Internal use

Audio Engineering Society
Convention Express Paper 194

Presented at the 156th Convention
2024 June 15-17, Madrid, Spain

This Express Paper was selected on the basis of a submitted synopsis that has been peer-reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This Express Paper has been reproduced from the author's advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. This paper is available in the AES E-Library (http://www.aes.org/e-lib) all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Measurement and analysis of spatial localization characteristics based on interaural time difference

Qiuxian Huang[1], Heng Wang[2], Xinyi Ma[3], Yuantao Han[4], and Qiuyang Liu[5]

[1] *Wuhan Polytechnic University, Wuhan, Hubei Province, China*

[2] *Wuhan Polytechnic University, Wuhan, Hubei Province, China*

[3] *Wuhan Polytechnic University, Wuhan, Hubei Province, China*

Correspondence should be addressed to Heng Wang (wh825554@163.com)

This work was supported in part by the Hubei Provincial Department of Education Research Foundation under Grant D20201601 and in part by the Hubei Key Laboratory of Intelligent Robots, Wuhan Institute of Technology, under Grant HBIR202101.

ABSTRACT

Given the significant impact of interaural time difference on horizontal localization, and based on existing research findings on binaural cue perception mechanisms, this paper studies the relationship between interaural time difference, azimuth, and frequency. It further explores the orientation perception characteristics of interaural time difference, providing foundational data and laying a theoretical basis for efficient compression in parametric encoding. Seven discrete interaural time difference values were non-uniformly selected based on the sensitivity of human ears to different horizontal directions for auditory perception experiments. Experimental results indicate that within the low-frequency range of 350-1370 Hz and the interaural time difference range of 0 μs to 500 μs, as interaural time difference increases, its perceptual effect is increasingly influenced by frequency. Specifically, the interaural time difference is more sensitive to sounds with frequencies below 1000 Hz compared to those above 1000 Hz. Additionally, when the interaural time difference is less than approximately 380 μs, the ability to accurately perceive interaural time difference stabilizes within the frequency range of approximately 1000 Hz to 1370 Hz. So the optimal frequency range for interaural time difference perception of azimuth appears to be between approximately 1000 Hz and 1500 Hz.

1 Introduction

In three-dimensional space, Interaural Time Difference (ITD) and Interaural Level Difference (ILD) primarily serve as cues for localizing sound sources on the horizontal plane [12-13]. They arise from the differences when sound is emitted from a source to the left and right ears of a listener, collectively termed binaural cues. ITD specifically manifests as differences in the time it takes for the sound to reach the left and right ears, and it is not only related to the azimuth of the sound source but also to its frequency. According to acoustic propagation theory, wavelength is inversely proportional to frequency. When the sound source emits a low-frequency signal, the wavelength is larger, sometimes exceeding the size of the human head. In such cases, the sound can bypass the head, resulting in minimal obstruction of the wavelength by the head, leading to nearly identical ILD but varying ITD [11]. Consequently, ITD plays a predominant role in localization at lower frequencies,

Internal use

and existing research suggests that ITD provides better localization cues at frequencies below 1500 Hz [5]. Due to the varying perceptual characteristics of the human ear for sound sources in different horizontal directions, non-uniform sound source points were selected to further investigate the directional perception characteristics of ITD. This guides the reduction of errors in parameter quantization in coding and alleviates azimuth distortion phenomena.

To investigate the localization effects of binaural cues in three-dimensional space, some scholars have researched the Just Noticeable Difference (JND) of ITD. In 1998, Mossop found that the JND of ITD increased slowly at first with increasing ITD values [2], then increased dramatically, and later showed no consistent pattern. In 2016, Hartmann et al. tested the duplex theory of low-frequency pure tones [9], demonstrating the crucial role of ITD itself in localization and lateralization, while ILD and frequency also influenced the localization effects of ITD. To obtain more precise JNDs of ITD, Thavam improved the experiment in 2019, conducting experiments on trained and untrained normal hearing listeners in the low-frequency range [7], ultimately showing that the auditory perception of trained listeners was better than untrained listeners. In the same year, Best measured the sensitivity of normal hearing and hearing-impaired individuals to ITD [8] using pure tones and single-word speech in quiet and noisy environments, revealing a smaller impact of hearing loss on single-word speech compared to pure tones. In 2020, Xie et al. investigated the dynamic changes of ITD and its experiment demonstrated the sound propagation theory of ITD [6], also indicating that ITD is an essential factor in the lateral localization of sound sources. To improve the frequency dependence of binaural cues, Rosen continued to measure the JND of ILD in 2022 [14], concluding that ILD plays an important role in high-frequency localization for tones. These experiments suggest that while ITD and ILD both contribute to sound localization, their roles may vary depending on the spatial characteristics of the sound.

Furthermore, extensive research has been conducted domestically and internationally on the influence of frequency on binaural cues, particularly ITD. In 2008, Chen investigated the relationship between binaural cues, Interaural Correlation (IC), and frequency using sinusoidal tone signals ranging from 20 to 15500 Hz [3]. The results indicated that binaural cues exhibited different degrees of frequency dependency, with ITD showing a stronger dependency on frequency compared to ILD and IC signals. To study the perceptual characteristics of ILD, Zhang conducted experiments in 2010 [1] and found that ILD varied insignificantly in the frequency range below 2 kHz but significantly varied in the frequency range above 2 kHz. In 2011, Dunai et al. tested and analyzed the relationship between ITD's JND and frequency [4]. It was found that in the low-frequency range, the perception effect of ITD is unstable, while in the high-frequency range, the JND sharply increased with the rise in frequency, with the lowest perceptual threshold of ITD occurring around 1 kHz. It was also analyzed that the JND of ITD exhibits a strong dependency on frequency. Building upon this foundation, Hu et al. further investigated the relationship between ITD's JND and frequency using narrowband Gaussian white noise ranging from 20 to 1720 Hz in 2014 [5]. The results demonstrated significant variations in ITD's JND with frequency, with a minimum value observed around 500 Hz. In the same year, Smith studied the sensitivity of humans to localize interaural time differences using low-frequency pure-tone sounds [16]. In 2015, Diedesch found that ITD is more sensitive than ILD in localizing sound directions at lower frequencies [10]. Subsequently, to further investigate the frequency dependence of ITD, Luo tested whether ITD has a low-frequency boundary in 2023 [15]. The experimental results revealed the existence of a low-frequency boundary for ITD, with a value of approximately 62.5 Hz.

These methods fully consider the influencing factors of ITD and its relationship with sound source localization, revealing the perceptual characteristics of binaural cues and supplementing the research work on spatial perception characteristics of the human ear. In recent years, both domestically and internationally, research on the perceptual characteristics of ITD has mainly focused on its localization stability and lateralization effects, providing data support for improving the head-related transfer function (HRTF) model. However, the experimental data and results regarding the perceptual characteristics of ITD are not detailed and comprehensive enough, which hinders the progress of improving HRTF models and the efficient encoding of multi-channel audio signals. Therefore, this experiment further explores the spatial localization characteristics of ITD, investigates the relationship between ITD and frequency and azimuth, and establishes a three-dimensional perceptual characteristic surface of ITD concerning perceived azimuth and frequency. This further complements the work on spatial perception

Internal use

characterization of the human ear, reduces spatial parameter redundancy in encoding, and improves encoding efficiency.

2 Experimental Methods

2.1 Participants

This experiment involved 7 participants, including 4 males and 3 females, aged between 19 and 26 years old, all of whom are students. The participants were able to perceive changes in azimuth fairly accurately and demonstrated sufficient patience to complete the experiment.

2.2 Experimental Equipment

To ensure the accuracy of the experimental data, professional audio equipment is selected for the listening experiment. As shown in Figure 1, sounds located in the horizontal spatial positions are captured with an artificial head in an anechoic chamber environment, with the position of the sound source precisely adjusted using a pulley system.

Participants wore headphones with an external sound card to reduce interference from the surrounding environment and to facilitate direct control of audio signals in both channels via a computer during the hearing test, thus expediting the experiment. To minimize listening errors, participants conducted individual listening tasks using desktop computers. They interacted with the listening software based on the perceived audio azimuth, completing the listening tasks.

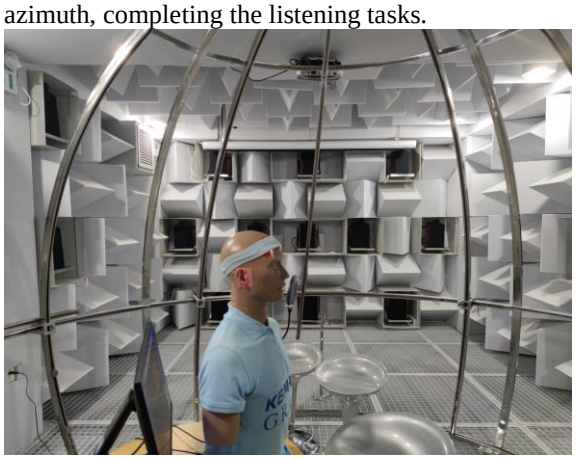


Figure 1. The artificial head collection environment.

The specific physical configuration is as follows:
Sound card: Innovative sound card, model no. SB1240;
Sound pressure meter: Sigma AR824;
Headphones: Sennheiser, model HD380 Pro.

2.3 Test Signal

2.3.1 Signal Type

Each set of test signals consisted of pure sine tones with a single frequency. These tones were sampled at 44.1 kHz with 16-bit precision and had a duration of 300 milliseconds.

2.3.2 Banding

To further investigate the relationship between ITD and azimuth as well as frequency, the paper measures the azimuth perceived by the human ear through ITD within the frequency range of 300-1480 Hz. Based on the definition of critical bands and the perceptual characteristics of the human ear, the ear exhibits a similar perceptual effect on sounds within a certain width of the frequency band. Therefore, the entire frequency range can be divided into critical bands with different bandwidths. Each sub-band can be represented by its center frequency to characterize the frequency range of that band. Hence, the experiment utilized the Bark band division method for the entire frequency range, selecting eight center frequencies for the bands: 350 Hz, 450 Hz, 570 Hz, 700 Hz, 840 Hz, 1000 Hz, 1170 Hz, and 1370 Hz.

2.3.3 Selection of Reference and Test sounds

The research employs a method of a fixed ITD value with varying azimuth at different audio frequencies. Specifically, this involves extracting ITD as the reference sound and using audio samples from a database as test sounds. The reference sound and test sounds constitute a set of test signals, with the test sounds being varied to examine the angle of the ITD reference sound.

The reference sound refers to the audio with the ITD value to be measured. The audio at the 0-degree frontal position is processed with ITD and saved to serve as the reference sound. Since the human ear is most sensitive to sound sources directly in front, as the angle corresponding to the ITD value increases, it becomes less sensitive as it approaches the left and right ears [5]. Therefore, in this experiment, seven discrete ITD reference values were non-uniformly selected. As the value increases, the interval also increases. The selected ITD values for the reference sound are 0 μs, 50 μs, 110 μs, 190 μs, 280 μs, 380 μs, and 500 μs. Test sounds are taken from a self-built test sound database, primarily used to test the perceived azimuth angle for different reference sounds.

2.3.4 Generation Method for a set of test signals

Internal use

Prepare eight segments of 300 ms pure tones at different frequencies as sound sources, with input and output sound pressure set to 75 dB. These sound sources will be played through speakers positioned on the horizontal plane of a spherical frame in an anechoic chamber. Using pulleys to fix the sound sources from 0 degrees to 90 degrees, with intervals set at 1 degree. Record the audio using an artificial head within the spherical frame. Save these recordings as a total of 90 WAV files, which will serve as the test sound database.

A set of test signals undergoes low-pass filtering to remove background noise interference, resulting in low-frequency sequences across eight frequency bands. Each low-frequency sequence has its right channel muted, and the sound pressure of the output audio signal is measured at the left channel output terminal. When the sound pressure meter indicates the preset 75 dB sound pressure level, the left channel signal is saved separately as a mono signal. This left channel signal is then duplicated onto the right channel signal to create a stereo signal.

Each set of test signals consists of two 300-millisecond audio segments: a reference tone and a test tone, both in stereo. The total duration is 1.1 seconds, with a 500-millisecond pause between the playback of the reference and test tones. The reference tone is based on a binaural perceptual audio signal located directly in front of a person (horizontal azimuth of 0 degrees), which is transformed into a stereo signal with different ITDs by altering the playback delay time of the left and right channel audio.

2.4 Methods of listening

2.4.1 Screening Methods for Subjects

To ensure that subjects perceive direction more accurately and complete the auditory perception experiment smoothly, Hu's auditory perception experiment method [5] was adopted. Training the participants' auditory perception abilities and selecting participants whose perceived azimuth corresponding to the reference tone falls within the acceptable error range, we ultimately selected 7 out of 12 participants.

During the training, a set of test signals, including a reference tone at 0 μs and a test tone initially set at 45 °, is generated and randomly played. The participants repeatedly choose the segment that seems more leftward from the two audio segments they hear, and the program records the accuracy of their choices. If the accuracy rate of selecting the correct segment exceeds 80 % within the specified number of trials, it indicates that the participants can perceive changes in the direction relatively accurately, with the participants passing the training. However, if the actual number of trials exceeds the specified number, it suggests training failure, and it is recommended to restart the training. After this round of screening, a total of seven participants successfully passed.

2.4.2 Auditory Methodlolgy Design

To ensure the authenticity of the experimental results, the two-alternative forced-choice method (2AFC) is employed to assess the perceived horizontal angle under different frequencies and ITD. In each trial, the system generates a set of test signals that are randomly played to the participant through headphones, consisting of a test tone and a reference tone, upon which the participant conducts hearing tests and makes subjective judgments. After listening to the two audio segments, the participant has one second to perceive and judge which sound source is relatively more to the left and selects the corresponding radio button on the interface. The "one up, two down" test method is also employed, involving multiple rounds of listening where the results of one round affect the stimulus signal of the next. The system fixes the ITD for the reference tone, while the test tone, with a specific directional value, is adjusted towards the reference tone based on the participant's judgments, following a step-size rule. Each cycle of the test tone decreasing and then increasing, or vice versa, is referred to as one reversal.

This experiment implemented a step-size adjustment rule to gradually approach a reference tone with the test tone. The step size of the test tone decreased progressively as it approached the reference tone to enhance auditory accuracy. During listening, to efficiently and accurately approach the reference tone, the step size of the test tone was linearly reduced (set to 4 degrees, 2 degrees, 1 degree in this experiment) when a certain number of reversals were reached (the reversal thresholds were successively set to 8 and 4 times in this experiment). The reduction in step size continued until the number of reversals reached 12, signaling the end of that trial. Subsequently, the average angle of the test tone during the final 4 reversals was calculated, representing the perceived directional value of the target ITD for that particular trial.

Based on the experimental scheme described above, a binaural time difference-based spatial perception auditory system software was designed using the Microsoft Foundation Class Library (MFC). It

Internal use

includes training modules, auditory condition preset modules, audio test signal generation modules, and auditory result-saving modules.

3 Experimental results and analysis

Seven participants completed the measurement of ITD spatial perception. Significant differences were observed in the ITD measurement results at individual frequency points. However, considering the participants' levels of listening training and auditory pattern characteristics, we selected the auditory data from participants with higher proficiency in listening and those who exhibited characteristics consistent with auditory patterns. Overall, the listening results for each participant were similar, with most of the listening data showing no significant differences. We took the average of their results as the final listening data.

3.1 Frequency and Perceived Azimuth Relationship

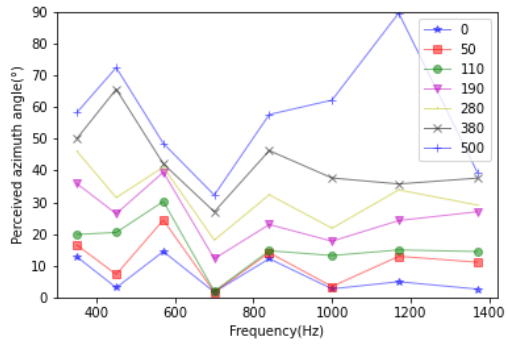


Figure 2. Plot of perceived azimuth versus frequency.

Based on the measured data, this paper plotted line graphs of the perceived azimuth of ITD at seven different levels against frequency to visually observe their relationship, as shown in Figure 2 above.

Overall, within the frequency range of 350 Hz to 1370 Hz, the curves depicting the perceived azimuth angle of ITD exhibit significant variations within certain ranges of perceived azimuth angle. As the frequency exceeds approximately 1000 Hz, changes in perceived azimuth angle with ITD values below 380 μs tend to occur more gradually with increasing frequency. These patterns can be theoretically analyzed and validated. Since ITD primarily influences the localization of sound sources below 1500 Hz, within the ITD range of 0 μs to approximately 380 μs, the perception of azimuth by the human ear appears to be relatively stable within the frequency range above approximately 1000 Hz and below 1370 Hz. Consequently, the optimal frequency range for perceiving spatial azimuth through ITD falls approximately between 1000 Hz and 1500 Hz.

From a local perspective, within the range of 350 Hz to 1370 Hz, different ITD curves exhibit a minimal perceived azimuth angle at around 700 Hz. Bilokon previously discussed the application of this phenomenon in his article [17]. It is concluded that ITD corresponds to the minimum azimuth angle at approximately 700 Hz within the low-frequency range of 350 Hz to 1370 Hz.

3.2 ITD and Perceived Azimuth Relationship

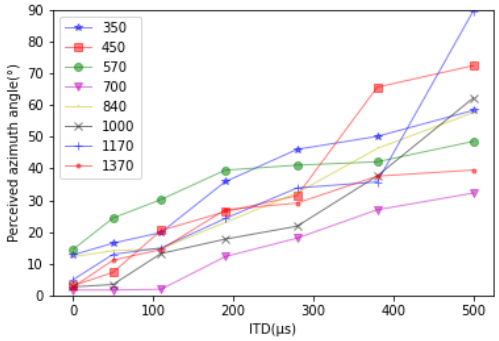


Figure 3. Plot of perceived azimuth versus ITD.

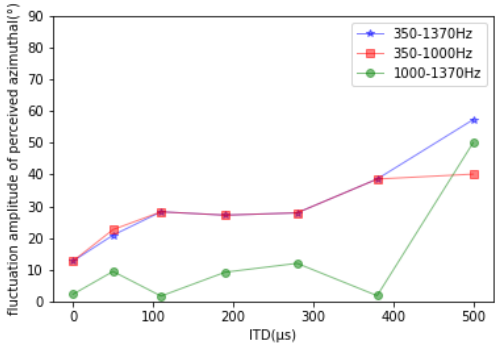


Figure 4. Curve of the amplitude variation of perceived azimuth with changing ITD values across different frequency ranges.

To analyze the relationship between ITD and perceived azimuth, this paper plots curves for eight frequency bands, representing the relationship between perceived azimuth and ITD variations, as shown in Figure 3. To visually observe the fluctuation amplitudes of perceived azimuth for measured ITD values in different frequency ranges, as well as their patterns, Figure 4 is constructed in this paper.

Internal use

In Figure 3, within the range of ITD values from 0 μs to 500 μs, the perceived azimuth angle of curves at different frequencies gradually increases from their initial values as the ITD value increases, demonstrating a consistent trend. Upon observing Figure 4, it is evident that the fluctuation amplitudes of the perceived azimuth angle for curves representing the frequency range of 350-1370 Hz increase with the increase in ITD values. This pattern can be theoretically analyzed and verified. As the ITD value increases, the corresponding azimuth angle of the sound source gradually increases from the respective initial angle. According to the ITD values of 0 μs, 350 μs, and 650 μs correspond to the sound source being at the median plane of the person, 45 °to the left, and at the left ear, respectively. Within the range of ITD values from 0 μs to 500 μs, the sound source approaches closer to the left ear as the ITD increases. When the frequency varies within the range of 350 Hz to 1370 Hz, the variation of perceived azimuth angle based on ITD becomes more unstable with increasing ITD values, indicating a decreased sensitivity in perceiving the azimuth based on ITD. Additionally, in Figure 3, the curve for 1170 Hz rises significantly after an ITD value of approximately 380 μs, suggesting that stabilized localization performance for frequencies above 1000 Hz can be observed only for ITDs below approximately 380 microseconds.

3.3 ITD versus Perceived Azimuth and Frequency

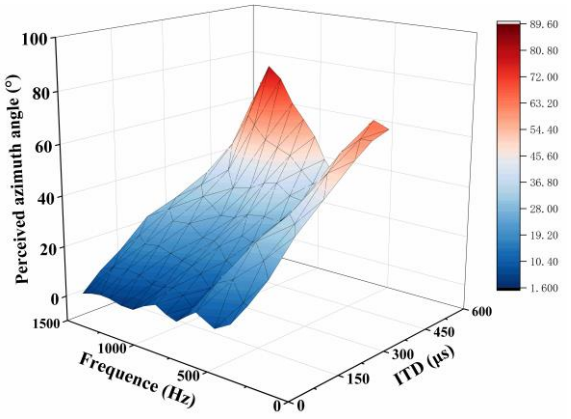


Figure 5. Fitted surface for ITD.

To investigate the relationship between ITD, azimuth, and frequency, given the limited amount of measured data, a simple and efficient method was chosen to complete the surface fitting. This paper adopts a matrix approach, where ITD values, frequency, and perceived azimuth angle are assigned as rows, columns, and corresponding matrix values, respectively. This facilitated surface fitting and smooth adjustment, resulting in the surface fitting graph depicted in Figure 5. From the graph, it is evident that within the range of ITD values from 0 μs to 500 μs and frequencies from 350 Hz to 1370 Hz, there is a clear proportional relationship between ITD and perceived azimuth angle. ITD does not show a direct relationship with frequency. However, the perceived azimuth angle fluctuates and increases with frequency variation, indicating an indirect relationship between ITD and frequency. Specifically, as frequency increases, the fluctuation in the perceived azimuth angle corresponding to ITD also increases.

4 Conclusions

To further investigate the spatial localization cues of ITD, measurements of perceived azimuth angles were conducted across 8 low-frequency bands and 7 different ITD values. Experimental results indicate that within frequencies below 1500 Hz and ITDs below 380 μs, the optimal frequency for perceiving azimuth via ITD falls between approximately 1000 Hz and 1500 Hz. In the frequency range of 350-1370 Hz and ITD ranging from 0 μs to 500 μs, the effect of ITD perception increases with increasing ITD, influenced by frequency. Within the frequency range of 350 Hz to 1370 Hz, 700 Hz represents a special frequency band where the azimuth angle perception of the sound source is relatively smaller. Finally, surface fitting of perceived azimuth angles with ITD and frequency, along with mathematical analysis based on their geometric relationship, can be utilized to explore the functional relationship between ITD, azimuth, and frequency, providing fundamental data and mathematical models for a comprehensive investigation into the spatial localization mechanism of binaural cues and theoretical support for the parameter encoding efficiency of multi-channel audio.

References

[1] J. Zhang, D. Mao, “Dependence of binaural loudness summation on interaural level difference and frequency for pure tones” *Science China Physics, Mechanics and Astronomy* vol 53, no 5, pp. 834—841 (2010).

[2] J.E. Mossop, J. Culling, “Lateralization of large interaural delays” *Journal of the Acoustical Society of America* vol 104, no 3, pp. 1574—1579 (1998).

Internal use

[3] S. Chen, R. Hu, Y. Liu. "Frequency Dependence of Spatial Cues and Its Implication in Spatial Stereo Coding" *2008 International Conference on Computer Science and Software Engineering* pp. 1066—1069 (2008).

[4] L.D. Dunai, W. Hartmann, "Frequency dependence of the interaural time diference thresholds in human listeners" *Journal of the Acoustical Society of America* vol 129, no 4, pp. 2485—2485 (2011).

[5] R. Hu, H. Wang, W. Tu, "Relative contribution of frequency and time difference to selectivity for interaural time difference" *Journal of Acta Acustica* vol. 39, no. 6, pp. 752—756 (2014).

[6] B. Xie, L. Liu, C. Zhang, "Virtual reproduction of surround sound in frontal space using four loudspeakers" *Journal of Acta Acustica* vol. 45, no. 6, pp. 885—896 (2020).

[7] S. Thavam, M. Dietz, "Smallest perceivable interaural time differences" *Journal of the Acoustical Society of America* vol. 145, no. 1, pp. 458—468 (2019).

[8] V. Best, J. Swaminathan, "Revisiting the detection of interaural time differences in listeners with hearing loss" *Journal of the Acoustical Society of America* vol. 145, no. 6, pp. EL508—EL513 (2019).

[9] W.M. Hartmann, B. Rakerd, Z.D. Crawford, "Transaural experiments and a revised duplex theory for the localization of low-frequency tones" *Journal of the Acoustical Society of America* vol. 139, no. 2, pp. 968—85 (2016).

[10] A.C. Diedesch, G.C. Stecker, "Temporal weighting of binaural information at low frequencies: Discrimination of dynamic interaural time and level differences" *Journal of the Acoustical Society of America* vol. 138, no. 1, pp. 125—33 (2015).

[11] B. Liu, Q. Liu, X. Guo, "Research progress of personalized head-related transfer function" *Journal of Electronic Measurement and Instrumentation* vol. 34, no. 11, pp. 155—165 (2020).

[12] X. Wu, D.S. Talagala, W. Zhang, "Individualized interaural feature learning and personalized binaural localization model" *Journal of Applied Sciences* vol. 9, no. 13, pp. 2682 (2019).

[13] W. Tan, D. Rao, Y. Wang, "The localization of virtual source reproduced by the crosstalk cancellation system under different direct-to reverberation energy ratios conditions" *Journal of the Acoustical Society of America* vol. 154, no. 4, pp. A271 (2023)

[14] B. Rosen, M. J. Goupell, "The effect of target and interferer frequency on across-frequency binaural interference of interaural-level-difference sensitivity" *Journal of the Acoustical Society of America* vol. 151, no. 2, pp. 924—938 (2022).

[15] J. Luo, F. Zeng, "Low-frequency limit for binaural timing cues" *Journal of the Acoustical Society of America* vol. 154, no. 4, pp. A153 (2023).

[16] R.C. Smith, S.R. Price, "Modelling of Human Low Frequency Sound Localization Acuity Demonstrates Dominance of Spatial Variation of Interaural Time Difference and Suggests Uniform Just-Noticeable Differences in Interaural Time Difference" *PloS one* vol. 9, no. 2, pp. e89033 (2014).

[17] A. Bilokon, M.J. Goupell, "The effects of aging on the across-frequency processing of interaural time differences" *Journal of the Acoustical Society of America* vol. 153, no. 3, pp. A158 (2023).