Y. Qiao and E. Y. 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 December).
DOI: https://doi.org/10.17743/jaes.2022.0109

The Effects of Individualized Binaural Room Transfer Functions for Personal Sound Zones

YUE QIAO, *AES Student Member,* **AND EDGAR Y. CHOUEIRI**
(yqiao@princeton.edu) (choueiri@princeton.edu)

3D Audio and Applied Acoustics Laboratory, Princeton University, Princeton, NJ

The extent to which the performance of personal sound zone (PSZ) reproduction systems is impacted by the individualization of Binaural Room Transfer Functions (BRTFs) and the coupling between the listeners' BRTFs was investigated experimentally. Such knowledge can be valuable for deriving rules for the design of high-performance, robust PSZ systems. The performance of a PSZ system consisting of eight frontal mid-range loudspeakers was objectively evaluated with PSZ filters designed using individualized BRTFs of a human listener and generic ones measured from a mannequin head, in terms of Inter-Zone Isolation, Inter-Program Isolation, and robustness against slight head misalignments. It was found that when no misalignments were introduced, Inter-Zone Isolation and Inter-Program Isolation are improved by an average of around 4 dB at all frequencies between 200 and 7,000 Hz by the individualized filters, compared to the generic ones. With constrained head misalignments, the robustness of both filters decreases as the frequency increases, and although the individualized filters maintain higher performance, their robustness above 2 kHz is lower than that of the generic ones. The evaluation also reveals an inter-listener BRTF coupling effect and a detrimental impact on the performance for both listeners when a single listener's BRTF is mismatched.

0 INTRODUCTION

Personal sound zone (PSZ) [1] (or personal audio) reproduction aims to deliver, using loudspeakers, individual audio programs to multiple listeners in the same physical space with minimum audio-on-audio interference between programs. For a particular audio program, two listening zones are usually generated, one "bright" zone (*BZ*) where the target program is rendered and one "dark" zone (*DZ*) where the sound pressure level corresponding to this program is minimized. When there are different programs delivered to each listener, by the principle of superposition, the *BZ* for one program is also at the same time the *DZ* for the other program, and vice versa.

In order to minimize the sound pressure level in *DZ* while preserving the audio quality in *BZ*, the Pressure Matching (PM) method [2] is usually applied, which involves specifying the target pressure at control points in both zones and minimizing the l^2-norm errors between the target pressure and the actual pressure generated by loudspeakers. In recent research, the PM method has been modified to better address the trade-off between audio quality and acoustic isolation [3–5] and to accommodate constraints on reconstruction error [6], filter impulse response [7], and choice of control points [8]. Although the PM formulation is usually cast in the frequency domain [3, 4, 7, 8], it has also been reformulated in the time domain [9–11].

Due to the nature of inverse filtering, the performance of a PSZ system based on PM heavily relies on the accuracy of the acoustic transfer functions (TFs) used for designing the filter (referred to as *setup* TFs in the paper). If mismatch exists between the setup TFs and those in the final evaluation (referred to as *playback* TFs), the PSZ performance (e.g., cancellation of interfering audio in *DZ*) is expected to be degraded. There are many factors that can potentially contribute to such TF mismatches, and previous studies have examined the effects of the presence/absence of the listener in the zone [5], mismatched sound speed [12], loudspeaker/microphone positions [1, 12–14], electro-acoustic responses of loudspeakers [14], and the existence of background noise [15]. In the context of automotive audio, which is one of the most common scenarios for PSZ applications, additional factors such as varying ambient temperature [16], number of passengers [17], and seat positions [18] are also investigated.

Although some aspects of PSZ systems can be evaluated with free-field [5, 7, 18] or mannequin head microphones [17, 19], the ultimate performance of such systems is best studied with actual listeners located in desired positions, often called *sweet spots*. If the filters are designed with

setup TFs different from those of an actual listener, the TF mismatch may result in degraded performance perceived by the listener. In [20], using mismatched filters, only around 10 dB of Acoustic Contrast was achieved with actual listeners, which is considered poor performance because the required non-distracting audio interference level is shown to be above 20 dB [21].

The importance of using individualized Head-Related Transfer Functions (HRTFs) has been well recognized in other spatial audio applications, and validated mainly in terms of localization accuracy [22, 23]. For instance, in binaural audio reproduction with loudspeakers using crosstalk cancellation (XTC) [24], which is conceptually similar to PSZ reproduction [12] and often utilizes the same inverse filtering methods, individualization effects have also been studied [25, 26] for localization performance. Although XTC research has shown benefits of HRTF individualization [25, 26], the performance metrics are different from those for PSZ systems, and the required XTC level is different from the isolation level in PSZ systems [21]. Moreover, a PSZ system differs from a typical XTC system in two other aspects: 1) more loudspeakers are required for the former, increasing the variability of TFs for a mismatched individual, and 2) the presence of multiple listeners can lead to coupling effects between listeners' HRTFs, especially in the near field.

In this paper, the authors experimentally investigate the effects of HRTF individualization on PSZ reproduction using an in-house PSZ system in a typical listening room. The control points are defined at the listeners' ears, resulting in compact, "ear-targeting" sound zones similar to those in [7, 17, 18] and those for XTC utilizing in-ear measurements [27, 28]. Such a PSZ setup allows direct measurements and comparison of the Binaural Room Transfer Functions (BRTFs), which consist of HRTFs convolved with room responses, of both the mannequin head and the human listener, as opposed to those in [3, 4, 8–10] in which larger zones with dense control points are specified. After the BRTFs are measured with in-ear binaural microphones and then used as setup TFs to generate generic and individualized PSZ filters, the performance of the PSZ system were experimentally evaluated with metrics the authors recently proposed in [29]: Inter-Zone Isolation (IZI) and Inter-Program Isolation (IPI), as well as robustness against slight head misalignments.

The objective evaluation was conducted with a setup of a human listener and reference mannequin head in two zones, with two sound zone configurations in which the *BZ* and *DZ* are exchanged. For each configuration, both generic and individualized filters were evaluated. The frequency range of interest was chosen as 200–7,000 Hz due to signal-to-noise ratio limitations and the working range of the transducers. Two cases were considered in order to evaluate both aspects of PSZ performance: one in situ case, in which the listener stays at the same position for both the capture of setup BRTFs and the measurement of the performance of generated individualized filters, and one ex situ case, in which the listener is instructed to leave the seat and sit back to the position where the setup BRTFs were measured.

This paper presents and extends the authors' prior work [30]: SEC. 1 introduces the PM method and its adaptation used for PSZ filter generation. SEC. 2 explains the PSZ system setup and the procedure for evaluating the performance of the implemented system. SEC. 3 shows the evaluation results from different performance perspectives. Discussion on the results and remarks on the choice of design parameters and evaluation methods are provided in SEC. 4. Lastly, in SEC. 5, the authors draw conclusions on the findings and suggest future directions.

1 METHODS FOR PSZ FILTER GENERATION

The authors consider a PSZ system for two listeners in two zones, having, in general, an array of L loudspeakers and M control points. In the frequency domain, each loudspeaker l has a complex gain of $g_l(\omega)$, $l = 1, \ldots, L$, and the resulting sound pressure at each control point m is $p_m(\omega)$, $m = 1, \ldots, M$. In this particular system, the control points are defined right at the ear positions; therefore, $M = 4$, for two listeners. The TF corresponding to the loudspeaker l and the control point m is denoted as H_{ml}. Then, the pressure at the control points is given by

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

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

1.1 Pressure Matching Method

In the PM method [2], given the specified target pressure $\boldsymbol{p}_T$ at the control points in *BZ* and *DZ*, the least-square cost function J is constructed as

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

and by minimizing J, the complex gains $\boldsymbol{g}^*$ corresponding to the optimal PSZ filter are given by

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

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

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

1.2 Optimal Filter Design

Apart from Eq. (4), a more common approach in practice to both the uniqueness of the solution and the numerical stability is through regularization, which consists of adding the loudspeaker energy term $\|\boldsymbol{g}\|^2$ to Eq. (2) as an additional cost with weighting β, yielding the modified optimal solution

$$\tilde{\boldsymbol{g}}^* = (\boldsymbol{H}^H\boldsymbol{H} + \beta\boldsymbol{I})^{-1}\boldsymbol{H}^H\boldsymbol{p}_T, \tag{5}$$

where β is often referred to as the regularization parameter and $\boldsymbol{I}$ is the identity matrix. Coleman et al. [12] showed that

the regularization parameter greatly affects the robustness performance, and it is also strongly frequency-dependent. However, because the type of TF mismatch considered in this paper is directly related to actual listeners and is difficult to be simulated, it would be time-consuming to find the optimal β through numerous measurements. Instead, the authors adopt a probabilistic approach similar to that used in [15], by assuming each TF as an independent and identically distributed random variable and minimizing the expected cost. More specifically, the TF H_{ml} is modeled as

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

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

$$\phi_{ml} \sim N(\hat{\phi}_{ml}, \sigma^2_{\phi,ml}), \tag{8}$$

where A_{ml} and ϕ_{ml} denote the amplitude and phase of the TF, $N(\cdot, \cdot)$ denotes the normal distribution, and the hat symbol and σ denote the mean and standard deviation, respectively. The corresponding cost function is expressed as

$$J_{prob} = \mathbb{E}\{\|\boldsymbol{H}\boldsymbol{g} - \boldsymbol{p}_T\|^2\}, \tag{9}$$

where $\mathbb{E}\{\cdot\}$ denotes taking the expectation, and $\boldsymbol{H}$ contains all the random variables H_{ml}. Its closed-form optimal solution is given by

$$\boldsymbol{g}^*_{prob} = (\hat{\boldsymbol{H}}^H \hat{\boldsymbol{H}} + \boldsymbol{\Sigma})^{-1} \hat{\boldsymbol{H}}^H \boldsymbol{p}_T, \tag{10}$$

where $\hat{\boldsymbol{H}}$ contains all the expected values of H_{ml}, and $\boldsymbol{\Sigma}$ is expressed as

$$\boldsymbol{\Sigma} = \sum_{m=1}^{M} \boldsymbol{\Sigma}_m, \tag{11}$$

$$\boldsymbol{\Sigma}_m = \mathrm{diag}\{\sigma^2_{A,m1}, \cdots, \sigma^2_{A,mL}\}. \tag{12}$$

It should be noted that only the standard deviation of the amplitude is included in the expression; therefore, it is sufficient to only consider the amplitude variation in obtaining the optimal PSZ filter. In [15], the variance is obtained through Bayesian inference, whereas here, the variance is determined empirically from multiple TF measurements.

2 EXPERIMENTAL DESIGN

2.1 System Setup

The evaluation experiment was conducted with a PSZ system built in a typical listening room ($RT_{60} \approx 0.24$ s in the range 1,300–6,300 Hz). Because no subjective evaluation is included within the scope of this paper, all BRTF measurements were taken with a single listener (male, 25 years old).

The testing system, shown in Fig. 1, is composed of an 8-driver linear, horizontal loudspeaker array working in mid-range (two 16-driver linear tweeter loudspeaker arrays shown in the figure were not used for the study). Two Brüel & Kjær Head and Torso Simulators (HATS, Type 4100) are used as the mannequin heads (with the built-in microphones removed), and two pairs of in-ear binaural microphones (Theoretica Applied Physics BACCH-BM Pro)

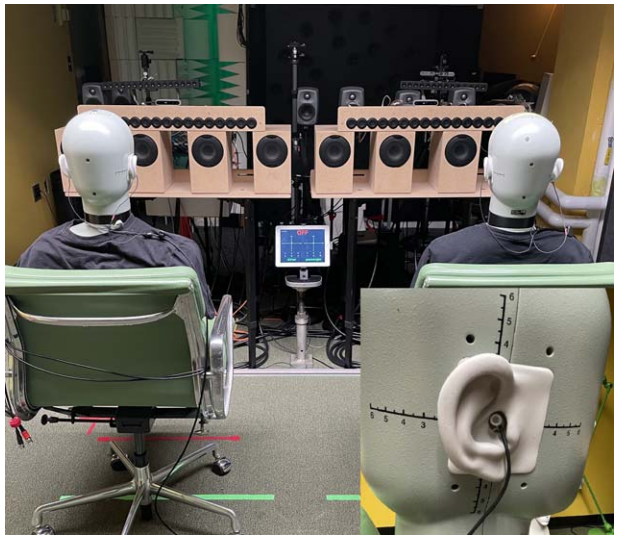


Fig. 1. Photographs of the PSZ reproduction system under study with two mannequin heads (the two tweeter arrays above the eight-driver array and the loudspeakers behind were not included in the study). Bottom right: the in-ear binaural microphone used in BRTF measurements.

are level matched and free-field equalized and used for measuring the BRTFs of both the human listener and mannequin head. The listener's head position is tracked using an infrared depth sensor (Intel RealSense D415), and both the head displacement (in *xyz* coordinates) and orientation (pitch/yaw/roll) are displayed in real time on a tablet screen to help the listener maintain the head position during the measurement. The listener and mannequin head are approximately 1 m away from the loudspeaker array.

All the BRTFs are measured using a series of exponential sine sweeps [31] at 48-kHz sampling frequency, with each sweep having a duration of 2 s. For PSZ filter design, the corresponding impulse responses are deconvolved and truncated to the first 2,048 samples with a Tukey window ($R = 0.05$). All filter impulse responses are centered and similarly truncated to 4,096 samples before being exported. For evaluation, the measured impulse responses are truncated to the first 8,192 samples.

The focus is to study the influence of mismatch in the BRTFs of a single listener. Because there are usually two listeners involved in the system, only one of them (right in the figure) is always the HATS as reference, whereas the other is interchange between the HATS and human listener. Two sound zone configurations (SZCs), as illustrated in Fig. 2, are considered in the evaluation:

1) The left listener (HATS or human) in *DZ* and the right listener (reference HATS) in *BZ*.
2) The right listener (HATS or human) in *BZ* and the right listener (reference HATS) in *DZ*.

For the human-HATS setup, two types of filters are evaluated within each SZC:

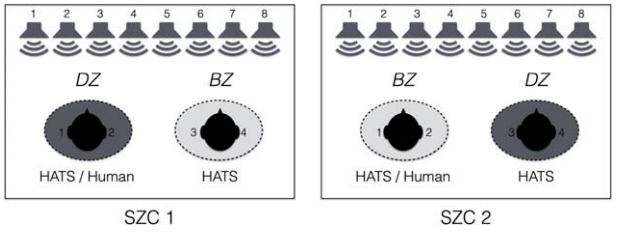


Fig. 2. Scheme of the two SZCs. The light/dark shaded area represents the designated bright/dark zone.

a) Generic Filter: The filter generated with the setup BRTFs measured from the HATS-HATS setup.
b) Individualized Filter: The filter generated with the setup BRTFs measured from the same setup.

To generate PSZ filters, the target pressure in *BZ* is chosen as the summed responses (truncated to first 2,048 samples) at two ears when a stereo pair of loudspeakers (for the listener, it is the first and fourth loudspeakers from the left, and for the HATS, it is the fifth and eighth loudspeakers) are driven in phase, and the target pressure in *DZ* is set to zero. In practice, this corresponds to a mono audio program, because it usually leads to higher isolation performance than other multichannel programs [29]. To determine the variance matrix in Eq. (10), the BRTFs are measured consecutively 20 times for the human-HATS case. For each measurement, the listener is instructed to leave the seat, reposition himself at the origin, and re-wear the binaural microphones, in order to generate PSZ filters that are robust against slight head movements. Then, the variance matrix is assembled by taking the empirical variance of this set of BRTFs.

2.2 Evaluation Procedure

2.2.1 System Calibration and Filter Preparation

First, the system is calibrated by placing two HATS at the specified geometry center and setting the *xyz* coordinates in the head-tracking display to zero. After a 4×8 matrix of BRTFs for the two HATS is measured, the left HATS is removed and replaced by the human listener. Then, the BRTFs for the human-HATS setup are measured consecutively 20 times to determine the variance matrix in Eq. (10). The generic filters for the HATS-HATS setup are generated based on the derived variance matrix.

2.2.2 In Situ Measurement

The BRTFs for the human listener and reference HATS are measured. Then, the filters corresponding to this set of setup BRTFs are generated offline and loaded in the rendering program. Then, with the listener remaining at the same position (in situ), two sets of overall TFs (BRTFs convolved with generated PSZ filters) are measured corresponding to the two filter configurations *a*, *b*, and each includes two filters for SZC 1 and 2. This way, the setup and playback BRTFs can be assumed to be near-identical, and the individualized filters are expected to achieve the best possible performance.

2.2.3 Ex Situ Measurement

To evaluate the robustness of the individualized filter against possible head misalignments, ten additional measurements of the overall TFs are taken. The generic filters and the individualized filters generated from the in situ measurement are used. The listener is asked to leave the seat and come back to the specified origin before the next measurement.

During each measurement, the listener is instructed to remain still around the specified origin until the measurement is finished. In practice, however, because no external head-supporting device is used, it is difficult for the listener's head to stay at the exact same position. Therefore, a range of maximum allowable head displacement/rotation is specified (displacement less than 1 cm in either *x*/*y*/*z* direction and rotation less than 10° around either axis), beyond which the measurement is aborted and a new one is restarted. For the ex situ measurement, the authors expect such an experimental setup to reasonably approximate a few "randomly sampled" head misalignments in practice. The body posture of the listener is not explicitly controlled in the study.

3 RESULTS

The authors first examine the differences in measured BRTFs and then evaluate the filter performance in terms of isolation between zones (IZI) and between programs (IPI) and filter robustness against slight head misalignments. All results shown below are processed with one-third–octave complex smoothing [32] for better visualization.

3.1 BRTF Variances

Fig. 3 shows the magnitude responses (mean and standard deviation) of a few selected BRTFs from the first set of 20 measurements under the human-HATS setup, normalized by those measured under the HATS-HATS setup. The order of indices m, l in H_{ml} is defined such that $m = 1, 4$ correspond to the listener's left ear and the reference HATS' right ear, and $l = 1, 8$ correspond to the leftmost and the rightmost loudspeaker (from the perspective of the listener in the figure) in the array.

From $\{|H_{11}|, |H_{18}|\}$ and $\{|H_{21}|, |H_{28}|\}$ in the figure, it is clear that the listener's BRTFs are different from those from the HATS, and the difference increases at higher frequencies. From the plots of $\{|H_{31}|, |H_{38}|\}$ and $\{|H_{41}|, |H_{48}|\}$, changes are observed in the reference HATS's BRTFs, even though the reference HATS is fixed the whole time. This demonstrates that varying one listener's BRTFs can also affect that of the other listener, most likely due to the scattering (both reflection and diffraction) of the sound off the hard torso and head of the HATS. This is further corroborated by observing the variation when the loudspeaker is closer to the human listener (comparing $|H_{31}|$, $|H_{41}|$ against $|H_{38}|$, $|H_{48}|$).

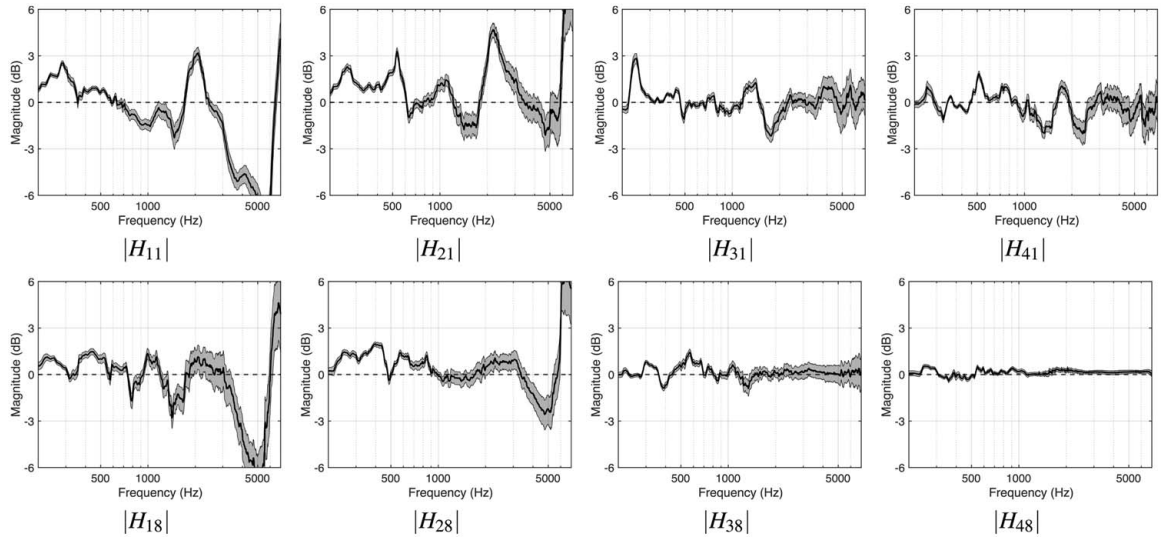


Fig. 3. Normalized magnitude responses of selected BRTFs measured from the human-HATS setup (20 measurements in total). The solid line and surrounding shaded area represent the mean and standard deviation. The BRTFs from the HATS-HATS setup are used as the reference, therefore showing a flat response of 0 dB (the dashed line).

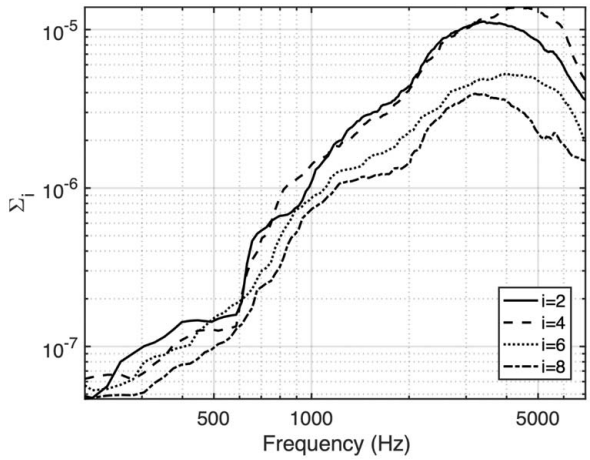


Fig. 4. Frequency-dependent values of selected elements of the matrix $\boldsymbol{\Sigma}$ in Eq. (10). Each index i corresponds to each loudspeaker index l.

Additionally, Fig. 4 shows the values of selected elements of the 8×8 diagonal matrix $\boldsymbol{\Sigma}$ in Eq. (10) used for filter generation, corresponding to the loudspeakers $l = 2, 4, 6, 8$. The values are determined by $\boldsymbol{\Sigma}_i = M \cdot \max_m \tilde{\sigma}^2_{A,ml}$, where $\tilde{\sigma}^2_{A,ml}$ is the empirical variance of $|H_{ml}|$ from the 20 consecutively measured, un-normalized TFs. Here, the matrix $\boldsymbol{\Sigma}$ works as frequency-dependent regularization that compensates the larger variance at higher frequencies. The authors also observe that the variances for loudspeakers closer to the replaced listener are noticeably higher at frequencies above 1 kHz than those of the further loudspeakers ($l = 2$, 4 compared against $l = 6$, 8).

3.2 The IZI and IPI Metrics

IZI and IPI indicate the isolation performance of a PSZ system from two complementary perspectives: the isolation between two *zones* given a single program and the isolation between two active *audio programs* in the same zone. IZI is conceptually similar to the well-known Acoustic Contrast [12] metric, which is proven to be a special case of IZI for single-channel programs [29]; IPI is more relevant to the perception of audio-on-audio interference [19, 33] than IZI and is suitable for scenarios in which multiple programs are simultaneously rendered. Following the general definition, the system setup in this paper corresponds to rendering two *mono* programs to two zones. To express the two metrics, the authors first define the PSZ filters corresponding to SZC 1 and 2 as $\boldsymbol{g}^*_{1,2}$, and the sub-matrices (i.e., the top/bottom two rows) of the TF matrix $\boldsymbol{H}$ corresponding to either the left or right listener because $\boldsymbol{H}_{L,R}$. IZI for SZC 1 and 2 can be expressed as

$$IZI_1 = \frac{\|\boldsymbol{H}_R\boldsymbol{g}^*_1\|^2}{\|\boldsymbol{H}_L\boldsymbol{g}^*_1\|^2}, \qquad IZI_2 = \frac{\|\boldsymbol{H}_L\boldsymbol{g}^*_2\|^2}{\|\boldsymbol{H}_R\boldsymbol{g}^*_2\|^2}, \tag{13}$$

in which case it is equivalent to the Acoustic Contrast metric. IPI for the left/right listener comparing two SZCs is given by

$$IPI_L = \frac{\|\boldsymbol{H}_L\boldsymbol{g}^*_2\|^2}{\|\boldsymbol{H}_L\boldsymbol{g}^*_1\|^2}, \qquad IPI_R = \frac{\|\boldsymbol{H}_R\boldsymbol{g}^*_1\|^2}{\|\boldsymbol{H}_R\boldsymbol{g}^*_2\|^2}. \tag{14}$$

3.3 PSZ Filter Performance for the In Situ Case

Fig. 5 shows the in situ–measured IZI and IPI with respect to the different SZCs and listeners, which also correspond to the best case scenario. The authors note that the use of individualized filters enhances both IZI and IPI at all frequencies of interest (200–7,000 Hz) by around 4 dB (logarithmically averaged across all four difference curves). The improved IZI and IPI achieve a minimum of 15 dB at most frequencies, with a logarithmic average (over the entire frequency band) of 19.9 (IZI_1)/19.5 (IZI_2)/18.7 (IPI_L)/21.2 (IPI_R) dB, approaching the established 25.6-

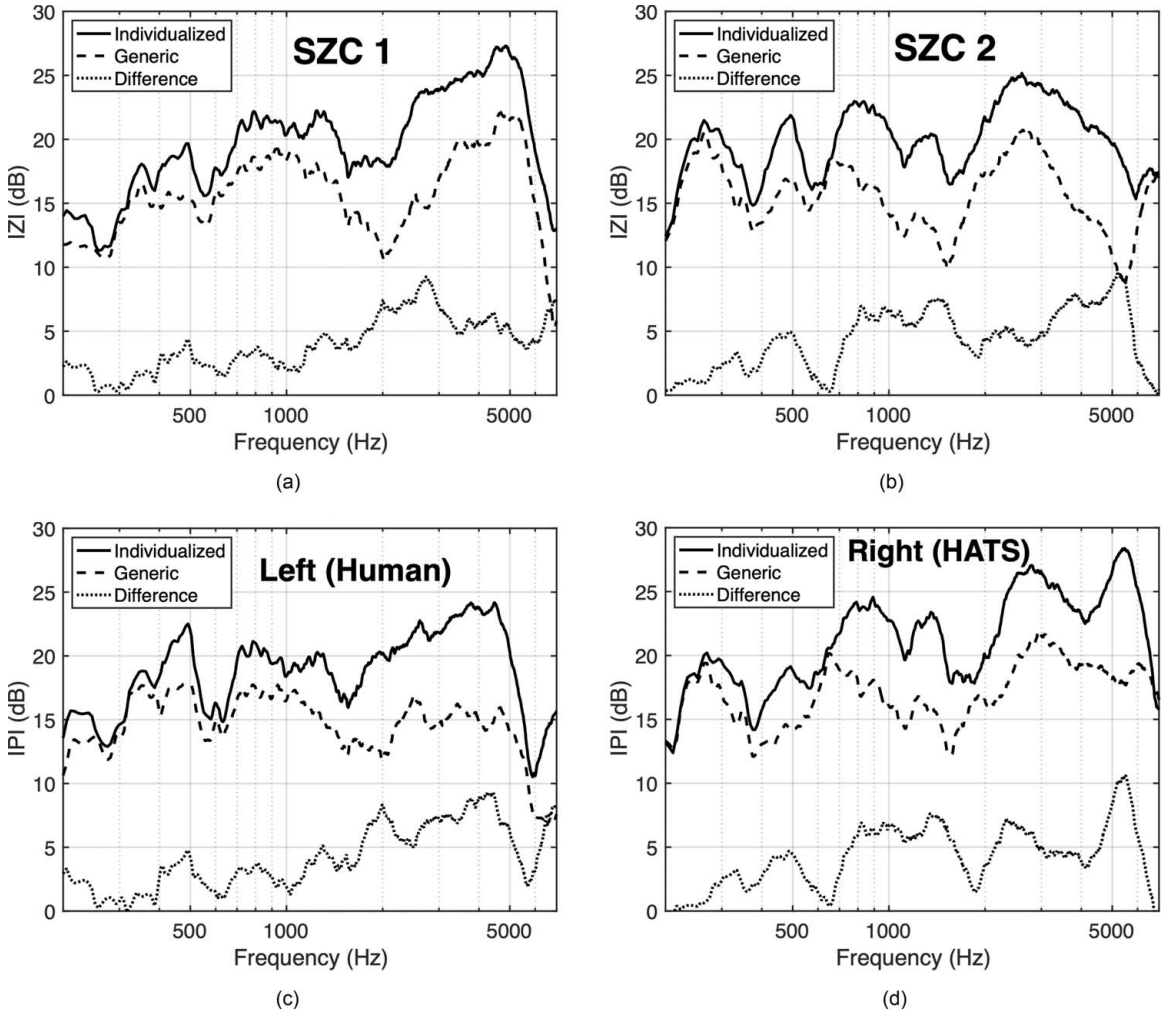


Fig. 5. In situ–measured Inter-Zone Isolation (top row) and Inter-Program Isolation (bottom row) for the generic filters (in dashed lines) and the individualized filters (in solid lines), along with their differences (in dotted lines). (a) and (b) represent sound zone configurations 1 and 2, respectively. (c) and (d) represent left (human) and right (HATS) listeners, respectively.

dB threshold for non-distracting audio interference [21]. Additionally, the improvement is observed to be generally larger as the frequency increases (e.g., the 10-dB peak seen at around 5 kHz for both IPI curves). This is because the optimization problem in PSZ filter design becomes less ill-conditioned as the wavelength decreases, and the TFs become more independent of each other.

Further insight can be obtained by comparing the IZI and IPI plots for different SZCs. First, by comparing the plots in each row, differences are noted between the metrics for two zones/listeners. Such differences are likely caused by the room/system asymmetry and the BRTF differences between the human listener and HATS. Next, comparing the column pairs, similar trends (especially at lower frequencies) are seen between IZI for SZC 1 and IPI for HATS, as well as between IZI for SZC 2 and IPI for the human listener, which is due to the same *DZ* being shared in the definition of both metrics.

It is worth noting that the improvement is observed in IZI and IPI for both SZCs/listeners even though only one listener was replaced. In other words, the change of one listener's BRTFs can equally affect the best possible performance for both listeners. Based on the results in SEC. 3.1, it can be assumed that the loudspeakers farther away from the listener play a more important role in canceling the interfering audio, because their BRTFs have the most variances compared to the loudspeakers closer to the listener. This also implies that the two listeners' BRTFs are closely correlated and that the individualization of PSZ filters should take into account both listeners, especially when they are close to each other.

3.4 PSZ Filter Performance for the Ex Situ Case

The IZI and IPI spectra from ten ex situ measurements are shown in Fig. 6. The boundaries of shaded areas in the plots represent the minimum and maximum values of all the measured results. Even though there are fluctuations in IZI and IPI due to slight head misalignments, the individualized filters still remain superior to the generic ones at most frequencies. More specifically, there is little overlap between the two shaded areas, indicating that the individ-

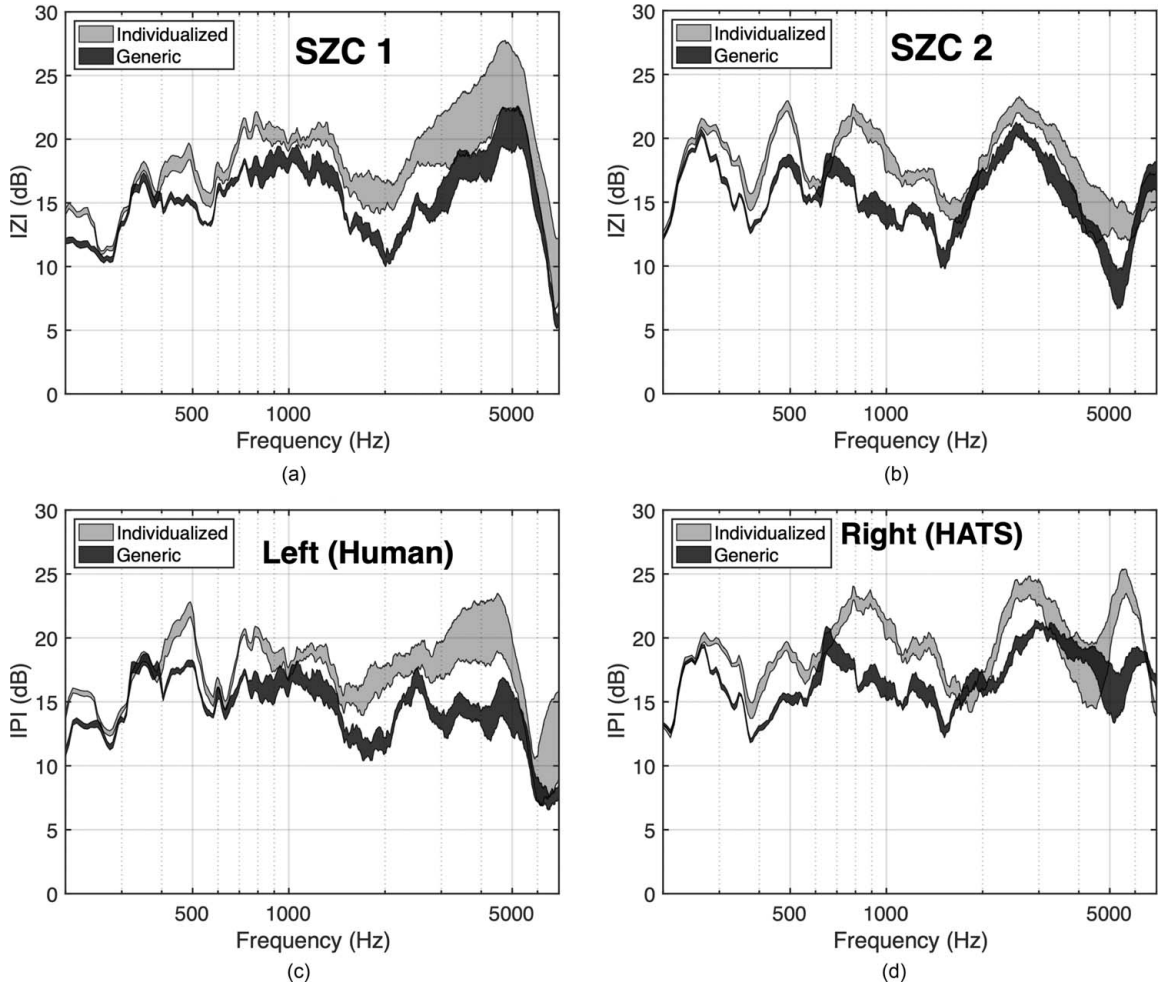


Fig. 6. Ex situ measured Inter-Zone Isolation (top row) and Inter-Program Isolation (bottom row) for the generic filters (in dark shading) and the individualized filters (in light shading). (a) and (b) represent sound zone configurations 1 and 2, respectively. (c) and (d) represent left (human) and right (HATS) listeners, respectively.

ualized filters always lead to better performance than the generic ones under the given constraints of head movements. It is observed that the overlapping mostly appears at frequencies above 2 kHz in IZI for SZC 2 and IPI for HATS in which the reference HATS is in *DZ*. This implies that at higher frequencies, the benefits of individualization (or matching BRTFs) for one listener can easily vanish with the other listener's slight head movements.

Furthermore, the robustness of both PSZ filters is examined against head misalignments by plotting the difference between the minimum and maximum IZI and IPI (or the width of the shaded areas in Fig. 6), as shown in Fig. 7. In general, a larger difference means IZI (or IPI) is more sensitive to head misalignments, and therefore, the corresponding PSZ filter is less robust. A clear trend present in all four plots is that the robustness of both filters decreases as the frequency increases. The robustness is the lowest at around 4.5 kHz, at which the variance between the human listener and HATS' BRTFs reaches local maxima (see Fig. 3).

The robustness of individualized filters is also observed to be worse than that of the generic ones, especially in SZC 1 in which the replaced listener is in *DZ*. In comparison, the robustness for the two listeners is more similar for the two filters designed in SZC 2. Although individualized filters yield better isolation compared to generic ones, it is not clear whether the resulting decrease of robustness (e.g., around 5 dB of Δ_{IZI} or Δ_{IPI} variation at high frequencies) can be perceived as an audible degradation in sound zone isolation.

4 DISCUSSION

From the results shown above, it is clear that using individualized BRTFs for PSZ filter generation leads to better isolation performance, even with slight misalignments of the listener. Nonetheless, the authors would like to point out that such an improvement can be affected by both the acoustic environment and the choice of filter design parameters, and it might be further increased under certain circumstances.

In terms of acoustic environment, the authors emphasize that the results obtained in this paper are based on measurements in a typical listening room. Consequently, the IZI and IPI levels are generally lower than those measured in an anechoic setting (for example, see [7]), where the impact of

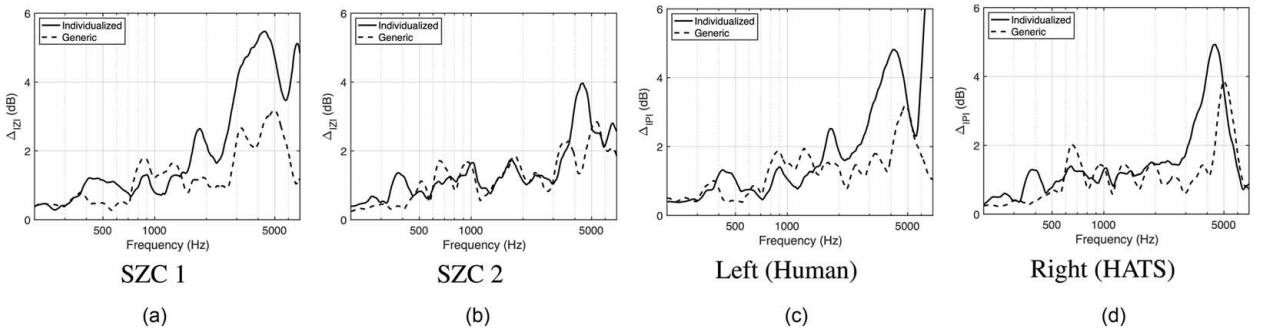


Fig. 7. Differences between the maximum and minimum IZI [(a) and (b)]/IPI [(c) and (d)] from the ex situ measurements of the generic (in dashed lines) and individualized (in solid lines) filters, extracted from Fig. 6. Higher value indicates lower filter robustness. The data is further processed with one-sixth–octave smoothing for better visualization.

room reflections is minimized. Larger differences can also be expected between the generic and individualized filters in an anechoic setting. However, the results under realistic listening conditions can better represent the highest isolation level one can achieve in practical applications. From the results, the authors deduce that it is feasible to implement a PSZ system that achieves a logarithmic average of 20 dB of isolation at listeners' ears.

The choice of filter design parameters also plays an important role in reaching the best possible performance. For example, the window length of the setup TFs used for filter generation, the target pressure specification, and the exported filter length all have significant impact on the effectiveness of the PSZ filters, as discussed in recent literature (e.g., [5, 6, 34]) on the optimization of such parameters. In this work, the parameters were chosen empirically with no further optimization, and therefore, increased improvement is expected once the parameters are optimized.

The authors also note that in order to derive the optimal PSZ filters, they only considered slight movements of the left listener and assumed the reference listener to be static, and thus, the modeled uncertainties of BRTFs would be less than the actual case in which movements of both listeners are possible. In practical use cases, a different uncertainty matrix $\mathbf{\Sigma}$ [in Eq. (10)] should be used, which may result in higher robustness but lower isolation. Such a trade-off is because the more BRTF modeling errors the uncertainty matrix takes into account, the less optimal the resulting PSZ filters are. In practice, the uncertainty should only include errors due to slight head misalignments and the head tracker accuracy, and the robustness against larger movements should be achieved by updating the PSZ filters with head tracking.

The results for SZC 2 imply that, unlike the case of a single-listener XTC system with individualized BRTFs, the individualization is generally more difficult for a two-listener PSZ system due to the scattering effects between listeners, unless the BRTFs of both listeners are captured at the same time. In other words, not only the listener but the surrounding acoustic environment also needs to be individually captured and matched during PSZ reproduction. This undoubtedly complicates the implementation of an individualized system and challenges its practical useability. One possible solution to mitigate this problem is to allow more uncertainty in the formulation of setup TFs to increase filter robustness; it is also possible to capture or estimate the scattered sound field from the other listener prior to filter generation.

The evaluation of other commonly-adopted metrics, such as Array Effort and Normalized Reproduction Error [7], are not discussed in the paper. Because the generic and individualized filters are generated with the same design method and parameters, the performance is expected to be nearly identical in Array Effort for both filters. The reproduction errors for the two filters are not comparable because 1) different target pressure is used to design each filter and 2) the discrepancy between the setup BRTFs (late-reverb tails truncated) and the playback BRTFs would result in large reproduction errors mostly due to phase differences, which are hard to gain much useful insight from. Therefore, a more suitable metric for evaluating the reproduced audio quality in reverberant conditions is required.

5 CONCLUSION

In this work, the authors presented an objective evaluation of a PSZ reproduction system using the generic and individualized BRTFs, under the setup of an eight-driver linear loudspeaker array for two listeners in a typical listening room. The BRTFs of a) two mannequin heads and b) a human listener and a mannequin head were measured, and the corresponding PSZ filters were generated using the PM method, with a statistical design approach for optimal robustness against constrained head misalignments. The resulting TFs convolved with different PSZ filters were measured both in situ and ex situ and evaluated in terms of two isolation performance metrics, IZI, IPI, and filter robustness against possible head misalignments.

The objective evaluation shows improvement in terms of higher IZI and IPI by replacing generic BRTFs with individualized ones in PSZ filter generation. From the in situ–measured results, the individualized filters improve IZI and IPI at all frequencies between 200 and 7,000 Hz when the human listener is in *DZ* by a logarithmic average of 4 dB, achieving around 20 dB of sound zone isolation. From the ex situ measurement, the individualized filters remain supe-

rior to the generic ones, but at a cost of degraded robustness at higher frequencies when the individualized listener is in *DZ*. A similar improvement in isolation is also observed when the mannequin head (the unchanged listener) is in *DZ*, implicating an inter-listener BRTF coupling effect and a great impact on the dark zone sound cancellation for one listener when the BRTFs of the other listener are mismatched due to listener movements or listener replacement. However, the filter robustness for the unchanged listener is not significantly affected.

Although this study is focused on small head movements (within 1 cm of displacement and/or 10° of rotation) in the setup, a natural extension would be to apply head tracking to compensate for larger head movements. Instead of modeling the movements as misalignment errors, the larger movements should be treated as the change in the parameter space for updating the corresponding PSZ filter coefficients. A few adaptive approaches can be utilized, e.g., [18, 27, 35]. Additionally, the decrease in the robustness of individualized filters implies that it is almost impractical to retain the best performance at higher frequencies. This can be potentially addressed by applying loudspeaker beamforming techniques with controllable directivity at higher frequencies (e.g., [36, 37]). Finally, as the next steps for this work, subjective listening tests are required to verify that the superiority of individualized PSZ filters confirmed by the objective evaluation is also perceptible to a large group of human subjects.

6 ACKNOWLEDGMENT

The authors wish to thank K. Tworek for the support on experimental system setup and many inspiring discussions throughout the study.

7 REFERENCES

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

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

[3] 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.

[4] 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 May). https://doi.org/10.1109/TASLP.2017.2700945.

[5] 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.

[6] M. Hu, H. Zou, J. Lu, and M. G. Christensen, "Maximizing the Acoustic Contrast With Constrained Reconstruction Error Under a Generalized Pressure Matching Framework in Sound Zone Control," *J. Acoust. Soc. Am.*, vol. 151, no. 4, pp. 2751–2759 (2022 Apr.). https://doi.org/10.1121/10.0010256.

[7] 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 Dec.). https://doi.org/10.17743/jaes.2020.0058.

[8] 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*, pp. 6812–6816 (Calgary, Canada) (2018 Apr.). https://doi.org/10.1109/ICASSP.2018.8461292.

[9] Y. Cai, M. Wu, and J. Yang, "Design of a Time-Domain Acoustic Contrast Control for Broadband Input Signals in Personal Audio Systems," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 341–345 (Vancouver, Canada) (2013 May). https://doi.org/10.1109/ICASSP.2013.6637665.

[10] 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 Jul.). https://doi.org/10.1109/TASLP.2015.2456428.

[11] 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.

[12] P. Coleman, P. J. Jackson, M. Olik, et al., "Acoustic 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.

[13] J. Mourjopoulos, "On the Variation and Invertibility of Room Impulse Response Functions," *J. Sound Vib.*, vol. 102, no. 2, pp. 217–228 (1985 Sep.). https://doi.org/10.1016/S0022-460X(85)80054-7.

[14] J.-Y. Park, J.-W. Choi, and Y.-H. Kim, "Acoustic Contrast Sensitivity to Transfer Function Errors in the Design of a Personal Audio System," *J. Acoust. Soc. Am.*, vol. 134, no. 1, pp. EL112–EL118 (2013 Jun.). https://doi.org/10.1121/1.4809778.

[15] 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 Jun.). https://doi.org/10.1109/TASLP.2019.2921151.

[16] 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.

[17] W.-H. Cho and J.-H. Chang, "Investigation of Effect on the Acoustic Transfer Function in a Vehicle Cabin According to Change of Configuration," in *Proceedings of the 23rd International Congress on Acoustics*, pp. 5190–5195 (Aachen, Germany) (2019 Sep.).

[18] 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.

[19] J. Rämö, S. Bech, and S. H. Jensen, "Validating a Real-Time Perceptual Model Predicting Distraction Caused by Audio-on-Audio Interference," *J. Acoust. Soc. Am.*, vol. 144, no. 1, pp. 153–163 (2018 Jul.). https://doi.org/10.1121/1.5045321.

[20] 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.

[21] 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.

[22] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization Using Nonindividualized Head-Related Transfer Functions," *J. Acoust. Soc. Am.*, vol. 94, no. 1, pp. 111–123 (1993 Jul.). https://doi.org/10.1121/1.407089.

[23] J. C. Middlebrooks, "Individual Differences in External-Ear Transfer Functions Reduced by Scaling in Frequency," *J. Acoust. Soc. Am.*, vol. 106, no. 3, pp. 1480–1492 (1999 Sep.). https://doi.org/10.1121/1.427176.

[24] E. Choueiri, "Binaural Audio Through Loudspeakers," in A. Roginska and P. Geluso (Eds.), *Immersive Sound: The Art and Science of Binaural and Multi-Channel Studio*, pp. 124–179 (Taylor & Francis, New York, NY, 2018).

[25] M. A. Akeroyd, J. Chambers, D. Bullock, et al., "The Binaural Performance of a Cross-Talk Cancellation System With Matched or Mismatched Setup and Playback Acoustics," *J. Acoust. Soc. Am.*, vol. 121, no. 2, pp. 1056–1069 (2007 Feb.). https://doi.org/10.1121/1.2404625.

[26] P. Majdak, B. Masiero, and J. Fels, "Sound Localization in Individualized and Non-Individualized Crosstalk Cancellation Systems," *J. Acoust. Soc. Am.*, vol. 133, no. 4, pp. 2055–2068 (2013 Apr.). https://doi.org/10.1121/1.4792355.

[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] M. Klunk, *Spatial Evaluation of Cross-Talk Cancellation Performance Utilizing In-Situ Recorded BRTFs*, Master's thesis, University of Salford, Salford, United Kingdom (2022 Apr.).

[29] 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.

[30] Y. Qiao and E. Choueiri, "The Performance of A Personal Sound Zone System With Generic and Individualized Binaural Room Transfer Functions," presented at the *152nd Convention of the Audio Engineering Society* (2022 May), paper 10579.

[31] 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.

[32] 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.

[33] K. Baykaner, P. Coleman, R. Mason, et al., "The Relationship Between Target Quality and Interference in Sound Zone," *J. Audio Eng. Soc.*, vol. 63, no. 1/2, pp. 78–89 (2015 Feb.). https://doi.org/10.17743/jaes.2015.0007.

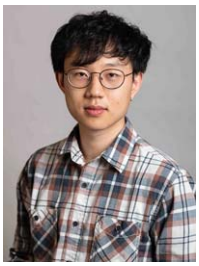
[34] W. Gallian, F. M. Fazi, C. Tripodi, N. Strozzi, and A. Costalunga, "Optimisation of the Target Sound Fields for the Generation of Independent Listening Zones in a Reverberant Environment," in *Proceedings of the Immersive and 3D Audio: From Architecture to Automotive (I3DA)*, pp. 1–10 (Bologna, Italy) (2021 Sep.). https://doi.org/10.1109/I3DA48870.2021.9610888.

[35] 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.

[36] 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.

[37] 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.

THE AUTHORS



Yue Qiao

Edgar Choueiri

Yue Qiao is a Ph.D. student in the 3D Audio and Applied Acoustics (3D3A) Laboratory at Princeton University where he is conducting research on personal sound zone reproduction 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.

•

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.