Audio Engineering Society

Conference Paper

Presented at the AES 2026 International Conference on
Audio for Virtual and Augmented Reality and Immersive Games
2026 June 30 – July 3, Sorbonne University/d'Alembert & IRCAM/STMS, Paris, France

Neural Regularization for Personal Sound Zones

Yazhou Li, Lin Wang, and Joshua Reiss

Center for Digital Music, Queen Mary University of London

Correspondence should be addressed to Yazhou Li (yazhou.li@qmul.ac.uk)

ABSTRACT

Pressure-matching (PM) for personal sound zone (PSZ) can achieve high contrast at nominal control points, but the performance may degrade when transfer functions are mismatched. We introduce a neural method that maps transfer functions to loudspeaker weights using a single-frequency input network with parameters shared across frequencies. We evaluate the robustness under position shifts, additive transfer-function noise, and added reflections, and compare against PM with Tikhonov regularization. Results show improved robustness to structured perturbations such as listener displacement, whereas regularized PM remains more resilient to unstructured random transfer-function noise and reverberation. We further explain these results using a singular value decomposition based perturbation projection. Finally, we analyze different regularization mechanisms induced by the network and derive practical guidelines for neural PSZ filter optimization. [1]

1 Introduction

Personal sound zones (PSZ) aim to create spatially separated bright and dark listening regions in a shared acoustic space using loudspeaker arrays. Typical methods consider two listening areas as the bright zone and the dark zone, where the bright zone's sound should be heard, but the dark zone should be silent. Design methods include pressure matching (PM) [1], acoustic contrast control (ACC) [2], and variable span trade-off filters (VAST) [3]. PSZ can also be viewed as a crosstalk cancellation (CTC) problem where the listeners' ears are tracked and controlled. However, traditional optimization methods are sensitive to room modeling errors [4]. Small perturbations in room transfer functions can cause large variations in the resulting loudspeaker filters and hence in the achieved contrast. The perturbations include microphone position shift [5], background noise [6], room acoustics [7], and so on.

Tikhonov regularization is usually used to improve the robustness of the system [8]. When the problem is ill-posed (e.g., rank-deficient or underdetermined), adding the regularization term makes the normal matrix positive definite, thereby yielding a unique solution. When the problem is ill-conditioned (i.e., small singular values and large condition number), Tikhonov regularization improves numerical conditioning and reduces sensitivity. At the same time, the explicit norm penalty limits loudspeaker effort, jointly enhancing robustness.

However, choosing the correct regularization parameter is an important issue in Tikhonov regularization. There are two kinds of methods to choose the parameter: spectral methods and non-spectral methods. Spec-

[1] The code is available at https://github.com/li630925405/NN_for_PSZ.

tral methods are based on singular value decomposition (SVD) analysis, including L-curve [9] and adaptive normalized Tikhonov [10]. SVD is usually used to calculate the condition number of the system and analyze the system robustness. SVD is also largely used in loudspeaker geometry optimization [11] and theoretical analysis of the multichannel CTC system [12]. Non-spectral methods use other targets like statistical error and physical constraints, including generalized cross-validation [13] and binary searches [14]. In our CTC-like PSZ setup where the number of loudspeakers is larger than the number of microphones and the microphones are relatively sparsely spaced, the plant matrix is already well conditioned, so condition-number-based criteria provide limited guidance.

Other methods to improve the robustness of PSZ include worst-case optimization and probability-model optimization, rooted in robust engineering and stochastic programming respectively [15]. Although these methods can be interpreted as regularization methods in form, their regularization terms are not heuristic but are derived explicitly from assumed transfer function uncertainty bounds or statistical error models, in contrast to Tikhonov regularization.

Recently, neural networks have been explored as solvers for PSZ filters. [16], [17], and [18] use neural networks for PSZ filter design or audio equalization. The network architecture in [16] uses room impulse responses (RIRs) as the network input and output loudspeaker filters. [18] simplifies the architecture by removing explicit network inputs, instead encoding the RIR information into the loss function. That work also designs IIR filters using neural networks, which is difficult using the traditional method. [17] compares different network architectures, including multi-layer perceptrons (MLP) and convolutional neural networks (CNN). [19] uses neural networks to predict filters from microphone positions directly, with training conducted across multiple room acoustics so that the resulting filters represent an average optimal solution. However, the acoustic contrast is limited if the room acoustics are not considered in the filter design. Overall, the robustness of neural-network-based PSZ solutions has not yet been systematically evaluated.

The generalization and robustness properties of neural networks have been extensively investigated in recent years. [20] shows that the stochastic gradient descent optimizer in neural networks tends to find the minimum norm or minimum rank solutions, which works as implicit regularization. [21] and [22] analyze the geometry of the loss landscape and report that the low-loss regions associated with different minima are connected. As a result, many local minima achieve comparable performance, and the solutions tend to lie in broad, flat valleys rather than in sharp isolated minima. Such flat minima are less sensitive to perturbations and result in solutions that are both more robust and better able to generalize.

Motivated by regularization effects in neural networks, we propose a neural network architecture for PSZ filter design. Our contributions are: (i) we introduce a frequency-domain neural network with one frequency input that maps transfer functions to loudspeaker filters, where the parameter sharing between frequency bins imposes an inductive bias toward cross-frequency coherent solutions, and can be interpreted as structured regularization; (ii) we benchmark the results against PM under three mismatch scenarios, and provide an SVD-based perturbation projection analysis that links the observed results to the structure of the mismatch; (iii) we investigate different regularization mechanisms in the neural network to clarify the impact of each component.

This paper is organized as follows. Section II reviews the background of PM for PSZ and the evaluation metrics. Section III presents the proposed neural network formulation and introduces an SVD-based robustness analysis framework. Section IV describes the experimental details. Section V reports and discusses the experimental results. Finally, Section VI concludes the paper and outlines directions for future work.

2 Background

2.1 Pressure Matching for PSZ

We consider a frequency-domain PSZ formulation with L loudspeakers and M control points [23]. We work independently per frequency bin f. For convenience, we also denote the angular frequency as $\omega = 2\pi f$. In implementation, frequency bins are sampled in f, and we use ω only for notation. Let $\mathbf{G} \in \mathbb{C}^{M\times L}$ denote the transfer-function matrix from loudspeakers to control points, and $\mathbf{w} \in \mathbb{C}^{L}$ the loudspeaker weights. All quantities in matrix form are implicitly dependent on frequency ω.

For each frequency ω, PM finds $\mathbf{w}$ by minimizing a Tikhonov-regularized least-squares objective [24]

$$J(\mathbf{w}) = \|\mathbf{G}\mathbf{w} - \mathbf{d}\|_2^2 + \beta\|\mathbf{w}\|_2^2, \tag{1}$$

where $\beta > 0$ is a Tikhonov regularization parameter, and $\boldsymbol{d} = [1, 1...0, 0]^T \in \mathbb{C}^M$ is the target pressure at M microphones, which is 1 at bright zone control points, and 0 at the dark zone points. The closed-form solution is

$$\mathbf{w} = \left(\mathbf{G}^{\mathrm{H}}\mathbf{G} + \beta\mathbf{I}\right)^{-1}\mathbf{G}^{\mathrm{H}}\mathbf{d}. \tag{2}$$

2.2 Evaluation metrics

We evaluate the PSZ system using acoustic contrast (AC) [25]. At each frequency ω, the sound pressures in the bright and dark zones are

$$p_b(\omega) = \frac{1}{M_b}\sum_{m\in\mathscr{B}}|p_m(\omega)|^2,\ p_d(\omega) = \frac{1}{M_d}\sum_{m\in\mathscr{D}}|p_m(\omega)|^2, \tag{3}$$

where $p_m(\omega)$ is the complex pressure at control point m, $\mathscr{B}$ and $\mathscr{D}$ are the sets of control points in the bright and dark zones, and M_b and M_d are the sizes of $\mathscr{B}$ and $\mathscr{D}$.

To obtain a perceptually meaningful representation, we compute $1/3$-octave-band averages of AC by:

$$\mathrm{AC}_{1/3}(\omega_j) = 10\log_{10}\left(\frac{1}{|\Omega_j|}\sum_{\omega\in\Omega_j}\frac{p_b(\omega)}{p_d(\omega)}\right), \tag{4}$$

where ω_j denotes the center frequency of the j-th one-third-octave band, Ω_j is the corresponding set of frequency bins within that band, and there are J one-third-octave bands in total. Finally, we report the average over the frequency range

$$\overline{\mathrm{AC}} = \frac{1}{J}\sum_{j=1}^{J}\mathrm{AC}_{1/3}(\omega_j). \tag{5}$$

For brevity, AC denotes $\overline{\mathrm{AC}}$ unless otherwise specified.

3 Method

3.1 Neural Network Formulation

A multi-layer perceptron (MLP) is used to learn the mapping from transfer functions to loudspeaker weights, as shown in Figure 1. We predict loudspeaker weights at each frequency bin ω using a single real-valued network shared across all bins. For each ω, the complex transfer matrix $\mathbf{G}$ is converted into a real input vector by stacking its real and imaginary parts:

$$\mathbf{x} = \begin{bmatrix}\Re\{\mathrm{vec}(\mathbf{G})\}\\ \Im\{\mathrm{vec}(\mathbf{G})\}\end{bmatrix} \in \mathbb{R}^{2ML}, \tag{6}$$

where $\mathrm{vec}(\cdot)$ flattens the matrix into a vector by concatenating its rows. The network outputs

$$\hat{\mathbf{y}} = f_\theta(\mathbf{x}) \in \mathbb{R}^{2L}, \tag{7}$$

which is interpreted as the complex loudspeaker weights via

$$\mathbf{w} = \hat{\mathbf{y}}_{1:L} + j\hat{\mathbf{y}}_{L+1:2L} \in \mathbb{C}^L. \tag{8}$$

The complex pressures at the setup microphones are

$$\mathbf{p} = \mathbf{G}\mathbf{w}. \tag{9}$$

We use a PM-inspired loss at each frequency bin:

$$\mathscr{L} = \|\mathbf{p} - \mathbf{d}\|_2^2 + \beta\|\mathbf{w}\|_2^2, \tag{10}$$

where β controls the amount of the Tikhonov penalty in the loss function. β is 0 by default, which means the Tikhonov penalty is not added to the loss function.

Although we do not explicitly optimize for robustness, the proposed predictor introduces several sources of regularization: (i) implicit regularization from the over-parameterized MLP and optimization dynamics, (ii) cross-frequency coupling due to the shared network parameters across frequency, and (iii) explicit training-time regularization such as Tikhonov penalty and dropout. We later relate these mechanisms to the observed robustness results.

3.2 SVD–based analysis of perturbations

We compute the thin SVD of $\mathbf{G}(\omega) \in \mathbb{C}^{M\times L}$ as

$$\mathbf{G}(\omega) = \mathbf{U}(\omega)\mathbf{\Sigma}(\omega)\mathbf{V}(\omega)^{\mathrm{H}}, \tag{11}$$

where $\mathbf{U}(\omega) \in \mathbb{C}^{M\times M}$, $\mathbf{V}(\omega) \in \mathbb{C}^{L\times M}$, and $\mathbf{\Sigma}(\omega) = \mathrm{diag}(\sigma_1(\omega), \ldots, \sigma_M(\omega))$, assuming $M \le L$.

With Tikhonov regularization, the solution is expressed as a weighted sum of singular modes, where the contribution of the k-th singular mode is scaled by a gain

$$g_k(\beta, \omega) = \frac{\sigma_k(\omega)}{\sigma_k(\omega)^2 + \beta}. \tag{12}$$

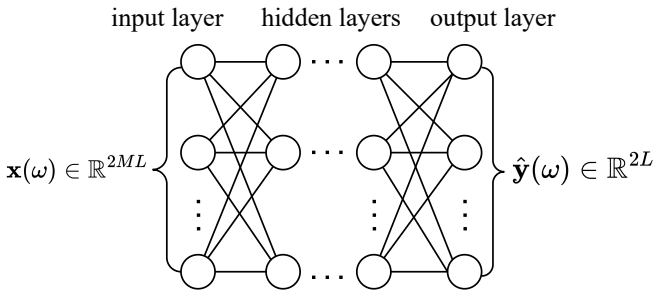

Fig. 1: A real-valued MLP that maps the transfer functions at frequency ω to the loudspeaker filters at ω.

For $\sigma_k(\omega)^2 \gg \beta$, $g_k \approx 1/\sigma_k$; for $\sigma_k(\omega)^2 \ll \beta$, $g_k \approx \sigma_k/\beta$. In the absence of regularization ($\beta = 0$), the gain reduces to $1/\sigma_k$, which strongly amplifies modes associated with small singular values. Therefore, introducing β suppresses these ill-conditioned modes and stabilizes the solution.

For each perturbation type p, we form the perturbed matrix $\mathbf{G}_p(\omega)$ and its deviation from the setup matrix $\mathbf{G}(\omega)$,

$$\Delta\mathbf{G}_p(\omega) = \mathbf{G}_p(\omega) - \mathbf{G}(\omega). \tag{13}$$

We then project $\Delta\mathbf{G}_p(\omega)$ onto the singular bases of $\mathbf{G}(\omega)$[26],

$$\Delta\mathbf{S}_p(\omega) = \mathbf{U}(\omega)^{\mathrm{H}}\,\Delta\mathbf{G}_p(\omega)\,\mathbf{V}(\omega) \in \mathbb{C}^{M\times M}, \tag{14}$$

whose (i,j)-th entry describes how the perturbation couples the j-th input singular vector to the i-th output singular vector, which correspond to the column vectors of $\mathbf{V}$ and $\mathbf{U}$, respectively. The coupling effect describes how an input singular direction excites multiple output singular directions under perturbation. To compare different perturbation types, we normalize the magnitude of each entry by the Frobenius norm of the perturbation,

$$a_{p,ij}(\omega) = \frac{\left|\left[\Delta\mathbf{S}_p(\omega)\right]_{ij}\right|}{\left\|\Delta\mathbf{G}_p(\omega)\right\|_F}. \tag{15}$$

Then we aggregate over frequencies by taking the median:

$$\tilde{a}_{p,ij} = \mathrm{median}_\omega\, a_{p,ij}(\omega). \tag{16}$$

4 Experiment

4.1 Experimental Setup

The experimental setup simulates a 2D rectangular room with dimensions 3 m × 5 m. The loudspeaker array consists of $L = 12$ loudspeakers positioned in a vertical array with a spacing of 0.1 m between each loudspeaker. The center of the loudspeaker array is [2.5, 2.5] m. There are two listeners: one in the bright zone with ears located at [2.0, 2.81] m and [2.0, 2.99] m, representing a head width of 18 cm, and the other in the dark zone with ears at [2.0, 2.01] m and [2.0, 2.19] m. The four ears of the two listeners are the $M = 4$ microphones, and the listeners face the loudspeaker array throughout the simulation.

For both the MLP and the PM methods, we report AC in three robustness tests, which are listener movement, transfer function (TF) noise, and reverberation. For the listener movement test, the evaluation microphones were shifted +3 cm in the y-direction from their training positions for both listeners. For the TF noise test, complex Gaussian noise with SNR = 20 dB is added to the transfer functions during evaluation. The reverberation test adds room reflections using the image source method, where the reflection order is 1, and the absorption coefficient is 0.5. Unless otherwise specified, evaluations are conducted using the listener movement test.

4.2 Baseline Methods

Three baseline methods are compared to the proposed MLP method: PM, a whole-frequency MLP, and CNN.

In theory, the unregularized case for PM corresponds to $\beta = 0$. In practice, we set $\beta = \varepsilon$ with $\varepsilon = 10^{-15}$ as a numerically stable approximation of $\beta = 0$ so that the matrix is not singular. We sweep different values of β and observe a broad optimal performance plateau for $\beta \in [10^{-13}, 10^{-4}]$, which will be shown in later results. We therefore use $\beta = 10^{-4}$ as the regularized PM baseline.

The whole-frequency MLP takes all frequency bins jointly as input, in contrast to the proposed single-frequency MLP, which processes one frequency bin at a time. Unless otherwise stated, the term MLP refers to the proposed single-frequency model. The input dimensionality of the whole-frequency MLP is $2MLN$, where N denotes the number of frequency bins, which results in a training dataset of size one. The whole-frequency MLP uses the same network architecture as the single-frequency MLP and is trained with dropout and a Tikhonov penalty ($\beta = 10^{-4}$) in the loss function. We also implement and evaluate the CNN architecture proposed in [16].

4.3 Datasets

Three different datasets are used: the RIR dataset, the RIR reverb dataset, and the binaural room impulse responses (BRIR) dataset. RIRs are simulated using the image source method using the pyroomacoustics library [27], and sound zone calculations are performed using a custom pyzone framework based on [28]. The RIR dataset is simulated with the free-field assumption. For the RIR reverb dataset, the image source method reflection order is set to 1 and the absorption coefficient of 0.5. The BRIR dataset is simulated using the RAZR toolbox [29]. We choose the small KEMAR head's head-related impulse responses from the CIPIC dataset [30]. The BRIRs are synthesized under a free-field assumption, where only the direct sound path is considered and no room reflections or late reverberation are included.

We use log frequency sampling for selecting frequency bins for the training dataset, which distributes 2000 points logarithmically from 20 Hz to 22050 Hz. This reduces the dominance of densely sampled high-frequency components compared to linear frequency sampling and makes the training process more stable. Therefore, the size of the training data is 2000.

4.4 Training Details

All experiments are implemented using PyTorch for neural network training and inference. To ensure reproducibility, all random number generators were seeded with a fixed value of 42. The network has the configuration [256, 512, 512, 512, 256], where the numbers represent the number of units in each hidden layer. Each hidden layer is followed by a ReLU activation function and a dropout layer with a probability of 0.2 to prevent overfitting. The network is trained for 5000 epochs with a batch size of 512. The Adam optimizer is used for training, and the learning rate is 10^{-4}. Training is performed on CPU (Apple Silicon) and completed in approximately 8 minutes.

5 Results

5.1 MLP vs PM

Figure 2 shows AC of PM in the listener movement test with different β values. The results confirm the statement in the experiment section that there is a plateau of AC with β values from 10^{-14} to 10^{-4}.

Figure 3 compares MLP to PM with and without Tikhonov regularization for the listener movement test. By comparing PM with regularization (the red line) and PM without regularization (the green line), we can see that Tikhonov regularization reduces the low frequency performance but improves the high frequency performance. The same applies to the comparison of MLP (the blue line) and PM with regularization (the red line). At low frequencies, all three methods reach contrasts of around 30 dB, beyond which further improvements have limited perceptual impact. At high frequencies, however, AC values fall below 10 dB, making contrast improvements perceptually meaningful. This motivates our use of high-frequency AC as a complementary evaluation metric to better reflect perceptual differences between the methods. Here we define high-frequency AC as AC above 1000 Hz.

Overall, MLP attains AC of 26.48 dB, compared to 24.85 dB for PM, corresponding to a 1.63 dB improvement. Although the overall improvement is limited, we can observe that the main improvement is in the high frequency. If we compare the high-frequency AC, MLP achieves AC of 16.08 dB compared to 9.66 dB for PM, with a 6.42 dB improvement. This suggests that the MLP provides a stronger regularization at high frequencies, mitigating the sensitivity of PM to listener movement.

5.2 Robustness to Perturbations

Table 1 shows AC of MLP and PM with or without Tikhonov regularization in different perturbation tests. MLP demonstrates superior robustness to listener position shifts. However, for transfer function noise and reverberation, regularized PM outperforms MLP. In

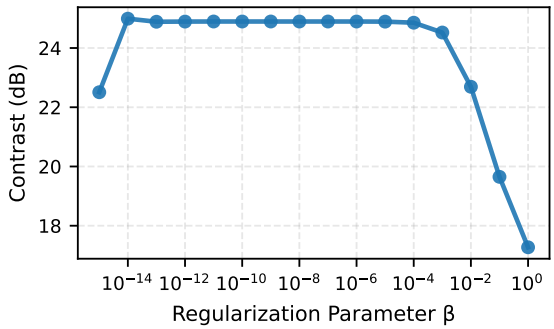

Fig. 2: Acoustic contrasts of PM using different β values.

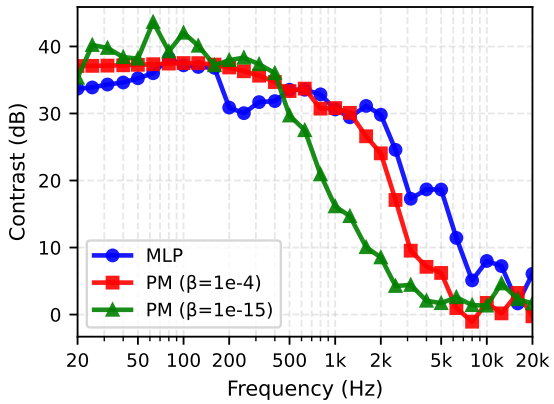

Fig. 3: Comparison between MLP and PM with or without regularization in the listener movement test.

summary, MLP excels at generalizing to systematic, learnable perturbations (spatial shifts), while regularized PM demonstrates inherent robustness to random disturbances (noise). To better understand how different perturbations affect the plant matrix and why Tikhonov regularization improves robustness in the reverberant and noisy cases, we analyze the transfer function perturbation in the singular vector basis.

Figure 4 shows the 4×4 matrix $\tilde{a}_{p,ij}$ as a heatmap. For listener movement, the diagonal entries dominate: the magnitude of the perturbation along modes 1 and 2 is around 0.4 and 0.35, while modes 3 and 4 are weaker. Off–diagonal terms are noticeably smaller than the diagonal. This indicates that head movements mainly perturb the dominant singular directions, with relatively little mixing between different modes. Reverberation distributes energy more evenly across the four modes. The diagonal magnitudes are similar, and the off–diagonal terms reach comparable levels. Hence, reverberation not only perturbs the dominant modes, but also injects energy into weaker modes and induces cross–couplings between them. In the TF–noise case, the heatmap is almost uniform: all entries are 0.12. This behaviour is consistent with additive Gaussian noise that is approximately isotropic.

Table 1: Full-band acoustic contrasts (dB) of MLP and PM with or without Tikhonov regularization in different robustness tests. Shift, Noise, and Reverb correspond to the listener movement test, the TF noise test, and the reverberation test.

Method	Shift	Noise	Reverb
MLP	**26.48**	25.77	17.22
non-regularized PM	22.51	16.21	12.30
regularized PM	24.85	**26.91**	**17.58**

As we mentioned before, Tikhonov regularization limits the filter gain in directions that are weakly excited by the setup plant. This explains why it is more beneficial for reverberation and TF noise than for listener movement. In the listener movement case, $\Delta\mathbf{G}$ is largely confined to the dominant modes. Regularization cannot compensate for such coherent distortions of the leading singular modes. In contrast, reverberation and TF noise both inject energy into weaker modes and off–diagonal couplings. These perturbations strongly excite directions associated with small singular values whose gain is suppressed by the Tikhonov term.

5.3 Network Architectures

To understand the regularization mechanisms of each part, we conduct an ablation study to compare dropout, Tikhonov penalties, implicit regularization. We also compare the network with the whole-frequency MLP and CNN to demonstrate the effectiveness of structured regularization from shared parameters of different frequencies.

We compare the results with and without Tikhonov regularization and with and without dropout. Without the Tikhonov penalty, high-frequency AC is 16.08 dB with dropout and 11.82 dB without dropout. When the Tikhonov penalty is included, the corresponding AC

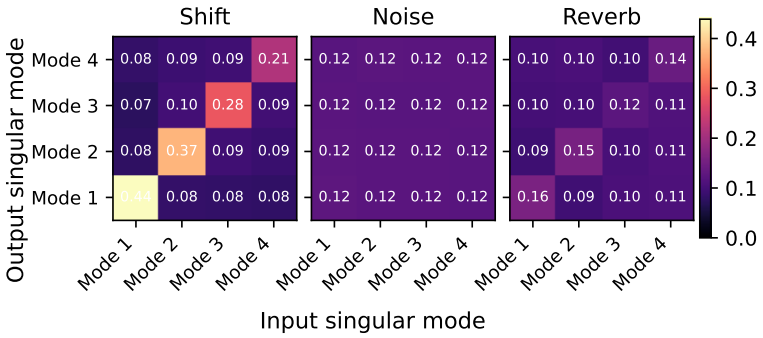

Fig. 4: Visualization of energy distribution of the perturbation in the singular vector basis.

values are 15.98 dB with dropout and 11.69 dB without dropout. Overall, the difference between using and not using the Tikhonov penalty is marginal: removing the penalty even leads to a slight increase in AC (about 0.1 dB) in both the dropout and non-dropout cases. In contrast, enabling dropout consistently improves AC by approximately 4 dB compared to the no-dropout setting.

Compared with regularized PM, MLP without dropout improves high-frequency AC by about 2 dB, likely due to implicit regularization, while applying dropout increases this improvement to more than 6 dB. This supports the view that dropout and implicit regularization bias the network toward flatter regions of the solution manifold, and play a more dominant role than Tikhonov penalties.

The whole-frequency MLP achieves AC of 19.58 dB over the full frequency band, while the high-frequency AC is only 3.82 dB, which is substantially lower than the performance of both PM and the single-frequency MLP. When all frequency bins are jointly provided as network inputs, different frequencies are effectively optimized independently, and the model lacks an inductive bias that enforces parameter sharing across frequencies.

CNN achieves AC of 22.10 dB, while the high-frequency AC is 4.84 dB. Although the performance is better than the whole-frequency MLP, it remains inferior to both PM and the single-frequency MLP. This suggests that, while the inductive bias introduced by convolutional layers helps exploit structured information and improves generalization compared to the whole-frequency MLP, it is still insufficient under severe data scarcity, leading to suboptimal performance.

These results suggest a clear hierarchy among different regularization mechanisms. Within training-time regularization, dropout plays a more important role than the Tikhonov penalty. In contrast, cross-frequency parameter sharing acts as a dominant form of structured regularization, without which neither implicit regularization nor training-time regularization alone is sufficient.

5.4 Different Datasets

Table 2: Acoustic contrasts (dB) above 1000 Hz using RIR, RIR reverb, and BRIR datasets for the listener movement test.

Method	RIR	RIR reverb	BRIR
MLP	**16.08**	**12.72**	**8.89**
PM	9.73	9.66	8.78

Table 2 is the high-frequency AC comparison between different datasets. Although MLP performs better than PM for all three datasets, the performance difference between MLP and PM is negligible for the BRIR dataset. The performance gap between the two methods is also smaller for RIR reverb datasets (3.06 dB) compared to the RIR datasets (6.35 dB), suggesting that the model performs better for smooth free-field data. At high frequencies, BRIRs exhibit sharp spectral notches/peaks and rapid spatial variation, making the mapping from the transfer function to the control filter highly non-smooth compared to the RIR case. MLP tends to impose a smooth structure, which can become overly restrictive for these high-frequency BRIR fine structures, leading to underfitting and reduced contrast.

6 Conclusion

We proposed a frequency-domain neural network solution for PSZ filter design, in which an MLP learns to map frequency-domain transfer functions to loudspeaker weights using a PM-inspired loss. We find that this single-frequency input network architecture can impose a structured regularization and perform more robustly than inputting all frequency bins. Also, dropout can act as a stronger regularizer than explicit Tikhonov penalties. The proposed method offers an alternative regularization mechanism that avoids the need for manual tuning of regularization parameters, as required in conventional Tikhonov-based approaches.

We also evaluate the performance in different perturbation settings and using different datasets. If dominant perturbations are structured (e.g., listener movement), learning-based predictors with shared cross-frequency parameters are advantageous. If perturbations are unstructured (e.g., TF noise, reverberation), traditional PM remains critical. For BRIR-based PSZ, high-frequency fine structure limits the effectiveness of smooth neural mappings. This suggests a fundamental limitation of learning smooth mappings for BRIR-based PSZ at high frequencies.

Future work will explore training across a wider range of listener positions and room configurations to improve generalization to unseen transfer functions, and will include perceptual experiments to assess whether the observed improvements in acoustic contrast translate into perceptual improvements for listeners.

References

[1] Vindrola, L. et al., "Pressure matching with forced filters for personal sound zones application," *Journal of the Audio Engineering Society*, 68(11), pp. 832–842, 2020.

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

[3] Lee, T., Nielsen, J. K., and Christensen, M. G., "Signal-adaptive and perceptually optimized sound zones with variable span trade-off filters," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28, pp. 2412–2426, 2020.

[4] Li, Y., Wang, L., and Reiss, J., "Impact of mismatched room acoustic modeling on transaural reproduction with loudspeaker arrays," in *Audio Engineering Society Convention 155*, 2023.

[5] 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.

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

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

[8] Golub, G. H., Hansen, P. C., and O'Leary, D. P., "Tikhonov regularization and total least squares," *SIAM journal on matrix analysis and applications*, 21(1), pp. 185–194, 1999.

[9] Nelson, P., "A review of some inverse problems in acoustics," *International journal of acoustics and vibration*, 6(3), pp. 118–34, 2001.

[10] Zhou, T., Yasueda, K., and Kataoka, A., "Efficient Tikhonov Regularization Parameter Selection for Multi-Zone Sound Field Reproduction," *Acoustical Science and Technology*, pp. e25–12, 2025.

[11] Zhu, Q. et al., "Robust personal audio geometry optimization in the SVD-based modal domain," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 27(3), pp. 610–620, 2018.

[12] Hamdan, E. C. and Fazi, F. M., "A modal analysis of multichannel crosstalk cancellation systems and their relationship to amplitude panning," *Journal of Sound and Vibration*, 490, p. 115743, 2021.

[13] Kim, Y. and Nelson, P., "Optimal regularisation for acoustic source reconstruction by inverse methods," *Journal of sound and vibration*, 275(3-5), pp. 463–487, 2004.

[14] Zhou, T., Yasueda, K., and Kataoka, A., "A comparative study of Tikhonov regularization parameter selection approaches for the pressure matching method," in *Proc. Autumn Meet. Acoust. Soc. Jpn*, pp. 167–170, 2024.

[15] Zhu, Q. et al., "Robust acoustic contrast control with reduced in-situ measurement by acoustic modelling," *Journal of the Audio Engineering Society*, 65(6), pp. 460–473, 2017.

[16] Pepe, G. et al., "Deep learning for individual listening zone," in *2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP)*, pp. 1–6, IEEE, 2020.

[17] Pepe, G. et al., "Deep optimization of parametric IIR filters for audio equalization," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 30, pp. 1136–1149, 2022.

[18] Pepe, G. et al., "Digital filters design for personal sound zones: A neural approach," in *2022 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, IEEE, 2022.

[19] Qiao, Y. and Choueiri, E., "SANN-PSZ: Spatially Adaptive Neural Network for Head-Tracked Personal Sound Zones," *IEEE Transactions on Audio, Speech and Language Processing*, 33, pp. 2735–2748, 2025.

[20] Zhang, C. et al., "Understanding deep learning (still) requires rethinking generalization," *Commun. ACM*, 64(3), p. 107–115, 2021, ISSN 0001-0782.

[21] Cooper, Y., "The loss landscape of overparameterized neural networks," *arXiv preprint arXiv:1804.10200*, 2018.

[22] Karhadkar, K. et al., "Mildly overparameterized ReLU networks have a favorable loss landscape," *arXiv preprint arXiv:2305.19510*, 2023.

[23] Jackson, P., Fazi, F., and Coleman, P., "Personalising sound over loudspeakers," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Brighton, UK, 2019.

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

[25] Coleman, P. et al., "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.

[26] Stewart, G. W. and Sun, J.-g., *Matrix perturbation theory*, Academic Press, 1990.

[27] Scheibler, R. et al., "Pyroomacoustics: A Python Package for Audio Room Simulation and Array Processing Algorithms," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Calgary, 2018.

[28] Coleman, P. et al., "Personal audio with a planar bright zone," *The Journal of the Acoustical Society of America*, 136(4), pp. 1725–1735, 2014.

[29] Wendt, T., Van De Par, S., and Ewert, S., "A computationally-efficient and perceptually-plausible algorithm for binaural room impulse response simulation," *Journal of the Audio Engineering Society*, 62(11), pp. 748–766, 2014.

[30] Algazi, V. et al., "The CIPIC HRTF Database," in *IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, New Paltz, 2001.