Y. Qiao, J. Luo, and E. Y. Choueiri,
"Spatial Sampling of Binaural Room Transfer Functions
for Head-Tracked Personal Sound Zones,"
J. Audio Eng. Soc., vol. 72, no. 7/8, pp. 455–466 (2024 Jul./Aug.).
https://doi.org/10.17743/jaes.2022.0144.

Spatial Sampling of Binaural Room Transfer Functions for Head-Tracked Personal Sound Zones

YUE QIAO,[1,*] *AES Student Member*, **JESSICA LUO,**[2] *AES Student Member* **AND**
(yqiao@princeton.edu) (jsl792@nyu.edu)
EDGAR Y. CHOUEIRI,[1] *AES Associate Member*
(choueiri@princeton.edu)

[1] *3D Audio and Applied Acoustics Laboratory, Princeton University, Princeton, NJ*
[2] *Music Technology, New York University, New York, NY*

The spatial sampling of binaural room transfer functions that vary with listener movements, as required for rendering personal sound zone (PSZ) with head tracking, was experimentally investigated regarding its dependencies on various factors. Through measurements of the binaural room transfer functions in a practical PSZ system with either translational or rotational movements of one of the two mannequin listeners, the PSZ filters were generated along the measurement grid and then spatially downsampled to different resolutions, at which the isolation performance of the system was numerically simulated. It was found that the spatial sampling resolution generally depends on factors such as the moving listener's position, frequency band of the rendered audio, and perturbation caused by the other listener. More specifically, the required sampling resolution is inversely proportional to the distance either between two listeners or between the moving listener and the loudspeakers and is proportional to the frequency of the rendered audio. The perturbation caused by the other listener may impair both the isolation performance and filter robustness against movements. Furthermore, two crossover frequencies were found to exist in the system, which divide the frequency band into three sub-bands, each with a distinctive requirement for spatial sampling.

0 INTRODUCTION

Personal sound zone (PSZ) [1, 2] is one sub-area of sound field control [3] that has received wide attention from the audio and acoustics community over the past two decades, with applications for mobile devices [4], automotive cabins [5, 6], and outdoor spaces [7]. Using loudspeaker arrays and digital signal processing techniques, two listening zones, often known as bright zone (*BZ*) and dark zone (*DZ*), are rendered in the same physical space, with the aim of providing listeners with minimally interfering individual audio programs (e.g., speech/music) or generating a quiet area for one listener while delivering audio to the other. Depending on the preferred/selected performance metric (e.g., isolation level between zones or reproduced audio quality at each zone), various filter design methods have been proposed, such as Pressure Matching (PM) [8–10], Acoustic Contrast Control (ACC) [11–13], and Variable Span Trade-Off Filtering [14, 15].

Up until recently, most PSZ studies have focused on rendering static sound zones (i.e., they are not compensated for time-varying changes in the environment), in which case, the system performance can easily degrade when the actual acoustic transfer functions (ATFs) significantly deviate from those used in plant modeling (i.e., estimating the plant ATFs based on previous observations) and filter design. Although regularization can be applied to the filter design to mitigate the effects of such ATF mismatches [16–18], the uncertainties in plant modeling can lead to sub-optimal performance. This is especially true for systems that directly control the pressure around or at the listeners' ears [10, 19] (as opposed to those targeting larger spaces [9, 12, 20–22]) due to head movements. Such systems create a so-called "virtual PSZ" through binaural reproduction, and they will be the main focus of this paper because they allow more precise control of the sound perceived by the listeners compared with those rendering larger zones; in this context, the plant ATF is equivalent to the binaural room transfer function (BRTF), which regards the room and listener as a whole "plant". These two terms (ATF and BRTF) will be used interchangeably hereafter.

In order to improve the robustness of the binaural reproduction-based PSZ system against head movements

*To whom correspondence should be addressed, email: yqiao@princeton.edu, Last updated: May 1, 2024

without sacrificing the isolation performance, recent studies have proposed solutions that apply head tracking to PSZ [23, 24]; the use of head tracking can also be found in other audio applications, e.g., loudspeaker crosstalk cancellation (XTC) [25–28] and loudspeaker equalization [29]. In the general case of moving sound zones, other studies have explored adaptive techniques such as moving horizon framework [30], Filtered-x Least-Mean-Square algorithm [6], and Recursive Least Squares algorithm [31] to update the PSZ filters in real time.

The practical implementation of a head-tracked PSZ system still remains challenging due to the following reasons: 1) Unlike applications that target a single listener (e.g., XTC), the presence of two or more listeners in a PSZ system significantly increases the efforts for plant measurements and/or estimation, considering the simultaneous movements of listeners and potential coupling effects between listeners [19]. 2) Compared to an XTC system, which operates on similar principles of plant estimation and filter generation, a PSZ system usually requires a higher level of performance index for perceptual acceptance [32], suggesting that the plant for rendering head-tracked PSZ needs to be estimated (or pre-measured) at a finer spatial resolution. Given these demanding requirements on the amount and resolution of plant estimation, it is necessary to study the spatial dependencies of ATFs (or BRTFs) and optimize the corresponding sampling process such that the amount of effort for plant estimation is minimized for rendering PSZ with high isolation over a large moving area.

This study seeks to investigate the effects of spatial sampling of BRTFs on the isolation performance of an experimental PSZ system and then derive the optimal spatial sampling resolution and rules for reducing the measurement effort given a target isolation level. Various factors are considered, such as the frequency band of the rendered audio, the type and extent of listener movement, whether the listener is moving or static, and whether the sound zone is *BZ* or *DZ*. The study is conducted by measuring the listeners' BRTFs along different moving trajectories and simulating the isolation performance at different sampling resolutions.

This study expands the authors' prior work [33] whose main focus was on the translational movements of listeners, by further investigating the case of rotational movements, as well as the influence of the other listener's perturbation in a two-listener setup. SEC. 1 introduces the PM method used for PSZ filter generation and the metrics for performance evaluation. SEC. 2 describes the PSZ system used in the study and the implementation of the experiments. SECS. 3 and 4 present the analysis of spatial sampling for head translations and head rotations, respectively. SEC. 5 examines the effects of one listener's perturbation on the sampling resolution. Discussion of the results is given in SEC. 6. Lastly, SEC. 7 gives conclusions on optimizing the BRTF spatial sampling and suggests future directions.

1 PSZ THEORY

A general PSZ system that consists of an array of L loudspeakers and M control points (for the system that targets the ears of two listeners, $M = 4$) is considered. Each loudspeaker l is assigned with a complex gain $g_l(\omega)$, $l = 1, \ldots, L$ and the resulting sound pressure at each control point m is $p_m(\omega)$, $m = 1, \ldots, M$, where ω denotes the frequency. The ATF corresponding to the loudspeaker l and the control point m is denoted as H_{ml}, which in matrix form (known as the plant matrix) is denoted as $\mathbf{H}$ and has the following relationship:

$$\mathbf{p} = \mathbf{H}\mathbf{g}, \tag{1}$$

where $\mathbf{p} = [p_1, \cdots, p_M]^T \in \mathbb{C}^{M\times 1}$, $\mathbf{H} = (H_{ml}) \in \mathbb{C}^{M\times L}$, and $\mathbf{g} = [g_1, \cdots, g_L]^T \in \mathbb{C}^{L\times 1}$. All quantities hereafter are implicitly dependent on the frequency ω.

1.1 PM With BRTF Modeling

The PM method formulated in the frequency domain [8–10] is used to generate the PSZ filters because PM has control over the phase of target audio programs, compared to other methods such as ACC. Given a target pressure vector $\mathbf{p}_T \in \mathbb{C}^{M\times 1}$ specifying the desired pressure at the control points, the original cost function J in PM is constructed as

$$J = \|\mathbf{p} - \mathbf{p}_T\|^2 = \|\mathbf{H}\mathbf{g} - \mathbf{p}_T\|^2, \tag{2}$$

and by minimizing J, the optimal loudspeaker gains $\mathbf{g}^*$ are given by

$$\mathbf{g}^* = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{p}_T, \tag{3}$$

where the $(\,\cdot\,)^H$ denotes taking the conjugate transpose. It should be noted that this form of solution only applies to cases where $L \leq M$. When $L > M$, there are infinitely many solutions. Among them, a solution that yields the "minimum energy" of loudspeaker gains is given by

$$\mathbf{g}^* = \mathbf{H}^H(\mathbf{H}\mathbf{H}^H)^{-1}\mathbf{p}_T. \tag{4}$$

A more common approach is to regularize the solution to ensure its robustness against a certain degree of plant ATF uncertainties. A probabilistic approach [18] is adopted by modeling the estimated H_{ml} as a random variable and minimizing the expectation of the resulting cost function:

$$J_{prob} = \mathbb{E}\{\|\mathbf{p} - \mathbf{p}_T\|^2\}. \tag{5}$$

Under the assumption that H_{ml} follows a complex normal distribution,

$$H_{ml} = A_{ml}e^{i\phi_{ml}}, \tag{6}$$

$$A_{ml} \sim N(\hat{A}_{ml}, \sigma^2_{A,ml}), \tag{7}$$

$$\phi_{ml} \sim U(0, 2\pi), \tag{8}$$

where A_{ml} and ϕ_{ml} denote the amplitude and phase of H_{ml}, respectively; $N(\cdot, \cdot)$ denotes the scalar normal distribution; the hat symbol and σ denote the mean and standard deviation, respectively; and $U(\cdot, \cdot)$ denotes the uniform distribution. The corresponding optimal solution is given by

$$\mathbf{g}^*_{prob} = (\hat{\mathbf{H}}^H\hat{\mathbf{H}} + \Sigma_{m=1}^{M}\Sigma_m)^{-1}\hat{\mathbf{H}}^H\mathbf{p}_T, \tag{9}$$

where $\hat{\mathbf{H}}$ denotes the expected value of $\mathbf{H}$, and Σ_m is expressed as

$$\Sigma_m = diag\{\sigma^2_{A,m1}, \cdots, \sigma^2_{A,mL}\}, \tag{10}$$

where $\sigma^2_{A,ml}$ is the amplitude variance of H_{ml} and can be determined experimentally [19].

1.2 Evaluation Metrics

Two metrics were adopted to evaluate the isolation performance of a PSZ system, as introduced in [34]: Inter-Zone Isolation (IZI), which represents the isolation of the sound *zones* given an audio program, and Inter-Program Isolation (IPI), which represents the isolation of the target *program* from the interfering *program* leaked into the same sound zone. These two metrics correspond to the perception of sound attenuation in *DZ* and audio-on-audio interference, respectively (see [34] for detailed explanations). Other metrics such as reproduction error are not adopted here because they are not directly related to the isolation performance (see the discussion in [19]).

The two zones are denoted as Z_1 and Z_2, the sub-matrices (i.e., the top/bottom two rows) of $\mathbf{H}$ corresponding to $Z_{1,\ 2}$ as $\mathbf{H}_{1,\ 2}$, and the PSZ filters corresponding to Z_1 (or Z_2) being the *BZ* as $\mathbf{g}_1^*$ (or $\mathbf{g}_2^*$). This study focuses on rendering mono audio programs (i.e., a single vector $\mathbf{p}_T$ that defines the *BZ* and *DZ*), in which case the definition of IZI is expressed as

$$IZI_1 = \frac{\|\mathbf{H}_1\mathbf{g}_1^*\|^2}{\|\mathbf{H}_2\mathbf{g}_1^*\|^2}, \qquad IZI_2 = \frac{\|\mathbf{H}_2\mathbf{g}_2^*\|^2}{\|\mathbf{H}_1\mathbf{g}_2^*\|^2}, \tag{11}$$

where the subscript 1 (or 2) of *IZI* refers to the case of rendering the target program in Z_1 (or Z_2) and silence in the other. In this particular case, IZI has also shown [34] to be equivalent to the commonly used Acoustic Contrast metric [16]. Correspondingly, IPI for two different *BZ*/*DZ* assignments is expressed as

$$IPI_1 = \frac{\|\mathbf{H}_1\mathbf{g}_1^*\|^2}{\|\mathbf{H}_1\mathbf{g}_2^*\|^2}, \qquad IPI_2 = \frac{\|\mathbf{H}_2\mathbf{g}_2^*\|^2}{\|\mathbf{H}_2\mathbf{g}_1^*\|^2}. \tag{12}$$

2 METHODOLOGY

The experiments in this study were conducted as follows: First, the BRTFs of two mannequin listeners were measured at a fine spatial resolution, with one listener moving along different trajectories and the other fixed as a reference. Then, the original plant BRTF grid was downsampled to sparser resolutions and used to generate the PSZ filter sets. Lastly, the isolation performance of the system was numerically simulated by convolving the full-resolution BRTFs with each of the filter sets and calculating the corresponding IZI and IPI metrics. By observing the variation in IZI and IPI with head movements, the lowest spatial sampling resolution was determined for achieving a specified isolation level threshold, and the dependency of isolation performance on various factors was analyzed. Although only the case of one moving listener is considered for simplicity, it will be shown that the findings can be extrapolated to the case of two moving listeners.

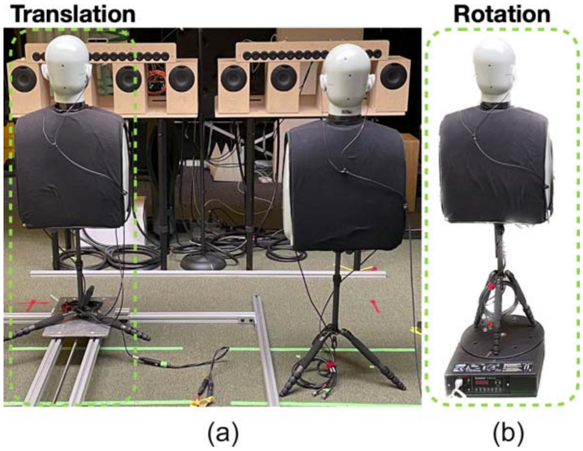

Fig. 1. Photo of the implemented system setup. The mechanical platform (a) is used for translational movements, while the turntable (b) is used for azimuthal rotations of the left listener.

2.1 System Implementation

The experimental PSZ system, which is identical to the one studied in [19], comprises a linear loudspeaker array with eight mid-range transducers (the tweeter loudspeaker arrays in Fig. 1 are not used). The two listeners were represented by two B&K Head and Torso Simulators (HATS Type 4100) with in-ear binaural microphones (Theoretica Applied Physics BACCH-BM Pro) placed in a typical listening room ($RT_{60} \approx 0.24$ s in the range 1,300–6,300 Hz). Two types of listener movements were implemented: the translational movements were realized with a custom-made mechanical translation platform, while the azimuthal rotations were achieved with a turntable (Outline ET250-3D), see Fig. 1. Synchronized exponential sine sweep signals were used [35] (a variant of the traditional exponential sine sweep [36] that correctly estimates higher harmonic frequency responses) to measure the BRTFs at 48-kHz sampling frequency, with each sweep having a duration of 0.5 s. All measured binaural room impulse responses (the time-domain counterpart of BRTFs) were truncated to the first 8,192 samples for subsequent processing (it was verified that further increase of the truncation window length had no noticeable effect on the results).

2.2 Filter Generation and Adaptation

The PSZ filters used in the evaluation were generated to maximize the isolation performance at matched listener positions while preserving the robustness against minor head misalignments that can occur in practical systems due to head tracking accuracy. Therefore, the variance matrix in Eq. (10) was determined by considering the BRTF fluctuations due to slight head movements, following the approach in [19]. The authors refer the readers to [19] for other details such as the choice of target pressure. As an improvement to the original approach, all the plant binaural room impulse responses used for filter generation (i.e., those that formed the plant matrix $\mathbf{H}$ and target pressure $\mathbf{p}_T$) were truncated to the first 4,096 samples (late reverb

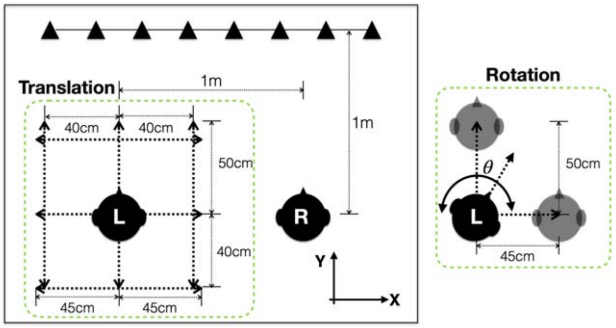

Fig. 2. Diagram of the measurement setup for head translations and rotations. For translation, the dotted lines around the left listener represent the moving trajectories along which the BRTFs were measured. For rotation, the two shadows of the left listener represent two additional positions where the BRTFs were measured. The black triangles represent the loudspeakers.

tails discarded) to improve the robustness, as suggested in [37, 38]. Additionally, when the filter is generated for a new head position, the target pressure is reselected corresponding to that position because reapplying the old target might degrade the isolation performance. The filters were computed with 8,192-size fast Fourier transform and exported to 8,192 taps to avoid any discrepancy in the performance due to using short filters designed with frequency-domain and time-domain methods [12].

As required by head-tracked rendering, the generated PSZ filters need to be adapted to new head positions. In the evaluation, the filter was re-generated for every new position for the full resolution of spatial sampling; for lower resolutions, the filters were only generated for a sparse grid of positions, and then they were assigned to the positions outside the sampling grid in a nearest-neighbor manner, i.e., no interpolation was performed between the filters when there was a mismatch between the plant BRTF and its assigned filter.

3 SPATIAL SAMPLING FOR TRANSLATIONAL MOVEMENTS

The authors first analyze the dependencies of BRTF spatial sampling on various factors for head translations. As shown in Fig. 2 on the left, the plant BRTFs of the two listeners were measured when the left listener was moving along six 90-cm–long trajectories: three in the X direction (left/right) and three in the Y direction (front/back). The full BRTF set was measured at a resolution of 1 cm between adjacent sampling points, which was further downsampled to resolutions of 3, 5, and 10 cm to form the sparser grids in the evaluation. The resulting performance is represented by IZI and IPI calculated at different sampling resolutions. Recall that there are two sets of IZI and IPI metrics defined in Sec. 1.2; here, only the results for IZI_2 and IPI_2 (the subscripts of which are hereafter neglected) are presented because the position-dependent target pressure for the left *BZ* also affects the sound energy in *BZ* and therefore complicates the analysis of the isolation performance. In this context, *IZI* and *IPI* can be regarded as the isolation performance for a moving *DZ* (with a static *BZ*) and that for a moving *BZ* (with a static *DZ*), respectively. All the IZI and IPI spectra to be presented were processed with 1/3-octave logarithmic smoothing [39] for better visualization and band-limited to 200–7,000 Hz due to signal-to-noise ratio limitations and the working range of the transducers.

The results of the "perfectly matched" case, where the filter was updated for every new position (at the highest resolution), are shown in Fig. 3 for both X and Y translations. The IZI and IPI are shown in the figures as functions of the position of the left listener (see Fig. 2) and the frequency of rendered audio, with black contour lines marking 20 dB of isolation, which is taken to be the threshold of isolation performance. It can be seen that above 500 Hz, IZI and IPI achieve above 20 dB irrespective of the listener position at most frequencies. This implies that, ideally, a desired level of isolation can be preserved over a relatively large rendering area. At lower frequencies, however, the isolation threshold is not always achieved due to the limitations of the system (e.g., room effects, loudspeaker array layout, and number of loudspeakers), and a dependency of IZI and IPI on the head position is also observed. Specifically, IZI and IPI levels start to decrease as the left listener moves closer to the right listener in the X direction, likely due to the occlusion and scattering effects of the other listener; however, such a trend is not observed for the Y translations, meaning that front/back movements have a lower impact on the low-frequency isolation performance. At frequencies roughly above 1 kHz, the authors notice some interference-like patterns for the IPI corresponding to X translations and a monotonic decrease in IPI as the left listener moves away from the loudspeakers in the Y direction. This is likely due to the changes in the target pressure for the left *BZ* as the left listener moves in both directions.

Next, the authors show how the expected isolation level is affected by applying filter sets that were generated from the downsampled plant BRTF grids. Figs. 4 and 5 show the IZI and IPI maps with sparse filter sets (updated every 3, 5, and 10 cm) for the same two trajectories as in the previous case. It can be seen that as the spatial sampling grid becomes sparser, the isolation level drops more rapidly between the two matched filters. The authors note that IPI is generally more robust than IZI, because the former is associated with the static *DZ*, which is indirectly affected by head movements of the moving *BZ*, and the latter corresponds to the moving *DZ* and, therefore, is less robust against head misalignments. Comparing X and Y translations, a lower robustness is found in the IZI for Y translations than that for X translations (e.g., by comparing the size of areas above 20 dB for the plots with 10-cm resolution), but the opposite is true for IPI. From these observations, it is concluded that Y translations can have a higher impact on the isolation of the moving *DZ* (corresponding to IZI) but a lower impact on that of the static *DZ* (corresponding to IPI), compared with X translations. However, this conclusion may not hold for PSZ systems with different array layouts.

The optimal resolution, which leads to a minimum of 20-dB isolation above 500 Hz for most head positions, not

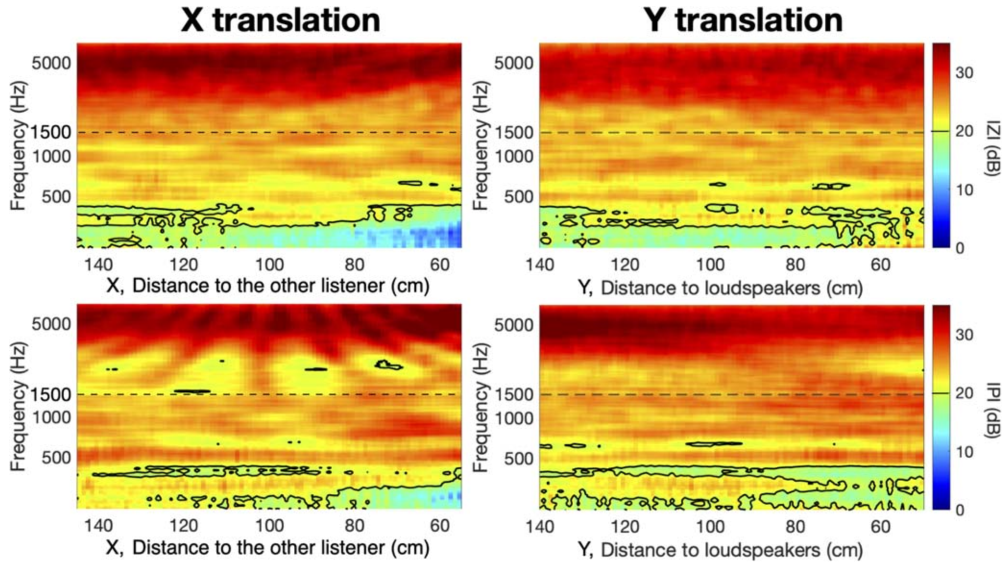

Fig. 3. Simulated IZI (top row) and IPI (bottom row) for the X (left column) and Y (right column) translations with PSZ filters generated at the highest resolution (1 cm), i.e., the "perfectly matched" case.

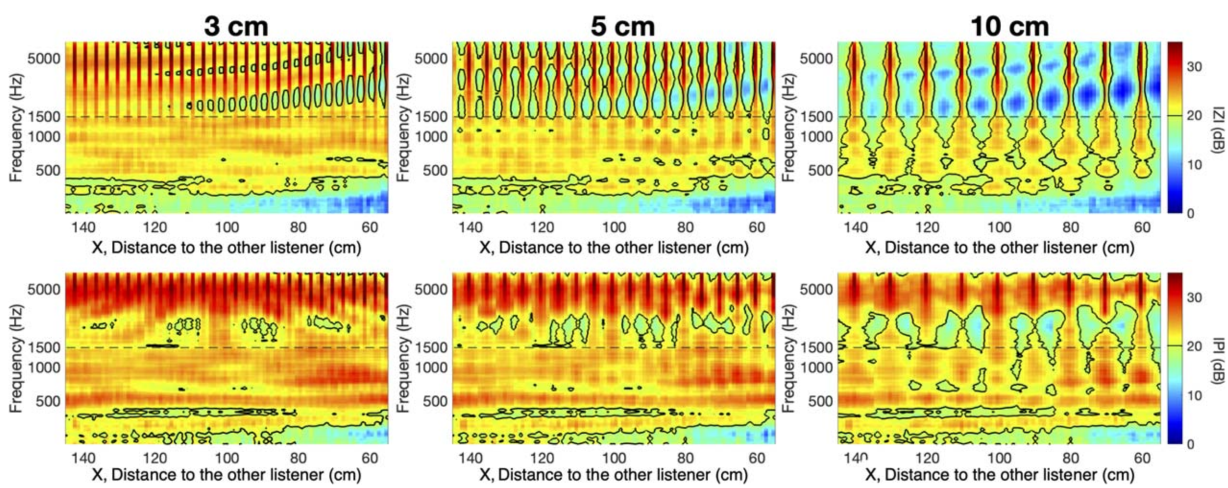

Fig. 4. Simulated IZI (top) and IPI (bottom) for X translations with filters generated using downsampled BRTFs at 3-cm (left column), 5-cm (middle column), and 10-cm (right column) resolutions.

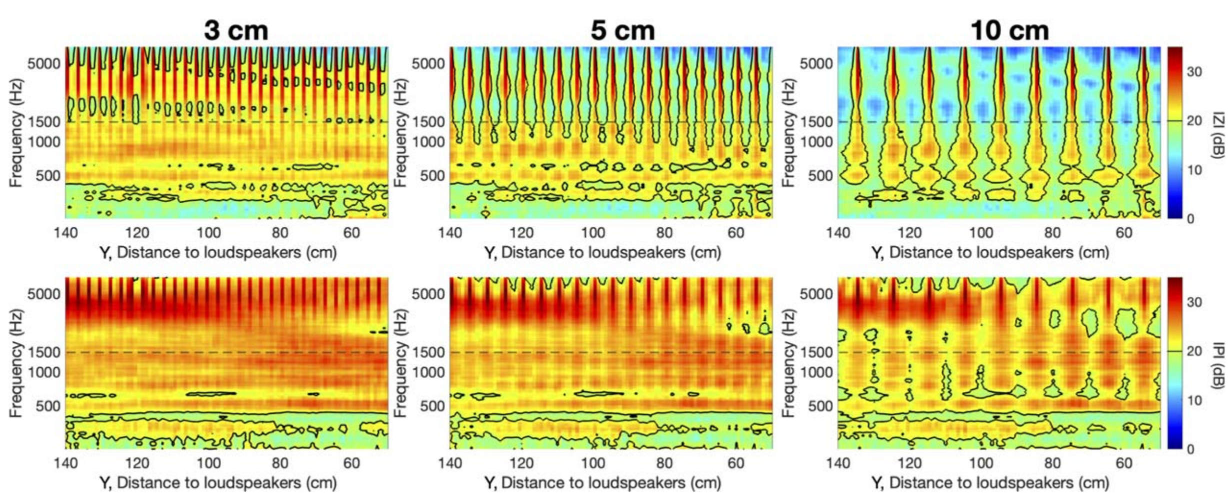

Fig. 5. Simulated IZI (top) and IPI (bottom) for Y translations with filters generated using downsampled BRTFs at 3-cm (left column), 5-cm (middle column), and 10-cm (right column) resolutions.

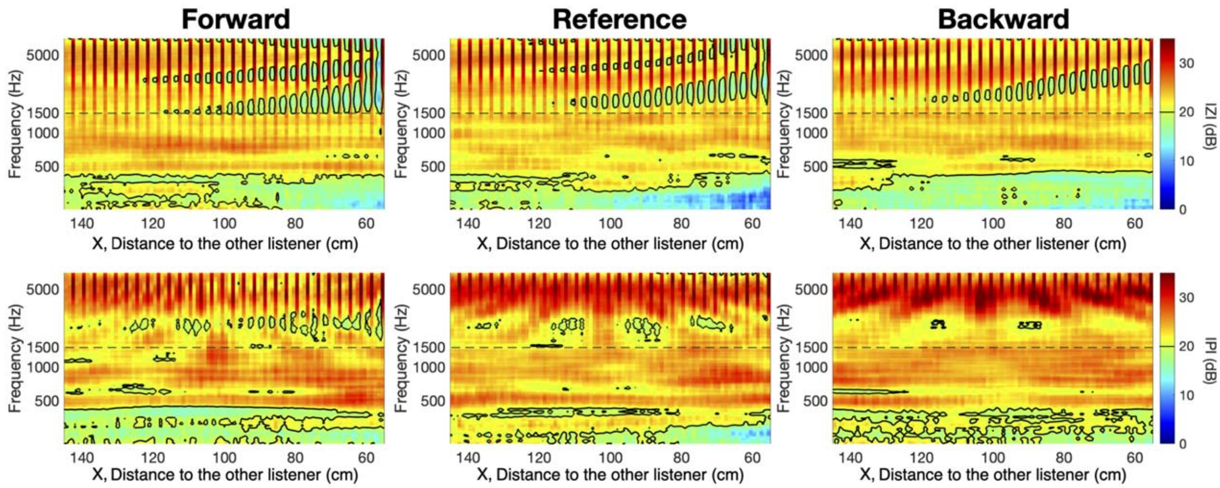

Fig. 6. Simulated IZI (top row) and IPI (bottom row) for X translations with three *y*-axis offsets of the left listener: 50 cm forward (left column), center reference (middle column), and 40 cm backward (right column), at 3-cm spatial sampling resolution.

only depends on the target zone/listener but is also strongly affected by the head position as well as the rendering frequency band. For example, from Fig. 4 it is noted that, while a resolution of 5 cm is sufficient for sampling both a moving *DZ* (indicated by IZI) and a moving *BZ* (indicated by IPI) in the X direction, for frequencies between 500 and 1,500 Hz (marked by the dashed horizontal line in the plots), it does not apply to higher frequencies above 1,500 Hz, unless the resolution is increased to 3 cm or higher. Particularly for IZI, the required resolution to maintain the isolation level increases as the left listener moves closer to the right one, meaning that the optimal spatial sampling grid would be non-uniformly distributed along the X direction. For Y translations (Fig. 5), the difference between the optimal sampling resolutions for IZI and IPI is also found: 5 cm is sufficient for IPI at most frequencies, while IZI requires 3 cm or higher. Compared to X translations, however, the spatial non-uniformity is not observed along the Y direction.

Lastly, the isolation performance corresponding to different position offsets is compared. Figs. 6 and 7 show the results for the three parallel trajectories in each moving direction, with a 3-cm resolution of plant spatial sampling. The authors note that although the robustness at frequencies below 1,500 Hz remains mostly unchanged for different moving trajectories, the higher-frequency robustness shows a clear dependency on the listener position. In Fig. 6, it can be seen that as the left listener moves away from the loudspeakers, the robustness regarding both IZI and IPI increases, meaning that the optimal resolution decreases by increasing the distance to the loudspeakers. This is expected as near-field BRTFs generally have more variations caused by head movements than far-field ones. In Fig. 7, a similar trend appears as the distance between the two listeners decreases. These observations further corroborate the finding that the optimal spatial sampling resolution is strongly dependent on the position of the listener.

4 SPATIAL SAMPLING FOR ROTATIONAL MOVEMENTS

Following the same approach as in the previous section, the authors now focus on the case of azimuthal (or yaw) head rotations. The BRTF set corresponding to the whole-body rotations of the left listener was measured at three different listener positions (center, 45 cm to the right, and 50 cm forward), with a span of 180° from left to right and a resolution of 1°, as illustrated in the right part of Fig. 2. In the evaluation, the sampling grid was downsampled to resolutions of 15°, 30°, and 45°.

Fig. 8 shows the IZI and IPI maps with filter sets updated at 15°, 30°, and 45° with both listener centered. It can be seen that for IZI, the shape of the "lobes" above 1,500 Hz remains the same across the rotations within each sub-figure, meaning that the filter robustness against head rotations is not dependent on the azimuth angle. Similar to the case of translations, IPI is more robust compared to IZI at all resolutions because the former is associated with the static *DZ*. Furthermore, the authors note that below 1,500 Hz, a resolution of 15° is sufficient for preserving the IZI level above 20 dB, while a higher resolution is needed for higher frequencies. For the IPI, however, a discontinuity is seen at large azimuth angles (e.g., around –50° and 75°), which is not present in IZI. This is likely due to the scattering effects from the left listener's body when moving toward or against the right listener.

The isolation performance is then compared for different positions of the left listener, for the case of 15° resolution, as shown in Fig. 9. As the left listener moves to the right, a large decrease is seen in IZI between lobes at extreme frequencies, similar to the case of *x*-axis offset regarding Y translation (see the right column of Fig. 7). The effects also resemble those in the case of translation at low frequencies as the two listeners move close to each other. Such degradation is less apparent in IPI for most frequencies above 500

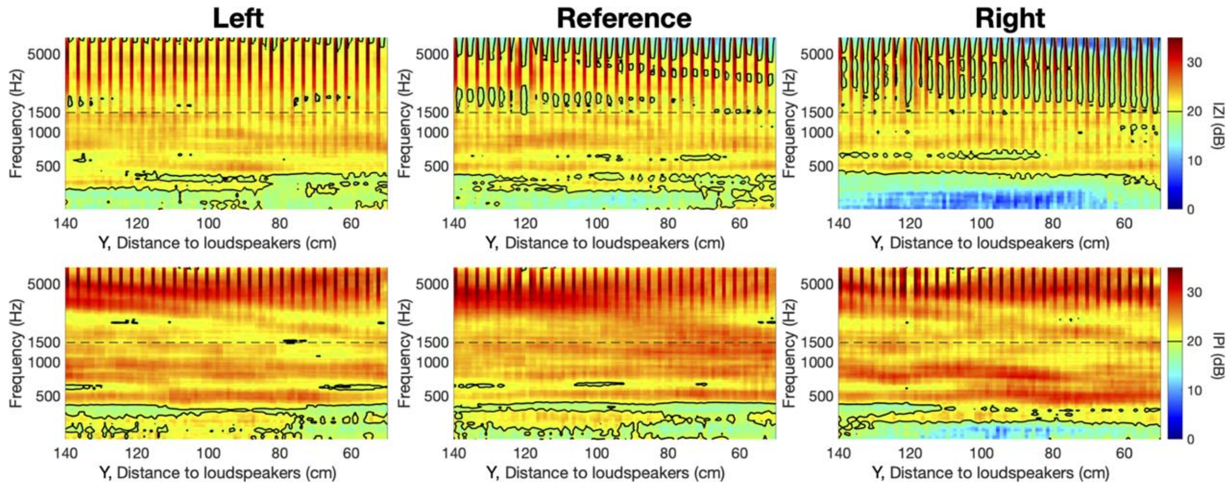

Fig. 7. Simulated IZI (top row) and IPI (bottom row) for Y translations with three *x*-axis offsets of the left listener: 45 cm to the left (left column), center reference (middle column), and 45 cm to the right (right column), at 3-cm spatial sampling resolution.

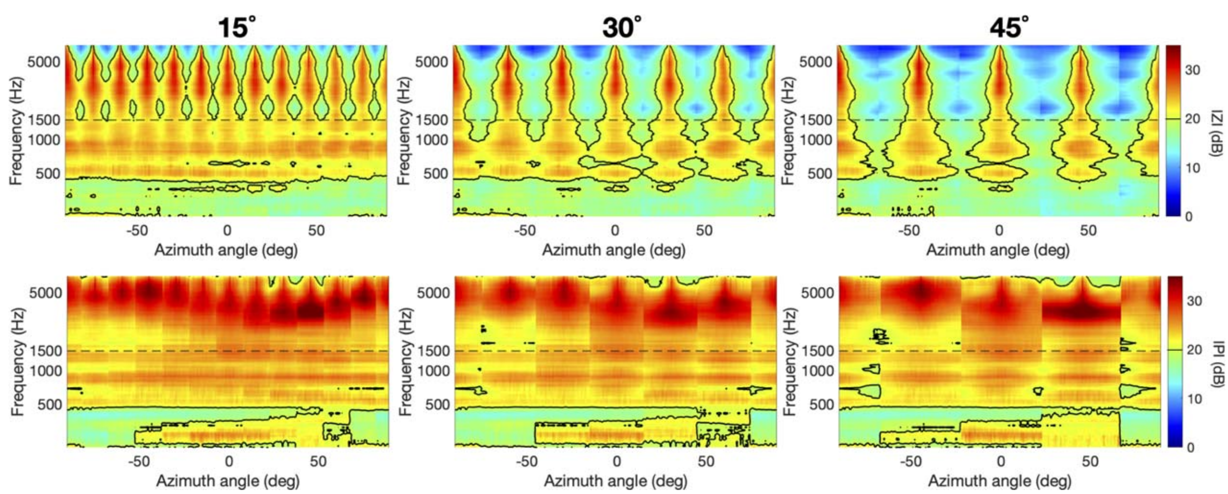

Fig. 8. Simulated IZI (top) and IPI (bottom) for azimuthal rotations with filters generated using downsampled BRTFs at 15° (left column), 30° (middle column), and 45° (right column) resolutions.

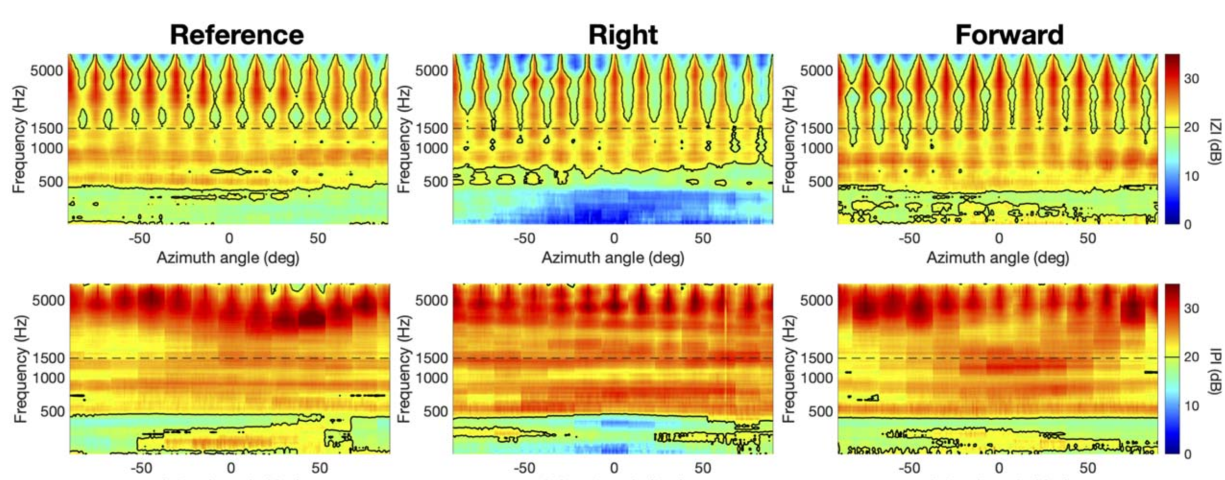

Fig. 9. Simulated IZI (top row) and IPI (bottom row) for azimuthal rotations at three positions of the left listener: center reference (left column), 45 cm to the right (middle column), and 50 cm forward (right column), at 15° spatial sampling resolution.

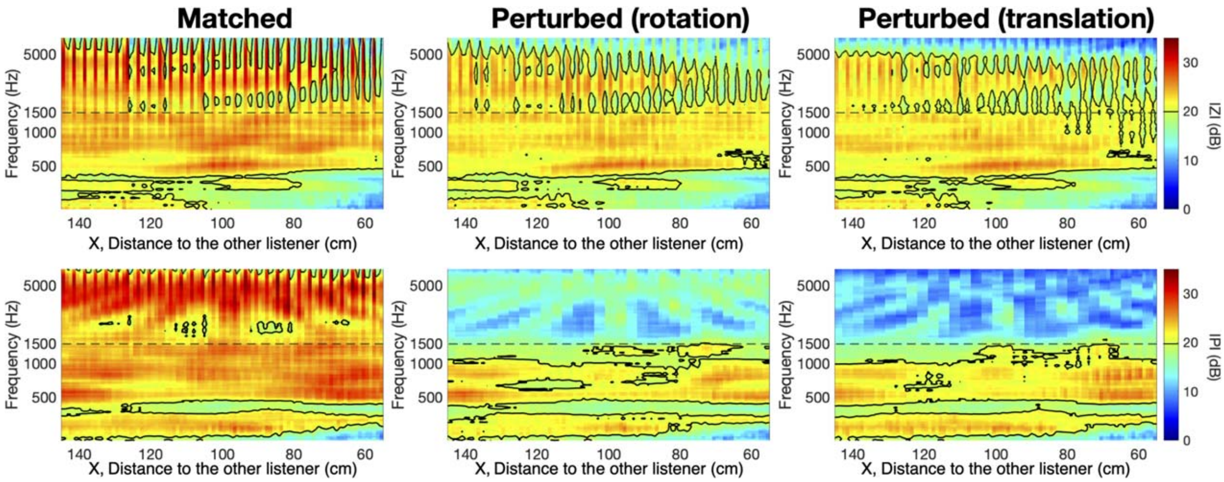

Fig. 10. Simulated IZI (top) and IPI (bottom) for X translations at three conditions: matched (left column), right listener rotated (middle column), and and right listener translated (right column), with 3-cm spatial sampling resolution.

Hz. When the left listener moves forward, a slight change is seen in the shape of the lobes above 1,500 Hz; however, the degradation in filter robustness is not as severe as that when the listener moves to the right. This is aligned with the findings for the case of translation with position offsets (i.e., Figs. 6 and 7). Furthermore, as the left listener moves closer to the right listener, IZI becomes more angle-dependent than that for the center reference position, because the lobes gradually move below 1,500 Hz as the listener moves to extreme angles. As is also pointed out in the analysis of Fig. 6, this is likely caused by larger BRTF variations as the listener moves into the near field.

5 EFFECTS OF PLANT BRTF PERTURBATIONS

In practical applications, both listeners can inevitably move at the same time. Therefore, it is important to analyze the effects of one listener's movement, which causes perturbation in plant BRTFs, on the isolation performance of both listeners (assuming no filter adaptation for the perturbation). The plant BRTFs corresponding to the left listener's X, Y translations and azimuthal rotations were remeasured, respectively, and the filters were regenerated at 3-cm resolution for translations and 15° resolution for rotations. Then, two additional sets of plant BRTFs were measured for the same movement types of the left listener as before but with the right listener's position displaced: one with the head slightly rotated and the other slightly translated, both by a small amount (around 5° in rotation and 5 cm in translation). The purpose of such displacement is to qualitatively analyze the influence of the isolation performance for the left listener. These two sets were convolved with the filters generated for the original set to evaluate the isolation performance under perturbation.

Figs. 10–12 show the IZI and IPI for the left listener's X, Y translations and rotations, respectively, each including three cases of filter/plant combinations (plant matched with the filters, plant with the right listener rotated, and plant

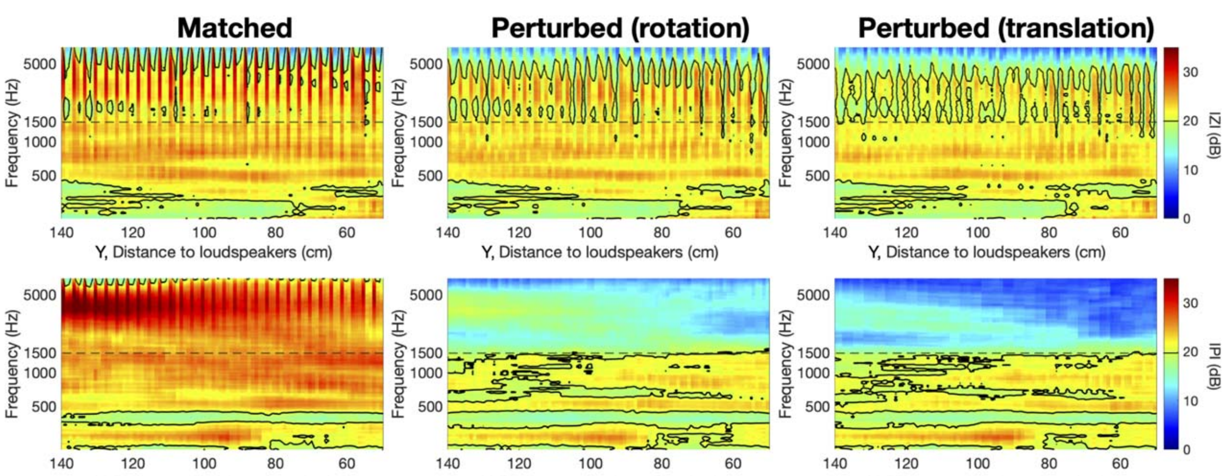

Fig. 11. Simulated IZI (top) and IPI (bottom) for Y translations at three conditions: matched (left column), right listener rotated (middle column), and and right listener translated (right column), with 3-cm spatial sampling resolution.

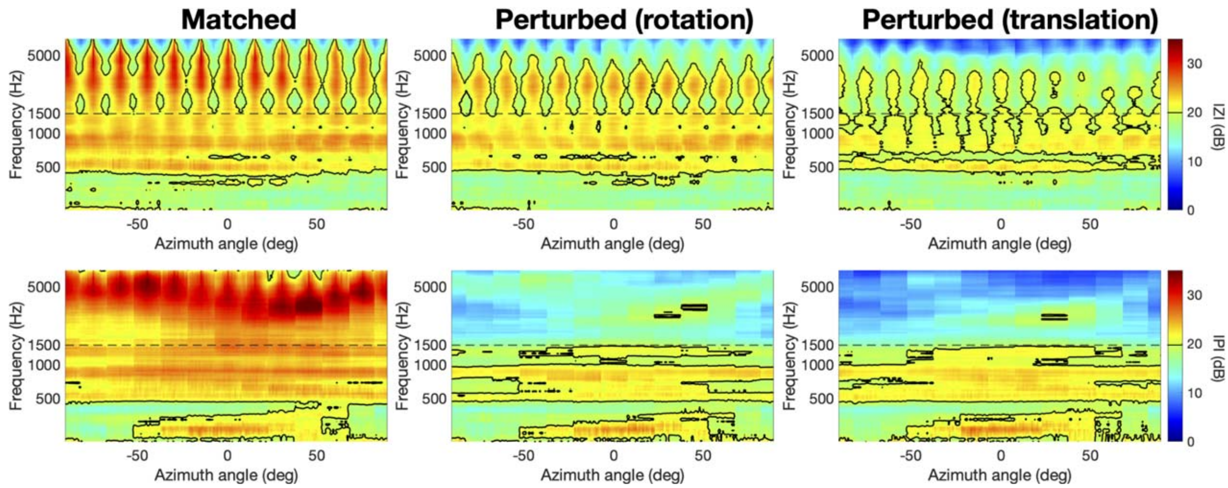

Fig. 12. Simulated IZI (top) and IPI (bottom) for azimuthal rotations at three conditions: matched (left column), right listener rotated (middle column), and right listener translated (right column), with 15° spatial sampling resolution.

with the right listener translated). Irrespective of the movement type, the authors note that, by comparing the matched case (the figures in the left columns) with the perturbed cases (the figures in the middle and right columns), the perturbations decrease the IZI and IPI levels significantly, especially at high frequencies. For IZI, the effects of perturbation mostly exist at frequencies above 1,500 Hz, where the isolation in the relatively high-frequency band (around 5,000 Hz) is impaired; they are more obvious under more challenging conditions, such as when the two listeners are very close for the X translation. The authors note that, from Fig. 12, the translation of the right listener also affects the filter robustness at most frequencies as the lobes become narrower and discontinuous, while such is not observed for the rotation perturbation.

Compared with IZI, IPI is more severely degraded by the perturbations: the isolation level above 1,500 Hz decreases to below 20 dB for all the cases, and for the translation perturbation, it is even below 10 dB in some moving areas. As pointed out in the previous sections, the filters need to be updated less than every 3 cm for translations and every 15° for rotations to maintain the isolation level at high frequencies, and such degradation is more likely because the perturbations have exceeded the movement range within which the filters are still robust, rather than due to the influence of the filter/plant mismatch caused by the left listener.

6 DISCUSSION

It is useful to discuss the extent to which the above findings can be relevant to other PSZ systems. First, different values of IZI and IPI are expected if the filters were designed with methods other than PM, e.g., ACC, because the latter can lead to higher isolation than the former but at a cost of phase distortion [2, 9]. Besides, other system-specific factors that may affect the results include filter design parameters (e.g., regularization level and target pressure specification), loudspeaker array setup (e.g., layout and the number of transducers), and room acoustics characteristics (e.g., Reverberation Time and Direct-to-Reverberant Ratio). Consequently, the filter robustness and the derived optimal spatial sampling resolution may be changed. Despite the fact that the findings were taken from a particular system, the authors expect that they apply to any PSZ system of similar dimensions, such as the position dependency of the optimal spatial sampling, because they are determined by the fundamental acoustic properties of the system, such as the coupling of two listeners' BRTFs in the near field [19]. Furthermore, in the case of two moving listeners, their corresponding plant BRTFs can also be sampled with different optimal resolutions as long as *BZ* and *DZ* are not exchanged.

In addition to optimizing plant spatial sampling for different scenarios, the results also shed light on improving the performance of head-tracked PSZ systems through other approaches. For example, the interpolation between plant BRTFs, although not investigated here, can be optimized by posing position-dependent accuracy constraints. Moreover, the real-time plant modeling approaches that utilize adaptive filtering [31] can also be improved by defining the required convergence time based on the listener position. Lastly, the observed transitional frequency at around 1,500 Hz can be used as a natural crossover, above which the isolation performance can be traded off for higher robustness by using more generalized plant modeling (e.g., the analytical plant models in [23, 24]). It should be noted, however, that the two crossover frequencies (500 and 1,500 Hz) discovered may change in practical PSZ systems due to the differences between the mannequin listener and real listeners, and therefore, onsite measurements are needed as calibration.

The authors emphasize that due to the limitation of the mannequin listeners used in the measurements, the listener's movement in this work implies the movement of both the head and torso. Therefore, some findings in the study may not directly apply to situations where head movements are more frequent than torso movements, especially those mainly affected by torso scattering at lower frequencies.

7 CONCLUSION

Head tracking is critical to improving the robustness of the PSZ system, which relies on controlling the sound pressure around listeners' ears. To achieve head-tracked rendering, the PSZ filters must be updated as a function of the changes in the plant transfer functions, either due to the simple movements of one listener or due to the perturbations and/or acoustic scattering from the other listener. Through measurements of the plant BRTFs under different conditions of listener movements and simulations of the isolation performance at multiple spatial sampling resolutions, the authors found that to achieve a given isolation level, the spatial sampling of plant BRTFs is dependent on multiple factors, including the type of listener movement (translations or rotations), the position of the listener in a system, the rendering frequency band, the targeted listener, and the effects of the other listener in *BZ*. More specifically, in the context of the PSZ system evaluated in this work, the effects of such factors can be summarized as follows:

- Type of movement: Different types of movement change the plant BRTFs in different ways, which in turn lead to different spatial sampling resolutions, especially at high frequencies.
- Position of the listener: In general, a denser spatial sampling of the plant BRTFs is required as the listener moves closer to either the other listener or the loudspeakers, due to the increased dominance of near-field effects in the plant BRTFs.
- Frequency of the rendered audio: The higher the frequency, the denser the spatial sampling is required to be. Furthermore, two natural crossover frequencies exist at around 500 and 1,500 Hz: the former determines a lower bound, below which the PSZ filters are not as effective due to the physical limitations of the system, and the latter determines an upper bound, above which the robustness of the filters is easily affected by slight listener movements.
- Targeted listener: A listener in *DZ* generally requires a denser spatial sampling than that in *BZ* to achieve a similar level of isolation.
- Perturbation caused by the other listener: The slight movements of the listener in *BZ*, when not compensated for by the filters, can degrade the isolation level in *DZ*. Small perturbations can decrease the isolation level at extreme frequencies but with minimal effects on the filter robustness, while larger ones can impair both the isolation and robustness.

Such rules can serve as guiding principles for the practical implementation of head-tracked PSZ systems, particularly in adopting a nonuniform plant sampling grid and optimizing the trade-offs between isolation performance and filter robustness. Future work will focus on incorporating listeners' subjective preferences for audio-on-audio interference into both filter design and performance analysis.

8 ACKNOWLEDGMENT

The authors wish to thank K. Tworek for the support of the hardware system used throughout the experiments. This work was supported by a research grant from Masimo Corporation.

9 REFERENCES

[1] W. Druyvesteyn and J. Garas, "Personal Sound," *J. Audio Eng. Soc.*, vol. 45, no. 9, pp. 685–701 (1997 Sep.).

[2] T. Betlehem, W. Zhang, M. A. Poletti, and T. D. Abhayapala, "Personal Sound Zones: Delivering Interface-Free Audio to Multiple Listeners," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 81–91 (2015 Feb.). https://doi.org/10.1109/MSP.2014.2360707.

[3] J. Yang, M. Wu, and L. Han, "A Review of Sound Field Control," *Appl. Sci.*, vol. 12, no. 14, paper 7319 (2022 Jul.). https://doi.org/10.3390/app12147319.

[4] S. J. Elliott, J. Cheer, J.-W. Choi, and Y. Kim, "Robustness and Regularization of Personal Audio Systems," *IEEE Trans. Audio Speech Lang. Process.*, vol. 20, no. 7, pp. 2123–2133 (2012 May). https://doi.org/10.1109/TASL.2012.2197613.

[5] J. Cheer, S. J. Elliott, and M. F. S. Gálvez, "Design and Implementation of a Car Cabin Personal Audio System," *J. Audio Eng. Soc.*, vol. 61, no. 6, pp. 412–424 (2013 Jun.).

[6] L. Vindrola, M. Melon, J.-C. Chamard, and B. Gazengel, "Use of the Filtered-x Least-Mean-Squares Algorithm to Adapt Personal Sound Zones in a Car Cabin," *J. Acoust. Soc. Am.*, vol. 150, no. 3, pp. 1779–1793 (2021 Sep.). https://doi.org/10.1121/10.0005875.

[7] F. M. Heuchel, D. Caviedes-Nozal, J. Brunskog, F. T. Agerkvist, and E. Fernandez-Grande, "Large-Scale Outdoor Sound Field Control," *J. Acoust. Soc. Am.*, vol. 148, no. 4, pp. 2392–2402 (2020 Oct.). https://doi.org/10.1121/10.0002252.

[8] M. Poletti, "An Investigation of 2D Multizone Surround Sound Systems," presented at the *125th Convention of the Audio Engineering Society* (2008 Oct.), paper 7551.

[9] J.-H. Chang and F. Jacobsen, "Sound Field Control With a Circular Double-Layer Array of Loudspeakers," *J. Acoust. Soc. Am.*, vol. 131, no. 6, pp. 4518–4525 (2012 Jun.). https://doi.org/10.1121/1.4714349.

[10] L. Vindrola, M. Melon, J.-C. Chamard, and B. Gazengel, "Pressure Matching With Forced Filters for Personal Sound Zones Application," *J. Audio Eng. Soc.*, vol. 68, no. 11, pp. 832–842 (2020 Nov.). https://doi.org/10.17743/jaes.2020.0058.

[11] J.-W. Choi and Y.-H. Kim, "Generation of an Acoustically Bright Zone With an Illuminated Region Using Multiple Sources," *J. Acoust. Soc. Am.*, vol. 111, no. 4, pp. 1695–1700 (2002 Apr.). https://doi.org/10.1121/1.1456926.

[12] M. F. S. Gálvez, S. J. Elliott, and J. Cheer, “Time Domain Optimization of Filters Used in a Loudspeaker Array for Personal Audio,” *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 23, no. 11, pp. 1869–1878 (2015 Nov.). https://doi.org/10.1109/TASLP.2015.2456428.

[13] M. B. Møller and M. Olsen, “Sound Zones: On Performance Prediction of Contrast Control Methods,” in *Proceedings of the AES International Conference on Sound Field Control* (2016 Jul.), paper P-2.

[14] T. Lee, J. K. Nielsen, and M. G. Christensen, “Signal-Adaptive and Perceptually Optimized Sound Zones With Variable Span Trade-Off Filters,” *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 28, pp. 2412–2426 (2020 Jul.). https://doi.org/10.1109/TASLP.2020.3013397.

[15] J. Brunnström, S. Koyama, and M. Moonen, “Variable Span Trade-Off Filter for Sound Zone Control With Kernel Interpolation Weighting,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1071–1075 (Singapore, Singapore) (2022 Apr.). https://doi.org/10.1109/ICASSP43922.2022.9746550.

[16] P. Coleman, P. J. B. Jackson, M. Olik, et al., â€œAcoustic Contrast, Planarity and Robustness of Sound Zone Methods Using a Circular Loudspeaker Array,” *J. Acoust. Soc. Am.*, vol. 135, no. 4, pp. 1929–1940 (2014 Apr.). https://doi.org/10.1121/1.4866442.

[17] M. Olsen and M. B. Møller, “Sound Zones: On the Effect of Ambient Temperature Variations in Feed-Forward Systems,” presented at the *142nd Convention of the Audio Engineering Society* (2017 May), paper 9806.

[18] M. B. Møller, J. K. Nielsen, E. Fernandez-Grande, and S. K. Olesen, “On the Influence of Transfer Function Noise on Sound Zone Control in a Room,” *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 27, no. 9, pp. 1405–1418 (2019 Sep.). https://doi.org/10.1109/TASLP.2019.2921151.

[19] Y. Qiao and E. Choueiri, “The Effects of Individualized Binaural Room Transfer Functions for Personal Sound Zones,” *J. Audio Eng. Soc.*, vol. 71, no. 12, pp. 849–859 (2023 Dec.). https://doi.org/10.17743/jaes.2022.0109.

[20] F. Olivieri, F. M. Fazi, S. Fontana, D. Menzies, and P. A. Nelson, “Generation of Private Sound With a Circular Loudspeaker Array and the Weighted Pressure Matching Method,” *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 25, no. 8, pp. 1579–1591 (2017 Aug.). https://doi.org/10.1109/TASLP.2017.2700945.

[21] A. Canclini, D. Markovic, M. Schneider, et al., “A Weighted Least Squares Beam Shaping Technique for Sound Field Control,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6812–6816 (Calgary, Canada) (2018 Sep.). https://doi.org/10.1109/ICASSP.2018.8461292.

[22] Y. Cai, M. Wu, and J. Yang, “Sound Reproduction in Personal Audio Systems Using the Least-Squares Approach With Acoustic Contrast Control Constraint,” *J. Acoust. Soc. Am.*, vol. 135, no. 2, pp. 734–741 (2014 Feb.). https://doi.org/10.1121/1.4861341.

[23] M. F. S. Gálvez, D. Menzies, and F. M. Fazi, “Dynamic Audio Reproduction With Linear Loudspeaker Arrays,” *J. Audio Eng. Soc.*, vol. 67, no. 4, pp. 190–200 (2019 Apr.). https://doi.org/10.17743/jaes.2019.0007.

[24] Y. Qiao and E. Choueiri, “Real-Time Implementation of the Spectral Division Method for Binaural Personal Audio Delivery With Head Tracking,” presented at the *151st Convention of the Audio Engineering Society* (2021 Oct.), e-Brief 657.

[25] X. Ma, C. Hohnerlein, and J. Ahrens, “Concept and Perceptual Validation of Listener-Position Adaptive Superdirective Crosstalk Cancellation Using a Linear Loudspeaker Array,” *J. Audio Eng. Soc.*, vol. 67, no. 11, pp. 871–881 (2019 Nov.). https://doi.org/10.17743/jaes.2019.0037.

[26] V. Bruschi, S. Nobili, F. Bettarelli, and S. Cecchi, “Listener-Position Sub-band Adaptive Crosstalk Canceller Using HRTFs Interpolation for Immersive Audio Systems,” presented at the *150th Convention of the Audio Engineering Society* (2021 May), paper 10474.

[27] T. Kabzinski and P. Jax, “An Adaptive Crosstalk Cancellation System Using Microphones at the Ears,” presented at the *147th Convention of the Audio Engineering Society* (2019 Oct.), paper 10307.

[28] B. Masiero and M. Vorländer, “A Framework for the Calculation of Dynamic Crosstalk Cancellation Filters,” *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 22, no. 9, pp. 1345–1354 (2014 Sep.). https://doi.org/10.1109/TASLP.2014.2329184.

[29] J. Lindfors, J. Liski, and V. Välimäki, “Loudspeaker Equalization for a Moving Listener,” *J. Audio Eng. Soc.*, vol. 70, no. 9, pp. 722–730 (2022 Sep.). https://doi.org/10.17743/jaes.2022.0020.

[30] M. B. Møller and J. Østergaard, “A Moving Horizon Framework for Sound Zones,” *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 28, pp. 256–265 (2019 Nov.). https://doi.org/10.1109/TASLP.2019.2951995.

[31] Z. Sipei and I. S. Burnett, “Adaptive Personal Sound Zones Systems With Online Plant Modelling,” presented at the *Proceedings of the 24th International Congress on Acoustics (ICA)* (Gyeongju, South Korea) (2022 Oct.).

[32] N. Canter and P. Coleman, “Delivering Personalised 3D Audio to Multiple Listeners: Determining the Perceptual Trade-Off Between Acoustic Contrast and Cross-Talk,” presented at the *150th Convention of the Audio Engineering Society* (2021 May), paper 10452.

[33] Y. Qiao and E. Choueiri, “Optimal Spatial Sampling of Plant Transfer Functions for Head-Tracked Personal Sound Zones,” presented at the *154th Convention of the Audio Engineering Society* (2023 May), paper 10659.

[34] Y. Qiao, L. Guadagnin, and E. Choueiri, “Isolation Performance Metrics for Personal Sound Zone Reproduction Systems,” *JASA Express Lett.*, vol. 2, no. 10, paper 104801 (2022 Oct.). https://doi.org/10.1121/10.0014604.

[35] A. Novak, P. Lotton, and L. Simon, “Synchronized Swept-Sine: Theory, Application, and Implementation,” *J. Audio Eng. Soc.*, vol. 63, no. 10, pp. 786–798 (2015 Oct.). https://doi.org/10.17743/jaes.2015.0071.

[36] A. Farina, “Simultaneous Measurement of Impulse Response and Distortion With a Swept-Sine Technique,” presented at the *108th Convention of the Audio Engineering Society* (2000 Feb.), paper 5093.

[37] M. Ebri, N. Strozzi, F. M. Fazi, A. Farina, and L. Cattani, "Individual Listening Zone With Frequency-Dependent Trim of Measured Impulse Responses," presented at the *149th Convention of the Audio Engineering Society* (2020 Oct.), paper 10409.

[38] V. Molés-Cases, S. J. Elliott, J. Cheer, G. Piñero, and A. Gonzalez, "Weighted Pressure Matching With Windowed Targets for Personal Sound Zones," *J. Acoust. Soc. Am.*, vol. 151, no. 1, pp. 334–345 (2022 Jan.). https://doi.org/10.1121/10.0009275.

[39] J. G. Tylka, B. B. Boren, and E. Y. Choueiri, "A Generalized Method for Fractional-Octave Smoothing of Transfer Functions That Preserves Log-Frequency Symmetry," *J. Audio Eng. Soc.*, vol. 65, no. 3, pp. 239–245 (2017 Mar.). https://doi.org/10.17743/jaes.2016.0053.

THE AUTHORS


Yue Qiao


Jessica Luo


Edgar Choueiri

Yue Qiao is a Ph.D. student in the 3D Audio and Applied Acoustics (3D3A) Laboratory at Princeton University where he conducts research on personal sound zone and binaural sound reproduction through loudspeakers and headphones. He received a B.S. degree in Physics at Peking University in 2019. Yue's research interests include sound field control, spatial audio reproduction, and array signal processing.

•

Jessica Luo is a graduate student in the Music Technology program in the Department of Music and Performing Arts Professions at New York University where she is focusing on immersive audio research and music production. She received her B.S. degree in Audio and Music Engineering in 2023 from the University of Rochester. Jessica's research interests include immersive audio perception and acoustics.

•

Edgar Choueiri is a professor of applied physics at the Mechanical and Aerospace Engineering department of Princeton University and associated faculty at the Department of Astrophysical Sciences. He heads Princeton's Electric Propulsion and Plasma Dynamics Lab and the 3D Audio and Applied Acoustics Lab. Edgar's research interests include plasma physics, plasma propulsion, acoustics, and spatial audio.