Audio Engineering Society

Convention Paper 10678

Presented at the 155th Convention
2023 October 25–27, New York, USA

Impact of mismatched room acoustic modeling on transaural reproduction with loudspeaker arrays

Yazhou Li, Lin Wang, and Joshua Reiss

Center for Digital Music, Queen Mary University of London

Correspondence should be addressed to Joshua Reiss (joshua.reiss@qmul.ac.uk)

ABSTRACT

Accurate room acoustic modeling plays an important role for designing a transaural reproduction system. If the room transfer functions used in system design are measured in an acoustic environment that differs from the evaluation environment, the performance will deteriorate. In this paper, we investigate the impact of mismatched room acoustic modeling on the performance of the system in reverberant environments. We model the room acoustic transfer functions with the image source method and incorporate the simulated room impulse response for the design of the loudspeaker array system. Simulation results show that the system performance will not be impacted by existence of reverberation when the transfer functions are matched. Additionally, the study finds that modeling only the early reflections is sufficient for satisfactory transaural reproduction.

1 Introduction

Personal sound zone (PSZ) reproduction from loudspeakers aims to provide different audio content with minimum interference to multiple listeners in the same physical space without the usage of headphones. This is achieved by creating a bright zone (where the content should be reproduced) and dark zones (where it should be cancelled). A multi-sound zone problem can be solved by linear superposition of several two-zone problems. Popular optimization methods include acoustic contrast control (ACC) [1, 2] and pressure matching (PM) [3, 4].

A similar principle lies in Crosstalk Cancellation (CTC), also known as transaural reproduction, which aims to provide binaural audio to the listener through loudspeakers. CTC can be seen as a PSZ problem by regarding the listener's two ears as two small sound zones [5]. However, PSZ is more challenging than CTC in that personal audio perception requires a higher energy ratio between the bright zone and the dark zone than spatial audio perception [6].

One challenge is that if we control the sound at the listener's ears precisely, the listener must stay stable and any head movements or rotations will result in performance degradation. Current works dealing with this problem include using head-tracking sensors, such as depth sensors [7] or cameras [8, 9, 10] to track the listener's position dynamically and update the filters in real time.

Moreover, accurate impulse responses (IRs), known as transfer functions (TFs) in the frequency domain, from the loudspeakers to the listeners are required to design filters. If mismatches exist between TFs used for designing the filters (referred to as *setup* TFs) and those during the evaluation (referred to as *playback* TFs), the system performance is expected to degrade.

Current research has shown that many factors can result in the TFs mismatch problem, including the mismatch between loudspeaker and microphone positions [11], variations of room temperature [12], the loudspeaker distortion [13, 14], the presence of background noise [15], listeners' scattering effects [15], and the problem of HRTFs individualization [7].

Room reverberation is another factor that causes TFs mismatch. Previous research has shown that if free-field TFs are used to design filters in reverberant rooms, the system performance will degrade. Using TFs measured in the same room results in a higher contrast between the bright zone and the dark zone than using free-field TFs or TFs measured in anechoic rooms [16]. Reverberation can also negatively impact an individual's directional perception in binaural reproduction [17].

However, measuring IRs in the real world is tedious and time-consuming. Therefore, current dynamic PSZ methods use free-field IRs [8, 10] for filter design, whose application is limited to the anechoic room. Alternatively, some studies measure IRs in advance [7], but this approach is only suitable for acoustically treated environments and cannot be generalized for home usage.

Current research that deals with the reverberation problem in sound field synthesis focuses on reducing the amount of reverberation and making the system robust to reverberation. Methods include using high-order directional loudspeakers or loudspeaker arrays to improve the direct-to-reverberation ratio of the sounds that arrive at the ears [18], designing a room compensation array to reduce the amount of lateral reverberation [19], and optimizing the number of sources in the array [20] and the position of the sources [21] to reduce reverberation. But the physical experiment results from the above-mentioned methods are still far from satisfactory. Wave-domain optimization [22] and impulse response reshaping [23] are also used to make the array robust to reverberation, but they still require measuring the IRs at a large number of microphone positions. These confirm the necessity to model the IRs accurately.

We hope to estimate the TFs in different room acoustics and at different positions, but whether it is possible to get satisfying results using modelled TFs for filter calculation remains unclear. This article examines how the system performance is affected by TFs mismatch by modeling the room reverberation and explores whether

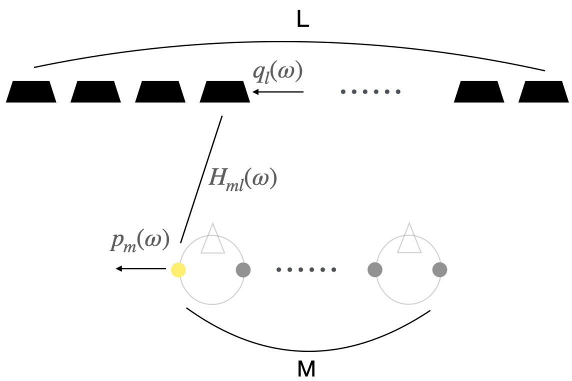


Fig. 1: There are L loudspeakers in the system and one microphone at each ear, M microphones in total. The yellow circle is the bright zone and the grey circles are the dark zones. There is only one bright zone in the system.

using the matched TFs can resolve the reverberation problem.

This paper is organized as follows. In Section 2, we introduce the formulation of CTC using a loudspeaker array and the room modeling method. In Section 3, we describe the simulation experiment implementation and setup details. In Section 4, we analyze the results of the experiments. Finally, we close this paper with conclusions in Section 5.

2 Background and Method

2.1 Problem formulation

PSZ aims to divide the listening area into different zones, where the audio content is audible in bright zones and silent in dark zones. CTC can be seen as a PSZ problem that one ear of one listener is regarded as the bright zone, and all other ears are regarded as the dark zones.

By controlling the sound at each ear of each listener, we are able to provide both spatial audio and personal audio to the listeners. By convolving the same audio content with head-related impulse response (HRIRs), known as head-related transfer functions (HRTFs) in the frequency domain, at the left ear and the right ear, we can provide two non-interfering sounds to the two ears, create a sound image at a specified angle and provide spatial audio to the listener. By delivering different audio content to different listeners, we can provide personal sound for multiple users.

As shown in Fig. 1, L loudspeakers are used for reproduction and M microphones are used for measuring the sound pressure levels (SPLs) when there are $M/2$ listeners. One microphone is placed in the bright zone and $M-1$ microphones are placed in the dark zones. The SPL in the bright zone is maximized and the SPLs in the dark zones are minimized. The sound pressure p_m at frequency ω measured by the m-th microphone is posed as

$$p_m(\omega) = \sum_{l=1}^{L} G_{ml}(\omega) q_l(\omega), \tag{1}$$

where $q_l(\omega)$ is the driving signal of the l-th loudspeaker at frequency ω, and $G_{ml}(\omega)$ is the transfer function from the l-th loudspeaker to the m-th microphone. The formulation in matrix form is

$$\boldsymbol{p} = \boldsymbol{G}\boldsymbol{q}, \tag{2}$$

where $\boldsymbol{p} = [p_1(\omega), ..., p_M(\omega)]^T \in \mathbb{C}^{M\times 1}$, $\boldsymbol{G} = (G_{ml}(\omega)) \in \mathbb{C}^{M\times L}$ and $\boldsymbol{q} = [q_1(\omega), ..., q_L(\omega)] \in \mathbb{C}^{L\times 1}$. All quantities in matrix form are implicitly dependent on ω.

In this paper, we use a pressure matching (PM) method [24] as the optimization algorithm to design filters to control the driving signals of loudspeakers. The loss function J is

$$J = ||\boldsymbol{p} - \boldsymbol{p}_T||^2 = ||\boldsymbol{G}\boldsymbol{q} - \boldsymbol{p}_T||^2, \tag{3}$$

where $\boldsymbol{p}_T = [0, ..., 0, 1, 0, ..., 0]^T \in \mathbb{C}^{M\times 1}$ is the target pressure at M microphones, which is specified as 1 from the loudspeaker to the microphone in the bright zone, and 0 in the dark zones. The optimal filter $\boldsymbol{q}_{opt} \in \mathbb{C}^{L\times 1}$ to apply to the loudspeakers is calculated by

$$\boldsymbol{q}_{opt} = \boldsymbol{G}^{-1}\boldsymbol{p}_T = (\boldsymbol{G}^H\boldsymbol{G})^{-1}\boldsymbol{G}^H\boldsymbol{p}_T. \tag{4}$$

When the number of loudspeakers L is larger than the number of microphones M, $\boldsymbol{G}^H\boldsymbol{G} \in \mathbb{C}^{L\times L}$ is a singular matrix and not invertible. Even if the number of microphones is larger than the number of loudspeakers, because the microphones are placed close to each other, the resultant filters will have a high gain. Therefore, Tikhonov regularization [24] is used to limit the filter gain and compensate for the singular value to ensure the matrix is invertible,

$$\boldsymbol{q}_{opt} = (\boldsymbol{G}^H\boldsymbol{G} + \beta\boldsymbol{I})^{-1}\boldsymbol{G}^H\boldsymbol{p}_T. \tag{5}$$

2.2 Room Modeling

To model the real room's reverberation, we use the Image Source Method (ISM) to estimate room impulse responses (RIRs) based on the room geometry and the wall materials. ISM is a geometrical method to estimate RIRs with the room's early reflections [25]. Each driver is mapped to an image source symmetrical to each wall. The image sources and original driver are together mapped against the wall again. The number of mapping is called image source order (ISO). Readers can refer to [26] for more implementation details. From RIRs we can get room transfer functions (RTFs) from L loudspeakers to M microphones $\boldsymbol{R} = (R_{ml}(\omega)) \in \mathbb{C}^{M\times L}$ by transforming RIRs to the frequency domain.

In order to control the sounds at two ears accurately, the head's effect on the sound field need to be considered. Therefore, HRIRs are used as the impulse responses from loudspeakers to microphones, in order to expresses the filtering effect of the body when the listener is present in the room without reverberation. Then HRTFs from L loudspeakers to M microphones $\boldsymbol{H} = (H_{ml}(\omega)) \in \mathbb{C}^{M\times L}$ can be calculated by transforming HRIRs to the frequency domain.

At last, we combine the modelling of room reverberation and head effects by multiplying RTFs $\boldsymbol{R}$ and HRTFs $\boldsymbol{H}$ at each frequency ω to get binaural room transfer functions (BRTFs) as the system's transfer functions $\boldsymbol{G} = (G_{ml}(\omega)) \in \mathbb{C}^{M\times L}$,

$$G_{ml}(\omega) = R_{ml}(\omega) \cdot H_{ml}(\omega). \tag{6}$$

2.3 Evaluation Metrics

In evaluation, there are one bright zone and $M-1$ dark zones, with one microphone in each zone. We add a perturbation (δ_x, δ_y) to the head position to test the robustness of the system.

The evaluation metric is acoustic contrast (AC), which describes the energy difference between the bright zone and the dark zone. The acoustic contrast $AC(\omega)$ at frequency ω is presented by

$$AC(\omega) = 20\log_{10}\frac{|p_b(\omega)|}{|p_d(\omega)|}, \tag{7}$$

where $p_b(\omega) = p_1(\omega)$ is the SPL of the evaluation microphone in the bright zone and $p_d(\omega)$ is the average SPLs of $M-1$ evaluation microphones

$[p_2(\omega), ..., p_M(\omega)]$ in the dark zones. Generally, AC from 15 to 20 dB at different frequencies is needed for satisfying spatial audio perception and more than 39 dB contrast is required for personal audio perception [6].

3 Simulation Experiment

3.1 System Setup

In the simulation, the room is a $4\,\mathrm{m} \times 5\,\mathrm{m}$ shoebox. An $L = 15$ sources line loudspeaker array with a spacing of 0.25 m is used and the drivers are modeled as free-field point sources. The loudspeakers' x-coordinates are 3.68 m. Only one listener ($M = 2$) is considered in this experiment. The center of the listener's head is positioned at $(2\,\mathrm{m}, 2\,\mathrm{m})$ and the listener is facing perpendicularly towards the array. We assume that the head width is $0.2\,\mathrm{m}$, the left ear is at $(2\,\mathrm{m}, 2.1\,\mathrm{m})$ and the right ear is at $(2\,\mathrm{m}, 1.9\,\mathrm{m})$. The left ear is specified as the bright zone and the right ear is the dark zone. A perturbation $(0\,\mathrm{m}, 0.01\,\mathrm{m})$ on the listener's position is added, which means the head is placed at $(2\,\mathrm{m}, 2.01\,\mathrm{m})$ in evaluation. The system setup is shown in Fig. 2.

We use the pyroomacoustics toolbox [27] to implement the image source method and estimate the RIRs from each loudspeaker to each microphone in the room. The four walls are seen as four reflectors. The broadband wall material absorption parameter is 0.1 for the whole frequency band, which means the sound magnitude reduces by 10% with each reflection for all frequencies. The simulation is done on the horizontal plane, and the ceiling and the ground are not considered. The sampling rate is 44100 Hz. The sound propagating speed is 343 m/s.

If the IR is longer than 0.2 s (8820 sampling points), it is truncated to only the first 8820 sampling points using a Tukey window with $R = 0.05$. If the IR is less than 8820 points, it is zero-padded to 8820 points. The processed RIRs are converted to the frequency domain using Fast Fourier Transform to get RTFs.

The small KEMAR head HRTFs from the CIPIC dataset [28] are used in the simulation. We assume the listener and the loudspeaker array are on the same horizontal plane and the elevation angles of HRTFs are always 0. After calculating the horizontal angle from the loudspeaker to the head, we choose the HRTFs from the dataset whose horizontal angle is the closest to the calculated angle. Then we multiply the RTFs at the left ear and the right ear with the left ear and the right ear's HRTFs respectively to get the BRTFs at two ears.

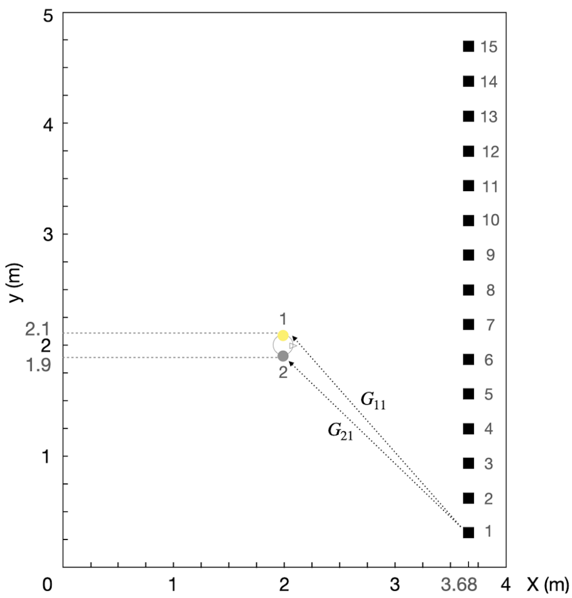


Fig. 2: Illustration of the simulation experiment setup. The black squares on the right represent 15 loudspeakers and the sound zones are indicated by two circle shadows, with the yellow one being the bright zone and the grey one being the dark zone. The dashed lines illustrate the indexed loudspeaker and microphone for the corresponding transfer function.

Then we calculate the filter weights of each loudspeaker using PM mentioned in Section 2 at these 8820 frequency points. The regularization parameter β in PM is 0.0001.

Finally, the sound pressures at the evaluation microphones in the bright zone and the dark zone are calculated to get AC of the system. All the results are smoothed with a 1/3 octave band square filter and are shown from 20 Hz to 20000 Hz to match the range of human hearing.

3.2 Experimental Design

Two experiments are conducted to investigate two questions as follows:

Table 1: ISO of playback TFs and setup TFs.

	Setup 1	Setup 2	Setup 3
playback TFs	0	3	3
setup TFs	0	0	3

- When the setup TFs match with the playback TFs, are the performances in the reverberant room and the anechoic room equivalent? In other words, if TFs are matched, will reverberation decrease the system robustness? Three experiment setups are shown in Table 1.

  Setup 1 means no reverberation in both setup TFs and playback TFs; Setup 2 means reverberation exists in playback TFs, but not in setup TFs; Setup 3 means reverberation exists in both playback TFs and setup TFs. In all 3 simulations, we use ISM with ISO = 3 to calculate TFs with reverberation, and ISM with ISO = 0 to calculate TFs without reverberation.

- When setup TFs and playback TFs are mismatched, how does the extent of mismatch between the two TFs affects the system performance? When ISO increases, higher-order reflections will result in longer reverberation time, representing a larger amount of reverberation in the room. We set ISO in playback TFs (referred to as the *playback* order) to 10, and set the ISO in setup TFs (referred to as the *setup* order) to 0, 3, 5, 9, 10 respectively.

4 Experiment Results

4.1 Matched TFs in different rooms

The results of the first experiment are illustrated in Fig. 3. By comparing AC of Setup 1 and Setup 2, it is evident that the presence of reverberation and TFs mismatch have a negative impact on the system performance, leading to an average AC of 11.8 dB from 20 Hz to 5000 Hz, which is below the 20 dB threshold and insufficient for a satisfactory binaural listening experience.

However, by comparing AC of Setup 1 and Setup 3, we can conclude that if setup TFs and playback TFs match with each other, there is not much difference when playback TFs are changed, which suggests that reverberation does not significantly affect AC between the two zones, as long as setup TFs are accurate.

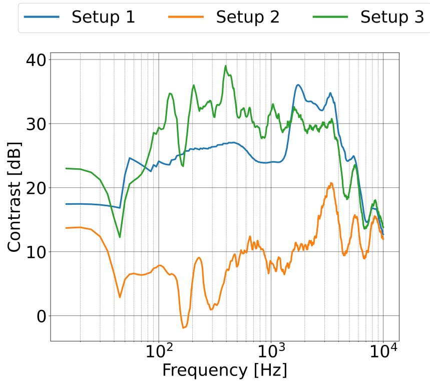


Fig. 3: Acoustic contrasts for three experiment setups.

In general, AC is the highest in mid frequencies because the beam width is large in low frequencies, so it is difficult to separate between the bright zone and the dark zone when they are near each other. At high frequencies, the system is less robust and degrades easily with perturbations to microphone positions due to spatial aliasing.

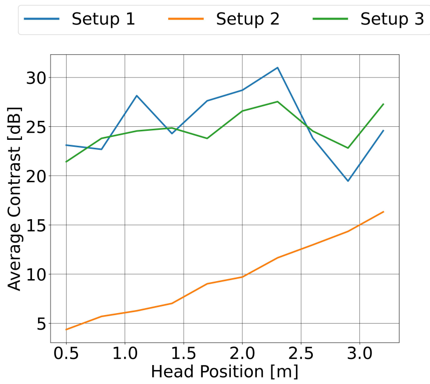


Fig. 4: Different head positions along the *x*-axis.

We test the system performance with different distances between the listener and the array. The listener's y-position remains 2 m, and the x-position changes from 0.5 m to 3.5 m at a step of 0.5 m, which indicates that the listener moves closer to the loudspeaker array. The result is shown in Fig. 4. When the TFs are matched (Setup 1 and Setup 3), AC remains similar when the listener moves. However, when the TFs are mismatched (Setup 2), AC increases when the listener is closer to the array, which suggests that the impact of reverberation becomes more pronounced when the listener is farther away from the loudspeakers.

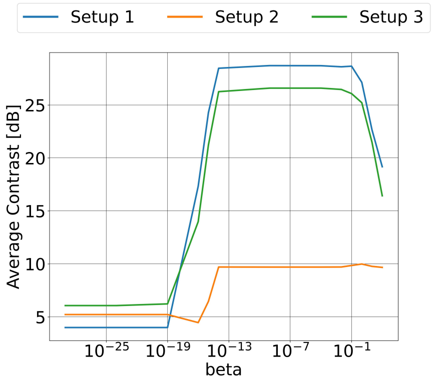


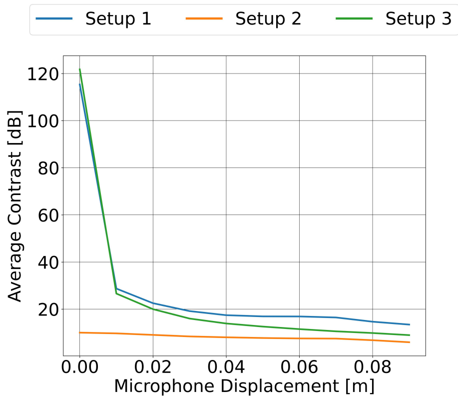
Fig. 5: Different regularization parameters β.

The effect of different regularization parameters β is also investigated. β is changed from 10^{-30} to 10^2 logarithmically at a step of 10^{-1}. As shown in Fig. 5, if the value of β is extremely high or low, the performance will degrade rapidly. The reason is that if β is too small, the regularization is not able to control the driver energy, resulting in poor system robustness. Conversely, if β is too large, the optimization performance will be constrained, leading to a decrease in AC. As long as β is chosen from 10^{-15} to 0.1, the system performance remains relatively consistent.

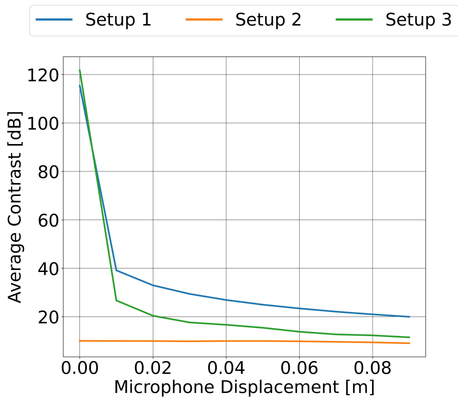
We also test the influence of the microphone perturbation, which is the difference between the position of the setup microphone and the playback microphone. The head is placed at $(2\,\text{m}, 2\,\text{m})$ in IRs calculation, and at $(2+\delta_x\,\text{m}, 2+\delta_y\,\text{m})$ in evaluation. For y-axis perturbation, $\delta_x = 0$, δ_y changes from 0 m to 0.1 m at a step of 0.01 m. For x-axis perturbation, $\delta_y = 0$, δ_x changes from 0 m to 0.1 m at a step of 0.01 m.

As shown in Fig. 6, for y-axis perturbation, AC of Setup 1 and Setup 3 are similar. However, for x-axis perturbation, AC in Setup 1 is higher than that in Setup 3 by more than 10 dB. This suggests that when reverberation exists in the evaluation, the system is less robust along the x-axis even if the TFs are matched. But as long as the perturbation is smaller than 0.02 m, the system can still provide AC larger than 20 dB in both situations.

The difference between the x-axis perturbation and the y-axis perturbation can be illustrated by the sound field visualization in Fig. 7. The visualization only considers reverberation modelled with the image source method, the head and torso diffraction is not considered, which means HRTFs are not added in the visualization. But the visualization can still provide some insights into the problem. When there is no reverberation, the sound field is more homogeneous along the x-axis. Therefore, the existence of reverberation will decrease the system robustness along the x-axis.



(a) y-axis perturbation

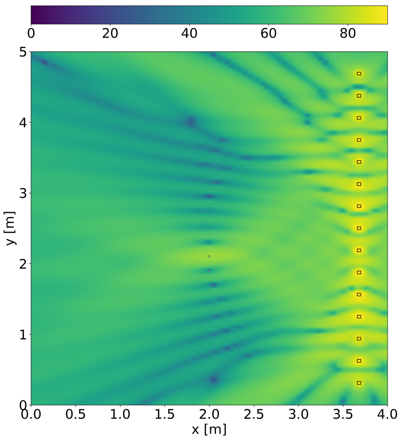


(b) x-axis perturbation

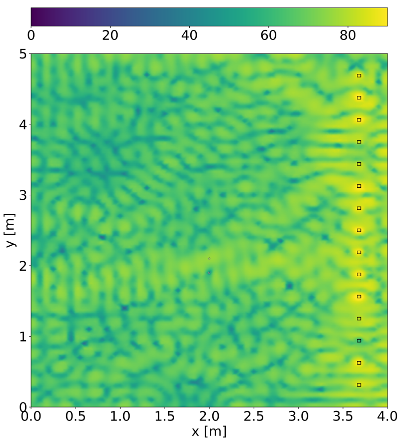
Fig. 6: Different perturbations of the microphone position.

4.2 Different setup TFs with the same playback TFs

Current simulated RIRs are perceptually satisfying, but they are still far from real RIRs and not accurate enough to be used as setup TFs for filter calculation. However,

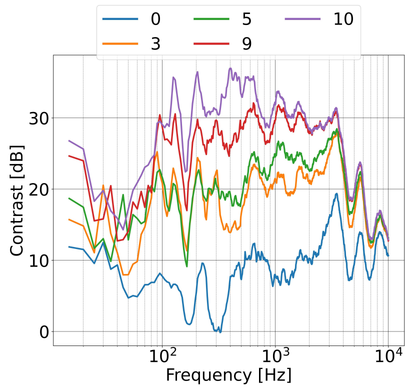


(a) free-field

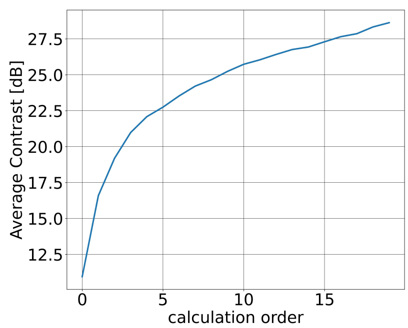


(b) the reverberant room

Fig. 7: Sound field visualization at 1000 Hz.



(a) Acoustic contrasts of setup orders (0, 3, 5, 9, 10).



(b) Average acoustic contrasts from 20 Hz to 5000 Hz.

Fig. 8: Acoustic contrasts when the setup order changes from 0 to 20 and the playback order is 20.

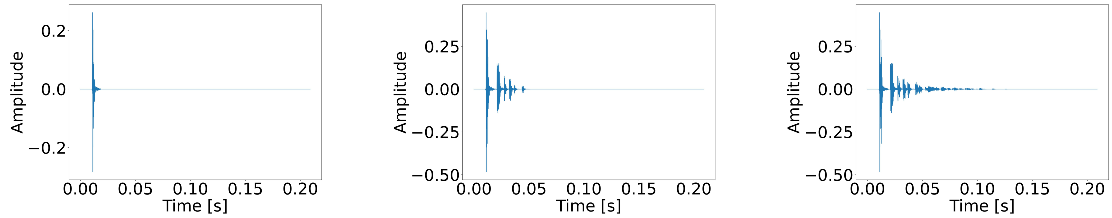
as shown in Fig. 8 (a), AC increases with the similarity between setup TFs and playback TFs.

When the playback order is 20 and the setup order changes from 0 to 20, the average AC from 20 Hz to 5000 Hz does not change linearly with the setup order, as shown in Fig. 8 (b). The average AC is 20.98 dB when the setup order is 3 and 28.63 dB when the setup order is 10. When the setup order is 3, it is already enough to provide AC larger than 20 dB. Therefore, the performance should be improved by just simulating the early reflections in the room using geometrical methods.

In order to further examine this result, BRIRs calculated with different image source orders are shown in Fig. 9. The reverberation time (RT_{60}) is also calculated from the BRIRs using the pyroomacoustics toolbox. The peaks in Fig. 9 (b) and 9 (c) are early reflections where the main energy in the RIR is contained, which are similar between 9 (b) and 9 (c), although their reverberation times are significantly different.

5 Conclusion

This paper looked into how using different transfer functions in filter calculation affects the result of Crosstalk Cancellation in the presence of reverberation. The simulation experiment showed that modeling only the early reverberation can lead to significant performance improvements. The finding suggests the



(a) ISO = 0. $RT_{60} = 0.001$ s. **(b)** ISO = 3. $RT_{60} = 0.024$ s. **(c)** ISO = 10. $RT_{60} = 0.099$ s.

Fig. 9: Binaural room impulse responses calculated with the different image source orders from the loudspeaker at (3.68 m, 2 m) to the left ear at (2 m, 2.01 m).

possibility to model the room using geometrical methods for filter calculation. It also highlights the need for more accurate room impulse responses modelling to calculate the filters. But the sound field in the reverberant room is less homogeneous and the robustness will be degraded along the vertical direction to the line array.

The work could be easily extended to multiple-user applications and provide both personal audio and spatial audio to the listeners. In real applications, late reverberation, loudspeaker noise, the loudspeaker's frequency responses and other factors could degrade the system performance. Therefore, further physical experiments and subjective experiments are needed to assess the approach's effectiveness in real-world applications.

References

[1] Choi, J.-W. and Kim, Y.-H., "Generation of an acoustically bright zone with an illuminated region using multiple sources," *The Journal of the Acoustical Society of America*, 111(4), pp. 1695–1700, 2002.

[2] Chang, J.-H. et al., "A realization of sound focused personal audio system using acoustic contrast control," *The Journal of the Acoustical Society of America*, 125(4), pp. 2091–2097, 2009.

[3] Poletti, M., "An Investigation of 2-D Multizone Surround Sound Systems," in *Audio Engineering Society Convention 125*, 2008.

[4] Betlehem, T. and Teal, P. D., "A constrained optimization approach for multi-zone surround sound," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 437–440, 2011.

[5] Jackson, P., Fazi, F., and Coleman, P., "Personalising sound over loudspeakers," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019.

[6] Canter, N. and Coleman, P., "Delivering Personalised 3D Audio to Multiple Listeners: Determining the Perceptual Trade-Off Between Acoustic Contrast and Cross-Talk," in *Audio Engineering Society International Conference on Audio for Virtual and Augmented Reality*, 2020.

[7] Qiao, Y. and Choueiri, E., "The Performance of A Personal Sound Zone System with Generic and Individualized Binaural Room Transfer Functions," in *Audio Engineering Society Convention 152*, 2022.

[8] Gálvez, M. F. S., Menzies, D., and Fazi, F. M., "Dynamic Audio Reproduction with Linear Loudspeaker Arrays," *J. Audio Eng. Soc*, 67(4), pp. 190–200, 2019.

[9] Song, M.-S. et al., "An Interactive 3-D Audio System With Loudspeakers," *IEEE Transactions on Multimedia*, 13(5), pp. 844–855, 2011.

[10] Simón Gálvez, M. F., Takeuchi, T., and Fazi, F. M., "Low-Complexity, Listener's Position-Adaptive Binaural Reproduction Over a Loudspeaker Array," *Acta Acustica united with Acustica*, 103(5), pp. 847–857, 2017.

[11] Park, J.-Y., Choi, J.-W., and Kim, Y.-H., "Acoustic contrast sensitivity to transfer function errors

in the design of a personal audio system," *The Journal of the Acoustical Society of America*, 134(1), pp. EL112–EL118, 2013.

[12] Olsen, M. and Møller, M. B., "Sound Zones: On the Effect of Ambient Temperature Variations in Feed-Forward Systems," in *Audio Engineering Society Convention 142*, 2017.

[13] Ma, X. et al., "Impact of loudspeaker nonlinear distortion on personal sound zones," *The Journal of the Acoustical Society of America*, 143(1), pp. 51–59, 2018.

[14] Ma, X. et al., "Personal Sound Zones: The Significance of Loudspeaker Driver Nonlinear Distortion," in *Audio Engineering Society International Conference on Sound Field Control*, 2016.

[15] Møller, M. B. et al., "On the Influence of Transfer Function Noise on Sound Zone Control in a Room," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 27(9), pp. 1405–1418, 2019.

[16] Olivieri, F. et al., "Loudspeaker Array Processing for Multi-Zone Audio Reproduction Based on Analytical and Measured Electroacoustical Transfer Functions," in *Audio Engineering Society International Conference: Sound Field Control - Engineering and Perception*, 2013.

[17] Sæbø, A., *Influence of reflections on crosstalk cancelled playback of binaural sound*, Ph.D. thesis, Norwegian University of Science and Technology, 2001.

[18] Poletti, M. A. and Abhayapala, T. D., "Spatial sound reproduction systems using higher order loudspeakers," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 57–60, 2011.

[19] Simón Gálvez, M. F. and Fazi, F. M., "Room compensation for binaural reproduction with loudspeaker arrays," in *European Acoustics Association Congress (Euroregio)*, 2016.

[20] Simón Gálvez, M. F., Blanco Galindo, M., and Fazi, F. M., "A study on the effect of reflections and reverberation for low-channel-count Transaural systems," in *INTER-NOISE and NOISE-CON Congress and Conference*, volume 259, pp. 6111–6122, 2019.

[21] Olik, M., *Personal sound zone reproduction with room reflections*, Ph.D. thesis, University of Surrey, 2015.

[22] Schneider, M. and Kellermann, W., "Adaptive listening room equalization using a scalable filtering structure in thewave domain," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 13–16, 2012.

[23] Jungmann, J. O. et al., "Combined Acoustic MIMO Channel Crosstalk Cancellation and Room Impulse Response Reshaping," *IEEE Transactions on Audio, Speech, and Language Processing*, 20(6), pp. 1829–1842, 2012.

[24] Kirkeby, O. and Nelson, P. A., "Reproduction of plane wave sound fields," *The Journal of the Acoustical Society of America*, 94(5), pp. 2992–3000, 1993.

[25] 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.

[26] Savioja, L. and Svensson, U. P., "Overview of geometrical room acoustic modeling techniques," *The Journal of the Acoustical Society of America*, 138(2), pp. 708–730, 2015.

[27] Scheibler, R., Bezzam, E., and Dokmanić, I., "Pyroomacoustics: A Python Package for Audio Room Simulation and Array Processing Algorithms," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 351–355, 2018.

[28] Algazi, V. et al., "The CIPIC HRTF database," in *IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, pp. 99–102, 2001.