Audio Engineering Society

Conference Paper

Presented at the 2022 International Conference on
Audio for Virtual and Augmented Reality
2022 August 15–17, Redmond, WA, USA

The influence of acoustic cues in early reflections on source localization

Song Li[1] and Jianyuan Feng[2]

[1]*Agora Inc., Hangzhou, China*
[2]*Agora Inc., Shanghai, China*

Correspondence should be addressed to Song Li (lisong@agora.io)

ABSTRACT

The image source method (ISM) is a widely applied approach for modelling early reflections in binaural rendering systems. Theoretically, each image source should be filtered by a pair of head-related transfer functions (HRTFs) to simulate its directional characteristic. However, from the perspective of perceived localization, it is unclear whether the complete acoustic cues need to be considered when modelling early reflections.

In this study, early reflections up to 2^{nd} order were generated with the ISM, and different monaural and binaural spectral information was removed from them to investigate the role of acoustic cues in early reflections on source localization. The results of the listening experiment showed that the 1^{st} order early reflections should be "correctly" simulated, and the importance of acoustic cues in early reflections reduces for nearby sound sources. Additionally, different acoustic cues related to sound localization were extracted and compared with the subjective results.

1 Introduction

Binaural technology plays an important role in the audio metaverse, providing listeners with an impression of being in a three-dimensional (3D) audio scene when listening with headphones [1].

Filtering a dry monaural signal by a pair of head-related transfer functions (HRTFs) is a common method for modelling the direction of arrival (DOA) of a virtual sound source [2]. The HRTF is highly individual, and is a directional function in the far-field. In the near-field, the HRTF depends not only on the direction, but also on the distance of the source from the listener. Head-related impulse responses (HRIRs) are the time domain representation of HRTFs. Typically, generic HRTFs from officially provided databases are used for binaural rendering applications, as accurate individual HRTFs are difficult to obtain, especially in consumer scenarios [3].

A virtual sound generated by HRTFs is the direct sound component from the source to the listener. To enhance the immersive experience when listening in a reverberant environment, reverberation should be added to the virtual sound [4, 5]. Reverberation can temporally be divided into early reflections and late reverberation. The density of reflections and the diffuseness of the decaying sound field increase with time. Early reflections, consisting of several distinct reflections from walls, ceilings, floors, and obstacles in the room, etc., can be observed within a few tens of milliseconds af-

ter the arrival of direct sound. They have an influence on source localization, source coloration, source width and speech intelligibility, etc. [6]. Late reverberation consists of high-density reflections, and contributes to the listener envelopment. The boundary point between the early reflections and the late reverberation is known as the mixing time [7].

The combination of the direct sound component and reverberation is actually the binaural room impulse response (BRIR), which can be measured or synthesized. Measuring BRIR sets for different head orientations and source-listener positions is time-consuming and storing high-density BRIR sets for each room requires a large amount of memory. Therefore, synthesizing BRIRs based on room information and HRTFs would benefit the applications of real-time binaural rendering.

Different approaches have been proposed to synthesize reverberation [8–13]. A fast and widely used approach is by using the image source method (ISM) [8] and the feedback delay network (FDN) [13] to simulate early reflections and late reverberation tails, respectively [14, 15]. For the early reflection part, each reflected sound source (image source) is treated as a virtual sound source that should theoretically be filtered by a pair of HRTFs. The number of image sources is usually limited up to 2^{nd} order, as the cost of rendering higher orders grows exponentially.

Hassager et al. [16] reported that the reduction in spectral information in the reverberant part had no significant effect on perceived externalization. However, in their study, the result was limited to a fixed distance test with visual cues presented. Catic et al. [17] and Leclère et al. [18] suggested that the binaural information contained in the reverberation is important for sound externalization. However, biases in the DOA were not reported in those studies. Variations in the acoustic cues contained in the reflections may affect not only the perceived distance but also the DOA. Both factors (DOA and distance perception) are critical to headphone-based virtual sounds, and deviations in each of them may lead to degraded immersive experience in virtual acoustic environments. Simulating source directivity in the 1^{st} order early reflections was sufficient without affecting the perception of the source orientation [19]. It is interesting to know whether this result holds for perceived sound localization as well.

In this study, we applied the ISM to simulate early reflections, and removed different monaural and binaural spectral information of 1^{st} and 2^{nd} order image sources to investigate the role of acoustic cues contained in early reflections on source localization. In addition, different acoustic cues related to sound localization, i.e., interaural level difference (ILD), spectral gradients (SGs), interaural coherence (IC), and direct-to-reverberant energy ratio (DRR), were extracted and compared with the subjective results. Better knowledge of the perceptual importance of acoustic cues contained in early reflections might help to simplify the acoustic simulation. It should be noted that the localization in this study includes both DOA and perceived distance.

2 Methods

2.1 Binaural rendering system

Figure 1 shows the block diagram of a binaural rendering system that we developed and used in this study.

The direct sound is attenuated according to the inverse-square law, and filtered by a distance-dependent high-shelving filter to simulate the effect of air absorption [20]. If the sound source is not omnidirectional, the orientation of the source can be modelled according to its directivity pattern. In the case of a rectangular room, the position of each image source $(p_x, p_y, p_z)^{\mathrm{T}}$ (Cartesian coordinate system) is calculated according to the geometry of the room and the position of the source [21]:

$$\begin{pmatrix} p_x \\ p_y \\ p_z \end{pmatrix} = \begin{pmatrix} (1-2u)s_x + 2nL_x \\ (1-2v)s_y + 2lL_y \\ (1-2w)s_z + 2mL_z \end{pmatrix}, \qquad (1)$$

where $\{L_x, L_y, L_z\}$ and $\{s_x, s_y, s_z\}$ represent the room dimensions and coordinates of the sound source in the room, respectively. $\{u, v, w\}$ and $\{n, l, m\}$ are integer vector triplets, where $u/v/w$ are restricted between 0 and 1, and the possible values of $n/l/m$ are based on the order of early reflections [21].

Each single early reflection is treated as a virtual sound source located at a different location. Thus, the effects of sound attenuation, air absorption and directivity are simulated for each reflection. In addition, low-pass filters with different cut-off frequencies and damping coefficients are used to simulate the absorption of walls, ceilings, floors, etc. The direct sound and early reflections are then filtered through the corresponding HRTFs with desired directions. In the binaural rendering system, the HRTFs are represented as minimum-phase

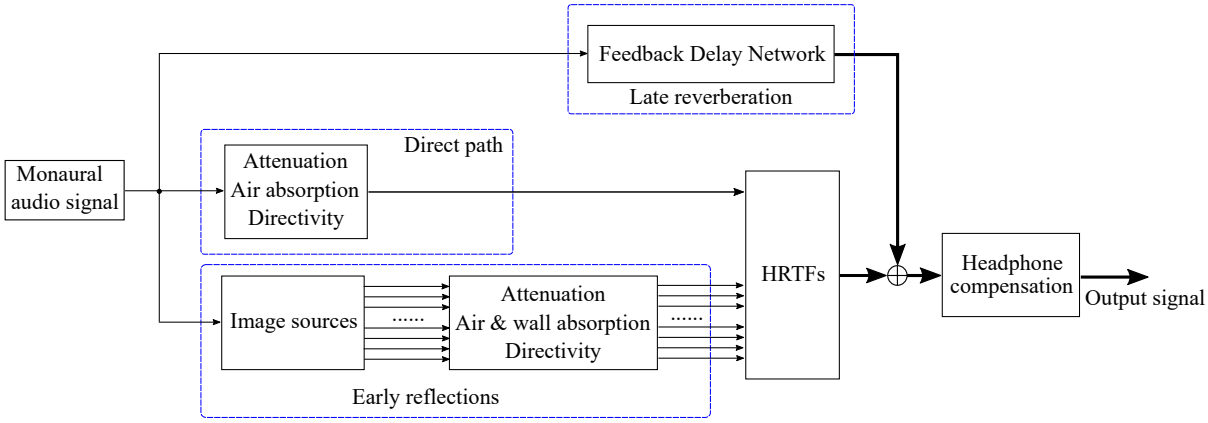

Fig. 1: Block diagram of binaural rendering system. Thin and bold lines represent mono and binaural signals, respectively.

systems with pure delays [15]. Because of the high-resolution of HRTFs (see Section 3), the update of the minimum-phase components was done by switching the HRTFs. The current and previous filters were cross-faded to eliminate audible artifices. The delays between the left and right channels represented the interaural time difference (ITD), and the changes in pure delays were achieved by linear interpolation of the tapped delay lines. The late reverberation tail is generated by using an FDN-based reverberator [13] with 16 internal feedback channels and decorrelated stereo outputs. The level of late reverberation was determined by the average level of early reflections around the mixing time. The frequency-dependent reverberation time was set according to a real conference room (see Section 3). The decorrelation of the late reverberation output was intended to enhance the spaciousness of the virtual sound.

The direct sound, early reflections and the late reverberation are then summed, and filtered through a pair of headphone compensation transfer functions (HcTFs) to compensate for the effects of the headphones [22].

2.2 Modification of early reflections

The 1^{st} and 2^{nd} order early reflections generated with the ISM were modified according to Table 1.

The first condition was the reference for this study, where the 1^{st} and 2^{nd} order early reflections were "correctly" modelled, i.e., filtered by the corresponding HRTFs ("2^{nd} HRTF" condition).

Table 1: Parameter setting for early reflections.

Condition	1^{st} order	2^{nd} order	Notation
1	HRTF	HRTF	2^{nd} HRTF
2	HRTF	/	1^{st} HRTF
3	Mono gain	Mono gain	2^{nd} Mono
4	ILD + ITD	ILD + ITD	2^{nd} Binaural
5	HRTF	Mono gain	HRTF+Mono
6	HRTF	ILD + ITD	HRTF+Binaural

In the second condition, the 2^{nd} order early reflections were not simulated, only the 1^{st} order early reflections were "correctly" generated ("1^{st} HRTF" condition).

In the third condition, a frequency-independent gain was used instead of applying a pair of HRTFs for each early reflection. The gain factor was calculated as the average of the maximum absolute values of left and right ear HRIRs ("2^{nd} Mono" condition). By this means, neither spectral information nor binaural cues were contained in the early reflections.

In the fourth condition, a pair of delayed impulses was applied for each early reflection, where the difference in delays between the left and right impulses was the ITD extracted from the HRIRs [23], and the gains of the impulses were chosen as the maximum absolute values of left and right ear HRIRs ("2^{nd} Binaural" condition). The level difference between the left and right impulses could be treated as the ILD. In this way, only the broadband binaural information (ITD and ILD) was contained in the 1^{st} and 2^{nd} order early reflections.

In the fifth condition, the 1^{st} order early reflections were "correctly" modelled, while frequency-

independent gain factors were applied for the 2^{nd} order early reflections ("HRTF+Mono" condition).

In the last condition, the 2^{nd} order early reflections contained only broadband binaural information ("HRTF+Binaural" condition).

Note that the effects of distance-related attenuation, air and wall absorption, and directivity for early reflections were simulated for all conditions. The late reverberation was calculated from the reference and was the same for all experimental conditions. The HRTFs used in this study were taken from the THK HRTF database for the dummy head Neumann KU 100, which has a spatial resolution of 2° [22].

3 Experimental setup

Figure 2 shows the geometry of the simulated virtual room (length × width × height: 7 m × 4 m × 3.2 m) with the locations of the listener and three omnidirectional virtual sound images (top view, the height information is not shown in the figure). The heights of sound sources and the listener were set to 1.1 m. Two virtual sound sources were far away from the listener (blue and red), and one was close to the listener (black). The blue sound source was in the frontal direction, while the blue and black sound sources were located in lateral positions relative to the listener. It should be noted that the three virtual sound sources were all in the far-field (source-listener distance is larger than 1 m).

The size of the virtual room is the same as a real conference room in our office. The parameters for the early reflection and late reverberation models were set according to the properties of the real room. The reverberation time of the room is about 0.5 s at 1 kHz. Three virtual sound sources with different conditions (see Section 2.2) were generated for the subjective listening test.

A multiple stimuli with hidden reference and anchor (MUSHRA) experiment was designed to assess the influence of acoustic cues in early reflections on perceived sound localization. Three types of audio content were presented in the test, namely music (8 s), speech (3 s), and pulsed pink noise (duration of 0.5 s with a 0.3 s-long pause and 0.02 s fade-in and fade-out) [24]. The music (Track ID 70, pop music) and speech (Track ID 50, male speech) signals were taken from the EBU Sound Quality Assessment Material

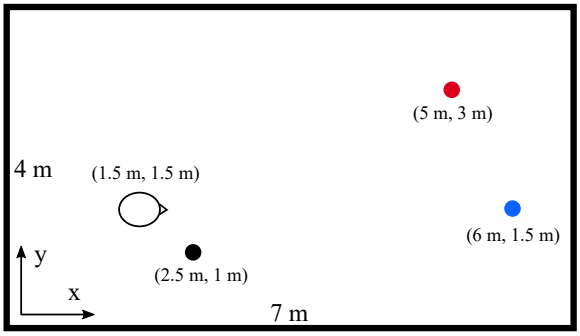

Fig. 2: The geometry of the simulated room with the positions of the listener and three virtual sound sources (top view).

(SQAM) [25]. Each test signal was played back in a loop.

The listening test comprised 18 sessions (three positions × three types of stimuli × two presentations/subject), each with six test signals including a hidden reference to be rated. An explicit anchor signal was not included in this test. In total, 108 stimuli should be rated. Eight subjects with normal hearing (one female, seven males, aged between 28 and 35), participated in the listening experiment. Five of them conducted the test in situ in the conference room, while the other three subjects performed the experiment remotely. No visual reference of the sound source location was provided to the participants. The listening experiment was developed in Python and shared with the remote participants. The local subjects listened to stimuli presented by a pair of Sennheiser HD650 headphones. The effect of headphones was compensated by applying a pair of HcTFs from the THK database [22]. The remote participants used their own headphones with a pair of corresponding HcTFs from the THK database [22]. They were asked to perform the test in a quiet environment. The reverberation of virtual sounds was designed based on the actual acoustics of the real conference room. Since the experiment aimed to detect the relative difference in sound localization between the reference and the test signals, the experimental results were not expected to be influenced by different listening environments.

At the beginning, subjects were instructed to familiarize themselves with the graphical user interface (GUI), the meaning of rating scale and the test audio signals.

Additionally, listeners were asked to choose an appropriate presentation level for the experiment. During the listening test, subjects were unable to adjust the sound level, and they need to rate the relative differences in perceived localization between the reference (the first condition in Table 1) and the test signals using a slider from 0 to 100 with a step-size of 1. A rating of 0 means no difference in perceived localization between the test audio signal and the reference, and a value upwards close to 100 denotes a large difference from the reference signal in terms of perceived localization. No head movement was allowed during the listening test, since no head tracking was available.

4 Results

The normal distribution of the results for all conditions was not confirmed by the Shapiro-Wilk test [26]. Therefore, non-parametric statistical analysis were applied to analyze the subjective data. The Friedman test was used for analyzing the significant effect of the test conditions on the results. In addition, the Wilcoxon test was performed to analyze significant differences in experimental results between pair of conditions [26].

Figure 3 shows the median differences in perceived localization with non-parametric 95 % confidence intervals (CIs). The top, middle and bottom panels display the results for music, speech and noise signals, respectively. In each panel, the circles with blue error bars represent the results for the distant frontal sound source. The squares with red error bars and the diamonds with black error bars show the results for distant and near lateral sound sources, respectively.

The Friedman test shows that the experimental conditions have a significant main effect on localization perception, especially for sound sources with large distances ($p \ll 0.01$). The p value for distant sources are much smaller than for the close source.

Large biases in sound localization can be observed for "1^{st} HRTF", "2^{nd} Mono", and "2^{nd} Binaural" conditions. In these three conditions, the biases are significantly lower for the close sound source compared to distant sound sources ($p \ll 0.01$), except for the lateral distant music and speech sound sources under the "1^{st} HRTF" condition ($p \approx 0.5$).

For distant sound sources, the monaural and binaural spectral information in the 1^{st} order early reflections is important for sound localization. A large bias in sound

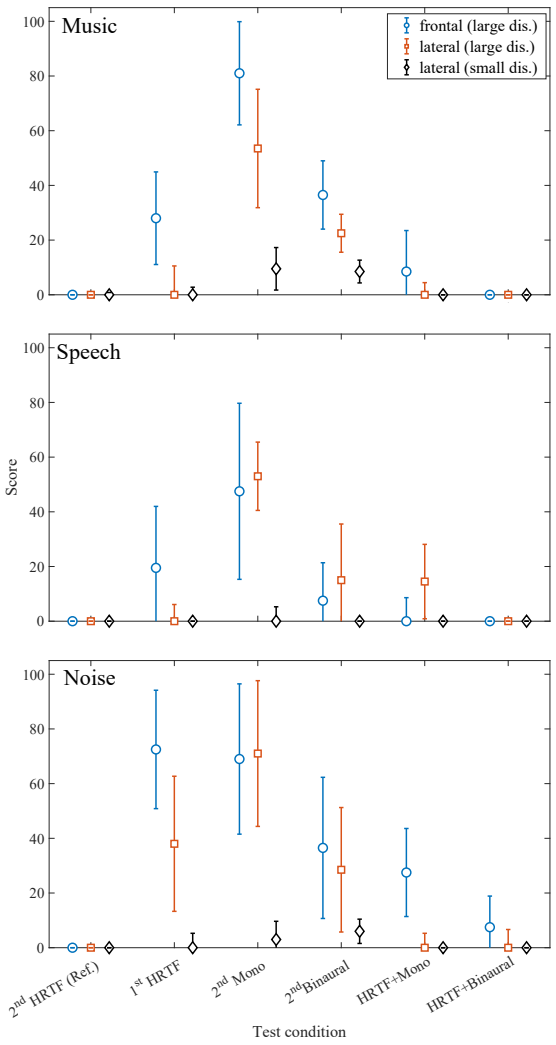

Fig. 3: Median differences in perceived localization with non-parametric 95 % confidence intervals (CIs) for different sound sources. A score of 0 and 100 means no difference and a large difference between the test audio signal and the reference signal, respectively.

localization can be observed when the 1^{st} order early reflections do not contain binaural cues and spectral information ("2^{nd} Mono" condition). When the 1^{st} order early reflections contain only binaural information (broadband ILD and ITD), a significant bias in localization can still be observed but less than that under the "2^{nd} Mono" condition.

If only the 1^{st} order early reflections are "correctly"

simulated ("1^{st} HRTF" condition), a large bias in source localization can be seen for the noise signal (frontal and lateral directions). In the case of speech and music signals, obvious biases can be found only for the distant frontal sound source, but significantly less than for the noise signal ($p < 0.02$), indicating that the type of sound source has an important influence on perceived sound localization.

The bias in sound localization reduces when the 2^{nd} order early reflections are simulated and the 1^{st} order early reflections are "correctly" modelled. Subjects could barely perceive biases in sound localization when the 1^{st} order early reflections are "correctly" simulated and the 2^{nd} order early reflections contain broadband binaural information ("HRTF+Binaural" condition).

In the case of the close sound source, clear biases can only be observed for music and noise signals under "2^{nd} Mono" and "2^{nd} Binaural" conditions, but significantly smaller than for distant sources ($p < 0.007$). It seems that the monaural and binaural spectral information contained in the early reflections has less influence on perceived localization for close sources than for distant sources.

5 Discussion

The combination of early reflections and artificial late reverberation is a beneficial and useful solution to generate reverberant parts for virtual sounds [19]. This study aimed to investigate whether the monaural and binaural spectral information contained in early reflection parts could be partially removed without affecting the sound localization (DOA and distance).

5.1 Analysis of subjective experimental results

Virtual sources located at three different positions (see Figure 2) were simulated with modified early reflections (see Table 1). To generate distant sound sources with desired positions, simulating only the 1^{st} order early reflections is not sufficient, especially for the noise signal.

The 2^{nd} order early reflections are required, but they need not contain the full acoustic cues as long as the 1^{st} order reflections are "correctly" simulated. In addition, the median bias in localization is lower when the early reflections contain broadband binaural information compared to broadband monaural information (see "2^{nd} Mono" vs. "2^{nd} Binaural" condition, and "HRTF+Mono" vs. "HRTF+Binaural" condition). This result corresponds well to the outcomes from [17], which reported that the binaural information contained in reverberation is important to externalization. The potential deviations in externalization lead to biases in perceived sound localization. A similar result can be found for perceiving the source orientation. Steffens et al. [19] reported that it was enough to simulate effects of source orientation in the 1^{st} order early reflections, while a fixed and average source-directivity filter could be applied for higher-order early reflections without affecting the perceived source orientation.

For a close sound source, a slight bias in perceived localization can be observed in "2^{nd} Mono" and "2^{nd} Binaural" conditions. This means that the monaural and binaural spectral information in the 1^{st} order early reflections is still important to perceived localization of close sound sources [27], but not as important as for distant sources. The results can be explained by the effect that for a near sound source, the energy of direct sound dominates, leading to a large DRR, and the monaural and binaural spectral information in early reflections can not be well perceived due to the masking effect. This outcome may explain the experimental results from [16], which showed that the spectral information contained in reverberation did not play an important role on externalization. In their study, frontal and lateral sound sources with a distance of 1.5 m were tested, and the result might changed with larger source-listener distances.

Based on the outcomes of this study, the early reflection model in real-time binaural rendering systems can be designed as distance-dependent, and the simulation of 2^{nd} order early reflections can be partially simplified without affecting the sound localization.

5.2 Objective analysis of acoustic cues contained in synthesized BRIRs

Four acoustic cues related to source localization (DOA and perceived distance), namely the ILDs, SGs, IC and DRR [28, 29], were calculated for comparison with subjective data under different experimental conditions. ILDs, SGs and IC are not only crucial for the DOA of sound sources [28, 30], but also for the perceived distance and externalization [17, 29]. DRR is known as a primary cue for perceived distance of sound sources in

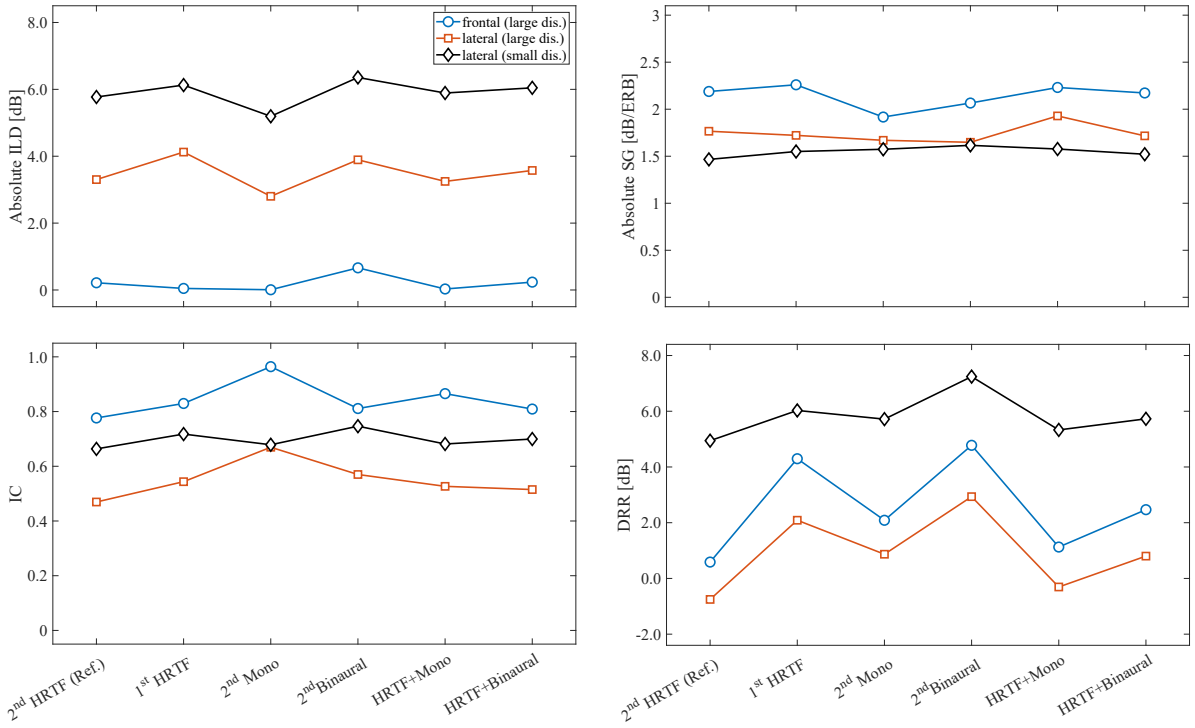

Fig. 4: Absolute ILDs (top left), absolute SGs (top right), IC (bottom left) and DRR (bottom right) for different experimental conditions and source positions.

reverberant environments [31]. Deviations in these four cues may lead to biases in perceived sound localization.

To calculate ILDs and SGs, BRIRs synthesized under different conditions were processed through an auditory periphery model, comprising a 4^{th} order gammatone filter bank [32] with a bandwidth of one Equivalent Rectangular Bandwidth (ERB) [33] and an inner hair cell model [34] (a half-wave rectifier followed by a 1^{st} order low-pass filter with a cut-off frequency of 1 kHz). The excitation patterns were calculated from the logarithm of the root mean square (RMS) energy in every frequency channel.

ILDs were calculated as the differences in the excitation patterns between the left and right ears. The average ILDs across frequency channels were used for the analysis. SGs for each ear were first calculated as the excitation differences between adjacent frequency bands, and then averaged over frequencies. The SGs for the left and right ear were further averaged for the analysis under different experimental conditions.

IC indicates the degree of coherence between the left and right ear signals, and is expressed as the maximum of the absolute value of the normalized interaural cross-correlation function [35]. DRR is defined as the ratio of direct and reverberant sound energy. The DRRs of the synthesized BRIRs were first calculated for each ear, and then averaged between the left and right ear. The broadband IC and DRR were used for the analysis in this study.

Figure 4 shows absolute ILDs, absolute SGs, IC, and DRR for different experimental conditions and source positions.

For the distant frontal sound source (circles), the overall ILD is close to zero. The variation in ILDs for different conditions is not obvious and is not well reflected in the subjective results. The experimental data can be partly explained by the deviations of SGs and IC. Clear deviations in SGs occur when no monaural and binaural spectral information ("2^{nd} Mono" condition) or only binaural information ("2^{nd} Binaural" condition) is contained in early reflections. In the case of IC, a noticeable increase in the IC values can be found in the "2^{nd} Mono" condition, but the values are similar

in the other conditions. High DRR deviations can be observed in "1^{st} HRTF" and "2^{nd} Binaural" conditions, which are in good agreement with the experimental results, but the DRR values can not explain the subjective data for the "2^{nd} Mono" condition.

For the distant lateral sound source (squares), the deviations in ILDs are partially reflected in the subjective data, although they are small across different conditions. The maximum bias in SGs can be found in the "HRTF+Mono" condition, but it does not match the experimental results. The increase in IC values across different conditions corresponds well to the subjective data, i.e., the highest deviation from IC can be observed in the "2^{nd} Mono" condition; high deviations from IC occur in "1^{st} HRTF" and "2^{nd} Binaural" conditions, but not as large as in the "2^{nd} Mono" condition. As with the distant frontal source, the DRR values can only partially explain the experimental results.

In the case of the close lateral sound source (diamonds), the deviations of the acoustic cues are not pronounced across experimental conditions. Relatively large biases from several acoustic cues can be found in "2^{nd} Mono" and "2^{nd} Binaural" conditions, which are well reflected in the subjective data. As with the distant sources, the bias of a single acoustic cue can not explain the subjective results for all conditions.

It can be seen that the variation in IC values can explain the subjective data under most conditions. However, a single acoustic cue is not sufficient to explain the experimental results for all conditions and sound sources with different positions. Each acoustic cue can only explain the subjective results under several conditions, suggesting that predicting the sound localization requires a combination of different acoustic cues. There exists a number of computational models that predict the DOA [28] or perceived externalization [16, 36] of virtual sound sources. They can be further extended to predict sound localization by combining different acoustic cues calculated in this study.

5.3 Limitations

This study tested the influence of acoustic cues in early reflections on source localization on the horizontal plane, including DOA and perceived distance. These two acoustic attributes were not assessed separately, as the deviations in each of them could lead to degraded immersive experience in acoustic environments. It would be interesting to investigate the influence of acoustic cues in early reflections on DOA and perceived distance separately. In addition, different virtual rooms, more source positions (azimuth and elevation positions), and more participants should be considered for listening experiments.

Further, the baseline (reference) used in this study may also produce biases in perceived localization compared to the real target positions, since non-individual HRTFs were used in the binaural rendering. In the experiment, the task for subjects was not to assess the absolute positions of the test signals, the outcome of this study is therefore valid. To further test the performance of the binaural rendering system with different settings in the early reflection model, individual HRTFs should be used.

Moreover, this study investigated the perceived localization of static binaural sounds without considering the dynamic conditions. Further experiments need to consider head and source movements as they may influence perceived externalization [37] and DOA [38] of virtual sounds.

Finally, the estimation of acoustic cues was performed by analyzing the synthesized BRIRs. Further analysis should take into account the sound characteristics, as the type of sound showed a significant effect on several conditions in the experiment.

6 Summary

This study evaluated the role of acoustic cues contained in early reflections on source localization in a reverberant room. Early reflections were simulated based on the ISM with different parameter settings. The experimental results suggested that the 1^{st} order early reflections should be "correctly" simulated, while the 2^{nd} order early reflections do not require complete monaural and binaural spectral information. The importance of the acoustic cues contained in early reflections reduces with decreasing sound distance in terms of perceived sound localization.

Four acoustic cues related to sound localization, i.e., ILDs, SGs, IC and DRR, were calculated and compared with the subjective results. Among them, the variation in IC values corresponds well to experimental data for most conditions. However, a single acoustic cue can not explain the experimental results for all conditions.

Future work is to study the effect of monaural and binaural spectral information in early reflections on

DOA and perceived distance, separately. Additionally, different virtual rooms, dynamic scenarios, and more participants should be taken into account. Moreover, a computational model to predict the localization of reverberant sound sources needs to be developed by combining different acoustic cues related to source DOA and perceived distance.

References

[1] Jot, J.-M., Audfray, R., Hertensteiner, M., and Schmidt, B., "Rendering Spatial Sound for Interoperable Experiences in the Audio Metaverse," in *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, pp. 1–15, IEEE, 2021.

[2] Wightman, F. L. and Kistler, D. J., "Headphone simulation of free-field listening. I: stimulus synthesis," *The Journal of the Acoustical Society of America*, 85(2), pp. 858–867, 1989.

[3] Li, S. and Peissig, J., "Measurement of head-related transfer functions: A review," *Applied Sciences*, 10(14), p. 5014, 2020.

[4] Best, V., Baumgartner, R., Lavandier, M., Majdak, P., and Kopčo, N., "Sound externalization: A review of recent research," *Trends in Hearing*, 24, p. 2331216520948390, 2020.

[5] Li, S., Schlieper, R., Tobbala, A., and Peissig, J., "The Influence of Binaural Room Impulse Responses on Externalization in Virtual Reality Scenarios," *Applied Sciences*, 11(21), p. 10198, 2021.

[6] Howard, D. and Angus, J., *Acoustics and psychoacoustics*, 5th Edition, Routledge, New York, 2017, ISBN 1317508297.

[7] Lindau, A., Kosanke, L., and Weinzierl, S., "Perceptual evaluation of model-and signal-based predictors of the mixing time in binaural room impulse responses," *Journal of the Audio Engineering Society*, 60(11), pp. 887–898, 2012.

[8] Allen, J. B. and Berkley, D. A., "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, 65(4), pp. 943–950, 1979.

[9] Krokstad, A., Strom, S., and Sørsdal, S., "Calculating the acoustical room response by the use of a ray tracing technique," *Journal of Sound and Vibration*, 8(1), pp. 118–125, 1968.

[10] Schroeder, M. R., "Natural sounding artificial reverberation," in *13th AES Annual Meeting*, Audio Engineering Society, 1961.

[11] Moorer, J. A., "About this reverberation business," *Computer music journal*, pp. 13–28, 1979.

[12] Gardner, W. G., "A realtime multichannel room simulator," *J. Acoust. Soc. Am*, 92(4), p. 2395, 1992.

[13] Jot, J.-M. and Chaigne, A., "Digital delay networks for designing artificial reverberators," in *90th AES Convention*, Audio Engineering Society, 1991.

[14] Wendt, T., Van De Par, S., and Ewert, S. D., "A computationally-efficient and perceptually-plausible algorithm for binaural room impulse response simulation," *Journal of the Audio Engineering Society*, 62(11), pp. 748–766, 2014.

[15] Li, S., Schlieper, R., and Peissig, J., "Externalization Enhancement for Headphone-Reproduced Virtual Frontal and Rear Sound Images," in *AES International Conference on Headphone Technology*, Audio Engineering Society, 2019.

[16] Hassager, H. G., Gran, F., and Dau, T., "The role of spectral detail in the binaural transfer function on perceived externalization in a reverberant environment," *The Journal of the Acoustical Society of America*, 139(5), pp. 2992–3000, 2016.

[17] Catic, J., Santurette, S., and Dau, T., "The role of reverberation-related binaural cues in the externalization of speech," *The Journal of the Acoustical Society of America*, 138(2), pp. 1154–1167, 2015.

[18] Leclère, T., Lavandier, M., and Perrin, F., "On the externalization of sound sources with headphones without reference to a real source," *The Journal of the Acoustical Society of America*, 146(4), pp. 2309–2320, 2019.

[19] Steffens, H., van de Par, S., and Ewert, S. D., "The role of early and late reflections on perception of source orientation," *The Journal of the Acoustical Society of America*, 149(4), pp. 2255–2269, 2021.

[20] Kronland-Martinet, R. and Voinier, T., "Real-time perceptual simulation of moving sources: application to the Leslie cabinet and 3D sound immersion," *EURASIP Journal on Audio, Speech, and Music Processing*, 2008, pp. 1–10, 2008.

[21] Brinkmann, F., Erbes, V., and Weinzierl, S., "Extending the closed form image source model for source directivity," in *Fortschritte der Akustik - DAGA*, 2018.

[22] Bernschütz, B., "A spherical far field HRIR/HRTF compilation of the Neumann KU 100," in *Proceedings of the 40th Italian (AIA) annual conference on acoustics and the 39th German annual conference on acoustics (DAGA) conference on acoustics*, AIA/DAGA Merano, 2013.

[23] Lindau, A., Estrella, J., and Weinzierl, S., "Individualization of dynamic binaural synthesis by real time manipulation of ITD," in *128th AES Convention*, Audio Engineering Society, 2010.

[24] Pelzer, R., Dinakaran, M., Brinkmann, F., Lepa, S., Grosche, P., and Weinzierl, S., "Head-related transfer function recommendation based on perceptual similarities and anthropometric features," *The Journal of the Acoustical Society of America*, 148(6), pp. 3809–3817, 2020.

[25] 3253, E., "Sound Quality assessment material Recordings for subjective tests," `https://tech.ebu.ch/publications/sqamcd`, 2008.

[26] Bech, S. and Zacharov, N., *Perceptual audio evaluation-Theory, method and application*, John Wiley & Sons, 2007.

[27] Shinn-Cunningham, B. G., Kopco, N., and Martin, T. J., "Localizing nearby sound sources in a classroom: Binaural room impulse responses," *The Journal of the Acoustical Society of America*, 117(5), pp. 3100–3115, 2005.

[28] Baumgartner, R., Majdak, P., and Laback, B., "Modeling sound-source localization in sagittal planes for human listeners," *The Journal of the Acoustical Society of America*, 136(2), pp. 791–802, 2014.

[29] Li, S., Schlieper, R., and Peissig, J., "The effect of variation of reverberation parameters in contralateral versus ipsilateral ear signals on perceived externalization of a lateral sound source in a listening room," *The Journal of the Acoustical Society of America*, 144(2), pp. 966–980, 2018.

[30] Strutt, J. W., "On our perception of sound direction," *Philosophical Magazine*, 13(74), pp. 214–32, 1907.

[31] Kolarik, A. J., Moore, B. C., Zahorik, P., Cirstea, S., and Pardhan, S., "Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss," *Attention, Perception, & Psychophysics*, 78(2), pp. 373–395, 2016.

[32] Lyon, R. F., "All-pole models of auditory filtering," *Diversity in auditory mechanics*, pp. 205–211, 1997.

[33] Glasberg, B. R. and Moore, B. C., "Derivation of auditory filter shapes from notched-noise data," *Hearing research*, 47(1-2), pp. 103–138, 1990.

[34] Dau, T., Kollmeier, B., and Kohlrausch, A., "Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers," *The Journal of the Acoustical Society of America*, 102(5), pp. 2892–2905, 1997.

[35] Faller, C. and Merimaa, J., "Source localization in complex listening situations: Selection of binaural cues based on interaural coherence," *The Journal of the Acoustical Society of America*, 116(5), pp. 3075–3089, 2004.

[36] Li, S., Baumgartner, R., and Peissig, J., "Modeling perceived externalization of a static, lateral sound image," *Acta Acustica*, 4(5), p. 21, 2020.

[37] Li, S., Schlieper, R., Peissig, J., et al., "The Impact of Trajectories of Head and Source Movements on Perceived Externalization of a Frontal Sound Source," in *144th AES Convention*, Audio Engineering Society, 2018.

[38] Kim, J., Barnett-Cowan, M., and Macpherson, E. A., "Integration of auditory input with vestibular and neck proprioceptive information in the interpretation of dynamic sound localization cues," in *Proceedings of meetings on acoustics ICA*, Acoustical Society of America, 2013.