Audio Engineering Society

Convention Express Paper 308

Presented at the AES 157th Convention
2024 October 8-10, New York, NY, USA

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 The complete manuscript was not peer reviewed.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.

The effects of interaural time difference and interaural level difference on sound source localization on the horizontal plane

Xinyi Ma[1] , Heng Wang[2], Qiuxian Huang[3] , Chuntao Liu[4] , Kangcheng Zhang[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*

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

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

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

ABSTRACT

Interaural Time Difference (ITD) and Interaural Level Difference (ILD) are the main cues used by the human auditory system to localize sound sources on the horizontal plane. To explore the relationship between ITD, ILD, and the perceived azimuth, a study was conducted to measure and analyze the localization effects on the horizontal plane by combining ITD and ILD. Pure tones were used as sound sources in the experiment. For each of the three different frequency bands, 25 combinations of ITD and ILD test values were selected. These combinations were used to process the perceived sound from directly in front of the listener (pure tone signals collected using an artificial head in an anechoic chamber). The tests were conducted using the 1up/2down and 2AFC (two-alternative forced-choice) psychophysical testing methods. The results showed that the perceived azimuth at 350 Hz and 570 Hz was generally higher than at 1000 Hz. Additionally, the perceived azimuth at 350 Hz and 570 Hz was similar under certain combinations. The experimental data and conclusions can provide foundational data and theoretical support for efficient compression of multi-channel audio.

1 Introduction

In three-dimensional space, the perception of sound source location on the horizontal plane involves multiple parameters, with ITD and ILD being two major ones. ITD and ILD are closely related to the azimuth and frequency of the sound source, among other factors [1-2]. In most cases, when a sound source propagates through the air to reach a person's left and right ears, there is a slight time difference, known as the interaural time difference. Additionally, the sound waves reaching the ears are attenuated due to head shadowing, leading to an interaural level difference.

At low frequencies (mainly below 1500 Hz), ITD plays a major role because low-frequency sound waves can diffract around the head to reach the farther ear. At high frequencies (mainly above 1500 Hz), ILD becomes more important because high-frequency sound waves (i.e., shorter wavelengths) cannot or rarely reach the farther ear. Therefore, the human ear's ability to distinguish direction at high frequencies primarily depends on ILD. During the transition from low to high frequencies, neither ITD

nor ILD can effectively function, and this phenomenon is most apparent when using pure tones as sound sources [3]. This is also the fundamental viewpoint of the classic duplex theory [4].

Research on spatial perception [5-8]mostly focuses on the just noticeable differences (JND) in ITD and ILD, as well as the minimum audible angle (MAA). Additionally, studies on the influence of ITD and ILD on azimuth perception [9] often use only one of these cues. Therefore, this paper aims to investigate the combined effects of ITD and ILD on the localization of sound sources on the horizontal plane through auditory experiments

2 Experimental method

2.1 Experimental Principle

Take the perceived sound from the front of the human body (the audio at the 0-degree position, referred to as 0.wav) and process it with ITD and ILD (i.e., by delaying and amplifying the audio data to produce ITD and ILD effects). Use the processed wav file as a reference sound. Theoretically, this audio contains certain spatial location information, meaning that the perceived position of this audio has shifted from the original position to some angle θ to the left (In this experiment, it was like this). Since θ is unknown, set the database's sound source as the test sound and invite subjects to participate in listening experiments to find the approximate angle of θ.

2.2 Methods for Generating ITD and ILD

The ITD processing method is as follows:First, calculate the number of sample points n needed for the delay based on the preset ITD value. Then, separate the audio data of the file 0.wav (which corresponds to the perceived sound directly in front of the listener) into left and right channels, and save these data separately. Next, process the right channel data by adding n zero sample points at the beginning to create a delay relative to the left channel.Since the audio data in the WAV file is stored in an interleaved format (i.e., the left and right channel data are alternated), you need to add n zero sample points at the end of the left channel data to match the length of the right channel data. After this adjustment, merge the left and right channel data and generate a new WAV file as the reference sound. After this processing, the perceived position of the sound in 0.wav will shift towards the left.

The ILD processing method is as follows: Assume the left and right channel gains are g_1 and g_2 respectively. To ensure that the overall audio signal's energy remains unchanged, the following constraint is applied:

$$g_1^2+g_2^2=1 \quad (1)$$

Additionally, the gain ratio should satisfy:

$$\frac{g1}{g2} = 10^{ild/20} \quad (2)$$

After calculating g_1 and g_2 based on these conditions, apply these gains to the left and right channel data respectively. By performing this adjustment, can achieve the desired interaural level difference in the processed audio signal.

2.3 Creating an Audio Signal Database

Using MATLAB, create three mono pure tones with durations of 300ms, a sampling rate of 44.1kHz, a bit depth of 16-bit, and frequencies of 350Hz, 570Hz, and 1000Hz, respectively, with a sound pressure level set at 75dB. Play these sound sources
through speakers fixed on the horizontal plane of a spherical frame (with a radius of 1.75 meters) in an anechoic chamber. Set the speaker interval to 1 degree. Place the KEMAR artificial head at the center of the spherical frame, ensuring that the artificial head's ears are on the same horizontal plane as the speakers. Sample at 44.1kHz and quantize at 16-bit using high-precision microphones placed inside the artificial head's ear canals. Collect the perceived sound range from the front of the human body to the left ear, i.e., from 0 degrees to 90 degrees, and save it as 90 wav files (named 0.wav, 1.wav, 2.wav, ...). This will serve as the audio database for one frequency band, and the audio databases for other frequency bands will be created similarly.

2.4 Preprocessing of Audio Databases

Apply a 10ms fade-in and fade-out (cosine squared) to the beginning and end of all audio files in the audio database to achieve a smooth crossfade transition between two audio segments during playback. Direct transitions can cause broadband clicks at non-zero crossing points, which may interfere with auditory perception and affect experimental results. The fade-in and fade-out transition method ensures a smoother transition between audio segments, avoiding interference.

2.5 Audiometric Program Design

The experiment employs a two-alternative forced-choice (2-AFC) method [10], where two audio segments are played to the participants through headphones (the playback order is random). Although headphone playback can lead to some degree of in-head localization effect, spatial orientation information can theoretically still be retained [11]. The two audio segments are a reference sound (which remains constant throughout the experiment) and a test sound (the test sound will initially be one of the audio files from the database, and will change based on the participant's judgment during the experiment).

The interval between the segments is 300ms. Considering that sound pressure levels can affect perceptual experiments [12], and that sound pressure levels between 60-80dB have minimal impact on binaural cue perception experiments [13], the preset sound pressure level for each frequency band in this experiment is 75dB. After listening to the two sounds, participants need to choose the one they perceive as more to the left. The program's judgment rule is as follows: selecting the test sound is recorded as result A, and selecting the reference sound is recorded as result B.

During the experiment, two situations may arise. In the first case (Figure 1), the degree of the reference sound's position is smaller than that of the test sound. Theoretically, the test sound will appear slightly more to the left. If the result is A for two consecutive selections, the test sound will be adjusted downwards based on the step size (i.e., in the next hearing test, the test sound will be replaced with the corresponding degree sound file from the database according to the step size; if the step size is 4 and the test sound is 45.wav, it will subsequently be changed to 41.wav) to approach the reference sound. If the result is B in any selection, the test sound will be adjusted upwards based on the step size (i.e., in the next hearing test, the test sound will be replaced with the corresponding degree sound file from the database according to the step size; if the step size is 4 and the test sound is 45.wav, it will subsequently be changed to 49.wav) to move away from the reference sound.

In the second case (Figure 2), the degree of the reference sound's position is greater than that of the test sound. Theoretically, the reference sound will appear slightly more to the left. If the result is B, the test sound will be adjusted upwards based on the step size to approach the reference sound. If the result is A for two consecutive selections, the test sound will be adjusted downwards based on the step size to move away from the reference sound.

A subject's judgment is considered a reversal when it changes from correct to incorrect or from incorrect to correct. The experiment ends after 12 reversals (with a step size of 4 degrees for reversals 1-4, 2 degrees for reversals 5-8, and 1 degree for reversals 9-12). The final result of the experiment is the average value of the last 4 reversal points. Based on the experimental results, we can determine the approximate position corresponding to a 0-degree sound source processed with ITD and ILD. The subject's judgment results are saved in real-time to an Excel file for subsequent analysis.

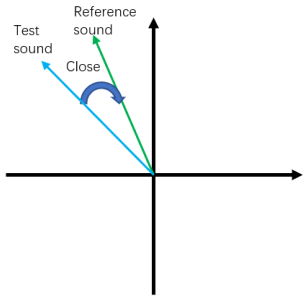


Figure 1. The test sound is positioned to the left of the reference sound

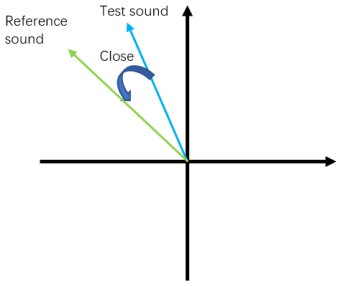


Figure 2. The test sound is positioned to the right of the reference sound

2.6 Selection of ITD and ILD test values

Before the experiment, different ITD delay treatments were preset (0us, 50us (delayed by 2.205 sample points, rounded down to 2 sample points), 110us (delayed by 4.851 sample points, rounded down to 4 sample points), 190us (delayed by 8.379 sample points, rounded down to 8 sample points), and 280us (delayed by 12.348 sample points, rounded down to 12 sample points)) and ILD gain treatments (0dB, 2dB, 4dB, 6dB, 9dB) for selection. During the experiment, a combination of one ITD delay and one ILD gain is selected (e.g., ITD of 0us and ILD of 0dB, or ITD of 0us and ILD of 2dB, etc.). The audio at 0 degrees is processed with these parameters and used as the reference audio for subsequent experiments. This approach allows for a systematic study of the impact of different parameter combinations on the results.

2.7 Experimental Equipment

To ensure the accuracy and validity of the experimental results, it is essential to minimize the influence of the external environment and equipment on the subjects. Therefore, a professional audio laboratory was chosen for the listening tests. The subjects wore headphones and used a hearing test program on a computer. The specific configuration is as follows: Computer: Windows 10 operating system, CPU i5-12400.Sound Card: Creative sound card, model Sound Blaster X-Fi HD. Headphones:

Sennheiser HD380 pro. Sound Pressure Meter: smart sensor AR824

2.8 Experimental subjects

The subjects consisted of 8 university students, aged between 18 and 24, including four males and four females. All subjects had normal hearing. Before the experiment, they underwent a period of auditory training to adapt to the experimental requirements. Additionally, considering that prolonged hearing tests may cause auditory fatigue, potentially affecting the results, subjects were required to take a break after 20 minutes of testing.

3 Experimental Results and Analysis

3.1 Experimental data

The experiment involved a total of 8 subjects. The auditory test results of the subjects were extracted from Excel sheets and averaged (results rounded to two decimal places) as shown in the following tables (Table 1, Table 2, Table 3).

ILD / ITD	0db	2db	4db	6db	9db
0us	12.75	17.5	26.5	47	90
50us	16.78	37.2	42.21	50.16	90
110us	20.15	43.46	44.5	84.4	90
190us	35.25	46.42	52.91	84.16	90
280us	42.28	50.95	53.8	88.46	90

Table 1: Corresponding Values of Different ITD and ILD Combinations to Horizontal Azimuth Angle θ at 350 Hz.

ILD / ITD	0db	2db	4db	6db	9db
0us	14.31	27.16	39.03	40.25	57.67
50us	24.31	36.9	43.37	50.43	75.56
110us	24.5	38.87	46.06	48.87	78.4
190us	37.78	42.96	49.34	54.83	74.78
280us	41.71	41.62	49.79	53.58	74.42

Table 2: Corresponding Values of Different ITD and ILD Combinations to Horizontal Azimuth Angle θ at 570 Hz

ILD / ITD	0db	2db	4db	6db	9db
0us	4.4	12.16	18.46	20.43	46.12
50us	8.25	13.75	17.08	21.1	50.46
110us	13.08	20.91	22.35	32.62	46
190us	18.55	20.95	23.54	30.25	43.9
280us	24.68	30.04	42.25	47.5	78.03

Table 3: Corresponding Values of Different ITD and ILD Combinations to Horizontal Azimuth Angle θ at 1000 Hz

From Tables 1, 2, and 3, it can be clearly seen that as ITD and ILD increase, the corresponding horizontal azimuth angle also increases, which is consistent with existing experimental conclusions. As ITD increases, it can be observed that under the same ILD conditions, the perceived angle generally shows an upward trend. Particularly at low ILD values, there is no abrupt increase in the perceived angle during the increase of ITD (in each segment). However, when ILD increases from 4 dB to 6 dB and from 6 dB to 9 dB, abrupt increases often occur, resulting in larger perceived angles and poorer fitting effects with the corresponding audio angles. This also confirms that ITD plays a major role at low frequencies.

3.2 The Relationship between Horizontal Azimuth Angle and ITD and ILD

To investigate the relationship between horizontal azimuth and ITD/ILD, curve fitting was performed on existing experimental data (Table 1, Table 2, Table 3) using interpolation methods to obtain additional horizontal azimuth values for various combinations of ITD and ILD. Trigonometric cubic interpolation was applied for surface fitting, and the fitted surfaces can be seen in Figures 3, 4, and 5.

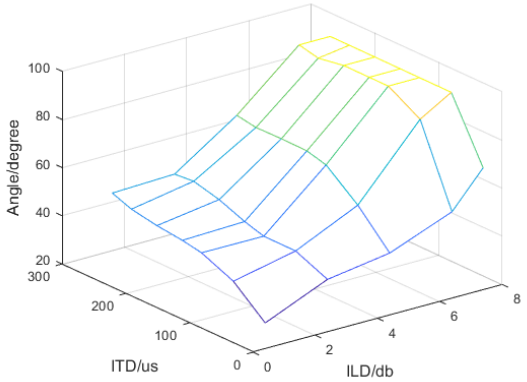


Figure 3. The horizontal azimuth corresponding to ITD and ILD combinations at 350 Hz

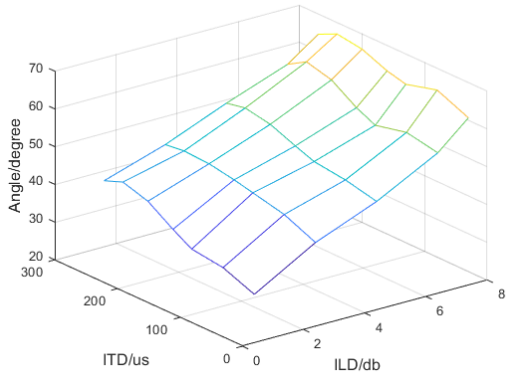


Figure 4. The horizontal azimuth corresponding to ITD and ILD combinations at 570 Hz

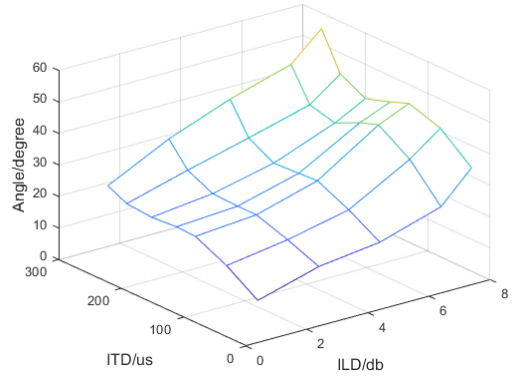


Figure 5. The horizontal azimuth corresponding to ITD and ILD combinations at 1000 Hz

Frequency is an important factor influencing the perceptual characteristics of binaural cues. To better explore the impact of frequency on the perceptual characteristics of binaural cues, this section plots the corresponding values of ITD and ILD combinations with horizontal azimuth θ at different frequencies, as shown in Figures 6 and 7. The x-axis represents combinations of ITD and ILD values, while the y-axis represents horizontal azimuth. Each figure includes three lines corresponding to three frequency bands. For ease of analysis, the data from the tables (Table 1, Table 2, Table 3) are unfolded by columns in Figure 6 and by rows in Figure 7.

From the figure6、7, it can be observed that the horizontal azimuth angle at 1000 Hz is consistently smaller than at 350 Hz and 570 Hz throughout (except at 280-9). This could be due to both ITD and ILD having limited effectiveness during the transition from low to high frequencies.

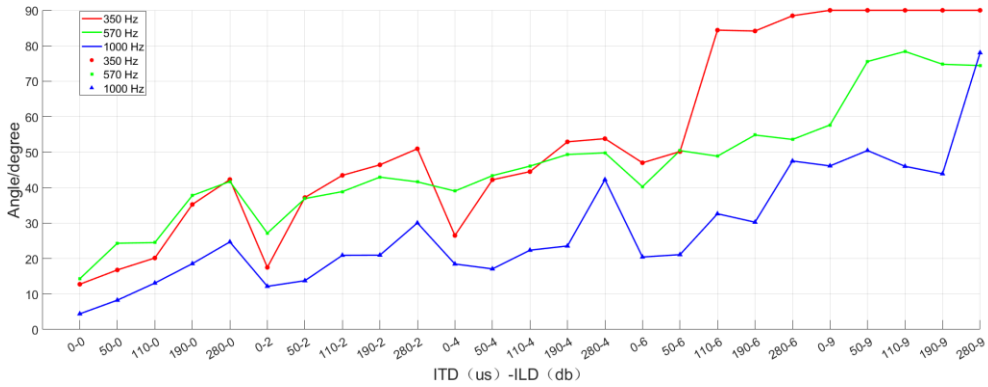


Figure 6. Corresponding values of horizontal azimuth angle θ for ITD-ILD combinations at different frequencies.

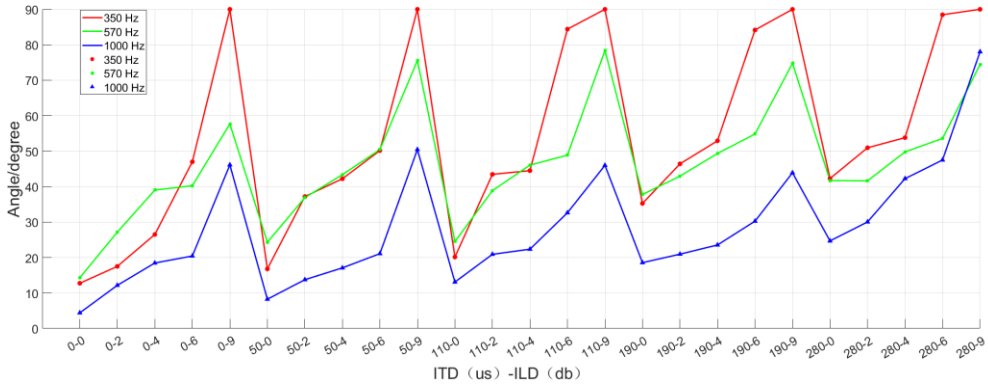


Figure 7. Corresponding values of horizontal azimuth angle θ for ITD-ILD combinations at different frequencies.

4 Conclusions

In the experimental process, we manipulated the perceived sound at 0 degrees by introducing delays and gains to induce effects such as ITD and ILD. Through auditory experiments, we obtained the perceived azimuth angles corresponding to different combinations of ITD and ILD, and through the above data analysis, we can clearly observe changes in horizontal perceived azimuth as ITD and ILD vary. Specifically, as ITD and ILD increase, the horizontal perceived azimuth gradually increases. Additionally, we observe the influence of different frequencies on horizontal perceived azimuth under the same ITD and ILD combinations, where at 350 Hz and 570 Hz, the horizontal perceived azimuth is generally greater than at 1000 Hz. Furthermore, under certain combinations, the horizontal perceived azimuth at 350 Hz and 570 Hz appears similar.

References

[1] Hu Ruimin,Wang Xiaochen,Zhang Maosheng,Li Dengshi,Wang Song,Gao Li,Yang Cheng,Yang Yuhong.Review on Three Dimension Audio Technology[J].Journal of Data Acquisition & Processing,2014,29(5):661-676 A. Camus, "The Art of Autoreferencing" Journal of Self-Absorption vol. 30, no. 2, pp. 1—2 (1998).

[2] YIN Fu-liang,WANG Lin,CHEN Zhe.Review on 3D audio technology[J].Journal on Communications,2011,32(2):130-138.

[3] Stevens S S, Newman E B. The localization of actual sources of sound[J]. The American journal of psychology, 1936, 48(2): 297-306.

[4] Rayleigh L. On our perception of the direotion of a source of sound[J]. Proceedings of the Musical Association, 1875, 2(1): 75-84.

[5] Shui-xian, Chen et al. "Frequency Dependence of Spatial Cues and Its Implication in Spatial Stereo Coding." *2008 International Conference on Computer Science and Software Engineering* 4 (2008): 1066-1069.

[6] Elena A ,J A K ,Monica G , et al.Comparison of auditory spatial bisection and minimum audible angle in front, lateral, and back space.[J].Scientific reports,2020,10(1):6279.

[7] Rummukainen S O ,Schlecht J S ,Habets P A E .Self-translation induced minimum audible angle[J].The Journal of the Acoustical Society of America,2018,144(4):EL340.

[8] Mills A W. On the minimum audible angle[J]. The Journal of the Acoustical Society of America, 1958, 30(4): 237-246.

[9] Yost, William A. "Lateral position of sinusoids presented with interaural intensive and temporal differences." The Journal of the Acoustical Society of America 70.2 (1981): 397-409.

[10] Bogacz R, Brown E, Moehlis J, Holmes P, Cohen JD. The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. Psychol Rev. 2006 Oct;113(4):700-65. doi: 10.1037/0033-295X.113.4.700. PMID: 17014301.

[11] HU Ruimin, WANG Heng, TU Weiping, ZHANG Cong, WANG Song. Relative contribution of frequency and time difference to selectivity for interaural time difference[J]. ACTA ACUSTICA, 2014, 39(6): 752-756. DOI: 10.15949/j.cnki.0371-0025.2014.06.027

[12] LONG Zhangcai, TANG Chaoqun, YAO Qiuping. Relations of intensity limen to intensity of tonal bursts[J]. ACTA ACUSTICA, 2000, 25(6): 486-491. DOI: 10.15949/j.cnki.0371-0025.2000.06.002

[13] R. M. Hershkowitz, N. I. Durlach; Interaural Time and Amplitude jnds for a 500-Hz Tone. *J. Acoust. Soc. Am.* 1 December 1969; 46 (6B): 1464–1467