Performance Optimization of Personal Sound Zones with Crosstalk Cancellation

Yue Qiao  
3D Audio and Applied Acoustics Laboratory  
Princeton University  
Princeton, USA  
qiao@princeton.edu

Edgar Choueiri  
3D Audio and Applied Acoustics Laboratory  
Princeton University  
Princeton, USA  
choueiri@princeton.edu

Abstract—Two optimization approaches are proposed to enhance the performance of personal sound zone (PSZ) systems with crosstalk cancellation (XTC). The two approaches adjust the trade-off between two important performance attributes of a system: acoustic isolation and crosstalk cancellation, by either modifying the cost function in the optimization problem (the direct approach) or controlling the amount of crosstalk in the target transfer functions (the indirect approach) in the filter generation process. The effectiveness of the two approaches is evaluated using metrics of inter-program isolation (IPI) and XTC level, through numerical simulations based on both a free-field system model and actual transfer function measurements. The results show that at low frequencies, the two approaches can effectively control the trade-off between the two attributes, but their effectiveness is reduced at high frequencies due to the small wavelength. Furthermore, the results demonstrate that the direct approach is more effective in manipulating the trade-off, whereas the indirect approach provides more precise control over the desired XTC performance.

Index Terms—personal sound zone, crosstalk cancellation, optimization

I. INTRODUCTION

Personal sound zone (PSZ) [1], [2] reproduction aims to deliver, through loudspeakers, independent audio programs to multiple listeners in the same physical space with minimum interference between programs. Over the past two decades, PSZ has received wide attention from both academic and industrial communities due to its potential applications in mobile devices [3], home entertainment [4], automotive cabins [5], [6], and outdoor spaces [7]. In general, a PSZ system generates two types of sound zones: a bright zone (BZ), in which a target audio program is reproduced with high fidelity, and a dark zone (DZ), in which the program is attenuated as much as possible. To render such two zones, digital audio filters are usually designed with various optimization methods and then convolved with the audio programs before they are reproduced by loudspeakers. Commonly used methods include Pressure Matching (PM) [8]–[10], Acoustic Contrast Control (ACC) [11]–[13], and Variable Span Trade-Off Filtering (VAST) [14], [15]. In PM, the filters minimize the difference between the actual sound pressure in sound zones and the target pressure specified in advance, but they can lead to lower acoustic isolation compared to those designed with ACC; in ACC, the filters maximize the energy difference between BZ and DZ, but they have no explicit control over the phase response of the reproduced audio program, therefore leading to relatively degraded audio quality [16]. Lastly, VAST subsumes both PM and ACC as special cases and allows flexible control of the trade-off between signal distortion and acoustic isolation.

PSZ has been known [2] to conceptually resemble interaural crosstalk cancellation (XTC) [17], which allows binaural audio reproduction over loudspeakers. Similar to PSZ systems that isolate sound zones, XTC systems “separate” the left and right channels of a binaural audio program to be delivered to the two ears of a listener by using XTC filters. In practice, XTC can be embedded in a PSZ system by using either PM-based filtering or loudspeaker array beamforming [18], [19], therefore enabling the delivery of personalized binaural audio programs. Such a system can be particularly useful for VR/AR applications. However, in such a system, as both acoustic isolation and crosstalk cancellation are important attributes, there exists a potential trade-off when achieving high performance in both aspects. For example, it has been shown that compared to mono audio programs, binaural audio programs can result in degraded acoustic isolation for PSZ systems with PM-based filters [20]. Therefore, it is crucial to understand the trade-off and optimize the system performance accordingly.

To investigate the relationship between acoustic isolation and crosstalk cancellation in terms of auditory perception, Canter and Coleman [21] conducted a series of headphone-based experiments. Two sets of listening tests were carried out: the first test determined the thresholds of the two attributes for each listener, from which it was concluded that the two attributes are independent of each other, and that acoustic isolation requires a higher performance threshold than crosstalk cancellation; the second test evaluated the listener’s preference of one attribute over the other, from which it was found that in general, higher acoustic isolation is preferred over better crosstalk cancellation. Such conclusions can be used to guide the optimization of PSZ systems with XTC.

Motivated by that subjective study, we propose two approaches that adjust the trade-off between the two attributes. Given the perceptual preferences, we can apply the two approaches accordingly to optimize the performance of the PSZ system. In the first approach (which we thereafter refer
to as the “direct” approach), we introduce a weighted PM method (similar to the one proposed in [22]) in which the relative importance of the two attributes is controlled by a weighting parameter; in the second “indirect” approach, we modify the target transfer functions by controlling the amount of crosstalk added to the two channels of the binaural audio program. We then evaluate the effectiveness of the two approaches (i.e., how effective one attribute can be improved by sacrificing the performance of the other) through numerical simulations, considering both a free-field model and a real listening environment. Finally, we discuss and compare the two approaches in terms of their advantages and limitations.

II. THE PSZ THEORY

We first describe a general PSZ system model and then explain the PM method for designing the PSZ filters. Lastly, we introduce metrics for quantifying the acoustic isolation and crosstalk cancellation in a PSZ system with XTC.

A. PSZ System Model

A PSZ system consists of an array of $L$ loudspeakers and $M$ control points defined in the two zones. In the frequency domain, we define each loudspeaker $l$ to have a complex gain of $g_l(\omega), l = 1, \ldots, L$, and the resulting sound pressure at each control point $m$ is $p_m(\omega), m = 1, \ldots, M$. The transfer function (TF) corresponding to the loudspeaker $l$ and the control point $m$ is denoted as $H_{ml}(\omega)$. Then, the pressure at the control points is given by

$$ p = Hg, $$

where $p = [p_1, \ldots, p_M]^T \in \mathbb{C}^{M \times 1}, H = (H_{ml}) \in \mathbb{C}^{M \times L}$, and $g = [g_1, \ldots, g_L]^T \in \mathbb{C}^{L \times 1}$. All quantities above and thereafter are implicitly dependent on the frequency $\omega$. Furthermore, considering an audio program with $I$ input channels, we denote the signal in the $i$-th channel as $s_i$, and its corresponding PSZ filter as $c_i \in \mathbb{C}^{L \times 1}$, then the loudspeaker gain vector can be expressed as

$$ g = c_is_i. $$

Combining the above two equations and considering all the input channels, we have

$$ p = HC_s, $$

where $s = [s_1, \ldots, s_I]^T \in \mathbb{C}^{I \times 1}$ denotes the vector of input channels, and $C = [c_1, \ldots, c_I] \in \mathbb{C}^{L \times I}$ denotes the filter matrix. Combining both $H$ and $C$, we denote the overall TF matrix as

$$ M = HC, $$

which represents the TF from the input channels of the audio program to the control points. In this paper, we consider a PSZ system that reproduces two binaural audio programs (four channels in total) to two zones ($Z_A, Z_B$), respectively, and there are two control points in each zone, located at the listener’s ears. The overall TF matrix of the system can then be expressed as

$$ M = [m_1, m_2, m_3, m_4] = \begin{bmatrix} M_{11} & M_{12} & M_{13} & M_{14} \\ M_{21} & M_{22} & M_{23} & M_{24} \\ M_{31} & M_{32} & M_{33} & M_{34} \\ M_{41} & M_{42} & M_{43} & M_{44} \end{bmatrix}, $$

and the target TF matrix $M_T$ for reproducing the two binaural programs is given by

$$ M_T = [m_{T,1}, m_{T,2}, m_{T,3}, m_{T,4}] = \begin{bmatrix} M_{A,L} & 0 & 0 & 0 \\ 0 & M_{A,R} & 0 & 0 \\ 0 & 0 & M_{B,L} & 0 \\ 0 & 0 & 0 & M_{B,R} \end{bmatrix}, $$

where the subscripts $A, B$ denote $Z_A$ and $Z_B$, respectively, and $L, R$ denote the left and right channels of a binaural program, respectively. We will focus on this model in the following sections.

B. Pressure Matching Method

The PM method essentially solves an optimization problem that minimizes the given cost function. Given the target pressure vector $p_T$, we can write the cost function $J$ for a single input channel as

$$ J = \| p - p_T \|^2 = \| Hg - p_T \|^2. $$

One can directly solve this minimization problem and obtain either a least square or minimum energy solution, depending on the nature of the problem (i.e., the relationship between $L$ and $M$). A more common choice is to add an extra filter energy term to the original cost function to ensure the numerical stability of the solution, leading to the following cost function

$$ J = \| Hg - p_T \|^2 + \lambda \| g \|^2, $$

where $\lambda$ is known as the regularization parameter. The solution to the modified optimization problem is uniquely given by

$$ g^* = (H^H H + \lambda I)^{-1} H^H p_T, $$

where $(\cdot)^H$ denotes taking the complex conjugate. Similarly, considering multiple input channels and assuming that the input signal is the Dirac delta function, we can replace the pressure terms $p, p_T$ with the corresponding TF matrices $M, M_T$ and replace the loudspeaker gain vector $g$ with the filter matrix $C$. Then, the corresponding filter matrix can be derived as

$$ C = (H^H H + \lambda I)^{-1} H^H M_T. $$

C. Performance Metrics

We use two metrics to quantify the two attributes of a PSZ system with XTC, respectively. For acoustic isolation, we adopt the Inter-Program Isolation (IPI) metric [20], which is defined as the ratio of the two averaged acoustic power spectra in the same zone, corresponding to the two different audio programs. Compared to the other metric, Inter-Zone Isolation...
and the subvectors of $\mathbf{H}$ where $\mathbf{h}$ in a binaural program, respectively. Then, for crosstalk can-
zone. Considering the previous PSZ system model, the IPI metric for $Z_A, Z_B$ is defined as

$$IPI_A = \min \left\{ \sum_{i=1}^{4} \left( \sum_{j=1}^{4} M_{ij}^2 \right) \right\},$$

$$IPI_B = \min \left\{ \sum_{i=3}^{4} \left( \sum_{j=1}^{4} M_{ij}^2 \right) \right\},$$

(11)

(12)

respectively. The two terms in the minimum function corre-
respond to the cases of correlated and uncorrelated channels in a binaural program, respectively. Then, for crosstalk can-
cellation, we define the amount of crosstalk cancellation for binaural reproduction in each zone as

$$\text{XTC}_A = \sqrt{\frac{\sum_{i=1}^{4} M_{ii}^2}{\sum_{i=1}^{4} M_{ii}^2}} = \frac{M_{11,11}}{M_{22,22}},$$

$$\text{XTC}_B = \sqrt{\frac{\sum_{i=3}^{4} M_{ii}^2}{\sum_{i=3}^{4} M_{ii}^2}} = \frac{M_{33,33}}{M_{44,44}}.$$  

(13)

(14)

All the quantities will be represented in the logarithmic scale ($10 \log_{10} \{ \cdot \}$) in Sec. IV.

III. PROPOSED OPTIMIZATION

In this section, we introduce the direct and indirect optimization approaches that control the trade-off between the two performance attributes.

A. Direct Approach

The direct optimization approach modifies the cost function used in the PM method by adding a weighting parameter $\alpha$, which controls the relative importance of the two attributes. First, we rewrite the cost function in Eq. 8 in terms of the PSZ filter vector $\mathbf{c}$ and the target TF vector $\mathbf{m}_T$ as

$$J = \| \mathbf{H} \mathbf{c} - \mathbf{m}_T \|^2 + \lambda \| \mathbf{c} \|^2$$

$$= \| \mathbf{H}_A \mathbf{c} - \mathbf{m}_{T,A} \|^2 + \| \mathbf{H}_B \mathbf{c} - \mathbf{m}_{T,B} \|^2 + \lambda \| \mathbf{c} \|^2,$$  

(15)

where $\mathbf{H}_A, \mathbf{H}_B$ and $\mathbf{m}_{T,A}, \mathbf{m}_{T,B}$ denote the submatrices of $\mathbf{H}$ and the subvectors of $\mathbf{m}_T$ corresponding to $Z_A$ and $Z_B$, respectively. Then, consider the case where the left input of $Z_A$ is the active channel, we have $\mathbf{m}_{T,A} = [M_{A,L}, 0]^T$ and $\mathbf{m}_{T,B} = [0, 0]^T$ from Eq. 6. The cost function can be rewritten as

$$J = \| \mathbf{h}_A \mathbf{c} - M_{A,L} \|^2 + \| \mathbf{h}_R \mathbf{c} \|^2 + \| \mathbf{H}_B \mathbf{c} \|^2 + \lambda \| \mathbf{c} \|^2,$$  

(16)

where $\mathbf{h}_A, \mathbf{h}_R$ denote the two row vectors of $\mathbf{H}_A$. We note that the first term corresponds to the reproducing error, while the second and the third term correspond to the crosstalk cancellation and the acoustic isolation, respectively. After adding the weighting term $\alpha$, the modified cost function is given by

$$J_{A,L} = 2\alpha \| \mathbf{h}_A \mathbf{c} \|^2 + 2(1-\alpha) \| \mathbf{H}_B \mathbf{c} \|^2$$

$$+ \| \mathbf{h}_R \mathbf{c} - M_{A,L} \|^2 + \lambda \| \mathbf{c} \|^2,$$  

(17)

where $0 < \alpha < 1$, and $\alpha = 0.5$ corresponds to the original cost function in which the two attributes are equally important. As $\alpha$ increases, the crosstalk cancellation becomes more dominant than the acoustic isolation, and vice versa. Minimizing the cost function yields the optimal filter vector

$$\mathbf{c}_{A,L} = (\mathbf{H}^H \mathbf{H} + \lambda \mathbf{I})^{-1} \mathbf{H}^H \mathbf{m}_{T,1}.$$  

(18)

where $\mathbf{A}_L = \text{diag} \{1, 2\alpha, (1-\alpha)\}$ is a 4-

B. Indirect Approach

In the indirect approach, we modify the target TF matrix $\mathbf{M}_T$ in Eq. 6 by introducing crosstalk to the contralateral channel of the binaural program and controlling the crosstalk amount with a weighting parameter, $\beta$. The modified target TF matrix $\mathbf{M}_T$ is given by

$$\mathbf{M}_T = \begin{bmatrix} \tilde{\mathbf{m}}_{T,1}, \tilde{\mathbf{m}}_{T,2}, \tilde{\mathbf{m}}_{T,3}, \tilde{\mathbf{m}}_{T,4} \end{bmatrix} \begin{bmatrix} \beta M_{A,L} & 0 & 0 & 0 \\ (1-\beta) M_{A,R} & \beta M_{A,R} & 0 & 0 \\ 0 & 0 & \beta M_{B,L} & (1-\beta) M_{B,R} \\ 0 & 0 & (1-\beta) M_{B,R} & \beta M_{B,R} \end{bmatrix},$$  

(19)

where $0.5 \leq \beta \leq 1, \beta = 1$ corresponds to the original case, and $\beta = 0.5$ corresponds to reproducing mono programs. Taking the left input channel of $Z_A$ as an example, the corresponding PSZ filter vector is given by

$$\mathbf{c}_{A,L} = (\mathbf{H}^H \mathbf{H} + \lambda \mathbf{I})^{-1} \mathbf{H}^H \tilde{\mathbf{m}}_{T,1}.$$  

(20)

By adding the weighted crosstalk terms, we equivalently transition from binaural programs to mono programs, therefore trading off crosstalk cancellation for potentially higher acoustic isolation.

IV. RESULTS

A. Free-Field Simulation

We first evaluate both approaches through numerical simulations in a free-field setting. The evaluated PSZ system (whose setup is illustrated in Fig. 1), based on a real system implemented in house, consists of a linear array of eight loudspeakers and two zones for two listeners, with a spacing of 1 m between the array and the listener, and the same spacing between the two listeners. The loudspeaker drivers are assumed to be omnidirectional and are modeled as circular baffled pistons in the simulation. To reproduce binaural programs,
we choose the natural responses of the four loudspeakers to the four ears as the target TFs on the diagonal of $M_\text{T}$, as shown in the figure. To simulate real-world disturbances (e.g., loudspeaker position inaccuracies and late reverberation) to the TFs, we model each TF $H_{\text{ml}}$ per frequency as a random variable whose amplitude and phase are normally distributed, with the same variance of $\sigma^2 = 10^{-4}$. From the same distribution $\mathcal{N}(0, \sigma^2)$, we take 20 samples of the TFs and use their mean value for filter generation. We choose $\lambda = 4\sigma^2 = 4 \times 10^{-4}$ to minimize the expectation of the cost function, as a result of the optimization approach described in Refs. [23], [24]. We then take another sample of the TFs and use it to calculate the resulting TF matrix $\text{M}$ and the corresponding metrics.

Fig. 2 shows the IPI and XTC metrics as a function of the weighting parameters $\alpha$, $\beta$, and the frequency between 100 and 1000 Hz. Due to the symmetry of the system setup, we only show results for IPI$_A$ and XTC$_A$, and therefore their subscripts will be ignored (same for the following subsection). The IPI and XTC levels are truncated up to 40 dB for better visualization. First, we note that the two weighting parameters have similar effects on the two attributes of the system performance: increasing either $\alpha$ or $\beta$ boosts XTC and therefore degrades IPI. Furthermore, for a fixed weighting parameter, both IPI and XTC have different frequency dependency in the two approaches. For example, XTC is generally higher at higher frequencies for a fixed $\alpha$ in the direct approach, while in the indirect approach, the XTC performance is relatively uniformly affected by changing $\beta$ across the frequencies.

In addition, the two attributes are not equally affected by changing the weighting parameters, especially at high frequencies. For example, although the XTC spectra is frequency-dependent over a wide frequency range, the change in the IPI spectra is only noticeable at low frequencies (more specifically, below around 300 Hz in Fig. 2). This observation indicates that the trade-off between the two attributes can only be effectively manipulated at low frequencies, and that at high frequencies, it is not worth sacrificing the XTC level for a small improvement in acoustic isolation.

To better understand how the weighting parameters in the two approaches affect the trade-off, we plot the distribution of the sound pressure levels (SPLs) in the area covering the two zones, at a frequency of 200 Hz and with different values of $\alpha$, $\beta$, as shown in Fig. 3. For simplicity, we only show the results for the case where only the left channel of the binaural program in $Z_A$ is active; in other words, only the leftmost control point is assigned with a nonzero target TF. The SPL in all the figures is normalized by that at the leftmost control point. First, we note that by decreasing either $\alpha$ or $\beta$, the dark region associated with the right ear of the left listener gradually moves toward the right listener until it is “absorbed” by the dark area in $Z_B$, indicating that the XTC performance degrades in $Z_A$. At the same time, the DZ in $Z_B$ gradually expands and becomes “darker”, indicating that the acoustic isolation improves. Furthermore, we note that although the general trends in the two approaches are similar, and that $\alpha = 0.5$ and $\beta = 1$ correspond to the same case, the direct approach leads to a larger DZ when $\alpha$ is close to 0, compared to the indirect approach when $\beta$ is close to 0.5. In other words, when the XTC performance is maximally compromised (i.e., the system approaches the mono case), the direct approach leads to better acoustic isolation than the indirect approach. The results at 200 Hz also give us a hint on why such a trade-off cannot be effectively manipulated at high frequencies: the dark region in $Z_A$ does not interfere with the DZ in $Z_B$ due to the small wavelength, and therefore changing the weighting parameters does not significantly affect the acoustic isolation.

B. Real-world Simulation

We further evaluate the two approaches using the actual TF measurements of the real PSZ system. The TFs were measured in a typical listening room with $RT_{60} \approx 0.24$ s in the range of 1300-6300 Hz. Two B&K Head and Torso Simulators with in-ear binaural microphones (Theoretica Applied Physics BACCH-BM Pro) were used as the listeners. We refer readers to Ref. [24] for details regarding the TF measurement. Two
Fig. 3. Calculated SPL distribution at 200 Hz for different values of $\alpha$ (top row, corresponding to the direct approach) and $\beta$ (bottom row, corresponding to the indirect approach). The triangles and cross markers correspond to the loudspeakers and control points, respectively, as in Fig. 1. The SPL in each figure is normalized by that at the leftmost control point, which is specified with the nonzero target TF.

Fig. 4. Photo of the actual PSZ system setup. Note that the two tweeter loudspeaker arrays on the top were not used in the TF measurement.

consecutively measured sets of TFs are used in the simulation: the first set, truncated to the first 4096 samples (at 48 kHz sample rate), is used for generating the PSZ filters; the second set, truncated to the first 8192 samples, is used for performance evaluation. The same constant regularization ($\lambda = 4 \times 10^{-4}$) as in the free-field simulation is applied to the filter generation. Fig. 5 shows the resulting IPI and XTC curves for the two approaches, in the range of 100-7000 Hz. We note that the general trends in the two approaches are similar to those in the free-field simulation, but the IPI and XTC levels are lower in the real-world case, due to room reflections and other disturbances in the real environment. Compared to the free-field simulation, it is clearer to see the differences between the two approaches in the real-world case: first, the direct approach can more effectively boost the acoustic isolation at low frequencies compared to the indirect approach; moreover, the indirect approach leads to a constant level of crosstalk cancellation at most frequencies, while the decrease in XTC in the direct approach is more frequency-dependent.

V. DISCUSSION

The two proposed optimization approaches, although both capable of manipulating the trade-off between the two attributes, have different advantages and limitations. The direct approach is more effective in controlling the trade-off, as it can achieve higher acoustic isolation at low frequencies with less sacrifice in crosstalk isolation, compared to the indirect approach. However, it has no control over the phase
response of the contralateral channel in a binaural program, and therefore may lead to degraded spatial audio quality, e.g., spatially distorted sound image, or inaccurate sound localization. In contrast, the indirect approach can fully control the sound image, as the weighting parameter works effectively as the panning of a sound source; however, it would require more compromise in crosstalk cancellation to achieve a certain level of acoustic isolation.

It is also important to consider the optimal selection of the weighting parameter when implementing such approaches in a practical PSZ system equipped with crosstalk cancellation, apart from their differences in performance. For example, as shown in the simulation results, the two approaches are less effective in improving the acoustic isolation at high frequencies, therefore the weighting parameter should be set to the default value (i.e., no optimization) in the corresponding frequency band; at low frequencies, the parameter can be determined according to a pre-defined threshold in IPI, such as that specified in [21]: once the threshold is met, the parameter will be fixed to avoid unnecessary compromise in crosstalk cancellation. An alternative option would be to allow the listeners to adjust the weighting parameter based on their preferences. Nevertheless, we note that the effectiveness of the optimization approaches is highly dependent on the specific PSZ system setup, therefore the optimization should be always based on the actual TF measurements.

VI. CONCLUSION

For PSZ systems with crosstalk cancellation, a potential trade-off exists between the performance in acoustic isolation and that in crosstalk cancellation. We proposed two optimization approaches that manage to manipulate such a trade-off by either directly modifying the cost function in the optimization problem or indirectly affecting the resulting pressure field with controlled crosstalk in the target program. Guided by the conclusion of a subjective study [21], that higher acoustic isolation is generally preferred over better crosstalk cancellation, the two approaches can be used to optimize the PSZ system performance. Through numerical simulations with both a free-field model and actual TF measurements, we showed that the direct approach is more effective in controlling the trade-off, while the indirect approach has better control over the reproduced sound image. Furthermore, the weighting parameters in the two approaches should be carefully selected based not only on the frequency range, but also on the performance requirements of the system. Future work will involve evaluating the proposed approaches with full measurements and conducting a comprehensive subjective study.

REFERENCES