Towards Perception-Informed Latent HRTF Representations

You (Neil) Zhang [1,2], *Andrew Francl* [2], *Ruohan Gao* [3], *Paul Calamia* [2], ***Zhiyao Duan*** [1], ***Ishwarya Ananthabhotla*** [2]

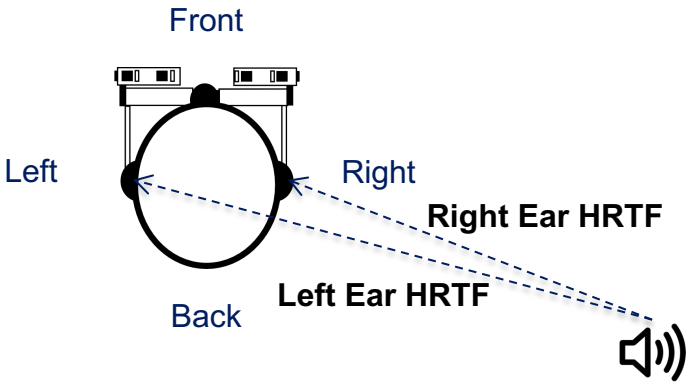
[1] University of Rochester
[2] Meta Reality Labs Research
[3] University of Maryland

IEEE WASPAA 2025

Tahoe City, CA - Oct 13, 2025

Head-Related Transfer Function (HRTF)

HRTF models the **acoustic filtering** effect of a listener's **head, ears, and torso** to enable 3D sound localization.



Left ear HRTF magnitudes (dB) of the midsagittal plane of one subject

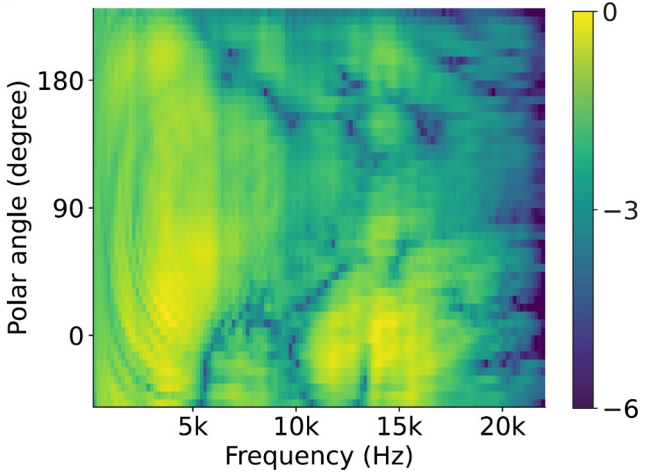


Figure from [Zhang+2023]

HRTF is unique to each person due to differences in ear, head, and torso shape.

Zhang, You, Yuxiang Wang, and Zhiyao Duan. "HRTF field: Unifying measured HRTF magnitude representation with neural fields." *ICASSP* 2023.

HRTF Applications: Virtual Spatial Audio Rendering

HRTFs encode human spatial cues to deliver immersive 3D sound.

Headphones

AR smart glasses

VR headsets

Figures generated by Gemini

Measure HRTFs

- An anechoic room
- Multiple loudspeakers on motorized arc
- Two microphones
- Head motion control

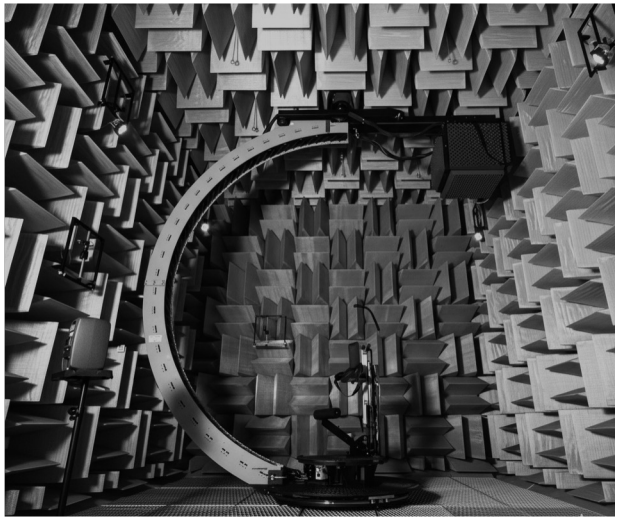


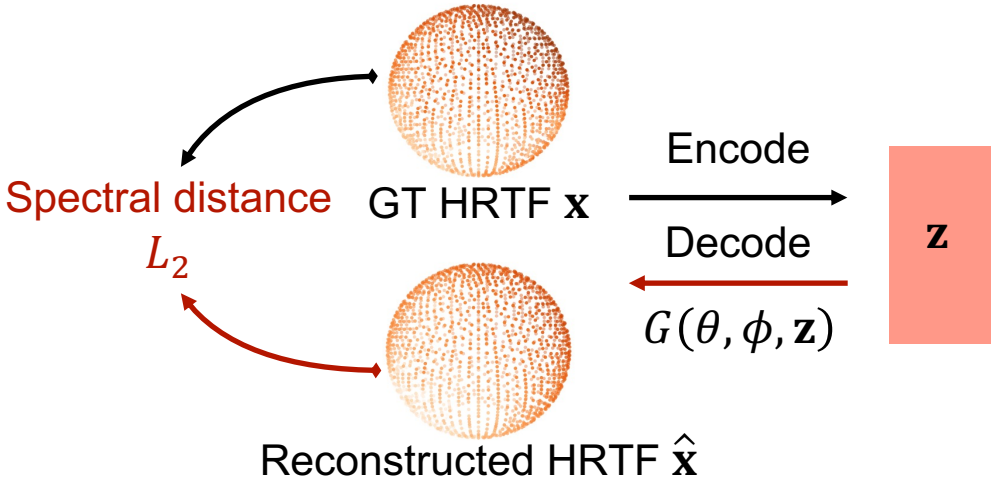
Figure from https://facebookresearch.github.io/SS2_HRTF/

Time-consuming & Resource-intensive!

Deep Learning for HRTF Representations

HRTF representation learning models

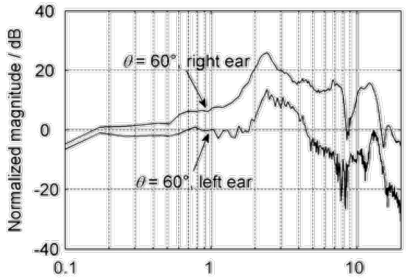
- Convolutional Autoencoder (CAE)
- Implicit Neural Representation (INR)



Zhao, Jiale, Dingding Yao, and Junfeng Li. “Head-related transfer function upsampling with spatial extrapolation features.” *IEEE TASLP* 2025.
Zhang, You, Yuxiang Wang, and Zhiyao Duan. “HRTF field: Unifying measured HRTF magnitude representation with neural fields.” *IEEE ICASSP* 2023.

Motivation

Most **existing** models are trained and evaluated with **spectral reconstruction**.



Spectral reconstruction

Perceptually plausible HRTF

Our contributions

Goal: Learn HRTF representations that more accurately reflect perceptual correlation, to enable better HRTF personalization for unseen users

- We study how well existing latent HRTF representations preserve perceptual relations, and introduce the benchmark for evaluating this.
- We propose a method for improving on this benchmark.
- We demonstrate practical utility for HRTF personalization.

1. How well do **existing** learned HRTF representations **preserve perceptual relations**?

HRTF Perception

Perceptual benefits of your *personal HRTF*:

- Reduced **Coloration** (less unwanted spectral distortion)
- Improved **Externalization** (sound appears outside the head)
- Enhanced **Localization** (accurately placing sounds in 3D space)

Majdak, Piotr, Bruno Masiero, and Janina Fels. “Sound localization in individualized and non-individualized crosstalk cancellation systems.” *JASA* 2013.
Brinkmann, Fabian, Alexander Lindau, and Stefan Weinzierl. “On the authenticity of individual dynamic binaural synthesis.” *JASA* 2017.
Jenny, Claudia, and Christoph Reuter. "Usability of individualized head-related transfer functions in virtual reality: Empirical study with perceptual attributes in sagittal plane sound localization." *JMIR Serious Games* 2020.

Coloration

Externalization

Localization

How do we mathematically model these?

Computational Auditory Modeling

Coloration: Predicted Binaural Coloration [McKenzie+2022]

Externalization: Auditory Externalization Perception [Baumgartner&Majdak2021]

Localization: Difference of Root Mean Square Error in Polar Angles [Barumerli+2023]

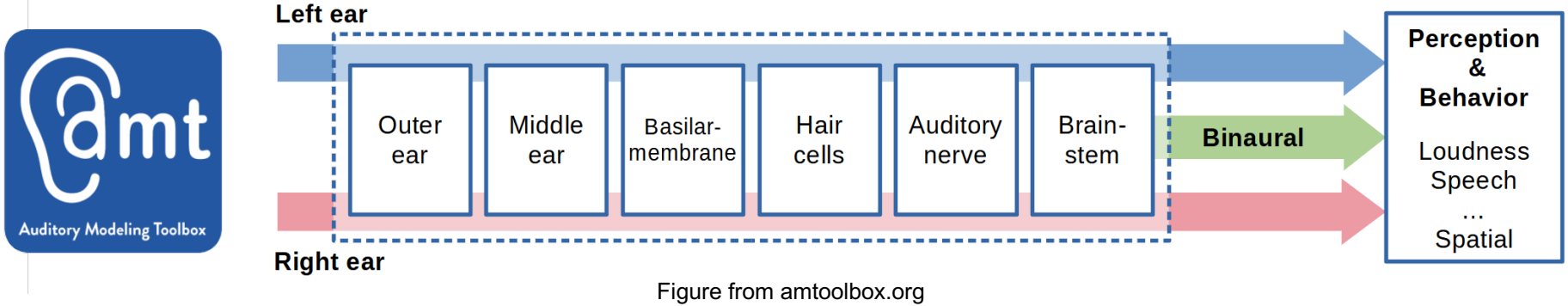


Figure from amtoolbox.org

McKenzie, Thomas, et al. "Predicting the colouration between binaural signals." *Applied Sciences* 2022.
Baumgartner, Robert, and Piotr Majdak. "Decision making in auditory externalization perception:model predictions for static conditions." *Acta Acustica* 2021
Barumerli, Roberto, et al. "A Bayesian model for human directional localization of broadband static sound sources." *Acta Acustica* 2023.

Computational Auditory Modeling

Coloration: **PBC** [McKenzie+2022]

Externalization: **AEP** [Baumgartner&Majdak2021]

Localization: **DRMSP** [Barumerli+2023]

Objective Perceptual Metrics

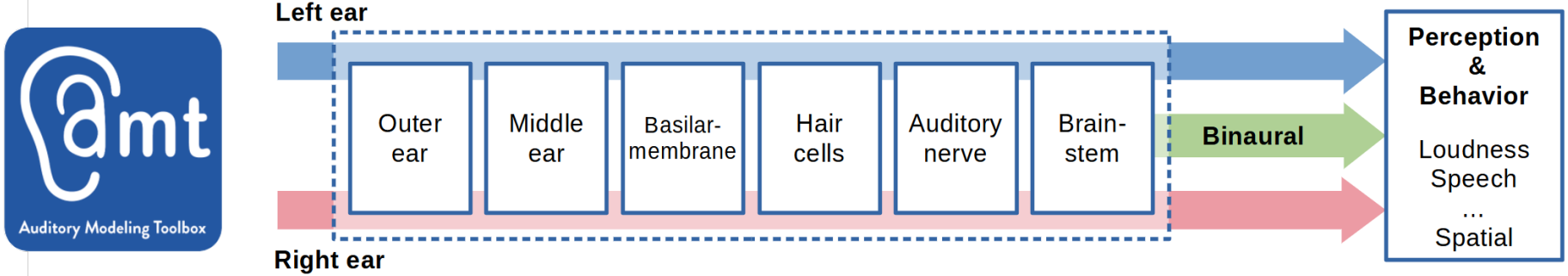


Figure from amtoolbox.org

McKenzie, Thomas, et al. "Predicting the colouration between binaural signals." *Applied Sciences* 2022.
Baumgartner, Robert, and Piotr Majdak. "Decision making in auditory externalization perception:model predictions for static conditions." *Acta Acustica* 2021
Barumerli, Roberto, et al. "A Bayesian model for human directional localization of broadband static sound sources." *Acta Acustica* 2023.

Experimental Setup

SS2 HRTF Database

- 1625 measurement locations
- 48 kHz sampling rate
- 78 subjects (65 for training, 13 for testing)

Train AI models with spectral reconstruction for HRTF data

↓

Compute pairwise latent distance across subjects

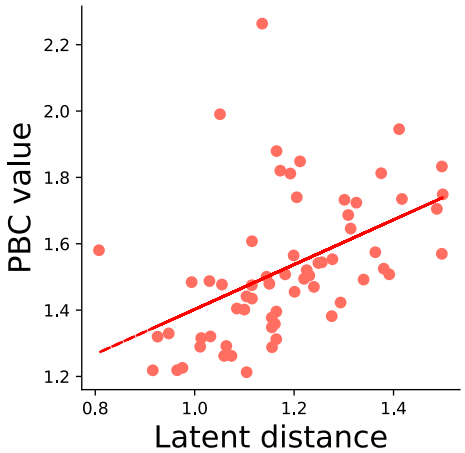
↓

Compute pairwise perceptual distance across subjects

Warnecke, Michaela, et al. "Sound sphere 2: A high-resolution HRTF database." *AES AVAR* 2024.

Alignment Between Latent Space and Perceptual Metrics

Pearson correlation (pairwise latent distances vs. perceptual distances)



$$\rho_{A,B} = \frac{\mathbb{E}[(A - \mu_A)(B - \mu_B)]}{\sigma_A \sigma_B}$$

A higher positive correlation indicates better alignment with human perception.

Partitions	**PBC**	**AEP**	**DRMSP**
train	0.60	0.60	0.40

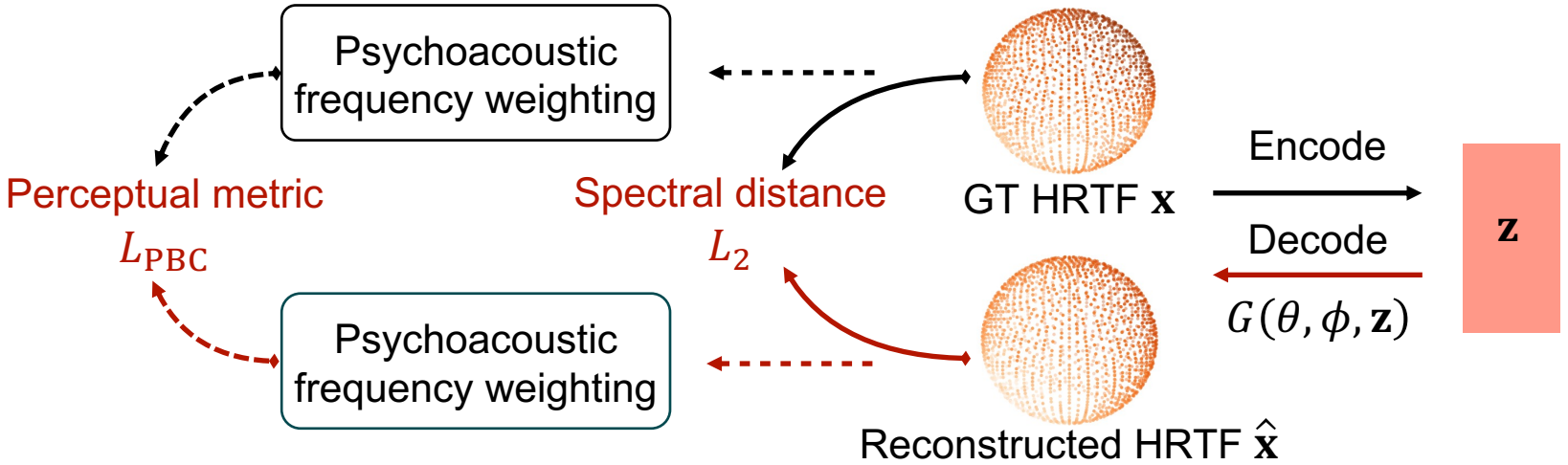
Minimizing spectral distances leads to limited perceptual correlation.

2. How do we **align** latent HRTF representations with **perception-informed space**?

Aligning with Perception-Informed Space

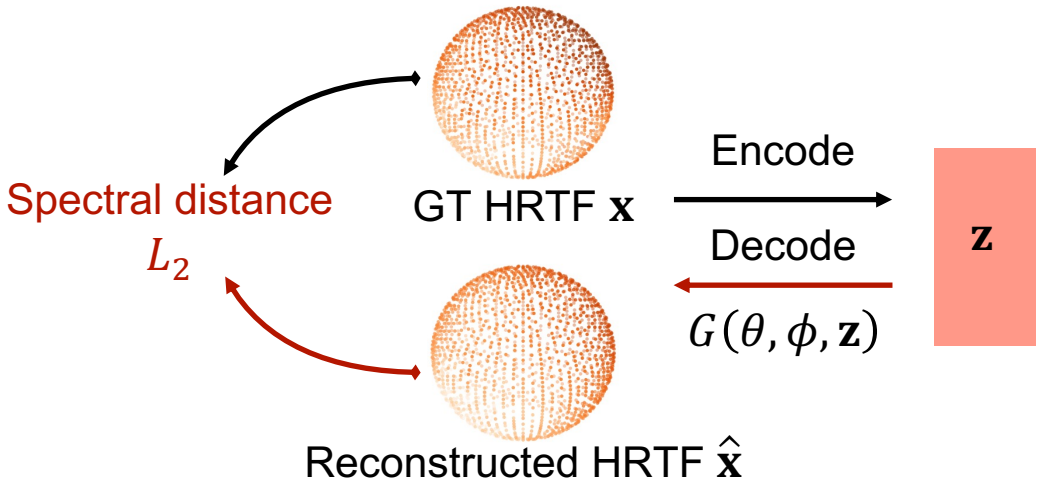
If the perceptual metric is differentiable, just add a straightforward perceptual loss.

- This only applies to PBC, which we reimplemented with PyTorch.



Aligning with Perception-Informed Space (Cont'd)

If the perceptual metric is not differentiable (AEP, DRMSP)



Metric multi-dimensional scaling (MMDS)

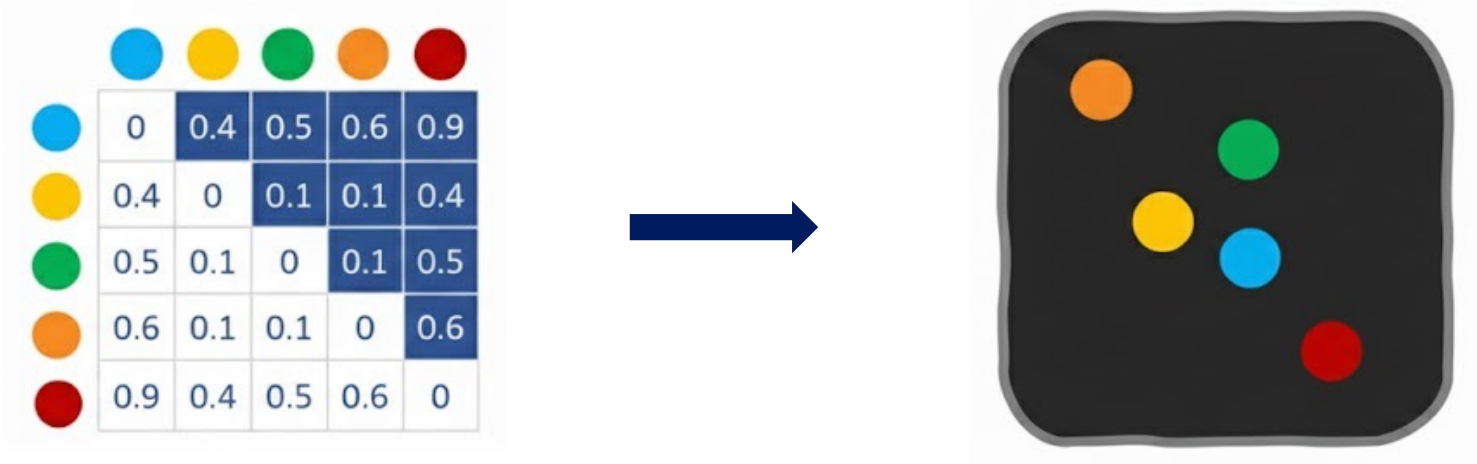
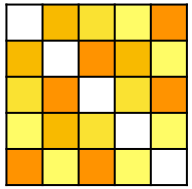


Figure from https://youtu.be/VKSJayDi_IQ post-processed by Gemini

Aligning with Perception-Informed Space (Cont'd)

If the perceptual metric is not differentiable (AEP, DRMSP)



Pairwise perceptual distance matrix $\mathbf{M}$

Metric multi-dimensional scaling (MMDS)

$\mathbf{z}_{\mathrm{MDS}}$

MMDS supervision

L_{Align}

This can also be applied to differentiable metrics.

GT HRTF $\mathbf{x}$

Spectral distance

L_2

Encode

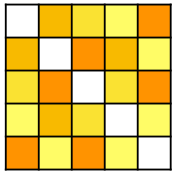
Decode

$G(\theta, \phi, \mathbf{z})$

$\mathbf{z}$

Reconstructed HRTF $\hat{\mathbf{x}}$

Overall Pipeline to Align Latent HRTF Representations



Pairwise perceptual distance matrix $\mathbf{M}$

Metric multi-dimensional scaling (MMDS)

$\mathbf{z}_{\text{MDS}}$

MMDS supervision

L_{Align}

Psychoacoustic frequency weighting

Perceptual metric

L_{PBC}

Spectral distance

L_2

GT HRTF $\mathbf{x}$

Encode

$\mathbf{z}$

Decode

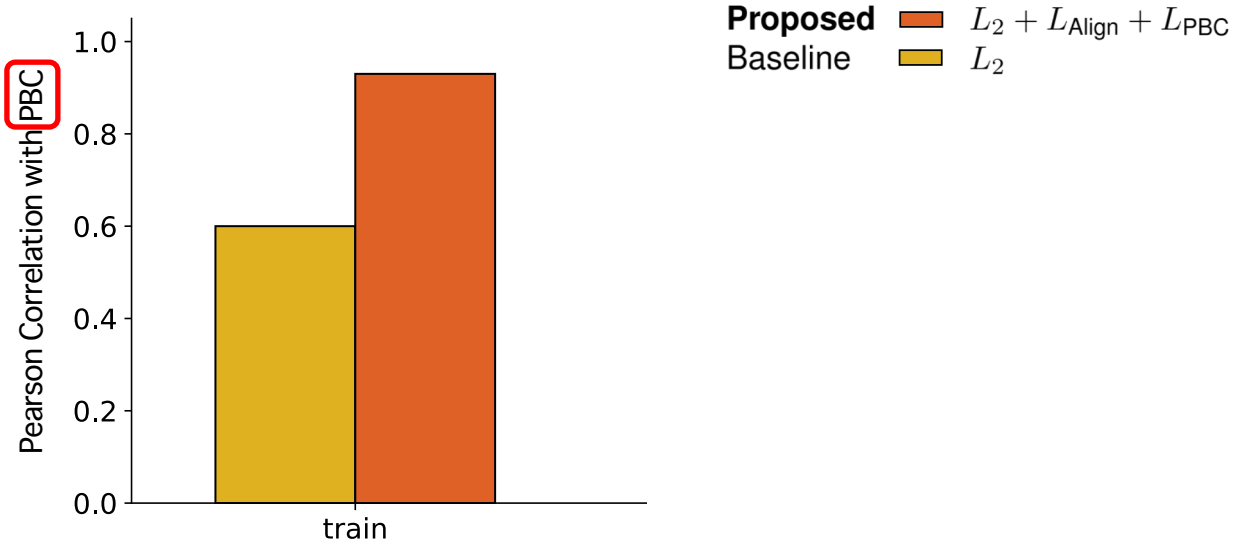
$G(\theta, \phi, \mathbf{z})$

Psychoacoustic frequency weighting

Reconstructed HRTF $\hat{\mathbf{x}}$

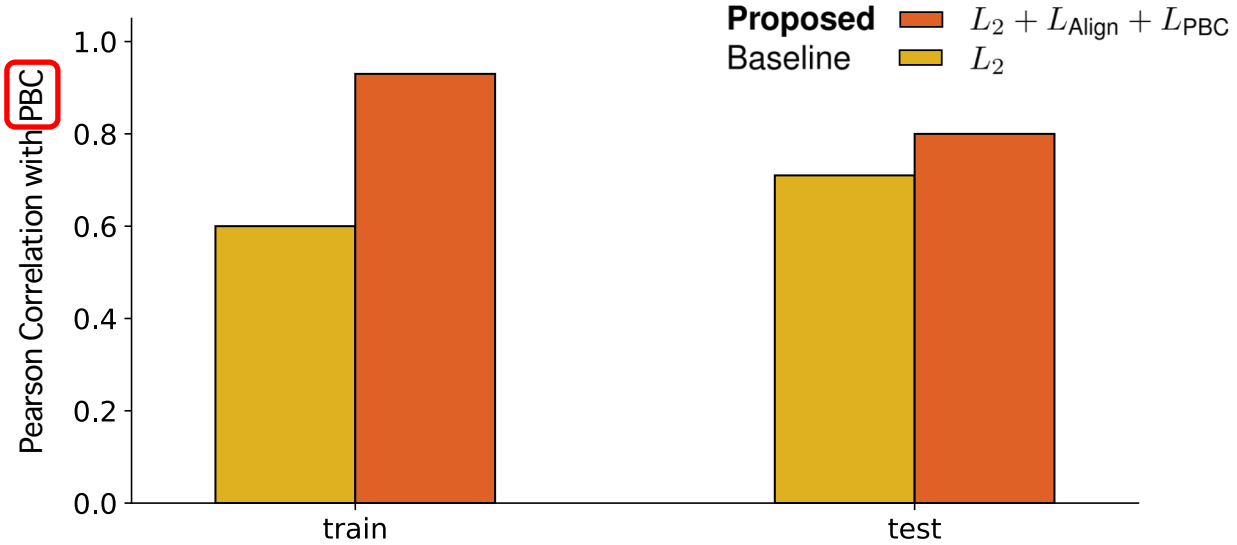
Results: Objective Perceptual Correlation Evaluation on PBC

- Our proposed method **achieves better alignment** with perception-informed space.
- The perceptual correlation learned in training transfer to test subjects (unseen).



Results: Objective Perceptual Correlation Evaluation on PBC

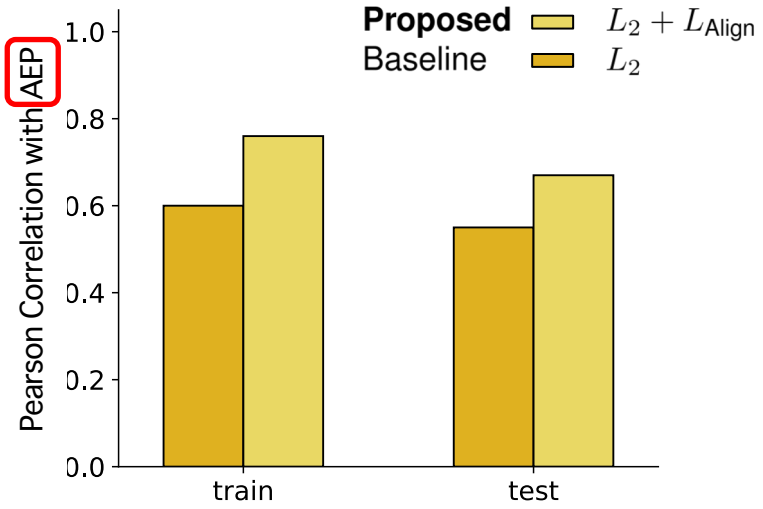
- Our proposed method **achieves better alignment** with perception-informed space.
- The perceptual correlation learned in training transfer to test subjects (unseen).

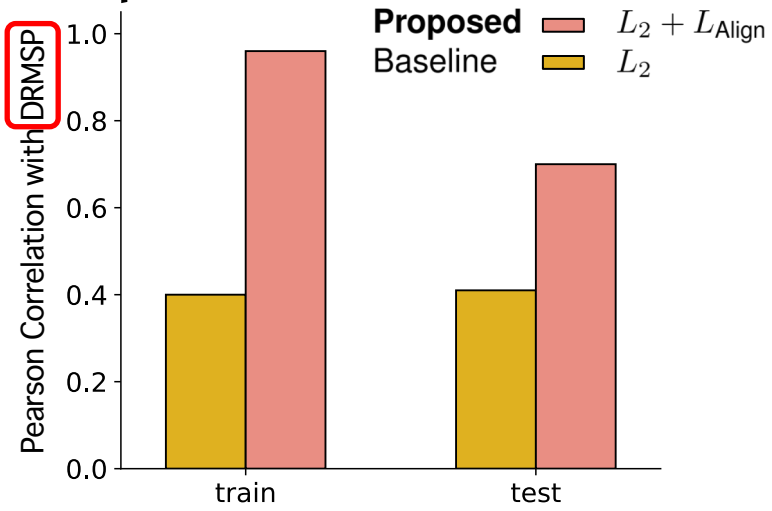


Generalization to AEP and DRMSP Metrics

- YES, our proposed correlation improvement method generalizes to externalization and localization.

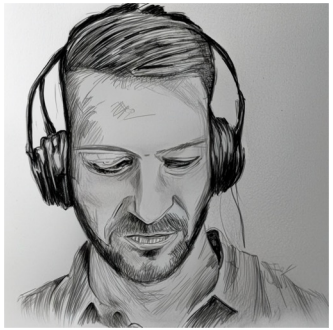
More results on the poster!





Application: Personalized HRTF Selection

For each of the 13 test (unseen) subjects, we select the nearest HRTFs from the 65 training subjects, based on the learned latent representations.



Methods	Best candidate	
	DRMSP↓	SDE (dB)↓
$L_2 + L_{\text{Align}}$	**3.20**	2.12
L_2	4.21	**2.07**

HRTFs selected by our proposed method yield lower perceptual distances with slightly higher SDE.

More results on the poster!

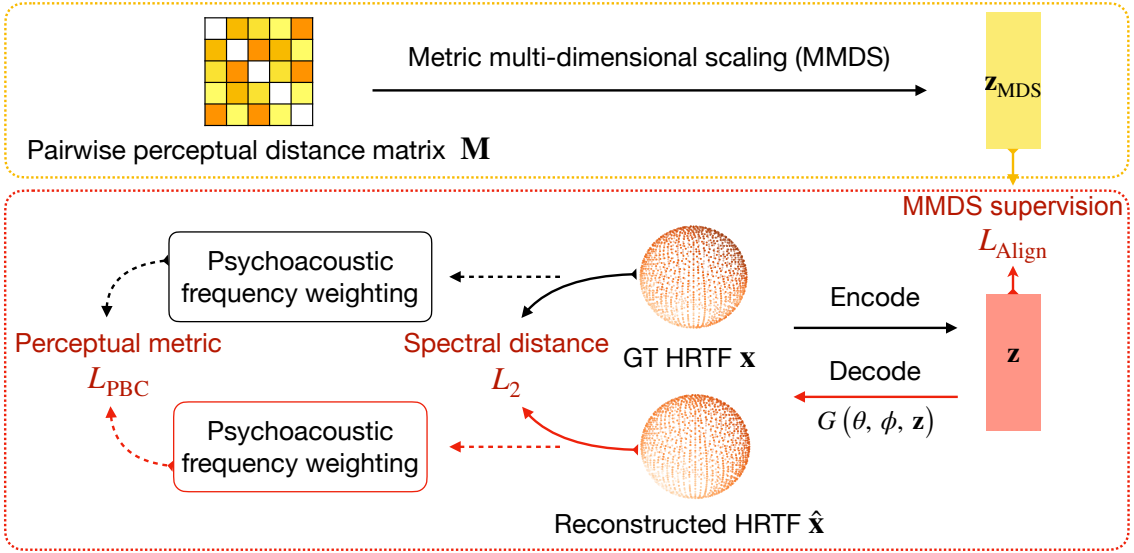
Limitations and Future Work

Limitations

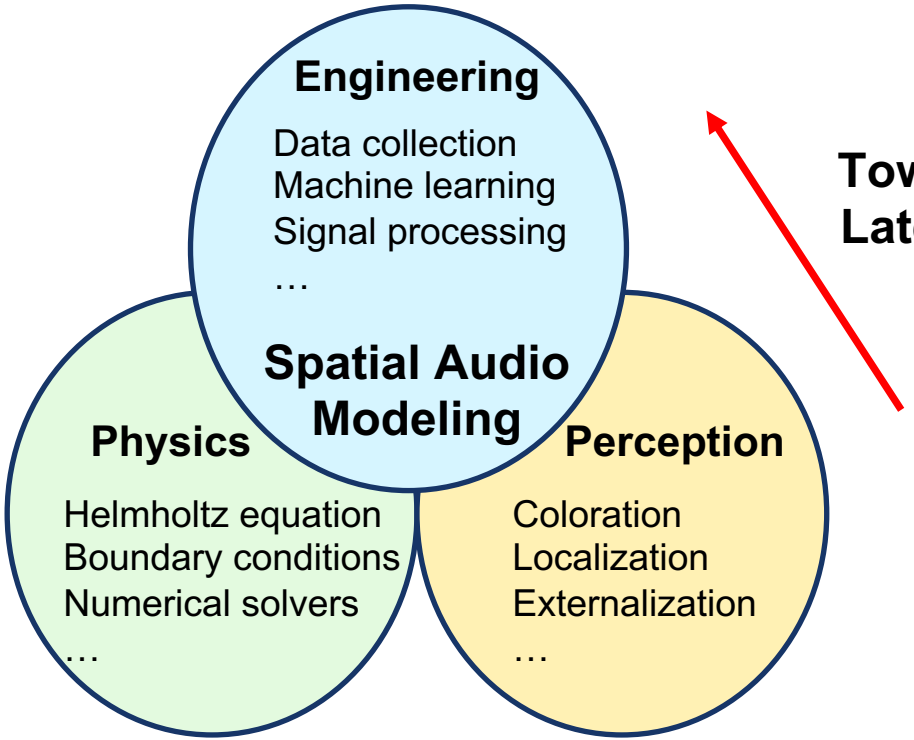
- Objective metrics vs. listening experience
- MMDS assumes symmetric dissimilarity
- Ignoring phase information

Future Work

- Subjective validation
- Extension to binaural synthesis



Future Vision



Towards Perception-Informed Latent HRTF Representations

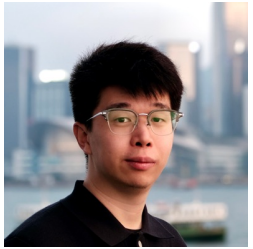
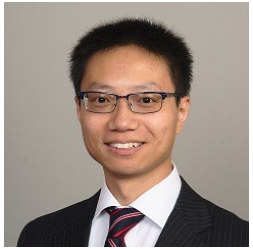
Takeaways

HRTF representation isn't just about spectral error —
perceptual distance is equally important!

Our proposed recipe: MMDS supervision + perceptual loss

Thank you! Questions?

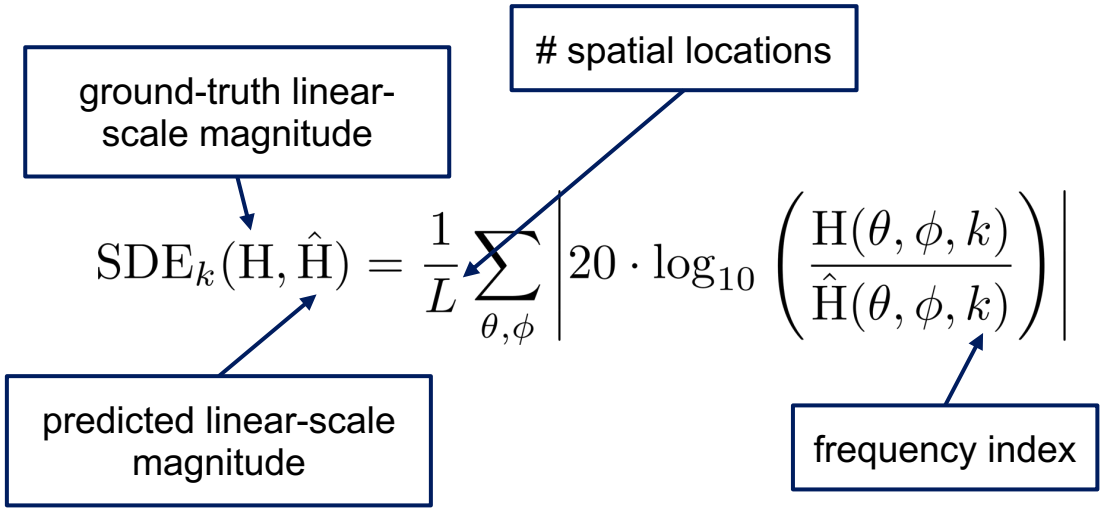
Check out our paper and poster for more, and let's discuss more!


<Backup Slides>

Spectral distance

- Spectral Difference Error (SDE)

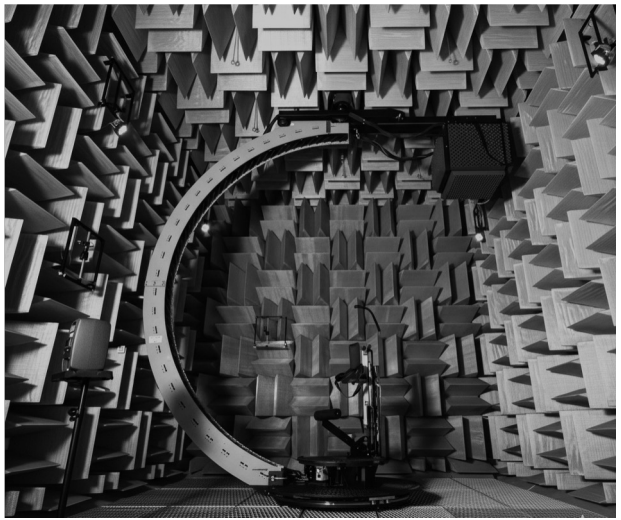


The median SDE across all frequency bins was computed to obtain a single SDE value.

Warnecke, Michaela, et al. "HRTF personalization based on ear morphology." *AES AVAR* 2022.

SS2 HRTF Dataset

- High-resolution HRTF database with 1625 measurement locations
- 48 kHz sampling rate
- 78 subjects



Warnecke, Michaela, et al. "Sound sphere 2: A high-resolution HRTF database." *AES AVAR* 2024.