Sound localization

Maria Pikhtina
Aalto Universtiy
Master's Programme CCIS / AAT

maria.pikhtina@aalto.fi

Abstract

This paper explores the fundamental auditory mechanisms utilized in sound localization, focusing on interaural time differences (ITD) and interaural level differences (ILD), along with monaural and dynamic cues. It discusses how these cues are integrated to enable precise localization of sounds in three-dimensional space, addressing their effectiveness across different frequencies and orientations. The paper also examines the impact of individual anatomy on these localization cues, particularly how variations in head-related transfer functions (HRTFs) influence the accuracy of sound localization.

1 Introduction

Sound localization is an essential auditory process that enables individuals to determine the origin of sounds in their environment. This capability is critical for navigating spaces and interacting within complex auditory landscapes. The human auditory system employs a sophisticated set of cues to decode the spatial attributes of sound, including interaural time differences (ITD), interaural level differences (ILD), monaural cues from the outer ears (pinnae), and dynamic cues resulting from head movements. Each of these cues plays a pivotal role in the perception of sound direction and has been the subject of extensive research to understand their mechanisms and implications for spatial hearing. This paper delves into each of these auditory cues, exploring their contributions to sound localization under various conditions and their integration by the auditory system.

2 Basics of Sound Localization

2.1 Interaural Time Differences

Interaural Time Differences (ITD) refer to the differences in time for a sound to reach each ear, crucial for the localization of sounds in the horizontal plane. This difference arises because sound waves take slightly different times to reach each ear, depending on the direction from which the sound originates. You can see the illustration of the effect from the Figure 1. The brain uses these differences in timing to determine the direction of the sound source. ITDs are particularly effective for low-frequency sounds where the head does not significantly affect the sound wave, allowing it to bend around the head. At these lower frequencies, the wavelengths are longer relative to the size of the human head, enabling the sound waves to bend without significant loss of energy or phase change. The research by Shinn-Cunningham et al. also discusses the influence of head and ear geometry on ITDs, noting that even slight differences in ear placement can impact the ITD cues, necessitating compensation by the auditory system to maintain sound localization accuracy. [3][12] [9][11]

2.2 Interaural Level Differences

Interaural Level Differences (ILD) arise from sound level differences at the two ears, created by the head casting an acoustic shadow. You can see the illustarion in the Figure 1. This effect is more pronounced for high-frequency sounds, where the wavelengths are shorter than the diameter of the head. As a result, the head blocks more sound energy from reaching the ear farther from the sound source, creating a noticeable level difference between the two ears. ILDs are used predominantly for localizing high-frequency sounds where the wavelengths are short enough relative to the size of the head to create significant sound level differences at each ear. According to Shinn-Cunningham et al., the study of ILDs reveals that these differences increase with frequency and are more complex in directional dependence, allowing for precise localization in three-dimensional space. The study also highlights the impact of the head's geometry on ILDs, demonstrating how these physical traits alter sound localization cues. [3][1] [9][11]

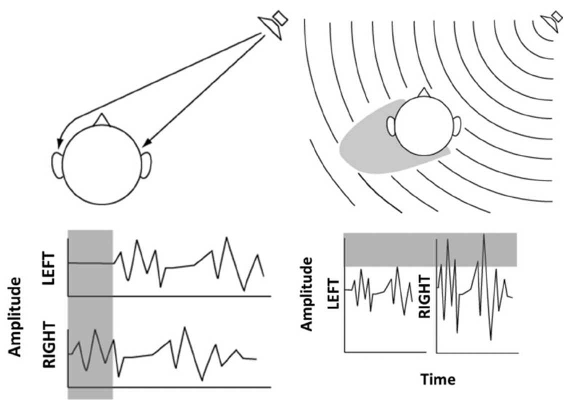

Figure 1: ITD and ILD

3 Horizontal and Vertical Sound Localization

3.1 Horizontal Localization

Horizontal Localization relies primarily on both ITD and ILD. The duplex theory, proposed by Lord Rayleigh, posits that ITDs are utilized for localizing low-frequency sounds, while ILDs are predominant for high-frequency sounds due to the acoustic shadow cast by the head. This shadowing means that the sound level at the ear nearer to a sound source would be greater than that at the farther ear, providing a cue to the sound source location. In contrast, for sounds of lower frequency, where the sound wavelength can be several times longer than the diameter of the human head, the sound wave "bends" around the head, and the difference in sound level at the two ears is negligible, making ITDs more reliable. The study by Makous and Middlebrooks, 1990, confirms that horizontal localization is generally more accurate and consistent across different frontal midline positions, with errors increasing as the sound source moves to more peripheral positions. [3] [2]

3.2 Vertical and Front/Back Localization

Vertical and Front/Back Localization utilize different cues compared to horizontal localization. In these dimensions, localization involves the spectral shape affected by the pinnae—the outer part of the ears. The pinnae interact with incoming sound waves, altering their spectral content based on the elevation and whether the sound source is in front or behind the listener. These changes, detected by the auditory system, allow the listener to identify the vertical position and front/back location of the sound source. The study by Makous and Middlebrooks, 1990, noted that while vertical localization showed increased errors with elevation, the performance was still better in the frontal half of space compared to the rear, with significant biases and larger errors observed for sounds originating from behind the listener. [3] [2]

4 Monaural and Dynamic Cues in Sound Localization

4.1 Monaural Cues in Sound Localization

Monaural cues, particularly important for vertical localization, involve changes in the spectral shape of sounds caused by interactions with the listener's anatomy (pinna, head, shoulders, and bust). These cues are characterized by the head-related transfer functions (HRTFs). HRTFs describe how sounds from different directions are filtered by the body's anatomy before reaching the ear canal. This filtering effect results in modifications of the sound spectrum, such as reinforcement (peaks) or attenuation (notches) at specific frequencies, which the auditory system uses to determine the sound source's elevation and distance. [10]

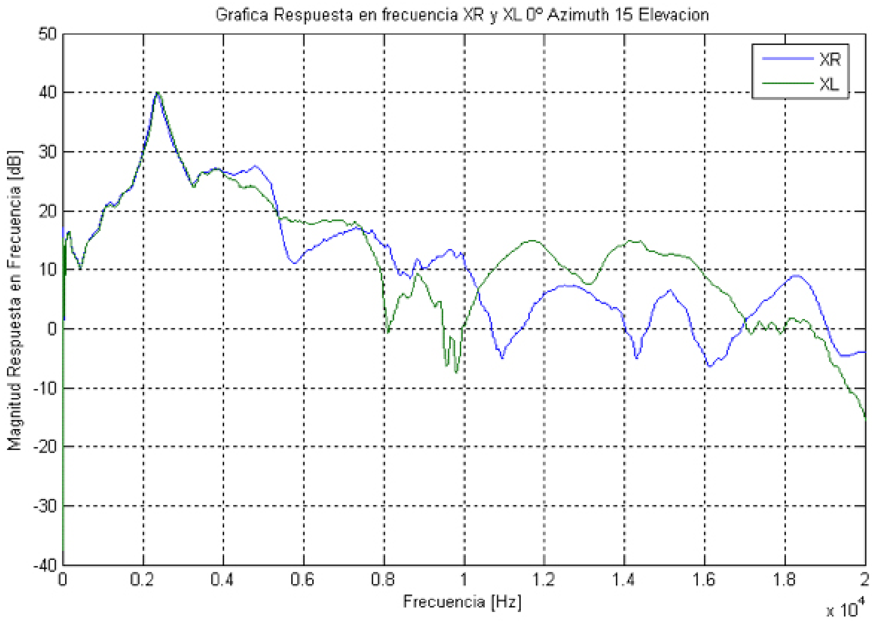

Figure 2: Sample of frequency response of ears: green curve: left ear XL(f), blue curve: right ear XR(f) for a sound source from upward front.HRTF H(f) is the Fourier transform of the HRIR h(t).

4.2 Dynamic Cues in Sound Localization

Dynamic cues refer to changes in perceived sound caused by either the movement of the listener or the sound source. These cues are crucial for resolving the "cone of confusion,", see Figure 2. a scenario where sounds from different locations may have similar ITD and ILD values, making them indistinguishable based solely on these binaural cues. Movement, particularly of the head or ears, alters the acoustic input in a way that can disambiguate these confusing signals. [9] [10]

Head movements adjust the phase and amplitude of the sounds reaching each ear, thereby providing additional information that can refine sound localization. For example, tilting or turning the head changes the alignment of the ears relative to the sound source, modifying the sound's path to each ear and helping to pinpoint the source's location more accurately. These dynamic cues enhance the auditory system's ability to detect changes in sound position, especially in complex acoustic environments. [9]

4.3 Integration of Monaural and Dynamic Cues

The auditory system combines these monaural and dynamic cues to create a detailed spatial map of the environment. This integration allows for the precise localization of sounds in three-dimensional space, not only horizontally but also vertically and in depth. Monaural cues, primarily through changes in spectral content, provide detailed information about sound source elevation, while dynamic cues help resolve ambiguities in the horizontal plane. [9]

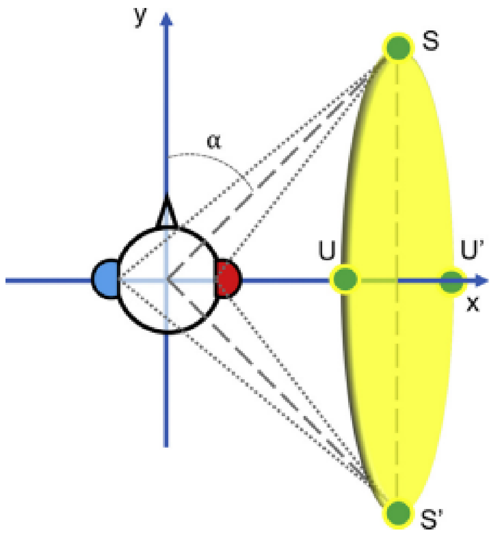

Figure 3: Depiction of the cone of confusion: When the subject keeps their head stationary, sources S and S' in the azimuthal plane demonstrate identical interaural time and level differences. Similarly, sources U and U' in the vertical plane exhibit the same characteristics. This ambiguity between front/back and high/low directions is applicable across the entire surface of the cone of confusion.

The ability to utilize these cues effectively is a result of both innate mechanisms and learned experiences. The brain adjusts to the specific filtering effects of an individual's anatomy over time, enhancing the ability to localize sounds based on subtle differences in spectral cues. Similarly, experience with moving in an environment helps refine the use of dynamic cues, making sound localization more accurate and robust.[9]

5 Detailed Exploration of HRTF-Based Sound Localization Accuracy

The study conducted by Masayuki Morimoto and Yoichi Ando at Kobe University meticulously explores the impact of individual head-related transfer functions (HRTFs) on sound localization accuracy using a sophisticated simulation setup. Their experiment focused on three male subjects, distinguished by their differing ear sizes, to investigate how these variations influence the effectiveness of sound localization in both the horizontal and median planes. This was achieved through a digital simulation of sound fields in an anechoic chamber, utilizing the measured HRTFs for both ears of the subjects.

The methodology involved using white noise processed through a two-channel loudspeaker system. Sound localization tests were carried out in separated horizontal and median planes, comparing the subjects' ability to localize sound using their own

HRTFs against using the HRTFs of the other subjects. Results showed that subjects localized sounds with high accuracy when their own HRTFs were used, closely mirroring the accuracy with real sound sources. However, significant deviations occurred when HRTFs not corresponding to their own anatomical features were used.

The statistical analysis highlighted a significant effect of the HRTF factor on localization errors, particularly in the median plane, where the individual differences in HRTF angle-dependency were prominent. Subjects with HRTFs showing high angle-dependency had better localization accuracy than those with lower angle-dependency in this plane.

This research underscores the crucial role of individual auditory anatomies in the perception of sound direction and emphasizes the need for personalized audio systems in virtual environments. The detailed statistical and visual representation of the results, including error margins and variance analysis, provided a comprehensive understanding of the implications of HRTF variations on sound localization accuracy.[5]

6 Minimum Audible Angle in Sound Localization

The minimum audible angle (MAA) is a measure used in psychoacoustics to determine the smallest angular separation between two sound sources from which an individual can identify two distinct sounds. This ability varies based on several factors, including sound frequency, the velocity of the sound source, environmental conditions, and the listener's dynamic relation to the sound.[7]

The MAA, studied by Mills in 1958, defines the smallest angular separation at which two sound sources can be perceived as distinct. Mills' work demonstrated that this ability varies with the frequency of the sound and its azimuthal position relative to the listener. Lower frequencies generally allowed for more accurate localization, which is attributed to better utilization of binaural cues like interaural time and level differences. [4] See illustration of the MAA in Figure 4.

6.1 Dynamic Sound Localization and Minimum Audible Movement Angle (MAMA)

The concept of Minimum Audible Movement Angle (MAMA) extends from MAA to scenarios where the sound source is in motion. Studies by Perrott and Musicant have shown that MAMA is affected by the velocity of the sound source. Their experiments indicated that the faster the movement, the larger the MAMA, suggesting reduced spatial resolution due to increased source velocity. This is because higher velocities provide less time for the auditory system to integrate sound cues effectively. For example, MAMA could range from as low as 8.3° at slow speeds to 21.2° at high speeds. See Figure 5 from the article. [6]

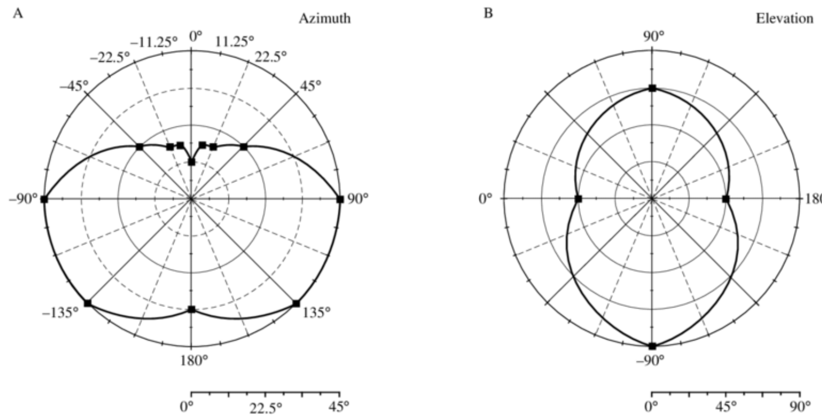

Figure 4: Minimum audible angles by RobertA Wyttenbach

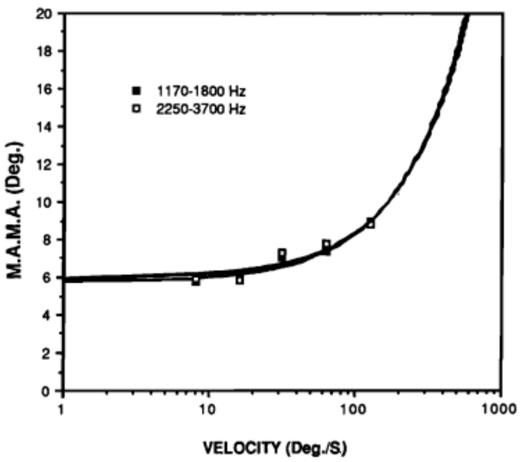

Figure 5: MAMA thresholds (in degrees) as a function of the velocity of the source

6.2 Influence of Frequency and Velocity on Dynamic Localization

The interaction between the frequency of the sound and the velocity of its source critically affects dynamic localization. Perrott and Tucker observed that spatial resolution is optimal for sounds below 1000 Hz. Above this threshold, particularly between 1300-2000 Hz, resolution deteriorates significantly. These findings suggest that the auditory system's mechanisms for processing static and dynamic sounds are fundamentally similar but are stressed differently depending on the sound's motion dynamics.[8]

The research further highlights that while minimal changes in spatial resolution are evident at velocities below approximately 32°/s, the resolution decreases significantly at higher velocities. This degradation at high velocities could be due to the auditory system's limited time to process the necessary spatial cues before the sound source has moved significantly. See Figure 6 from article.[8]

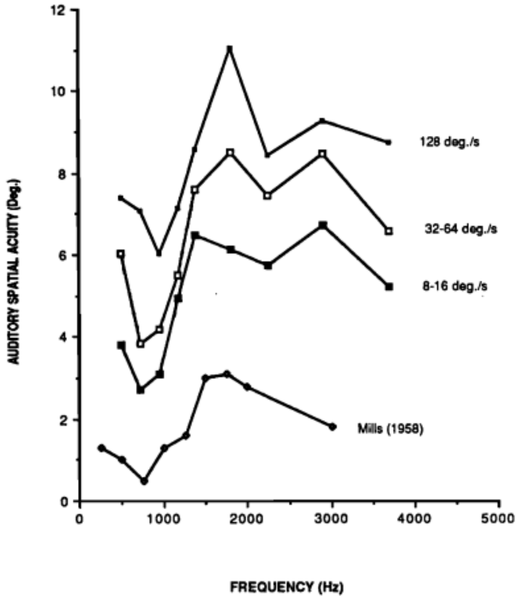

Figure 6: MAMA thresholds (in degrees) for three velocity ranges (8°- 16°, 32°-64°, and 128°/s) as a function of the frequency of the signal localized

6.3 Experimental Insights into MAMA

Perrott and Tucker's detailed experiments used a single-interval, forced-choice paradigm to explore how both the rate of displacement of a sound source and its frequency affect dynamic spatial resolution. Their findings replicated earlier observations about the inverse relationship between spatial resolution and the rate of travel, with a significant drop in resolution as velocity increased. These experiments also confirmed the significant impact of frequency on dynamic localization, aligning with static localization principles where frequency plays a crucial role in spatial cue processing.

The study of minimum audible angles, both MAA and MAMA, provides critical insights into human auditory spatial awareness under various conditions. The transition from research on static sound sources to dynamic ones highlights the adaptability and limitations of the auditory system, emphasizing the need for ongoing research to further decipher complex auditory processing mechanisms in dynamic settings such as developing auditory systems in vehicles, improving hearing aids, and designing more effective public address systems.

7 Conclusions

This seminar paper elucidates the sophisticated auditory mechanisms critical for sound localization, integrating insights on interaural time differences (ITD), interaural level differences (ILD), monaural cues, dynamic cues, and their cumulative effects as observed in MAA and MAMA measurements. ITDs and ILDs, fundamental in horizontal localization, vary in effectiveness based on sound frequency—ITDs are more efficient at low frequencies as sound waves circumvent the head, while ILDs become predominant at higher frequencies due to the head casting an acoustic shadow. MAA measurements help illustrate these phenomena by quantifying the minimum angular separation at which sound sources can be distinctly localized, emphasizing the influence of frequency on spatial resolution.

For vertical and front/back localization, the role of the pinnae is crucial as it modifies spectral cues depending on the direction of sound waves, a fact underscored by the study of MAMA, which extends the principles of MAA to dynamic scenarios where the sound source or listener is in motion. MAMA findings highlight how the auditory system integrates dynamic changes in the acoustic environment, enhancing spatial awareness and localization precision under movement.

The research extensively covers the importance of monaural cues derived from head-related transfer functions (HRTFs) and dynamic cues from head movements, which significantly refine the accuracy of sound localization in three-dimensional spaces. These cues enable individuals to navigate and interact within their environments more effectively, showcasing the complexity and precision of the human auditory system. This understanding not only highlights the natural capabilities of auditory perception but also underscores the potential for advancements in audio technologies, such as virtual reality systems and sophisticated hearing aids, that could benefit from mimicking these natural processes. The integration of auditory cues not only sheds light on the inherent sophistication of the human auditory system but also presents significant opportunities for technological innovations that aim to enhance or replicate these natural auditory processes, especially through the application of MAA and MAMA in designing more effective auditory interfaces.

8 References

The references will be numbered in order of appearance [3] [1] [12] [9] [10].

References

[1] Ashida, G., and Carr, C. E. Sound localization: Jeffress and beyond. *Current Opinion in Neurobiology 21*, 5 (2011), 745–751.

[2] Makous, J. C., and Middlebrooks, J. C. Two-dimensional sound localization by human listeners. *Journal of the Acoustical Society of America 87*, 5 (1990).

[3] Middlebrooks, J. C. *Handbook of Clinical Neurology: Sound Localization*, vol. 129 of *Neuropsychology and Neurology*. Elsevier, 2015.

[4] Mills, A. W. On the minimum audible angle. *Journal of the Acoustical Society of America 30*, 4 (1958).

[5] Morimoto, M., and Ando, Y. On the simulation of sound localization. *J. Acoust. Soc. Jpn. (E) 1*, 3 (1980), not specified in the text. Faculty of Engineering, Kobe University, Nada, Kobe, 657 Japan.

[6] Perrott, D. R., and Musicant, A. Minimum auditory movement angle: Binaural localization of moving sound sources. *Journal of the Acoustical Society of America 62*, 6 (1977), 1463–1466.

[7] Perrott, D. R., and Saberi, K. Minimum audible angle thresholds for sources varying in both elevation and azimuth. *Journal of the Acoustical Society of America 87*, 4 (1990).

[8] Perrott, D. R., and Tucker, J. Minimum audible movement angle as a function of signal frequency and the velocity of the source. *Journal of the Acoustical Society of America* (1988).

[9] Raspaud, M., Viste, H., and Evangelista, G. Binaural source localization by joint estimation of ild and itd. *IEEE Transactions on Audio, Speech, and Language Processing 18*, 1 (2010), 68–77.

[10] Risoud, M., Hanson, J.-N., Gauvrit, F., Renard, C., Lemesre, P.-E., Bonne, F.-X., and Vincent, C. Sound source localization. *European Annals of Otorhinolaryngology, Head and Neck Diseases 135* (2018), 259–264.

[11] Shinn-Cunningham, B. G., Santarelli, S., and Kopco, N. Tori of confusion: Binaural localization cues for sources within reach of a listener. *Journal of the Acoustical Society of America 107*, 3 (2000).

[12] Wightman, F. L., and Kistler, D. J. The dominant role of low-frequency interaural time differences in sound localization. *JASA 91*, 3 (1992), 1648–1661.