Audio Engineering Society

Convention Paper 10579

Presented at the 152nd Convention
2022 May, In-Person and Online

The Performance of A Personal Sound Zone System with Generic and Individualized Binaural Room Transfer Functions

Yue Qiao[1] and Edgar Choueiri[1]

[1] *3D Audio and Applied Acoustics Laboratory, Princeton University, Princeton, New Jersey, 08544, USA*

Correspondence should be addressed to Yue Qiao (`yqiao@princeton.edu`)

ABSTRACT

The performance of a two-listener personal sound zone (PSZ) system consisting of eight frontal mid-range loudspeakers in a listening room was evaluated for the case where the PSZ filters were designed with the individualized BRTFs of a human listener, and compared to the case where the filters were designed using the generic BRTFs of a dummy head. The PSZ filters were designed using the pressure matching method and the PSZ performance was quantified in terms of measured Acoustic Contrast (AC) and robustness against slight head misalignments. It was found that, compared to the generic PSZ filters, the individualized ones significantly improve AC at all frequencies (200-7000 Hz) by an average of 5.3 dB and a maximum of 9.4 dB, but are less robust against head misalignments above 2 kHz with a maximum degradation of 3.6 dB in average AC. Even with this degradation, the AC spectrum of the individualized filters remains above that of their generic counterparts. Furthermore, using generic BRTFs for one listener was found to be enough to degrade the AC for both listeners, implicating a coupling effect between the listeners' BRTFs.

1 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 the program is minimized. 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 resulting audio quality and acoustic isolation [3, 4], as well as other constraints such as filter response [5] and choice of control points [6]. While the PM formulation is usually cast in the frequency domain [3, 4, 6, 5], it has also been adapted to time-domain problems [7, 8, 9].

Due to the nature of inverse filtering, the performance of a PSZ system based on PM heavily relies on the accuracy of 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 during the final evaluation (referred to as *playback* TFs), the system performance, such as 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 mismatched sound speed [10], loudspeaker/microphone positions [1, 10, 11, 12], electro-acoustic responses of loudspeakers [12], and the existence of background noise [13]. In the context of automotive audio, which is one of the most common scenarios for PSZ applications, additional factors such as varying ambient temperature [14], number of passengers [15], and seat positions [16] are also investigated.

While some aspects of PSZ systems can be evaluated with free-field or dummy head microphones [5, 15, 16, 17], the ultimate performance of such systems is best studied with actual listeners located in desired positions, often called *sweet spots*. If the system is designed with setup TFs different from those of an actual listener, the TF mismatch may result in degraded performance perceived by the listener. In Ref. [18], around 10 dB of Acoustic Contrast was measured with actual listeners, which is considered poor performance as the required non-distracting audio interference level is shown to be above 20 dB [19].

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 [20, 21]. For instance, in binaural audio reproduction with loudspeakers using crosstalk cancellation, which is conceptually closely-related to PSZ reproduction [10] and often utilizes the same inverse filtering methods, such effects have also been studied [22, 23] for localization performance. While the research on crosstalk cancellation has shown benefits of HRTF individualization [22, 23], the evaluation metrics are different from those for PSZ systems, and different level of thresholds are observed [19]. Moreover, a PSZ system differs from a typical crosstalk cancellation system in two other aspects: *1)* more loudspeakers are required, increasing the variability of TFs for a mismatched individual; and *2)* multiple listeners are present, leading to possible coupling effects between listeners' HRTFs, especially in the near field.

In this paper, the effects of HRTF individualization on PSZ reproduction were experimentally evaluated in terms of 1) frequency-domain Acoustic Contrast (AC) [10] and 2) robustness against slight head misalignments, with a system consisting of a horizontal array of eight frontal loudspeakers in a typical listening room, and the reproduction is limited to zones at the listeners ears (like in Ref. [5, 15, 16], as opposed to larger zones in Ref. [3, 4, 6, 7, 8]). The Binaural Room Transfer Functions (BRTFs), which consist of HRTFs convolved with room responses, of both a dummy head and a human listener were measured with binaural microphones, and then used as setup TFs to generate generic and individualized PSZ filters.

The performance evaluation was conducted with a setup of a human listener and a reference dummy head in two zones, with two sound zone configurations assigning *DZ* to each zone separately. For each configuration, both generic and individualized filters were evaluated. The frequency range of interest was chosen as 200 - 7000 Hz due to the working range of the loudspeaker units. Two cases were considered to evaluate both aspects of PSZ performance: one *in situ* case, where the listener stays at the same position for both the capture of setup BRTFs the measurement of the AC performance of the generated individualized filters; and one *ex situ* case, where the listener is instructed to leave the seat and reposition his head approximately in the sweet spot where the setup BRTFs have been measured.

2 Methods for PSZ Filter Generation

We consider a PSZ system for two listeners in two zones, having an array of L loudspeakers and M control points in total. In the frequency domain, each loudspeaker l has a complex gain of $g_l(\omega), l = 1, \cdots, L$, and the resulting sound pressure at each control point m is $p_m(\omega), m = 1, \cdots, M$. In our 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}, which, in matrix form is

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

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 are implicitly dependent on the frequency ω.

2.1 Pressure Matching Method

In the PM method [2], given the specified target pressure $\mathbf{p}_T$ at the control points in *BZ* and *DZ*, the least-square cost function J 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 overdetermined problems where $L < M$.

2.2 Optimal Filter Design

In most practical implementations, regularization is applied to the solution in Eq. 3 in order to improve both the numerical stability and the robustness against potential TF mismatches. Regularization consists of adding the loudspeaker energy term $\|\mathbf{g}\|^2$ to Eq. 2 as an additional cost with weighting β, yielding the modified optimal solution

$$\tilde{\mathbf{g}}^* = (\mathbf{H}^H\mathbf{H} + \beta\mathbf{I})^{-1}\mathbf{H}^H\mathbf{p}_T, \tag{4}$$

where β is often referred to as the regularization parameter and $\mathbf{I}$ is the identity matrix. Coleman et al. [10] showed that the regularization parameter greatly affects the robustness performance, and it is also strongly frequency-dependent. However, as 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, we adopt a probabilistic approach similar to that used in Ref. [13], by assuming each TF as an independent and identically distributed (i.i.d.) random variable and minimizing the expected cost. More specifically, the TF H_{ml} is modeled as

$$H_{ml} = A_{ml}e^{i\phi_{ml}}, \tag{5}$$

$$A_{ml} \sim N(\hat{A}_{ml}, \sigma^2_{A,ml}), \tag{6}$$

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

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. The corresponding cost function is expressed as

$$J_{prob} = \mathbb{E}\{\mathbf{H}\mathbf{g} - \mathbf{p}_T\}, \tag{8}$$

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

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

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

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

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 optimal loudspeaker gains. In Ref. [13], the variance is obtained through bayesian inference, while here, we determine the variance empirically from multiple TF measurements.

3 Experimental Evaluation

The evaluation experiment was conducted with a PSZ system built in a typical listening room ($RT_{60} \approx 0.2s$). As no subjective evaluation is included within the scope of this paper, all BRTF measurements were taken with a single listener (male, 25 years old).

3.1 System Setup

The testing system, shown in Fig. 1, is composed of an 8-unit linear, horizontal loudspeaker array working in mid-range (two 16-unit linear tweeter loudspeaker arrays working shown in the figure were not used for the study). Two Brüel & Kjær Head and Torso Simulators (HATS, Type 4100) are used as the dummy heads (with the built-in microphones removed), and two pairs of in-ear binaural microphones (Theoretica Applied Physics BACCH-BM Pro) are calibrated and used for measuring the BRTFs of both the human listener and the dummy head. The listener's head position is tracked using a 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 the dummy head are approximately 1 meter away from the loudspeaker array.

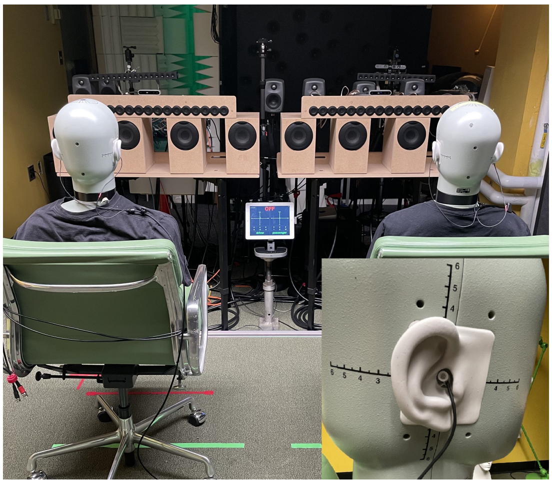


Fig. 1: Photographs of the PSZ reproduction system under study with two dummy heads. Bottom right: the in-ear binaural microphone used in BRTF measurements.

All the BRTFs are measured using a series of exponential sine sweeps [24] at 48 kHz sampling frequency, with each sweep having a duration of 2 seconds. For PSZ filter design, the corresponding impulse responses are truncated to the first 2048 samples with a Tukey window ($R = 0.05$). All filter impulse responses are centered and similarly truncated to 4096 samples before being exported. For evaluation, the measured impulse responses are truncated to the first 8192 samples.

3.2 Experimental Design

Our focus is to study the influence of mismatch in the BRTFs of a *single* listener. As there are usually two listeners involved in the system, only one of them (right in the figure) is always the HATS as reference, while the other is interchanged from the HATS to the human listener. For the human-HATS setup, two sound zone configurations (SZCs) are considered:

1) The human listener in *DZ* and the reference HATS in *BZ*;

2) The human listener in *BZ* and the reference HATS in *DZ*.

Within each sound zone configuration, two filters are evaluated:

a) Generic Filter: The filter generated with the setup BRTFs measured from the two HATS;

b) Individualized Filter: The filter generated with the setup BRTFs measured from the human listener and the reference HATS.

To generate PSZ filters, the target pressure in *BZ* is chosen as the responses (truncated to first 2048 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. To determine the variance matrix in Eq. 9, 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 to slight head movements. Then, the variance matrix is assembled by taking the empirical variance of this set of BRTFs.

The procedures of the experiment are explained as follows:

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-by-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 for 20 times to determine the variance matrix in Eq. 9. The generic filters for the HATS-HATS setup are generated based on the derived variance matrix.

In situ measurement. The BRTFs for the human listener and the 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 SCZ 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 performance.

Ex situ measurement. To evaluate the robustness of the individualized filter against possible head misalignments, 10 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 recenter his head 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, as no external supporting device is used, it is difficult for the listener to keep his head at the exact same position. Therefore, a range of maximum allowable head displacement/rotation is specified (displacement less than 1cm in either X/Y/Z direction and rotation less than 10 degrees around either axis), beyond which the measurement is aborted and a new one is restarted. The body posture of the listener is not explicitly controlled in the study.

3.3 Results

We first examine the differences in measured BRTFs, and then evaluate the filter performance in terms of AC and filter robustness against slight head misalignments. All results shown below are processed with 1/3-octave complex smoothing [25] for better visualization.

3.3.1 BRTF Differences

Fig. 2 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 $|H_{11}|$ and $|H_{18}|$ 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 $|H_{41}|$ and $|H_{48}|$, we observe changes 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_{41} against H_{48}).

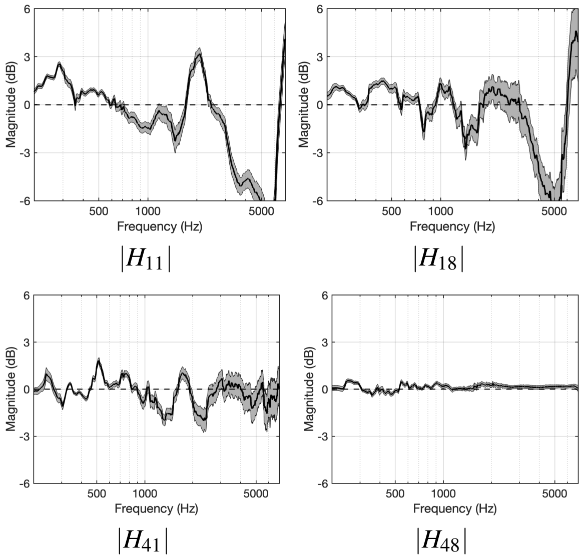


Fig. 2: Normalized magnitude responses of selected BRTFs measured from the human-HATS setup (20 measurements in total). The solid line and the surrounding shaded area represent the mean and standard deviation. The BRTFs from the HATS-HATS setup are used as the reference, therefore showing a flat response of 0 dB (the dashed line).

3.3.2 PSZ Filter Performance in Best-Case Scenario

The Acoustic Contrast (*AC*), as the main evaluation metric for PSZ filter performance, represents the ability of cancelling interfering audio in *DZ* and is defined in Ref. [10] as

$$AC = \frac{\mathbf{g}^H \mathbf{H}_B^H \mathbf{H}_B \mathbf{g}}{\mathbf{g}^H \mathbf{H}_D^H \mathbf{H}_D \mathbf{g}}, \tag{11}$$

where the variables with subscript *D*/*B* denote the submatrices that correspond to *DZ*/*BZ*. As the *DZ* and *BZ* are interchangeable, they correspond to $m = 1,2$ and $m = 3,4$ respectively in SZC 1, and vice versa in SZC 2.

Fig. 3 shows the measured AC for the *in situ* case. The left plot shows a significant increase in AC at all frequencies (200-7000 Hz), especially at higher frequencies starting from 2 kHz, which represents an improvement in interfering audio cancellation when the human listener is in *DZ*. The improvement has an average of 5.3 dB and a maximum of 9.4 dB at 2.7 kHz. This enhancement is expected as the playback BRTFs are matched with the setup BRTFs. A similar trend is also observed in the right plot of the figure, where the reference HATS is in *DZ*. Given that the BRTFs are different, this implies that even small changes in playback BRTFs can lead to performance degradation, especially for the BRTFs corresponding to loudspeakers that are further away. A possible explanation is that those loudspeakers have more impact on cancelling the interfering audio, while the closer ones are more responsible for reproducing the target audio. In addition, in the case where the two listeners are close to each other, the best PSZ performance can only be achieved when the BRTFs of both listeners are matched with those used for filter generation.

3.3.3 PSZ Filter Robustness

The robustness of both PSZ filters is examined with *AC* spectra obtained from the *ex situ* measurements. Fig. 4 shows the range between the minimum and maximum *AC* from the 10 measurements. As the AC does not necessarily follow a simple normal-like distribution, the mean and standard deviations are less insightful than the best and worst cases. We note that, in both configurations, despite the variation, the individualized filters are still superior to the generic filters at all frequencies.

Given the specified constraints on the head misalignment, the individualized filter is less robust than the generic one for frequencies above 2 kHz in SZC 1, with the average AC reduced by 3.6 dB compared to the *in situ* case. This suggests that above 2 kHz, where the two shaded areas begin to overlap, the benefits of using individualized filter can vanish easily if the head is slightly moved. In SZC 2, however, both filters show similar robustness as the BRTF variance in the *ex situ* measurements is minor for the other listener.

4 Summary and Concluding Remarks

In this paper, we evaluated the performance of a PSZ system using the generic and individualized BRTFs, under the setup of an 8-loudspeaker head-tracked PSZ system for two listeners in a typical listening room. Two sets of BRTFs of a) two dummy heads and b) a human listener and a dummy head were measured, and the corresponding PSZ filters were generated using the PM method, with a statistical design approach for optimal robustness against slight head misalignments. The resulting TFs convolved with different PSZ filters were measured both *in situ* and *ex situ*, in order to evaluate PSZ performance in terms of AC and robustness against possible head misalignments.

From the *in situ* measurement, the individualized filters improve the AC at all frequencies between 200 and 7000 Hz when the human listener is in *DZ* by an average of 5.3 dB and a maximum of 9.4 dB. From the *ex situ* measurement, although less robust compared to their generic counterparts above 2 kHz, the individualized filters remain superior to the generic ones in terms of measured AC. Furthermore, similar findings were also observed when the dummy head is in *DZ*, implicating a great impact on the dark zone sound cancellation for one listener by using mismatched BRTFs for the other listener. However, the filter robustness for the non-targeted listener is not significantly affected.

We emphasize that the results obtained in this paper are based on measurements in a real room. Consequently, the AC level is generally lower than that measured in an anechoic setting (for example, see Ref. [5]), where the degradation due to room reflections is minimized. We also expect more differences between the generic and individualized filters in an anechoic setting. However, the results under realistic listening conditions may be of more practical interest, as they better represent the highest AC level one can achieve in realistic situations.

The evaluation of other commonly-adopted metrics, such as the Array Effort and the Normalized Reproduction Error [5], are not discussed in the paper. Since the generic and individualized filters are generated with the same design method, we expect near-identical performance in Array Effort for both filters. The reproduction errors for the two filters are not comparable as 1) different target pressure is used to design each filter, and 2) the discrepancy between the setup BRTFs (late reverb 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. A more suitable metric for evaluating the reproduced audio quality in reverberant conditions is required.

While in our study, only slight head misalignments are represented as uncertainties, the effects of larger head movements can be treated separately as another source of TF mismatch. However, even so, the observed robustness of the individualized filters indicates that it is almost impractical to retain the best performance at high frequencies. This lack of robustness can be potentially addressed by using beamforming techniques with loudspeaker arrays and/or updating PSZ filters with respect to the tracked head position of the listener. In addition, an accurate head tracking system in 6 degree-of-freedoms is also necessary.

Lastly, despite the significant improvement in AC, subjective evaluation is needed to verify that the measured improvement from the individualized filters is perceptible to a large group of human subjects.

5 Acknowledgements

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

References

[1] Druyvesteyn, W. and Garas, J., "Personal sound," *Journal of the Audio Engineering Society*, 45(9), pp. 685–701, 1997.

[2] Poletti, M., "An investigation of 2-d multizone surround sound systems," in *Audio Engineering Society Convention 125*, Audio Engineering Society, 2008.

[3] Chang, J.-H. and Jacobsen, F., "Sound field control with a circular double-layer array of loudspeakers," *The Journal of the Acoustical Society of America*, 131(6), pp. 4518–4525, 2012.

[4] Olivieri, F., Fazi, F. M., Fontana, S., Menzies, D., and Nelson, P. A., "Generation of private sound with a circular loudspeaker array and the weighted pressure matching method," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25(8), pp. 1579–1591, 2017.

[5] Vindrola, L., Melon, M., Chamard, J.-C., and Gazengel, B., "Pressure Matching With Forced Filters for Personal Sound Zones Application," *Journal of the Audio Engineering Society*, 68(11), pp. 832–842, 2020.

[6] Canclini, A., Markovic, D., Schneider, M., Antonacci, F., Habets, E. A. P., Walther, A., and Sarti, A., "A weighted least squares beam shaping technique for sound field control," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6812–6816, IEEE, 2018.

[7] Cai, Y., Wu, M., and Yang, J., "Design of a time-domain acoustic contrast control for broadband input signals in personal audio systems," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 341–345, IEEE, 2013.

[8] Gálvez, M. F. S., Elliott, S. J., and Cheer, J., "Time domain optimization of filters used in a loudspeaker array for personal audio," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 23(11), pp. 1869–1878, 2015.

[9] Møller, M. B. and Olsen, M., "Sound zones: On performance prediction of contrast control methods," in *Audio Engineering Society Conference: 2016 AES International Conference on Sound Field Control*, Audio Engineering Society, 2016.

[10] Coleman, P., Jackson, P. J., Olik, M., Møller, M., Olsen, M., and Abildgaard Pedersen, J., "Acoustic contrast, planarity and robustness of sound zone methods using a circular loudspeaker array," *The Journal of the Acoustical Society of America*, 135(4), pp. 1929–1940, 2014.

[11] Mourjopoulos, J., "On the variation and invertibility of room impulse response functions," *Journal of Sound and Vibration*, 102(2), pp. 217–228, 1985.

[12] Park, J.-Y., Choi, J.-W., and Kim, Y.-H., "Acoustic contrast sensitivity to transfer function errors in the design of a personal audio system," *The Journal of the Acoustical Society of America*, 134(1), pp. EL112–EL118, 2013.

[13] Møller, M. B., Nielsen, J. K., Fernandez-Grande, E., and Olesen, S. K., "On the influence of transfer function noise on sound zone control in a room," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 27(9), pp. 1405–1418, 2019.

[14] Olsen, M. and Møller, M. B., "Sound zones: on the effect of ambient temperature variations in feed-forward systems," in *Audio Engineering Society Convention 142*, Audio Engineering Society, 2017.

[15] Cho, W.-H. and Chang, J.-H., *Investigation of effect on the acoustic transfer function in a vehicle cabin according to change of configuration*, Universitätsbibliothek der RWTH Aachen, 2019.

[16] Vindrola, L., Melon, M., Chamard, J.-C., and Gazengel, B., "Use of the filtered-x least-mean-squares algorithm to adapt personal sound zones in a car cabin," *The Journal of the Acoustical Society of America*, 150(3), pp. 1779–1793, 2021.

[17] Rämö, J., Bech, S., and Jensen, S. H., "Validating a real-time perceptual model predicting distraction caused by audio-on-audio interference," *The Journal of the Acoustical Society of America*, 144(1), pp. 153–163, 2018.

[18] Ebri, M., Strozzi, N., Fazi, F. M., Farina, A., and Cattani, L., "Individual Listening Zone with Frequency-Dependent Trim of Measured Impulse Responses," in *Audio Engineering Society Convention 149*, Audio Engineering Society, 2020.

[19] Canter, N. and Coleman, P., "Delivering Personalised 3D Audio to Multiple Listeners: Determining the Perceptual Trade-Off Between Acoustic Contrast and Cross-Talk," in *Audio Engineering Society Convention 150*, Audio Engineering Society, 2021.

[20] Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L., "Localization using nonindividualized head-related transfer functions," *The Journal of the Acoustical Society of America*, 94(1), pp. 111–123, 1993.

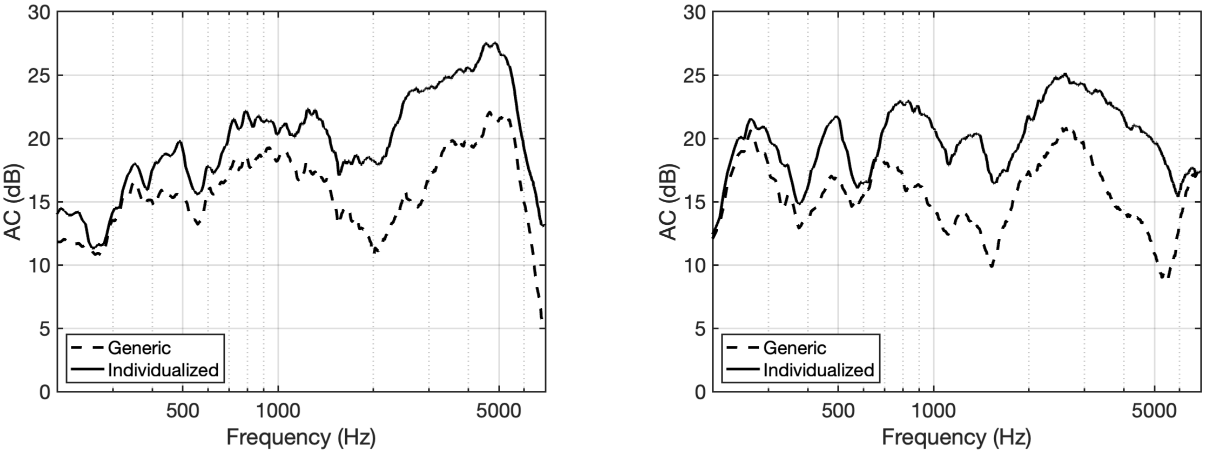
[21] Middlebrooks, J. C., "Individual differences in external-ear transfer functions reduced by scaling in frequency," *The Journal of the Acoustical Society of America*, 106(3), pp. 1480–1492, 1999.

[22] Akeroyd, M. A., Chambers, J., Bullock, D., Palmer, A. R., Summerfield, A. Q., Nelson, P. A., and Gatehouse, S., "The binaural performance of a cross-talk cancellation system with matched or mismatched setup and playback acoustics," *The Journal of the Acoustical Society of America*, 121(2), pp. 1056–1069, 2007.

[23] Majdak, P., Masiero, B., and Fels, J., "Sound localization in individualized and non-individualized crosstalk cancellation systems," *The Journal of the Acoustical Society of America*, 133(4), pp. 2055–2068, 2013.

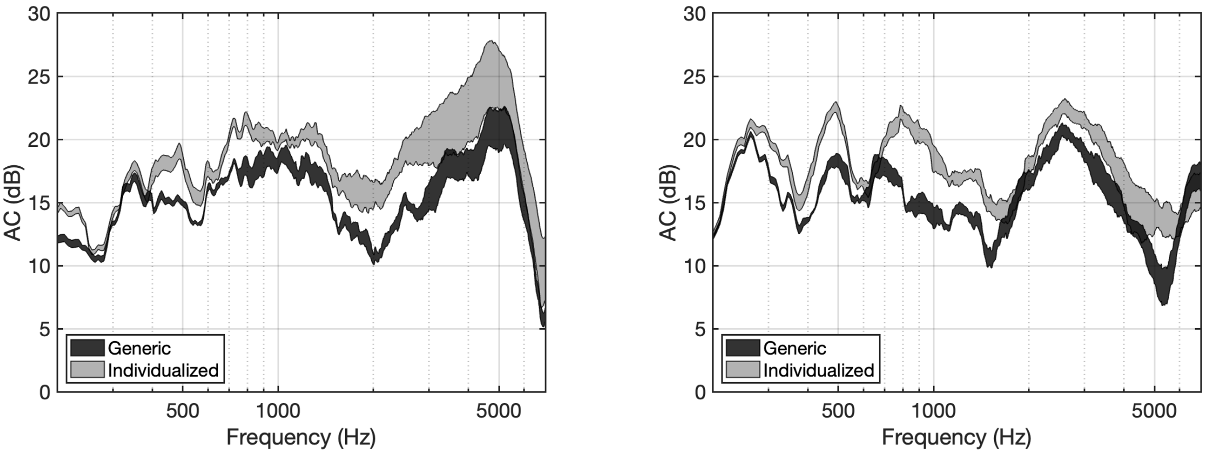
[24] Farina, A., "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Audio Engineering Society Convention 108*, Audio Engineering Society, 2000.

[25] Tylka, J. G., Boren, B. B., and Choueiri, E. Y., "A Generalized Method for Fractional-Octave Smoothing of Transfer Functions that Preserves Log-Frequency Symmetry," *Journal of the Audio Engineering Society*, 65(3), pp. 239–245, 2017.



Sound Zone Config 1 Sound Zone Config 2

Fig. 3: Measured Acoustic Contrast (*in situ*) for the generic filters (in dashed line) and the individualized filters (in solid line). The left and right plots represent two different sound zone configurations.



Sound Zone Config 1 Sound Zone Config 2

Fig. 4: Measured Acoustic Contrast (*ex situ*) for the generic filters (dark shading) and the individualized filters (light shading). The shaded area represents the range between the minimum and maximum AC from 10 repositioned measurements. The left and right plots represent two different sound zone configurations.