: generated by head shadow created by the head.
Interaural Phase Difference

The path length difference (i.e., distance) between the two ears with respect to the azimuth of source introduces an ongoing *interaural phase difference (IPD)* in the waveform (“fine structure”) of the sound.

For tones, IPD *varies with frequency*. Thus, sounds of different frequencies coming from same location have the *same ITD*, but *different IPD*.
IPDs and Phase Ambiguity

IPDs depend on stimulus frequency.

**Phase ambiguity**: A given IPD can occur at multiple ITDs, and therefore, potentially at multiple azimuths. For example:

- If ITD matches the period of the stimulus (0.6 ms), then...
- ...after one cycle, IPD is zero

So sound might be perceived as straight ahead, even though it is 90° lateral (in this case to the right).
ITD Cues

ITD varies with azimuth. The human range: 0 μs at midline (0°) to about 660 μs at ± 90° (“ecological”, or “physiological” range).

The maximum ITD correlates to the wavelength of a sound input with a frequency of 1500 Hz. For tone frequencies below this limit, there is an unambiguous interaural phase difference (IPD) between the sound waves entering the ears providing acoustic localization cues. For higher frequencies, the IPD is an ambiguous cue.
Many signals can be characterized by an envelope and the fine-structure waveform that falls under the envelope (e.g., sinusoidal amplitude modulated (SAM) tones, bandlimited noises). In fact, most waveforms can be described by the following formula:

\[ x(t) = e(t)f(t) \]

where \( e(t) \) is the envelope function, \( f(t) \) is the fine-structure waveform, and \( x(t) \) is the complex waveform.

The envelope provides timing cues for localizing broadband sounds.
The human head ~20 cm (0.2 m) in diameter. **Interaural level (intensity) differences** will be created by two mechanisms: spreading loss and diffraction.

Re: diffraction, recall that $\lambda = \frac{c}{f}$. Thus, significant sound shadow (and ILDs) are generated for frequencies greater than ~1720 Hz ($= \frac{344 \text{ m/s}}{0.2 \text{ m}}$).

**ILD:**
- Varies with azimuth.
- Grows with frequency
- Maximum human ILD is about **40 dB** at 60° and 120° azimuth at high frequencies.

Feddersen et al. 1959
Based on their reliability, the duplex theory states that interaural time differences (ITDs) are used to localize low frequency sounds, while interaural level differences (ILDs) are used to localize high frequency sounds.

ITDs: most reliable for **low frequencies** because phase ambiguity is a problem for high frequencies with periods shorter than max ITD produced by the head (~0.65 ms).

ILDs: most reliable for **high frequencies**, because head shadow generates perceivable ILDs only when wavelengths are less than head diameter.

Localization is less accurate around 1-3 kHz, i.e., at the transition between ITD and ILD cues.
The duplex theory states that ITDs are used to localize the azimuth low frequency sounds, whereas ILDs are used to localize the azimuth high frequency sounds.

However, a given ITD or ILD can arise from more than one location in space. In particular, the same ITD or ILD can arise from sounds in front of or behind a person. Thus, front/back confusions are common if sounds are narrowband (sinusoidal).
Interaural Cues: Cones of Confusion

More generally, for a given IPD or ILD, there will be a ~conical surface extending out of the ear that will produce identical IPDs and ILDs, making precise sound source localization over this surface difficult.

Localization errors can be reduced by:
- head movement (takes ~500ms, thus only helpful for long duration sounds)
- monaural cues
Pinna and external canal are \textit{frequency dependent directional filters}. In particular, the pinna generates \textit{spectral cues} for vertical localization in the form of high-frequency (6-8 kHz) \textit{notches} in “directional transfer functions”. The notches shift systematically to higher frequencies with increasing elevation.

Vertical localization is most accurate for \textit{broad-band} and \textit{high-frequency complex sounds}, which can convey the spatial cues in their spectra.
In addition to providing cues to elevation, monaural spectral cues also resolve localization ambiguities on cones of confusion. HRTFs on cones of confusion show a variety of salient features (peaks and notches), the frequency of which change systematically over a range of an octave or more as the source moves from front to back.
Spectral Cues

More generally, the first-notch frequency is a potential cue for both azimuth and elevation. In particular (panel C; cat), the locations in space having the same first-notch frequencies extend from high contralateral to low ipsilateral positions. Thus, sounds can be localized over a significant region of space based on a knowledge only of the first-notch frequencies in the two ears.
Distance

The human auditory system uses a variety of cues to determine the distance of a sound source. Three main cues are:

Intensity: distant sound sources have a lower loudness than close ones. This aspect can be evaluated especially for well-known sound sources.

Direct/Reflection ratio: sounds have direct and indirect components, more distant sound sources are likely to exhibit more reflections than near ones. Thus, the D/R ratio decreases with distance. May be computed as part of the precedence effect during the localization dominance period.

Head movement: nearby sound sources will appear to move faster than distant sound sources.

Fluitt et al. 2014
Eye movements re-map the location of auditory targets (only azimuth and elevation studied) on a global scale, wherein the center of the auditory field is moved towards the direction the eyes are looking.
Eye Position and Sound Localization 2

The re-mapping of the center of the auditory system is quick.

Adaptation
- begins immediately after eye movement
- grows exponentially over about 1-2 minutes
- ultimately reaches 40-50% of eye eccentricity, depending on subject.

Razavi et al. (2007)
Summary

Sound localization refers to a listener's ability to identify the location or origin of a detected sound in direction (azimuth and elevation) and distance.

Different physical cues contribute to our ability to localize sounds:

• azimuth: IPD (freq < 1500 Hz), ILD (freq > 2000), spectral cues
• elevation: spectral cues
• distance: intensity, D/R, motion parallax