Archives of Acoustics Vol. 50, No. 4, pp. 465–476 (2025), https://doi.org/10.24425/aoa.2025.154827

Research Paper

Localization of Virtual Sound Source Reproduced by the Crosstalk Cancellation System Under Different Reflective Conditions

Wei TAN, Guangzheng YU, Jun ZHU, Dan RAO*

Acoustic Laboratory, School of Physics and Optoelectronics, South China University of Technology
Guangzhou, China

*Corresponding Author e-mail: phdrao@scut.edu.cn

Received November 10, 2024; revised July 28, 2025; accepted August 30, 2025; published online October 9, 2025.

This study explores the localization of virtual sound source reproduced by the crosstalk cancellation system under different reflective conditions in virtual rooms and analyzes the localization results with binaural cues. Binaural room impulse responses are generated using the high-order image source method. By modifying the acoustic parameters of the virtual room to manipulate the intensity and temporal structure of the reflection, psychoacoustic experiments were conducted using headphone reproduction. The experimental results indicate that, changes in reflection intensity within a certain range by altering the room reverberation time (RT) do not cause noticeable variations in virtual source localization. Increasing the loudspeaker–listener distance (changing temporal structure of reflections) deteriorates localization performance. The primary distinction between variations in the loudspeaker–listener distance and RT lies in whether the temporal structure of the reflection component changes. Overall, the study highlights the importance of the reflection temporal structure in the virtual source localization. The analysis of binaural cues indicates that, even in reverberant environments, the interaural time difference exhibits greater consistency with localization than the interaural level difference.

Keywords: sound localization; crosstalk cancellation; reflection environment; binaural cues.

1. Introduction

Binaural reproduction attempts to accurately reconstruct the desired auditory events in the listener's ear. Through binaural reproduction, listeners can perceive the spatial impression of acoustic scenes that have been recorded elsewhere or synthesized. This technique is commonly employed in immersive virtual reality (Lentz, 2008; Villegas, 2015). Both headphones and loudspeakers can be used to reproduce binaural signals. For loudspeaker reproduction, the crosstalk phenomenon inevitably occurs between the loudspeaker and the listener's ears. Crosstalk is defined as the sound transmitted from one loudspeaker to the opposite ear, which always deteriorates the sound source localization performance and timbre (Masiero *et al.*, 2011). Crosstalk cancellation (CTC) filters eliminate or reduce the contributions from crosspaths (Gardner, 1998). These filters are typically derived by inverting the head-related transfer function (HRTF) matrices, indicating that the CTC system is best suited to anechoic environments. However, in practical applications, CTC systems are routinely employed in various acoustic environments, such as listening rooms, offices, living rooms, etc. The reflections in actual reproduction environments may disrupt binaural cues, i.e., interaural time difference (ITD) and interaural level difference (ILD).

Regarding the localization performance, researchers have mainly focused on the influence of low-order reflections on the direction localization of the CTC system. By simulating the low-order reflection from an wall with different distances, it is possible to investigate the impact on virtual sources reproduced by the CTC system, with the results explained in terms of binaural cues (Kosmidis *et al.*, 2014; Sæbø, 2001; Tan *et al.*, 2023). Other studies have also investigated the localization performance of reflections on the CTC system through loudspeaker experiments under multiple reflective surfaces (Bahri, 2019; Sæbø, 1999).

In general, existing studies have been restricted to simplified reflection situations, investigating only a limited order of reflections without considering the realistic situation of a full sequence of reflections. On the other hand, despite many studies that have investigated the localization of real sources in reflective environments (Blauert, 1997; Brown *et al.*, 2015; Hartmann, 1983; Rakerd, Hartmann, 2010), differences exist in the sound field generated by virtual and actual sound sources. For a virtual source under reflective conditions, the reflections are determined by the position, input signal intensity, and phase of the loudspeakers reproducing the virtual source (Tan *et al.*, 2023). However, in the case of a real source, the reflections are only determined by the real source itself. Therefore, the binaural signal received by the listener differs significantly in the two cases, potentially resulting in localization disparities. Given this, a systematic study on the localization performance of virtual sources reproduced by CTC systems under different reflective conditions is essential.

This study aims to examine the localization of virtual sources reproduced by the CTC system under varying reflective conditions within enclosed spaces. Although the geometric dimensions of rooms, absorption boundary conditions, and other parameters are complex and varied, reflections can still be characterized by their temporal structure and intensity. Therefore, we explore the effect of reflections with varying temporal structures and intensities on localization of the virtual sound source reproduced by the CTC system, where we manipulate the reflection intensity and temporal structure by changing reverberation time (RT) of the virtual room and the distance between the listener's position and the loudspeakers. Considering that modifying the acoustic parameters in a real room and conducting virtual source localization experiments using loudspeakers are laborious and time-consuming tasks, and implementing such tests poses significant challenges. Thus, the research objectives are achieved using virtual reproduction technology (auralization) based on headphones. The crucial aspect for virtual sound reproduction based on headphones is to produce the correct binaural room impulse response (BRIR). There are two main approaches for obtaining BRIRs in different reflective environments: binaural measurement (Genuit, 1992; Li, Peissig, 2020; Møller, 1992) and simulation (Lehnert, Blauert, 1992; Møller, 1992). The measurements are relatively accurate, but it can be challenging to alter the acoustic parameters of the room. This difficulty can be solved by simulation methods, if the simulation methods are validated by numerical simulations and experimental measurements, such as the validation of the reverberation room model simulated in the ODEON program (Nowoświat, Olechowska, 2022) and room-acoustics diffusion theory (Visentin *et al.*, 2013). The image source method (ISM) is commonly used for acoustic simulations. The ISM is prevalent in architectural acoustics and provides a valuable method for evaluating a room's acoustic quality (Allen, Berkley, 1979; Habets, 2010). The localization performance of sound sources based on the BRIRs generated by the ISM and the stochastic scattering method has been validated via headphones, revealing that it is generally equivalent to the measured BRIR (Rychtáriková *et al.*, 2009).

In this study, the high-order ISM is employed to simulate the spatial room impulse responses (SRIR) in empty rectangular rooms of different sizes under various RTs and loudspeaker distances. The BRIRs under different acoustic conditions are then synthesized by the combination of SRIRs and the corresponding HRTFs. Furthermore, the BRIRs are processed by a series of CTC filters, followed by a synthesis of binaural signals at different target azimuths and conditions. The subjective experiments via headphones are conducted to examine the localization of virtual sound sources generated by the CTC system under the above acoustic conditions. The localization results are analyzed in terms of the ITD and ILD, which are calculated based on a binaural auditory model that accounts for the precedence effect, and discussed from the perspective of psychoacoustics.

The rest of this paper is organized as follows: Sec. 2 introduces the CTC system and the method of generating BRIRs; Sec. 3 conducts the experiment about virtual source localization under different reflective conditions and analyze the experimental results; Sec. 4 analyses the localization cues for experimental results; Sec. 5 conducts a discussion for the results, and finally Sec. 6 presents the conclusions to this study.

2. Simulation of the CTC system in a virtual room

2.1. CTC system

For the two-loudspeaker CTC system in an anechoic room, the transmission of sound signals is shown in Fig. 1. When the loudspeakers of the CTC system emit sound, one of the listener's ears can simultaneously receive signals from both the left and right loudspeakers. To reduce directional distortion caused by crosstalk, binaural signals should be processed through a series of CTC filters before being delivered to the loudspeakers. For the CTC system in the frequency domain, the transmission of sound signals is given by

$$\begin{bmatrix} P_L \\ P_R \end{bmatrix} = \begin{bmatrix} H_{LL} & H_{RL} \\ H_{LR} & H_{RR} \end{bmatrix} \begin{bmatrix} C_{LL} & C_{RL} \\ C_{LR} & C_{RR} \end{bmatrix} \begin{bmatrix} H_L \\ H_R \end{bmatrix} E_0, \quad (1)$$

or

$$\mathbf{P} = \mathbf{H} \cdot \mathbf{C} \cdot \mathbf{E}, \quad (2)$$

where P_L and P_R are binaural pressures in an anechoic room, respectively. H_{IJ} are the elements of $\mathbf{H}$, representing the HRTF of the I-th source to the J-th ear, where I and J denote L or R. The elements C_{IJ} of $\mathbf{C}$ are the corresponding CTC filters. The HRTF of the target virtual source are denoted as H_L and H_R. E_0 is the monaural signal and $\mathbf{E}$ represents the binaural signal.

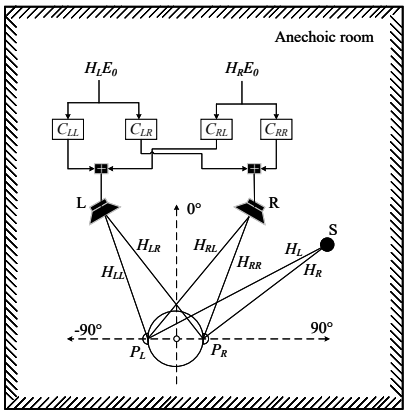

Fig. 1. CTC system in an anechoic room. The coordinate system is represented in the diagram. The virtual source is denoted by 'S'; 'L' and 'R' represent the left and right loudspeakers for reproduction.

To eliminate crosstalk, the product of $\mathbf{H}$ and $\mathbf{C}$ should equal the identity matrix $\mathbf{I}$, that is,

$$\mathbf{HC} = \mathbf{I}. \tag{3}$$

In consequence, $\mathbf{C}$ is the inverse matrix of $\mathbf{H}$. Due to HRTFs are nearly singular and cannot be inverted at some frequencies. To enhance the robustness of the solution, we utilized a regularization method to compute the matrix $\mathbf{C}$ (Kirkeby, Nelson, 1999). Consequently, $\mathbf{C}$ can be computed as the pseudoinverse of $\mathbf{H}$:

$$\mathbf{C} = \left(\mathbf{H}^{\mathrm{T}}\mathbf{H}+\lambda\mathbf{I}\right)^{-1}\mathbf{H}^{\mathrm{T}}, \tag{4}$$

where the superscript T represents the conjugate transpose, λ is the regularization parameter. In Eq. (4), λ can be adjusted to 0.001 to balance accuracy and the stability of virtual source. The HRTFs used in the present work were obtained from the simulated KEMAR (Knowles Electronics Manikin for Acoustics Research) artificial head HRTF database. The spectral range of the HRTF starts at 50 Hz and extends up to 22.5 kHz with a spatial resolution of 1° and an increment of 50 Hz between each step, which was computed by the boundary element method (Katz, 2001; Rui *et al.*, 2013) as executed in Mesh2HRTF (Ziegelwanger *et al.*, 2015).

2.2. BRIR simulations

To obtain BRIRs in rectangular empty rooms with different reverberations, the high order ISM was used to generate spatial room impulse responses (SRIR). The ISM is based on the principle that a wavefront arriving from a point source and reflections from an infinite plane can be modeled as emanating from an image source. This image source can therefore be visualized as a mirror source. Consider a rectangular room with dimensions of $\{L_x, L_y, L_z\}$ and a sound source positioned at $\{s_x, s_y, s_z\}$. The relative positions of the image sources with respect to the receiver position can be written as

$$(x_i, y_i, z_i) = \begin{pmatrix} (1-2u)s_x + 2nL_x, (1-2v)s_y + 2lL_y, \\ (1-2w)s_z + 2mL_z \end{pmatrix}, \tag{5}$$

where $\{u, v, w\}$ and $\{n, l, m\}$ are integer vector triplets; u, v, and w can take values of 0 or 1, whereas the possible values of n, l, and m are based on the order of the reflections.

For simplification, only omnidirectional sound sources are considered here, and RT are used to replace the variation of sound absorption boundary conditions. Energy absorption by the walls of the room and attenuation over distance for sound propagation (Ocheltree, Frizzel, 1989) are integrated into the calculations of the impulse responses of different order image sources. In this study, the precise materials corresponding to the given absorption coefficients were not specified. For the sake of simplification, a uniform absorption coefficient was assigned to all surfaces in the simulation, thereby enabling a focused investigation of the impact of reverberation time and the delay and intensity of reflected sound.

To incorporate enough reflections in the simulation, the order of the image sources is configured to be sufficiently high, ensuring that the energy attenuation of the image source exceeds 60 dB at that order. After performing the inverse discrete Fourier transform (IDFT) on the corresponding HRTFs, the corresponding head-related impulse responses (HRIRs) are obtained. Next, the corresponding HRIRs were convolved with the impulse represented by the direct source and each image source, and the resulting responses were summed to obtain the BRIR. The process of obtaining BRIR is shown in Fig. 2.

Considering the loudspeaker angles in Fig. 1 in a virtual room, we use the binaural room transfer function (BRTF) to replace the HRTF matrix in Eq. (1). Finally, the binaural sound pressure produced by the CTC system in a virtual room can be expressed as

$$\begin{bmatrix} P'_L \\ P'_R \end{bmatrix} = \begin{bmatrix} B_{LL} & B_{RL} \\ B_{LR} & B_{RR} \end{bmatrix} \begin{bmatrix} C_{LL} & C_{RL} \\ C_{LR} & C_{RR} \end{bmatrix} \begin{bmatrix} H_L \\ H_R \end{bmatrix} E_0, \tag{6}$$

where P'_L and P'_R are the binaural sound pressures at each ear in a virtual room and B_{IJ} is the transfer function for the I-th loudspeaker to the J-th ear.

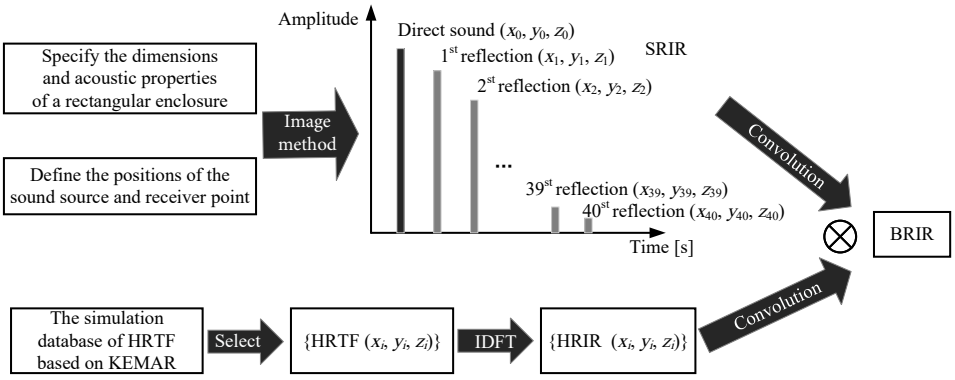

Fig. 2. Process of obtaining BRIRs.

3. Experiment: Different temporal structures and intensities reflections

The change in the reflection environment (condition) is essentially a variation in the temporal structure and intensity of the reflection. Therefore, in this section, we attempt to explore the effect of reflections of different temporal structure and intensity on the localization of virtual source reproduction by the CTC system by varying the acoustic parameters of the room and loudspeaker arrangement.

3.1. Experimental design

3.1.1. Experimental condition

Due to the fact that variations in RT and loudspeaker distance will respectively alter the intensity and temporal structure of the reflections (with intensity changing concurrently), both will also change the direct-to-reverberant energy ratio (DRR), which could potentially affect the localization of virtual sound sources. Therefore, in the present experiment, we consider controlling the RT and the loudspeaker distance to modify the intensity and temporal structure parameters of the reflections.

Although the actual room types, acoustic parameters within the rooms, and other factors are numerous and highly complex, in order to qualitatively analyze the impact of reflection intensity and temporal structure parameters on the localization of virtual sound sources reproduced by the CTC system, we selected two representative acoustical spaces of different scales for the experiments.

The empty room ① 6.4 m (length) × 5.6 m (width) × 2.7 m (height), and the empty room ② measures 8.0 m (length) × 7.0 m (width) × 3.5 m (height). The two rooms and their loudspeaker arrangements are shown in Fig. 3. The center position of the listener's head is set at a nonspecial location in the central area of the room. The initial distance between the sound source and the wall is maintained at 2 m or more to avoid the occurrence of reflections with a small delay. The sound source is positioned at a height of 1.2 m, aligning with the center of the listener's head. The arrangement angle of sound source is 60°.

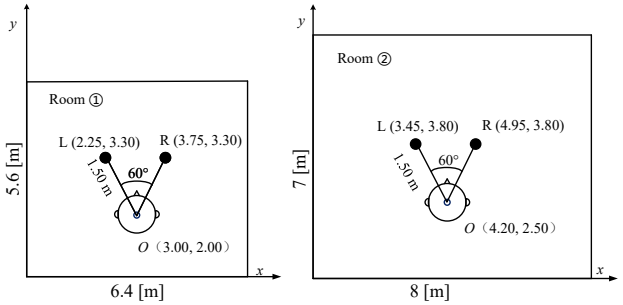

Fig. 3. Room and loudspeaker layouts. The distance between the sound source and the listeners is 1.5 m. Coordinate positions of both the listeners and the loudspeakers are shown for each listening scenario.

Experiment condition 1: different intensities of reflections. In this experiment condition, the listener position and loudspeaker layout are identical to those in the room ①. RT values of 0.3 s, 0.8 s, and 1.2 s are configured, encompassing the typical RTs of the acoustic environments used by CTC systems. Under these conditions, only the intensity of the reflections will change (as shown in Table 1, all reflections parameters are calculated relative to the direct sound). In this scenario, the minimum delay of the reflections remains

Table 1. Variation in reflection parameters due to RT changes.

RT [s]	Minimum delay [ms]	Intensity of the minimum delay reflection [dB]	Total intensity of early reflections [dB]	Total intensity of late reverberation sound [dB]
0.3	3.90	−4.8	1.4	−0.8
0.8	3.90	−3.4	5.5	4.7
1.2	3.90	−3.2	6.4	8.0

Table 2. Variation in reflection parameters due to loudspeaker distance changes.

Loudspeaker distance [m]	Minimum delay [ms]	Intensity of the minimum delay reflection [dB]	Total intensity of early reflections [dB]	Total intensity of late reverberation sound [dB]
1.50	3.90	−6.5	2.1	0.5
2.50	2.80	−3.9	6.3	4.4
3.50	2.20	−2.7	9.4	7.7

around 3.90 ms, which is within the suppression range of the precedence effect. Therefore, the temporal structure of the reflections does not change, while the intensity of the early reflections increases by approximately 5.0 dB and the late reverberation increases by about 8.8 dB.

Experiment condition 2: different temporal structures of reflections. We employ the method of changing loudspeaker distance to control temporal structures of reflections. Under this experimental condition, the size of the virtual room, the loudspeaker arrangement, and the listener's position are consistent with room ② in Fig. 3, and the RT is set to 0.7 s. The loudspeaker distances are set at 1.50 m, 2.50 m, and 3.50 m, respectively, while the loudspeaker span angle remains at 60°. As the distance of the loudspeaker increases, the minimum delay of the reflection decreases from 3.90 ms to 2.20 ms, shifting from the suppression range of the precedence effect (usually greater than 3 ms) to the range where the precedence effect begins to take effect. Additionally, the intensity of the reflection increases accordingly, as shown in Table 2. Unlike changing the RT, altering the loudspeaker distance simultaneously changes both the temporal structure and the intensity of the reflections.

3.1.2. Subjects

The experiment involved eight participants, comprising five males and three females, with ages ranging from 20 to 26 years old. All participants were Master's degree candidates. They self-reported as having typical hearing abilities and had previously engaged in sound localization studies. Compensation was provided for their involvement in the experiment.

3.1.3. Experimental procedure

The BRIRs in the virtual rooms were obtained using the method described in Sec. 2, where the image source order was set to 40. The calculations were implemented in MATLAB on a personal computer. Three stimuli were chosen: music (from Blue Danube), speech (from a Chinese corpus read by a baritone), and a 6-second duration of pink noise processed with fade-in and fade-out. The pink noise was passed through a 10 kHz low-pass finite impulse response filter and reproduced using the Etymotic Research (ER-2) insert earphone. The ER-2 earphones are inserted into the ear canal and bypass the pinna's acoustic effects, their corresponding headphone transfer function does not include pinna coloration. Given that the flat frequency response of the ER-2, no additional headphone equalization was applied. Each stimulus was presented randomly and repeated three times. The average binaural sound pressure level in the condition of the room ① was calibrated to approximately 65 dB(A). The virtual source's target azimuths were categorized into seven distinct directions, ranging from −90° to 90°, with each direction separated by 30° intervals.

Listening tests were performed in an isolated control room. Participants initially engaged in a training session, where they listened to the test stimuli, being clearly informed that the stimuli could emanate from any location within the frontal plane. Feedback on responses was not given throughout the training stage. The azimuth of the virtual source was determined using the Polhemus Fastrak G4™, a portable and mobile wireless electromagnetic tracker that achieves full 6-degrees-of-freedom localization. Each subject held a lightweight carbon fiber rod in their hand with a sensor attached at the end. When the subject heard a stimulus, they pointed the sensor towards the perceived location of the sound source. The sensor recorded the position information and transmitted it to a personal computer. After real-time processing, the subject's perceived angle was determined and recorded. The experiments were divided into three groups, i.e., different room types, different RTs, and different loudspeaker distances. Subjects are required to take a break every 15 to 20 minutes.

3.2. Experimental results

Figure 4 shows the virtual source localization results of the CTC system in rooms with different intensities of reflections (different RTs). At a ±90° target azimuth, the average perceived azimuth (absolute value) under the RT condition of 0.3 s is slightly larger than the average perceived azimuth under other RT conditions. At other target azimuths, there is no significant difference in the average perceived azimuth under different RT conditions (i.e., 0.8 s and 1.2 s). Additionally, the standard deviation of the lateral perceived azimuth under RT conditions of 0.8 s and 1.2 s is slightly higher than that under the RT condition of 0.3 s, with a difference ranging from about 1° to 4°. A multifactor repeated measures analysis of variance (ANOVA) showed that the main effects of RT and signal type

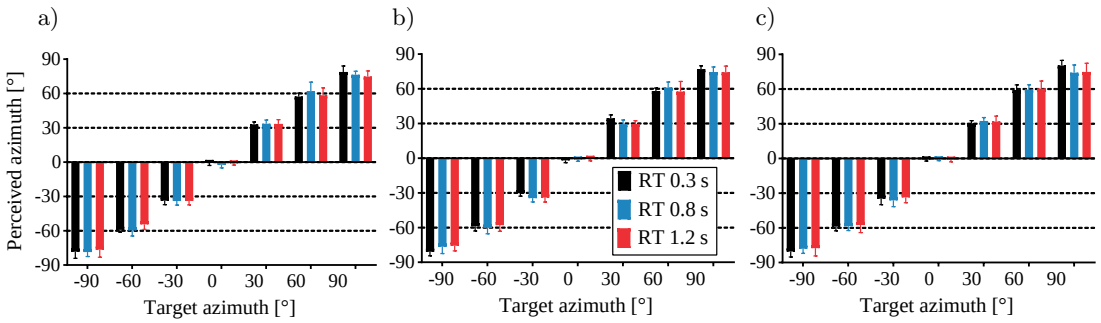

Fig. 4. Localization results at different RTs (different intensities of reflections): a) speech; b) music; c) pink noise.

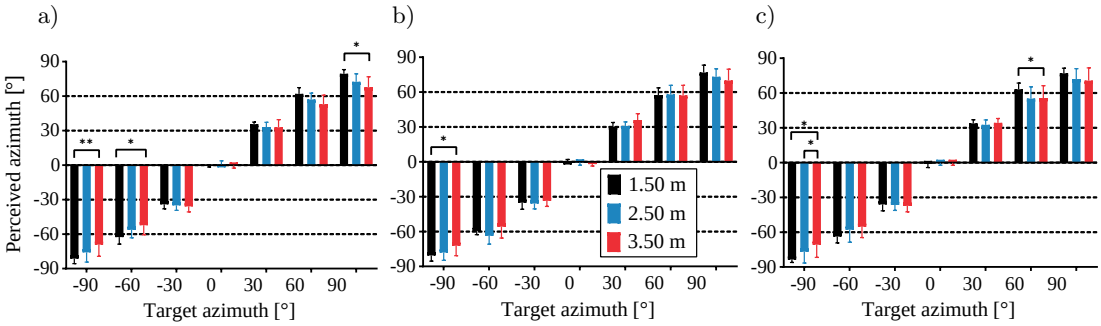

Fig. 5. Localization results of different distance condition (different temporal structures of reflections): a) speech; b) music; c) pink noise.

were not significant. Overall, the localization results indicate that even with a significant increase in the intensity of reflections, when the reproduction loudspeakers are positioned away from the wall (more than 2 m), the subjects can still locate the virtual sound source. The localization accuracy of the CTC system's reproduced virtual sound sources does not significantly decline. Therefore, when the loudspeakers are relatively far from the wall, changes in RT within a certain range, that is, changes in the intensity of the reflections (without altering the temporal structure of the reflections), do not affect the localization of the virtual sound source.

Figure 5 displays the localization results for the cases of different loudspeaker distances (different temporal structures of reflections). For front target sound sources (0°, 30°, and −30°), there is little difference in the perceived azimuth under different loudspeaker distance conditions. However, for lateral target sound sources (±60° and ±90°), the perceived azimuth (absolute value) tends to decrease with the increasing loudspeaker distance. For example, in the case of the speech signal and the 90° target azimuth, the perceived azimuths at loudspeaker distances of 1.5 m, 2.5 m, and 3.5 m are 80°, 72°, and 67°, respectively. In addition, as the loudspeaker distance is raised, there is a noticeable increase in the SD of the lateral perceived azimuths (±60° and ±90°). For instance, with a 1.5 m loudspeaker distance, the SD of lateral perceived azimuths ranges from 6° to 8°, whereas at larger loudspeaker distances, this range increases to 8° to 13°. This indicates that participants experience an increase in localization variability when localizing virtual sources at larger distances.

The perceived azimuths were subjected to multifactor repeated measures ANOVA. No significant main effects are found for either distance or signal type. However, pairwise comparisons with Bonferroni corrections show that, for the −90° target azimuth, a significant difference exists between the localization for distances of 1.5 m and 3.5 m (with different stimuli, all $p < 0.05$, refer to the asterisks in Fig. 5 for more details). For 90° and ±60° target azimuths, significant differences exist for some signals between the localization for distances of 1.5 m and 3.5 m, e.g., for speech at 90°, $p = 0.017$.

The ANOVA analysis results confirmed the previously described trends in localization changes. Specifically, at lateral target angles, the perceived azimuths are smaller under conditions of greater loudspeaker distances compared to smaller loudspeaker distances. This indicates that the temporal structure of the reflections affects the localization of virtual sound sources.

4. Localization cues analysis

4.1. Binaural auditory model

To analyze the changes in binaural cues under different reflection conditions, a binaural auditory model was introduced. The model architecture considered throughout this section is shown in Fig. 6. The binaural signal (right and left channels) was obtained by simu-

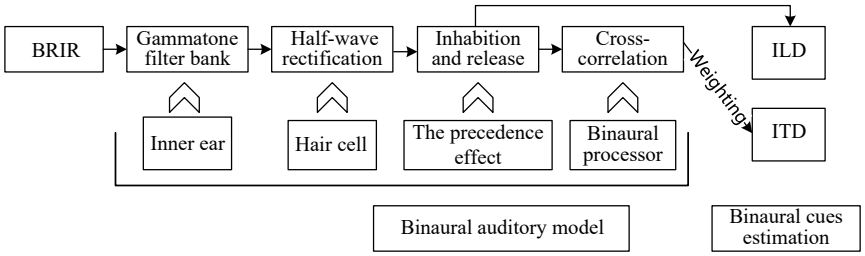

Fig. 6. Binaural auditory model structure of the cross-correlation-based precedence effect.

lation, as described in Sec. 2. The peripheral components contain the middle and inner ear. The influence of the middle ear on the localization is typically omitted, and its effect on the signal is uniform for both ears, thus leaving the ITD and ILD unaffected. The inner ear frequency selectivity was modeled using a gammatone filter bank (Slaney, 1993) of 42 bandpass equivalent rectangular bandwidth (ERB) channels. The center frequencies of the filter bank varied from 100 Hz to 10 kHz, because the main energy of the stimuli was below 10 kHz. A gammatone filter bank is often used as the front end in cochlea simulations, converting intricate sounds into multi-channel activity patterns akin to those observed in the auditory nerve. The nonlinear behavior of the hair cell was then simulated by applying half-wave rectification to the output of the gammatone filters (Braasch, 2013; Cooke, 2005).

To account for the precedence effect, suppression and release mechanisms for reflections were employed. A segmented function was adopted to fit the original function proposed in (Martin, 1997; Yost, Gourevitch, 1987). Figure 7 shows the delay-varying function of the precedence effect on the lag component in localization. In the first few milliseconds, the influence of the delayed sound diminishes as the delay increases. When the delay reaches about 3 ms, the weight is approximately 0, and this value is maintained until the delay is 15 ms. This stage corresponds to the inhibition process. As the delay continues to increase, inhibition slowly releases, and the weight gradually increases until the delay reaches 35 ms. For delays greater than 35 ms, the weights remain unchanged.

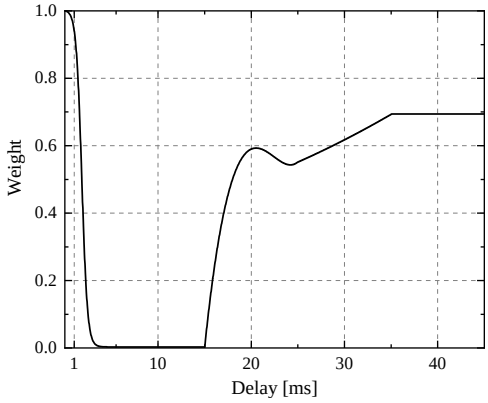

Fig. 7. Function simulating the precedence effect on lag sounds.

In this model, the stage of binaural processing occurs after the precedence effect. The binaural processor was simulated using a cross-correlation model, with the following cross-correlation function employed to obtain the ITD:

$$\Phi_{LR}(\tau) = \frac{\int_{-\infty}^{+\infty} B_{L,N}(t+\tau) B_{R,N}(t)\,\mathrm{d}t}{\left\{\left[\int_{-\infty}^{+\infty} B_{L,N}^2(t)\,\mathrm{d}t\right]\left[\int_{-\infty}^{+\infty} B_{R,N}^2(t)\,\mathrm{d}t\right]\right\}^{1/2}}, \quad (7)$$

where $B_{L,N}$ and $B_{R,N}$ represent the binaural signals of the N-th ERB channel. The range of $\Phi_{LR}(\tau)$ is from 0 to 1. This equation gives the maximum value of $\Phi_{LR}(\tau)$ in the case of $|\tau| \le 1$ ms, which represents the interaural cross-correlation coefficient (IACC). Lower IACC values (greater than 0) typically indicate a larger auditory source width, potentially resulting in an increased localization variability (SD of perceived azimuth) (Blauert, 1997; Morimoto, Iida, 1995).

4.2. Modified binaural localization cues

As described in Eq. (7), under anechoic conditions, $\tau = \tau_{\max}$ corresponding to this maximum value is defined by the ITD (Xie, 2013). Under reflective conditions, the interference between the reflected sound and the direct sound causes severe fluctuations in binaural factors with frequency variations (Kosmidis *et al.*, 2014; Tan *et al.*, 2023). This also leads to apparent multi-peak situations, where the ITD obtained from the peak corresponding to the maximum value usually has difficulty matching the actual perceived direction of the sound source. Therefore, we calculated the delay values corresponding to all peaks of the cross-correlation function and selected the one closest to the ITD value under anechoic conditions as the ITD in the reflective sound environment (i.e., choosing a reasonable ITD value) (Tollin, Henning, 1998).

The ILD is defined as

$$\mathrm{ILD}(f) = 20\log_{10}\left|\frac{P_R(f)}{P_L(f)}\right|, \quad (8)$$

where $P_R(f)$ and $P_L(f)$ represent the binaural sound pressures at frequency f.

Drawing on the auditory system's mechanism of amalgamating spatial cues across different frequency ranges, we calculated the average value and SD of the ITD below 1500 Hz (corresponding to ERB channels 1 to 21) and the ILD from 1.5 kHz to 10 kHz (corresponding to ERB channels 22 to 42). Moreover, the sensitivity of observers to the ITD in the frequency range centered around 700 Hz is widely recognized (Bilsen, 1973; Folkerts, Stecker, 2022; Zwislocki, Feldman, 1956); this frequency band is described as 'the dominance region'. Here, we set up an empirical frequency weighting function to simulate this phenomenon (Stern *et al.*, 1988). For frequencies below 1200 Hz, this function is fitted as a cubic polynomial, and for frequencies above 1200 Hz, the weight coefficients are equal to the value at 1200 Hz. The weighting function is shown in Fig. 8.

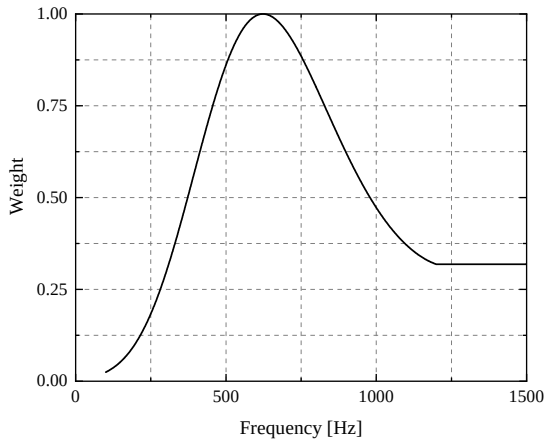

Fig. 8. Empirical frequency weighting function of ITD. The data were gathered by Raatgever (1980).

The weighted average ITDs under the different experimental conditions are shown in Fig. 9. Compared with the localization results of Figs. 4 and 7, the ITDs under these conditions exhibit analogous trends. First, regardless of the experimental conditions, the ITD increases with the target azimuth. Second, as shown in Fig. 9a, the ITD does not change with RT or the intensity of the reflections at most target azimuths. In Fig. 9b, with increasing distance, the delay of the reflection decreases, the intensity of the reflection increases, and the ITD of lateral target azimuths decreases. The above ITD trends generally align with the trends observed in localization results. However, in some cases, the ITD results do not match the localization results (e.g., at −30° under RT conditions of 0.8 s or 1.2 s). This discrepancy may be due to the general binaural auditory model not being applicable to all experimental conditions. Generally, even under larger RT conditions, ITD factors can provide relatively accurate localization information.

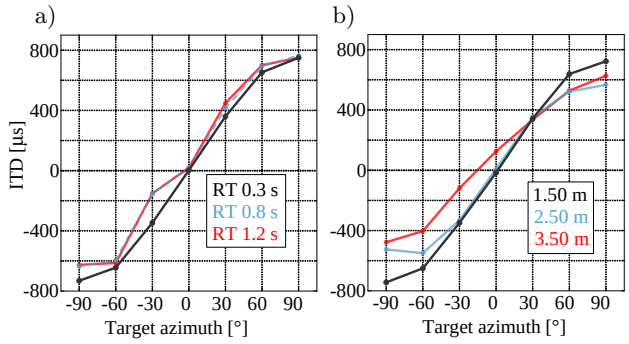

Fig. 9. Weighted average of ITDs under different: a) RTs (reflection intensities); b) distances (temporal structures).

The average ILDs at different azimuths under the different experimental conditions are shown in Fig. 10. The absolute values are significantly smaller in the higher reverberation condition than in the low reverberation condition. For example, the ILDs with RT = 0.3 s are larger than those for RT = 1.2 s or 0.8 s. This is because the late reverberant reflections can come from any direction, causing both ears to receive late reverberant energy of equal intensity. Consequently, ILD (absolute value) decreases towards zero as the DRR decreases, making it less reliable (Shinn-Cunningham *et al.*, 2005). A comparison between the results for the average ILDs and the localization results shows that there are almost no similar trends. This validates the unreliability of ILDs under low-DRR conditions.

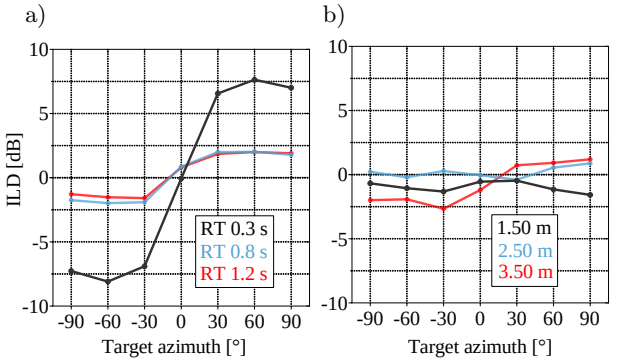

Fig. 10. Average ILD under different: a) RTs (reflection intensities); b) distances (temporal structures).

5. Discussion

5.1. Effects of reflection intensities (change RTs)

An increase in reflection intensities will decrease both the DRR and the ILD. The ILD cues (high-frequency cues) indicate that the perceived direction tends to be biased toward the front as the RT increases. However, the localization results in Sec. 3 show that the perceived azimuths are largely unaffected by changes in RT within the range of our experiments (i.e., 0.3 s–1.2 s). Similar findings have been observed in

the localization of real sound sources, where altering the RT (within the moderate reverberation range) alone did not markedly reduce the subjects' ability to localize sound sources (Rakerd, Hartmann, 2010; Rychtáriková *et al.*, 2009; 2011). This indicates that the ILD is not reliable under low-DRR conditions, as reported in a previous study (Shinn-Cunningham *et al.*, 2005). In contrast, the results in Fig. 9a demonstrate the robustness of the ITD against changes in reflection intensities (RT), which agrees with the localization results. Owing to the large distance from the loudspeaker to the wall, the earliest delay exceeds 11.40 ms (calculated based on geometric distance). In this situation, early reflections are largely suppressed by the precedence effect, resulting in little effect on localization cues. Furthermore, late reverberation adds uncorrelated signals with approximately equal amplitudes into two ears, which decreases the IACC but has little influence on the ITD. The IACC and ITD are calculated using the maximum peak value within a certain delay range of the cross-correlation function and the position at which this maximum peak value occurs, respectively. The cross-correlation function, i.e., Eq. (7), can be rewritten as

$$\Phi_{LR} = \frac{(B_{L,\mathrm{dir}} + B_{L,\mathrm{rev}}) \otimes (B_{R,\mathrm{dir}} + B_{R,\mathrm{rev}})}{|B_{L,\mathrm{dir}} + B_{L,\mathrm{rev}}||B_{R,\mathrm{dir}} + B_{R,\mathrm{rev}}|}, \qquad (9)$$

where $B_{L,\mathrm{dir}}$ and $B_{L,\mathrm{rev}}$ represent the direct sound and reverberation sound of the left impulse response, respectively, and similarly for the right impulse response. The symbol $\otimes$ denotes the correlation operation.

We hypothesize that the role of early reflection in localization is largely suppressed, and the late reverberation creates an ideal diffuse sound field. Hence, the correlation between direct and late reverberation sound, as well as the correlation with binaural late reverberation, is zero. Equation (11) can then be further simplified as

$$\Phi_{LR} = \frac{B_{L,\mathrm{dir}} \otimes B_{R,\mathrm{dir}}}{(B^2_{L,\mathrm{dir}} + B^2_{L,\mathrm{rev}})^{1/2}(B^2_{R,\mathrm{dir}} + B^2_{R,\mathrm{rev}})^{1/2}}. \qquad (10)$$

Considering our experimental conditions, the late reverberation increases with increasing RT, and so the denominator in Eq. (10) becomes larger. Moreover, the maximum peak value of the cross-correlation function decreases, indicating a decrease in the IACC (this implies a slight increase in the SD of the perceived azimuths with increasing RT). However, the position of the maximum peak remains unchanged, resulting in an unchanged ITD.

Based on the above analysis, the possible reason for the slight effect of the reflection intensities (RTs) on the localization of virtual sources are that listeners are more reliant on the ITD (low-frequency cues) than the ILD (high-frequency cues) for the localization in a reverberant environment. Previous studies have shown that subjects struggle to rely on the ITD for localization when stimuli lack transient information (Hartmann, 1983). Although the pink noise in our study was subjected to fade-in and fade-out processing, its localization does not differ significantly from other transient signals. This can be attributed to the fact that pink noise is composed of a series of small impulses, which have random amplitude fluctuations. These fluctuations are transient, meaning that the subjects are still able to utilize the ITD information within it for localization.

5.2. Effects of temporal structures of reflections (change loudspeaker distances)

For a virtual room with constant acoustic parameters, changes in loudspeaker distance will alter the temporal structure and intensity of the reflection. Under the condition of a 3.50 m loudspeaker distance, the minimum delay of the reflection is approximately 2.20 ms. This delay falls within the range where the precedence effect is actively suppressing (below 3 ms to 5 ms). At this point, the relatively high-energy early reflections are not completely suppressed by the precedence effect. A series of partially unsuppressed reflections interfere with the direct sound, causing the ITD to fluctuate with frequency. The average ITD changes with the loudspeaker distance (as shown in Fig. 9), and this interference also leads to a decrease in IACC (Gourevitch, Brette, 2012; Rakerd, Hartmann, 2010; Shinn-Cunningham, Kawakyu, 2003; Tan *et al.*, 2023). Moreover, the localization results in Sec. 3 demonstrate the same trend as the ITDs, that is, as the distance increases, the localization performance (including the SD and deviation of localization) of lateral virtual sources deteriorates. According to the auditory mechanism that merges locational data throughout various frequency bands (Hancock, Delgutte, 2004; Xia, Shinn-Cunningham, 2011), the degraded localization performance of the virtual source may arise from fluctuations in the ITD with frequency and the deviation of the mean ITD.

Variations in both the RT and loudspeaker distance change the DRR (as illustrated in Fig. 11), but only the loudspeaker distance affects the localization of virtual sources (as shown in Figs. 4 and 5). Even when the DRR is similar under different conditions, e.g., an RT of 1.2 s and loudspeaker distance of 3.5 m, there may be significant differences in the localization results. This indicates that the DRR alone may not adequately predict the localization performance in rooms. The temporal structure of reflections, i.e., the time distribution of reflection sequences, may indeed play a crucial role in the localization of the sound source. It is also reasonable to believe that reflections with small delay have a more disruptive effect on the localization of virtual sound sources compared with later reverberations with

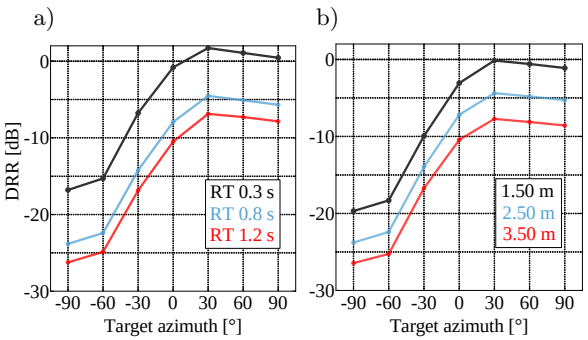

Fig. 11. DRR of the right ear under different: a) RTS (reflection intensities); b) distances (temporal structures). DRR is calculated as the ratio of the sound pressure levels between direct and reflected sound. As the DRR is obtained from the impulse response of the right ear, the DRR of the target azimuth on the right side (30° to 90°) is expected to be greater than that on the left side (−30° to −90°).

higher intensity. These findings provide the following guidance for CTC applications: even in rooms without acoustic decoration, placing the loudspeaker far enough from the wall (ensuring that the earliest delay is much longer than 1 ms) enables good localization performance of the CTC system.

6. Conclusions

This paper has investigated the influence of different temporal structures and intensity of reflections on the localization of virtual sources reproduced by a two-loudspeaker CTC system. The reasons for the variations in localization under different reflective conditions have also been revealed.

The principal conclusions derived from this study can be encapsulated in the following points:

- when the reproduction loudspeaker is located far from the wall (larger than 2 m in this work), in the RT variation range of our experiments (0.3 s to 1.2 s), the increase in the intensity of reflections does not significantly affect the localization performance of virtual sound sources due to the suppression of the precedence effect;
- when the reproduction loudspeaker distance increases (moving away from the listener and closer to the wall), the delay of early reflections decreases, and the temporal structure of the reflection changes. This results in a series of early reflections that are not fully suppressed interfering with the direct sound. This interference causes localization deviation and an increase in the degree of variation in the localization of lateral target angles of the virtual sound source;
- the DRR alone seems inadequate for determining the localization performance of virtual sources in reverberant environments. The temporal structure of reflections may play an important role in sound source localization. Compared to the late reverberation, early reflections with short delays (especially for that not fully suppressed by the precedence effect) have a greater impact on the localization of virtual sound sources;
- the average weighted ITD calculated based on the binaural auditory model accounting for the precedence effect can qualitatively explain the experimental results to some extent, but the average ILD does not.

It should be noted that, in this study, headphone-based binaural reproduction was adopted, and an acoustic simulation based on the ISM was employed. In reality, due to the material properties and geometric irregularities of room surfaces, complex absorption and diffuse reflection occur. While the ISM simplifies the modeling process and improves computational efficiency, it does not fully capture the acoustic response of real environments. Therefore, the results and conclusions presented in this study are limited to the specific experimental conditions (purely specular reflections in the room simulation and headphone reproduction) adopted herein.

Fundings

This research was funded by the National Natural Science Foundation of China with grant number 12074129 and 12474465, as well as by the Natural Science Foundation of Guangdong Province through grant number 2024A1515011446.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Authors' contributions

These authors made equal contributions to this work. All authors reviewed and approved the final manuscript.

Acknowledgments

We express our appreciation to all participants for their contribution to the study.

References

1. Allen J.B., Berkley D.A. (1979), Image method for efficiently simulating small-room acoustics, *The Journal of the Acoustical Society of America*, **65**: 943–950, https://doi.org/10.1121/1.382599.
2. Bahri K. (2019), *Loudspeaker directivity and playback environment in acoustic crosstalk cancelation*, Msc. Thesis, Charles University of Technology, Gothenburg.

3. Bilsen F.A. (1973), Spectral dominance in binaural lateralization, *Acustica*, **28**: 131–132.

4. Blauert J. (1997), *Spatial Hearing: The Psychophysics of Human Sound Localization*, 2nd. ed., The MIT Press, Harvard MA.

5. Braasch J. (2013), A precedence effect model to simulate localization dominance using an adaptive, stimulus parameter-based inhibition process, *The Journal of the Acoustical Society of America*, **134**: 420–435, https://doi.org/10.1121/1.4807829.

6. Brown A.D., Stecker G.C., Tollin D.J. (2015), The precedence effect in sound localization, *Journal of the Association for Research in Otolaryngology*, **16**: 1–28, https://doi.org/10.1007/s10162-014-0496-2.

7. Cooke M. (2005), *Modelling Auditory Processing and Organisation*, Cambridge University Press, London.

8. Folkerts M.L., Stecker G.C. (2022), Spectral weighting functions for lateralization and localization of complex sound, *The Journal of the Acoustical Society of America*, **151**: 3409–3425, https://doi.org/10.1121/10.0011469.

9. Gardner W.G. (1998), *3-D Audio Using Loudspeakers*, Springer Science & Business Media.

10. Genuit K. (1992), Standardization of binaural measurement technique, *Le Journal de Physique IV*, **2**: 405–407, https://doi.org/10.1051/jp4:1992187.

11. Gourevitch B., Brette R. (2012), The impact of early reflections on binaural cues, *The Journal of the Acoustical Society of America*, **132**: 9–27, https://doi.org/10.1121/1.4726052.

12. Habets E.A. (2010), *Room impulse response generator*, Technische Universiteit Eindhoven, Technical Report.

13. Hancock K.E., Delgutte B. (2004), A physiologically based model of interaural time difference discrimination, *Journal of Neuroscience*, **24**: 7110–7117, https://doi.org/10.1523/JNEUROSCI.0762-04.2004.

14. Hartmann W.M. (1983), Localization of sound in rooms, *The Journal of the Acoustical Society of America*, **74**: 1380–1391, https://doi.org/10.1121/1.390163.

15. Katz B.F. (2001), Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation, *The Journal of the Acoustical Society of America*, **110**: 2440–2448, https://doi.org/10.1121/1.1412440.

16. Kirkeby O., Nelson P.A. (1999), Digital filter design for inversion problems in sound reproduction, *Journal of the Audio Engineering Society*, **47**(7/8): 583–595.

17. Kosmidis D., Lacouture-Parodi Y., Habets E.A. (2014), The influence of low order reflections on the interaural time differences in crosstalk cancellation systems, [in:] *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2873–2877, https://doi.org/10.1109/ICASSP.2014.6854125.

18. Lehnert H., Blauert J. (1992), Principles of binaural room simulation, *Applied Acoustics*, **36**(3–4): 259–291, https://doi.org/10.1016/0003-682X(92)90049-X.

19. Lentz T. (2008), Binaural technology for virtual reality, *Journal of the Audio Engineering Society*, **124**(6): 3358–3359, https://doi.org/10.1121/1.3020604.

20. Li S., Peissig J. (2020), Measurement of head-related transfer functions: A review, *Applied Sciences*, **10**(14): 5014, https://doi.org/10.3390/app10145014.

21. Martin K.D. (1997), Echo suppression in a computational model of the precedence effect, [in:] *Proceedings of 1997 Workshop on Applications of Signal Processing to Audio and Acoustics*, https://doi.org/10.1109/ASPAA.1997.625622.

22. Masiero B., Fels J., Vorländer M. (2011), Review of the crosstalk cancellation filter technique, [in:] *Proceedings of ICSA 2011*.

23. Morimoto M., Iida K. (1995), A practical evaluation method of auditory source width in concert halls, *Journal of the Acoustical Society of Japan (E)*, **16**(2): 59–69, https://doi.org/10.1250/ast.16.59.

24. Møller H. (1992), Fundamentals of binaural technology, *Applied Acoustics*, **36**(3–4): 171–218, https://doi.org/10.1016/0003-682X(92)90046-U.

25. Nowoświat A., Olechowska M. (2022), Experimental validation of the model of reverberation time prediction in a room, *Buildings*, **12**(3): 347, https://doi.org/10.3390/buildings12030347.

26. Ocheltree K.B., Frizzel L.A. (1989), Sound field calculation for rectangular sources, *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, **36**(2): 242–248, https://doi.org/10.1109/58.19157.

27. Rakerd B., Hartmann W.M. (2010), Localization of sound in rooms. V. Binaural coherence and human sensitivity to interaural time differences in noise, *The Journal of the Acoustical Society of America*, **128**(5): 3052–3063, https://doi.org/10.1121/1.3493447.

28. Raatgever J. (1980), *On the binaural processing of stimuli with different interaural phase relations*, Ph.D. Thesis, Technische Universiteit Delft, Netherlands.

29. Rui Y., Yu G., Xie B., Liu Y. (2013), *Calculation of individualized near-field head-related transfer function database using boundary element method*, Presented at the Audio Engineering Society Convention, paper 8901.

30. Rychtáriková M., van den Bogaert T., Vermeir G., Wouters J. (2009), Binaural sound source localization in real and virtual rooms, *Journal of the Audio Engineering Society*, **57**: 205–220.

31. Rychtáriková M., van den Bogaert T., Vermeir G., Wouters J. (2011), Perceptual validation

of virtual room acoustics: Sound localisation and speech understanding, *Applied Acoustics*, **72**(4): 196–204, https://doi.org/10.1016/j.apacoust.2010.11.012.

32. Sæbø A. (1999), *Effect of early reflections in binaural systems with loudspeaker reproduction*, Presented at the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp. 552–556, New York.

33. Sæbø A. (2001), *Influence of reflections on crosstalk cancelled playback of binaural sound*, Ph.D. Thesis, Norwegian University of Science and Technology, Trondheim.

34. Shinn-Cunningham B., Kawakyu K. (2003), Neural representation of source direction in reverberant space, [in:] *2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (IEEE Cat. No.03TH8684)*, pp. 79–82, https://doi.org/10.1109/ASPAA.2003.1285824.

35. Shinn-Cunningham B.G., Kopco N., Martin T.J. (2005), Localizing nearby sound sources in a classroom: Binaural room impulse responses, *The Journal of the Acoustical Society of America*, **117**(5): 3100–3115, https://doi.org/10.1121/1.1872572.

36. Slaney M. (1993), *An efficient implementation of the Patterson–Holdsworth auditory filter bank*, Apple Computer Technical Report #35, Perception Group – Advanced Technology Group.

37. Stern R.M., Zeiberg A.S., Trahiotis C. (1988), Lateralization of complex binaural stimuli: A weighted-image model, *The Journal of the Acoustical Society of America*, **84**(1): 156–165, https://doi.org/10.1121/1.396982.

38. Tan W., Yu G., Rao D. (2023), Influence of first-order lateral reflections on the localization of virtual source reproduced by crosstalk cancellation system, *Applied Acoustics*, **202**: 109165, https://doi.org/10.1016/j.apacoust.2022.109165.

39. Tollin D.J., Henning G.B. (1998), Some aspects of the lateralization of echoed sound in man. I. The classical interaural-delay based precedence effect, *The Journal of the Acoustical Society of America*, **104**(5): 3030–3038, https://doi.org/10.1121/1.423884.

40. Villegas J. (2015), Locating virtual sound sources at arbitrary distances in real-time binaural reproduction, *Virtual Reality*, **19**: 201–212, https://doi.org/10.1007/s10055-015-0278-0.

41. Visentin C., Prodi N., Valeau V., Picaut J. (2013), A numerical and experimental validation of the room acoustics diffusion theory inside long rooms, [in:] *Proceedings of Meetings on Acoustics*, **19**(1): 015024, https://doi.org/10.1121/1.4798976.

42. Xia J., Shinn-Cunningham B. (2011), Isolating mechanisms that influence measures of the precedence effect: Theoretical predictions and behavioral tests, *The Journal of the Acoustical Society of America*, **130**(2): 866–882, https://doi.org/10.1121/1.3605549.

43. Xie B. (2013), *Head-Related Transfer Function and Virtual Auditory Display*, 2nd ed., J. Ross Publishing.

44. Yost W.A., Gourevitch G. (1987), *Directional Hearing*, Springer.

45. Ziegelwanger H., Kreuzer W., Majdak P. (2015), *Mesh2HRTF: An open-source software package for the numerical calculation of head-related transfer functions*, Presented at the 22nd International Congress on Sound and Vibration, https://doi.org/10.13140/RG.2.1.1707.1128.

46. Zwislocki J., Feldman R.S. (1956), Just noticeable differences in dichotic phase, *The Journal of the Acoustical Society of America*, **28**(5): 860–864, https://doi.org/10.1121/1.1908495.