Neural Regularization for Personal Sound Zones

Yazhou Li, Lin Wang, Joshua Reiss,

Center for Digital Music, Queen Mary University of London

Queen Mary University of London

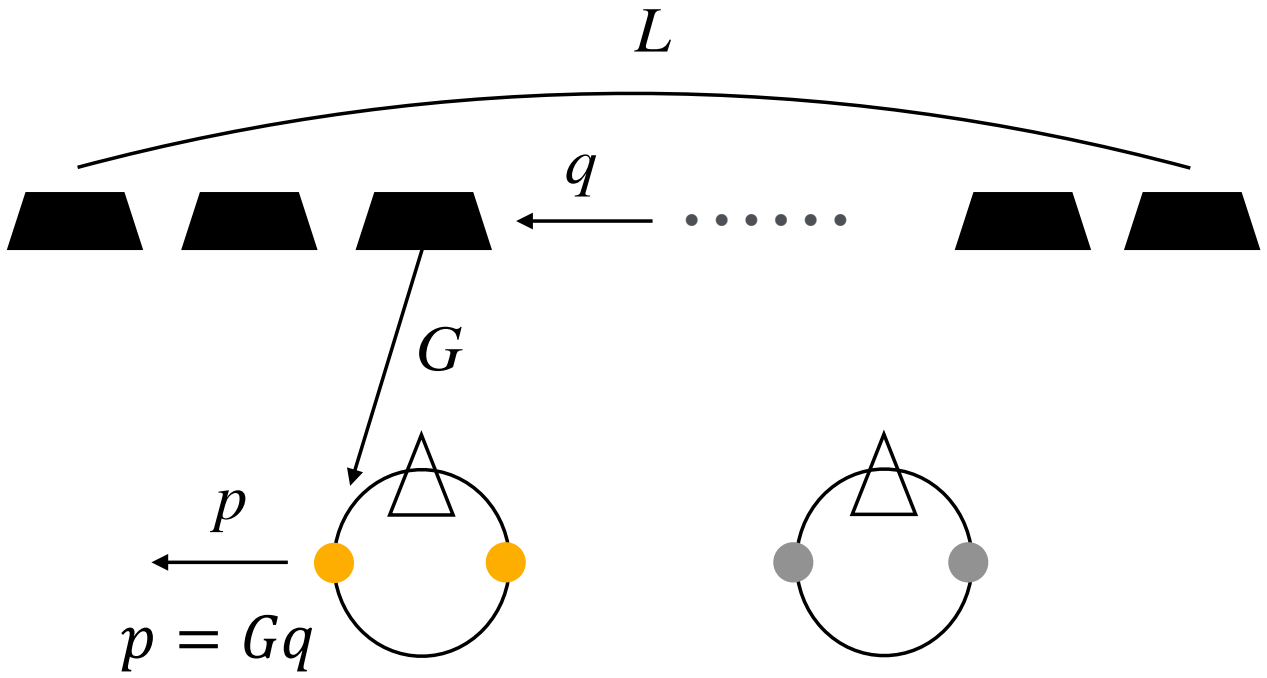
UK Research and Innovation

AI + MUSIC

centre for digital music

Background: Personal Sound Zone (PSZ)

Formulation:



Metric:

Acoustic Contrast $\frac{p_b}{p_d}$

Sound pressure level ratio between bright zone and dark zone

Optimization Method:

Pressure Matching (PM)

$$\min_q \|Gq - p_T\|^2 + \beta\|q\|^2$$

$$q = (G^{\mathrm{H}}G + \beta I)^{-1}G^H p_T$$

Background: Tikhonov regularization

- **Problem: Robustness**
  - Pressure matching is sensitive to mismatch.
    listener movement / noise / room reflections
  - ill-conditioning:
    Small singular value, large condition number

- **Evaluation: Singular value decomposition (SVD)**
  $\mathbf{G}(\omega) = \mathbf{U}(\omega)\,\mathbf{\Sigma}(\omega)\,\mathbf{V}(\omega)^{\mathrm{H}},$

- **Solution: Tikhonov regularization**
  $g_k(\beta, \omega) = \dfrac{\sigma_k(\omega)}{\sigma_k(\omega)^2 + \beta}.$
  introducing β suppresses these ill-conditioned modes

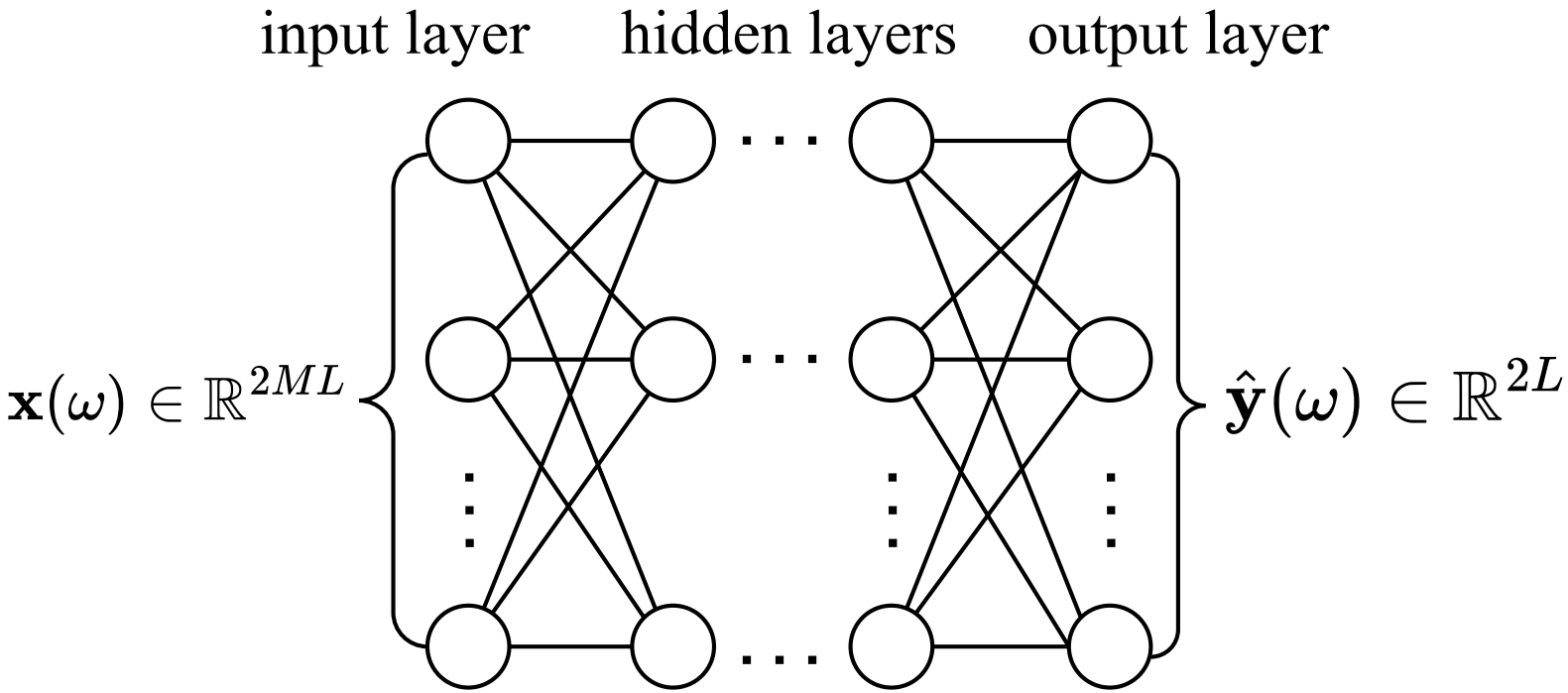
Motivation

- **Problem:**
  Tikhonov regularization helps, but β tuning is not trivial[1]
- **Current literature:**
  Neural networks have been used for PSZ optimization[2], but their robustness is not evaluated.
- **Question:**
  Can a neural network provide a different form of regularization and improve the robustness of PSZ?

[1] Tong, Zhou, “A Comparative Study of Tikhonov Regularization Parameter Selection Approaches for the Pressure Matching Method”, ASJ, 2024

[2] Pepe, Giovanni, "Deep optimization of parametric IIR filters for audio equalization." *IEEE/ACM TALSP*, 2022

Proposed Method: single-frequency MLP



pressure-matching loss $\mathcal{L} = \|p - p_T\|_2^2 + \beta\|q\|_2^2$

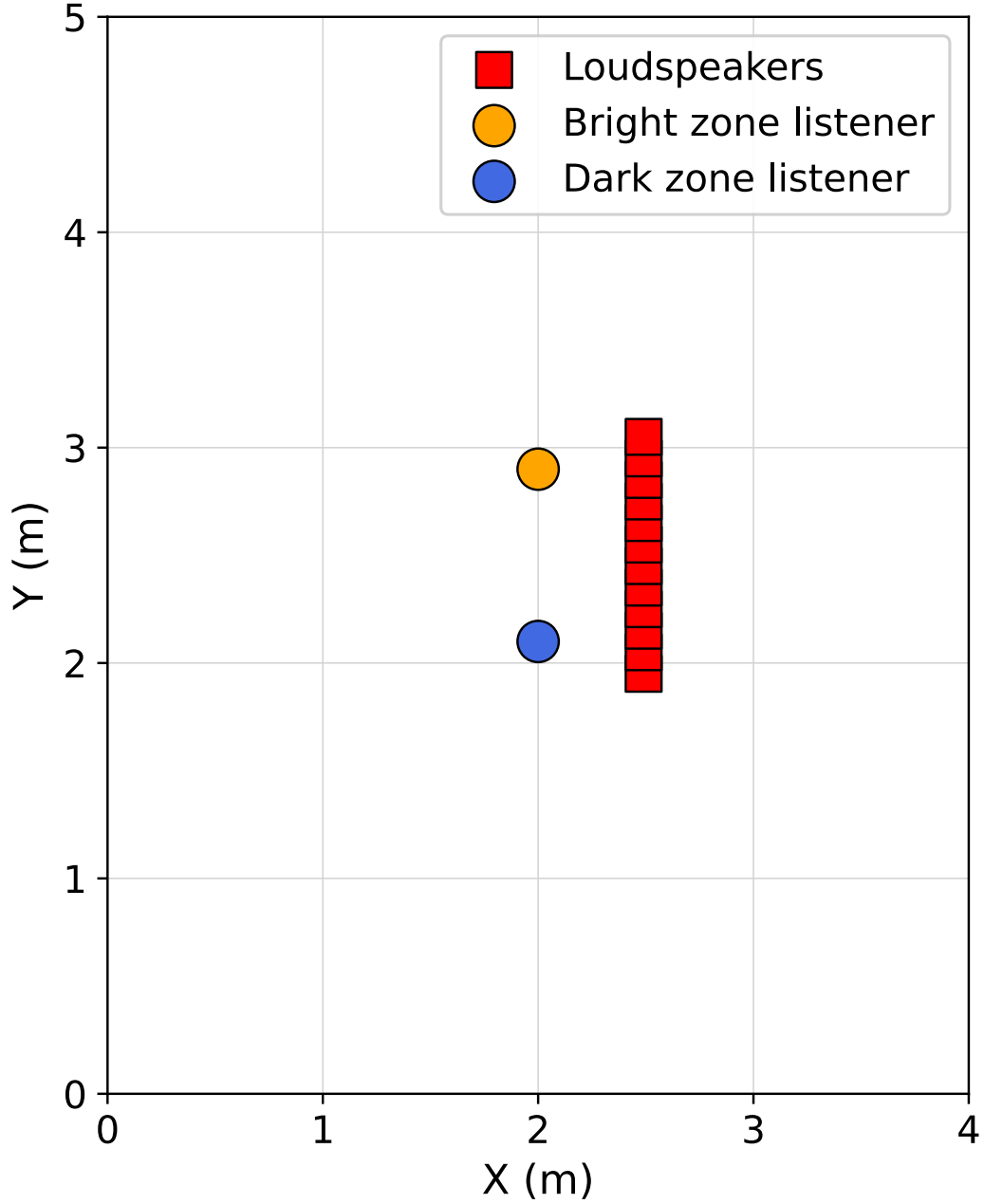
Regularization mechanisms

- **Implicit regularization**[1]
  over-parameterized MLP and optimization dynamics
- **Network architecture**
  shared network parameters across frequency
- **Explicit training-time regularization**
  - Tikhonov penalty
  - dropout

[1] Cooper, Yaim. “The loss landscape of overparameterized neural networks.” *arXiv,* 2018

Simulation

- **Setup transfer function:**
  free-field RIR (pyroomacoustics)
- **Evaluation transfer function:**
  three perturbation types:
  - Shift: 3cm to the left
  - Noise: 20dB SNR
  - Reverb:
    image source method order=1

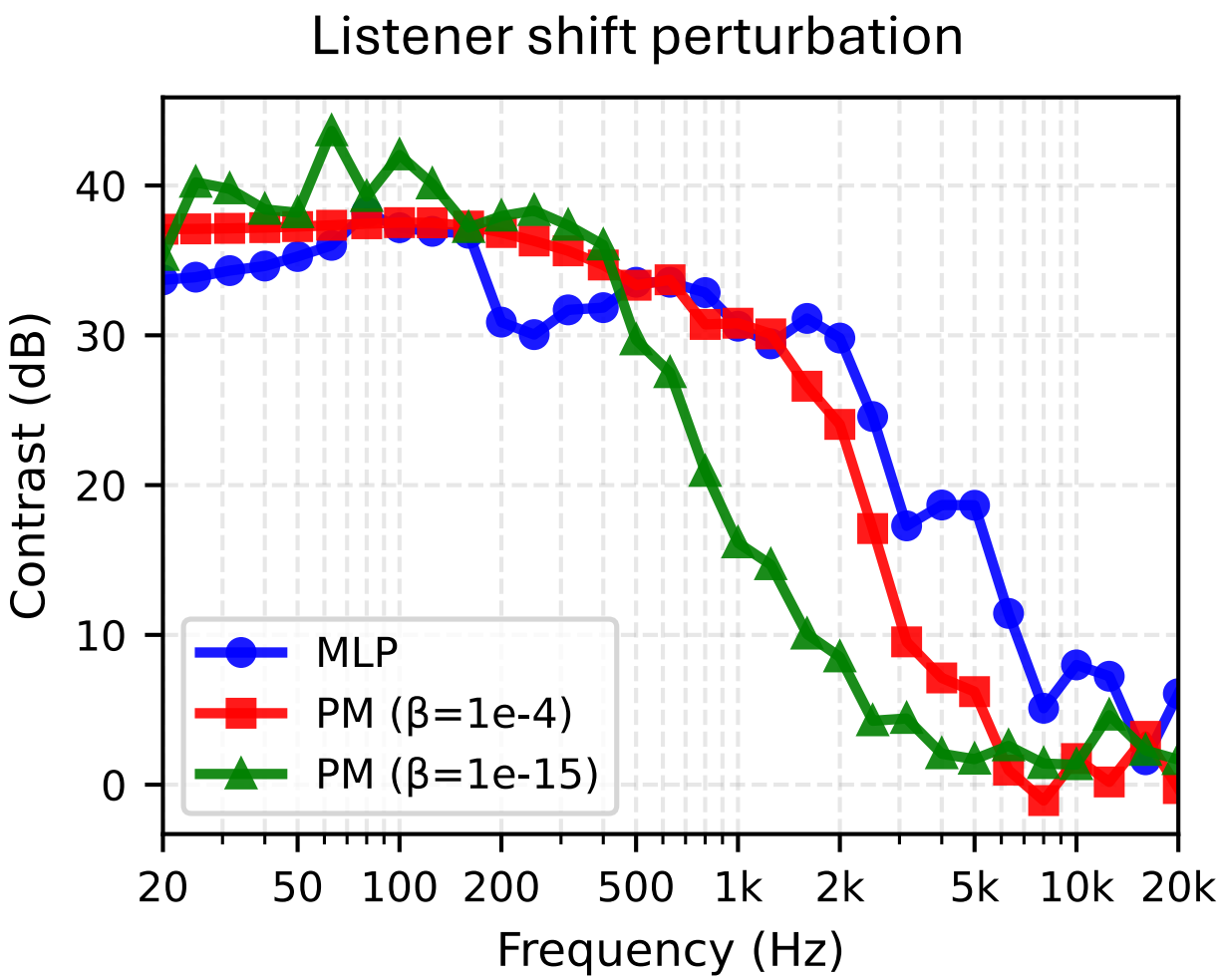


Simulation

- **Methods:**
  - Single-frequency MLP (proposed)
  - non-regularized PM
  - Regularized PM
  - whole-frequency MLP
  - Convolutional neural network (CNN)
- **Training:**
  - Dropout (0.2)
  - Tikhonov penalty (1e-4)
- **Dataset** (for the shift perturbation)**:**
  - Free-field RIR
  - RIR (image source method order=1)
  - Free-field BRIR (RAZR toolbox[1], CIPIC HRTF)

[1] Torben, W. “A Computationally-Efficient and Perceptually-Plausible Algorithm for Binaural Room Impulse Response Simulation” *JAES,* 2014

Result



Method	Shift	Noise	Reverb
single-frequency MLP (proposed)	**26.48**	25.77	17.22
non-regularized PM	22.51	16.21	12.30
regularized PM	24.85	**26.91**	**17.58**

MLP performs best for listener shift,
while regularized PM is better for noise and reverb.

Method: SVD-based analysis

- Calculate the perturbation of transfer matrix:

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

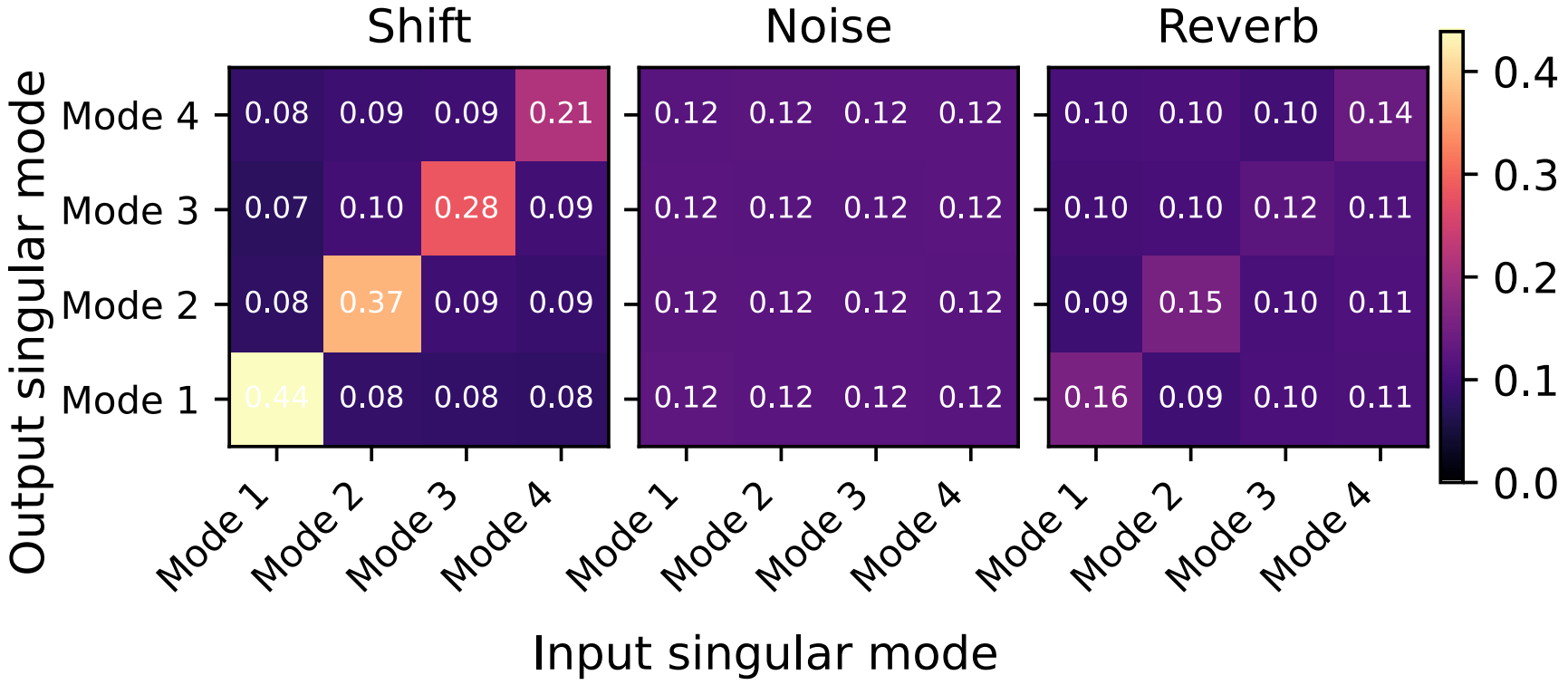
- Project each perturbation into the singular-vector basis:

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

- Interpretation:
  - each entry:
    how a perturbation in one input singular mode affects an output singular mode
  - diagonal dominant:
    perturbation stays in same modes and is more structured

Result: SVD-based explanation

Why Single-frequency MLP performs better for shift, not for noise and reverb?



Ablation study - Architectures

Result of listener shift

Model	Full-band AC (dB)	HF AC (dB)
single-frequency MLP (proposed)	**26.48**	**16.08**
whole-frequency MLP	19.58	3.82
CNN	22.10	4.84
regularized PM	24.85	9.66

Tikhonov	dropout	HF AC (dB)
✅	❌	11.69
❌	✅	**16.08**
❌	❌	11.82
✅	✅	15.98

Ablation study - Datasets

Result for the shift perturbation

Method	RIR	RIR reverb	BRIR
single-frequency MLP (proposed)	**16.08**	**12.72**	**8.89**
regularized PM	9.73	9.66	8.78

The benefit decreases for BRIRs, likely because high-frequency BRIR fine structures are less smooth and harder for the MLP to regularize.

Conclusion

- A single-frequency MLP with shared parameters acts as structured regularization.
- It improves robustness to structured perturbation, such as listener position shift.
- Dropout and implicit regularization both work as regularization mechanisms; Tikhonov penalty in the loss function is not necessary.
- Future work:
  - Compare with other neural network architectures.
  - Physical experiment and perceptual evaluation.