From Neural Fields to Personalized Spatial Audio

A Hands-on Tutorial on HRTF Modeling

You (Neil) Zhang

Senior Researcher, Dolby Laboratories

Assistant Professor, Telecom Paris (Incoming)

Yoshiki Masuyama

Visiting Research Scientist,

Mitsubishi Electric Research Laboratories

https://bit.ly/4qFhGSy

DAFx Tutorial

September 1, 2026

Time Allocation

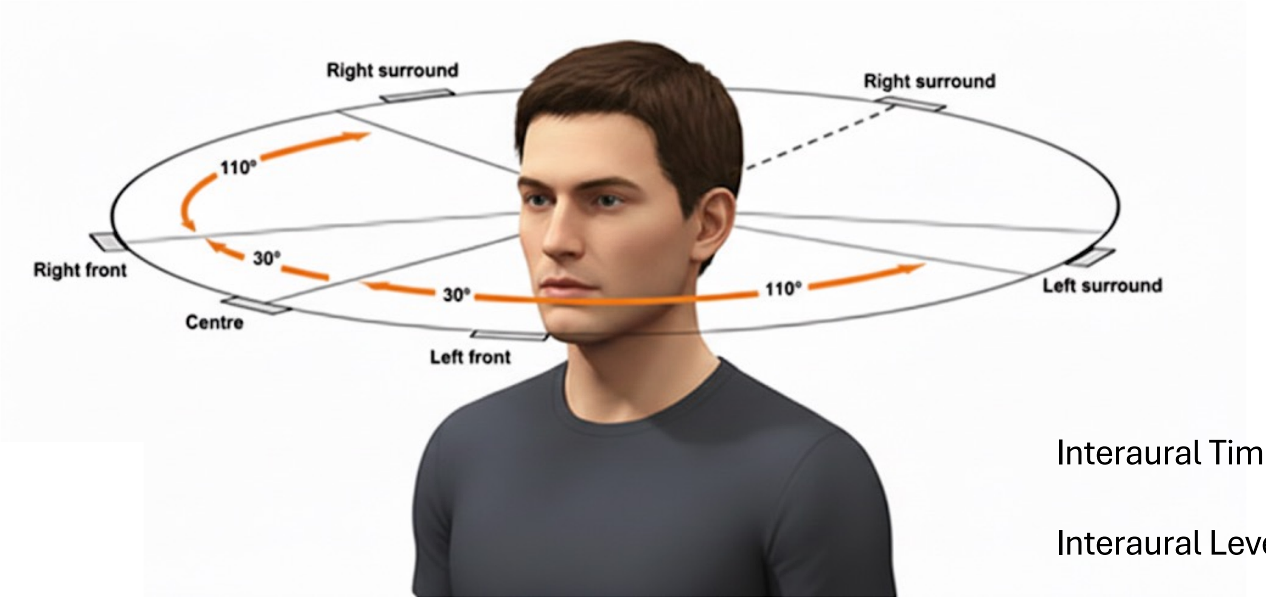
Section	Time (min)
Intro and background	15
Basics [ICASSP'23]	10
Hands-on 1	10
Intermediate [ICASSP'24]	10
Hands-on 2	10
Advanced	30
[ICASSP'25]	10
[OJSP'25]	10
[WASPAA'25]	10
Q&A	5
Total	**90**

https://bit.ly/4qFhGSy

https://github.com/yzyouzhang/dafx2026-hrtf-tutorial

Spatial Effects and Sound Localization

Humans localize sound sources by processing the differences between sounds received by their two ears.



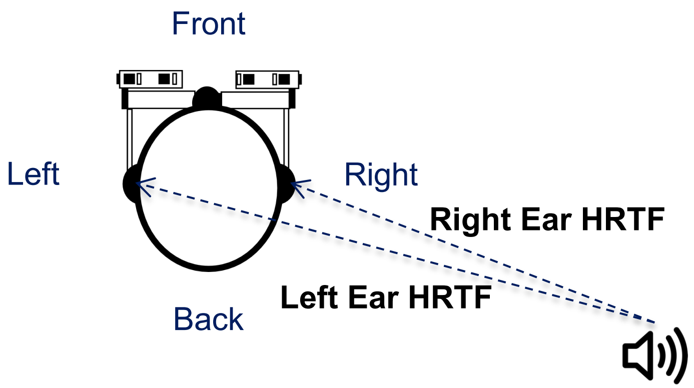
Interaural Time Difference (ITD)

Interaural Level Difference (ILD)

Figure adapted from https://www.soundonsound.com/reviews/mp3-surround

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

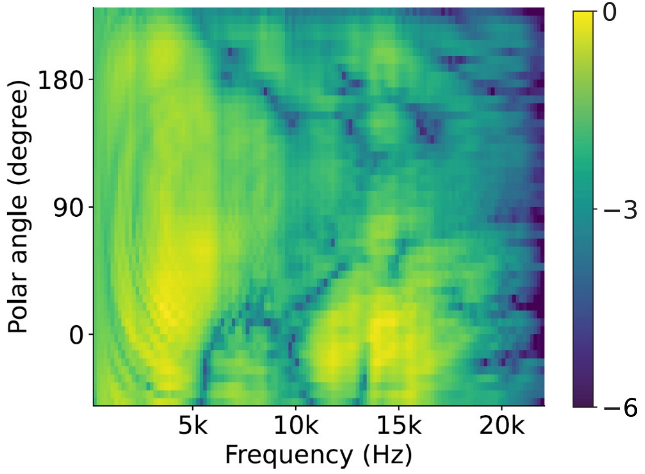


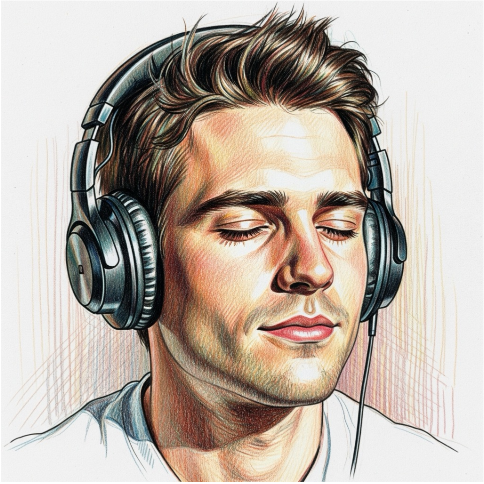
Figure from [Zhang+2023]

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

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

Application: Virtual Spatial Audio Generation

Personalized HRTFs have wide applications to deliver immersive 3D sound.



Headphones

AR smart glasses

VR headsets

Measure HRTFs

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

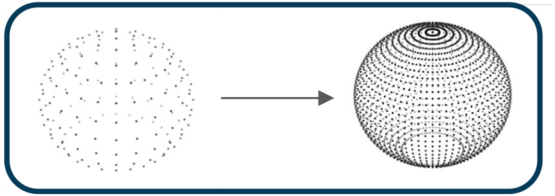
Time-consuming & Resource-intensive!

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

Research Tasks

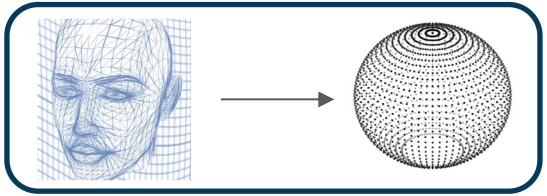
- ➢ HRTF Upsampling / Interpolation

  (use sparse locations to estimate HRTF at unseen directions)



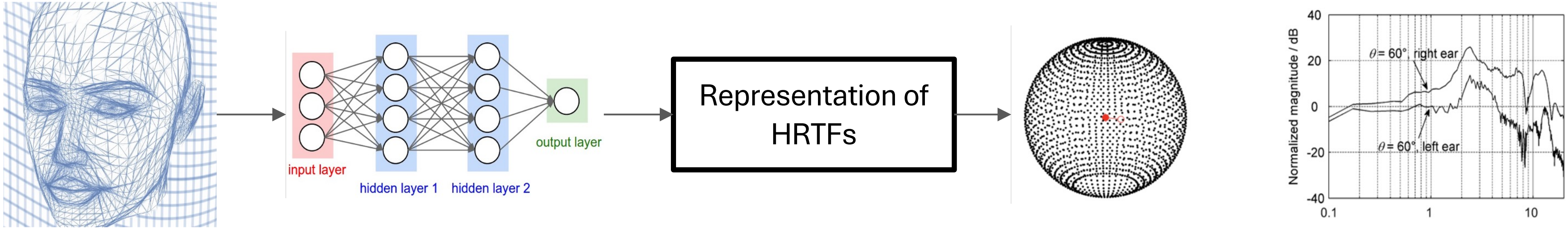
- ➢ HRTF Personalization from Human Input

  (anthropometry, ear shape, head mesh)



Personalizing HRTF with Machine Learning

Leverage measured data for personalized HRTF prediction



Human physical geometry

HRTF prediction network

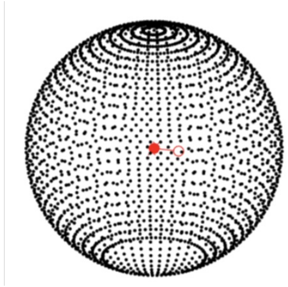
HRTFs at various spatial locations (of arbitrary spatial sampling schemes)

HRTFs at a particular position

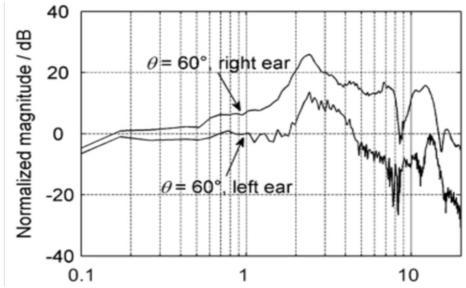
Assumption: Many characteristics are shared across individuals and captured by the **model**, while personalized effects are addressed by adapting the **input**.

Challenge 1: HRTF is High-Dimensional

For each subject, ear, and spatial location, HRTF is a function of frequency.



HRTFs at various spatial locations (of arbitrary spatial sampling schemes)



HRTFs at a particular position

$$\boldsymbol{x} \in \mathbb{R}^{L \times F \times 2}$$

L: number of locations (~1000)

F: number of frequency bins (~128)

2: left and right ear

1000 x 128 x 2 = 256,000. *A huge number!*

Challenge 2: Small Datasets and Different Setups

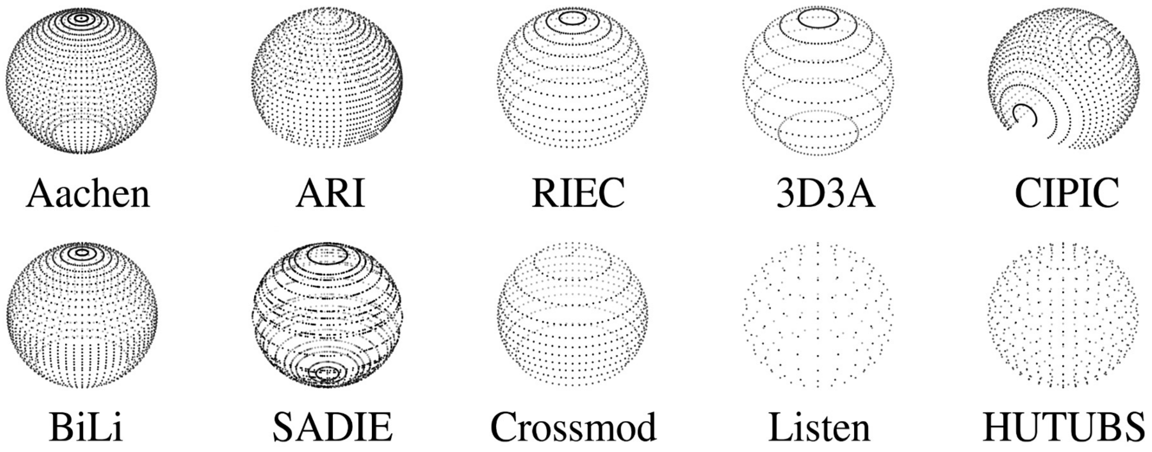
Existing measured HRTF databases each contain only dozens of subjects.

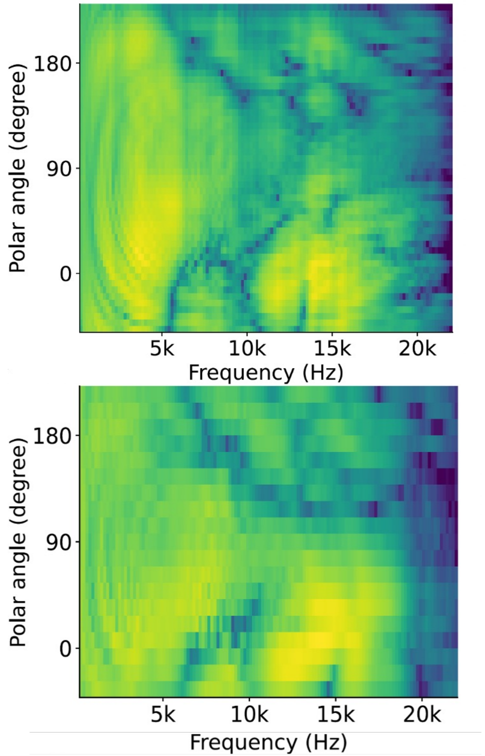
Statistics of public measured far-field HRTF databases.

Name	# Subjects	# Locations
3D3A [Sridhar+2017]	38	648
Aachen [Bomhardt+2016]	48	2304
ARI	97	1550
BiLi [Carpentier+2014]	52	1680
CIPIC [Algazi+2001]	45	1250
Crossmod	24	651
HUTUBS [Fabian+2019]	96	440
Listen	50	187
RIEC [Watanabe+2014]	105	865
SADIE II [Armstrong+2018]	18	2818
SONICOM [Engel+2023]	120	793
SS2 [Warnecke+2024]	78	1625
SONICOM v2 [Poole+2025]	**300**	793

Challenge 3: Different Spatial Sampling Schemes

The source location grids used in HRTF databases differ from one to another, making cross-dataset learning difficult.





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

From Neural Fields to Personalized Spatial Audio

Basics

HRTF Modeling with Neural Fields

[HRTF Field, ICASSP'23]

Intermediate

Neural IIR Filter Field

[NIIRF, ICASSP'24]

Advanced

Various Approaches for Improved Modeling

[ICASSP'25] [OJSP'25] [WASPAA'25]

From Neural Fields to Personalized Spatial Audio

Basics

HRTF Modeling with Neural Fields

[HRTF Field, ICASSP’23]

Intermediate

Neural IIR Filter Field

[NIIRF, ICASSP’24]

Advanced

Various Approaches for Improved Modeling

[ICASSP’25] [OJSP’25] [WASPAA’25]

Preliminary: Neural Fields

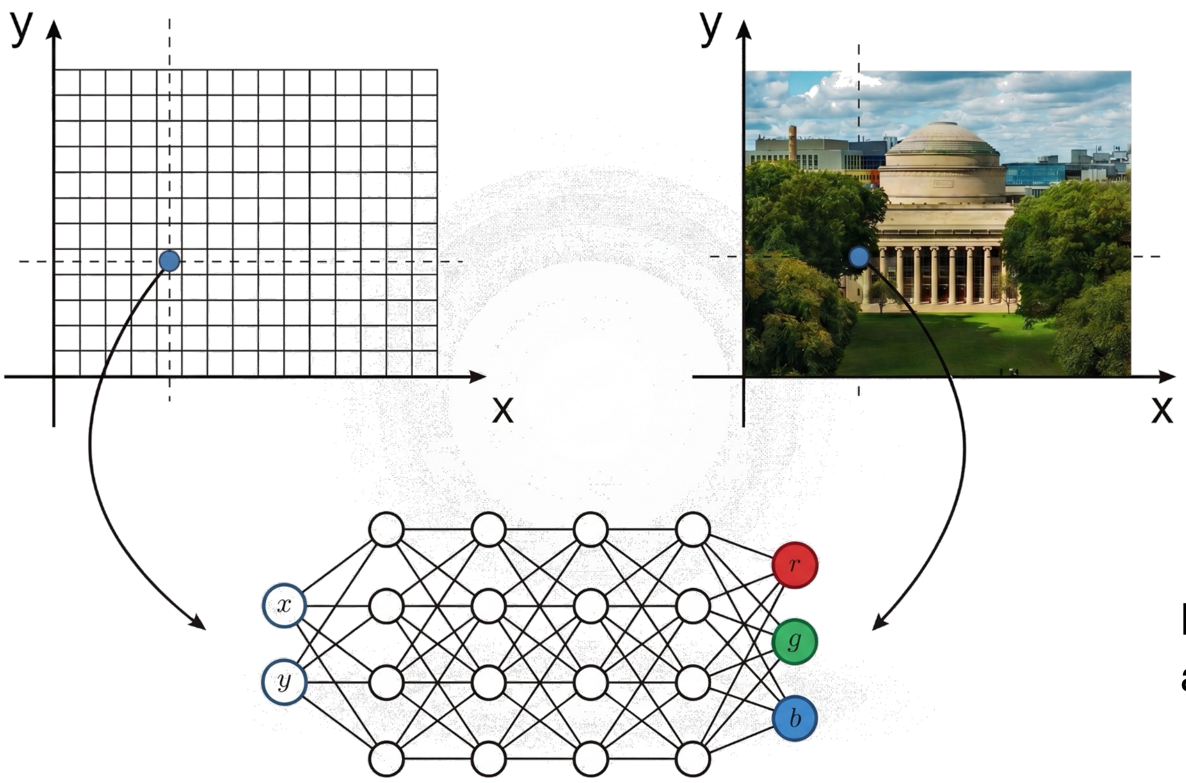


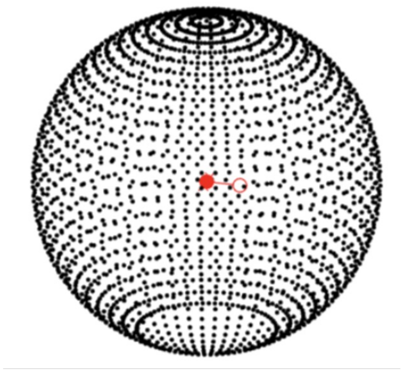
Figure adapted from [Skorokhodov+2021]

Representing discrete data as a continuous function

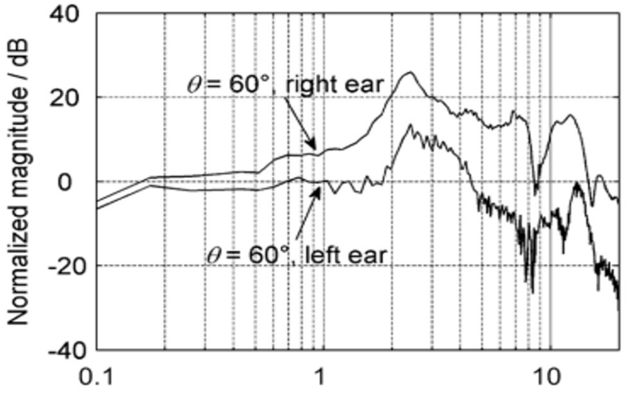
Ivan Skorokhodov, Savva Ignatyev, and Mohamed Elhoseiny. “Adversarial generation of continuous images.” *CVPR* 2021.

Preliminary: Revisit HRTF Data

Can we represent HRTFs the same way, as a **continuous function** of spatial location, rather than a fixed discrete grid?



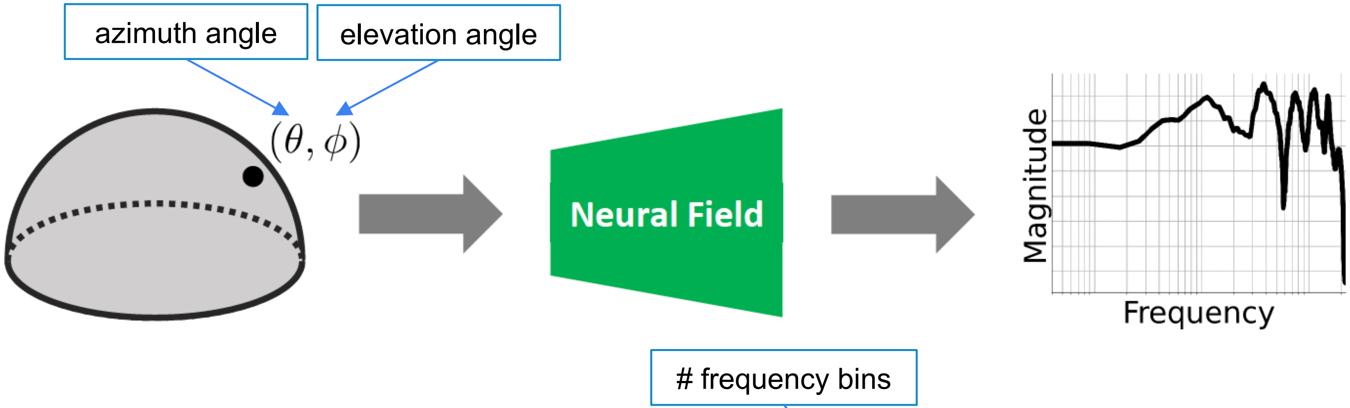
HRTFs at various spatial locations (of arbitrary spatial sampling schemes)



HRTFs at a particular position

Method: HRTF Field

Represent a single subject's HRTFs with a neural field

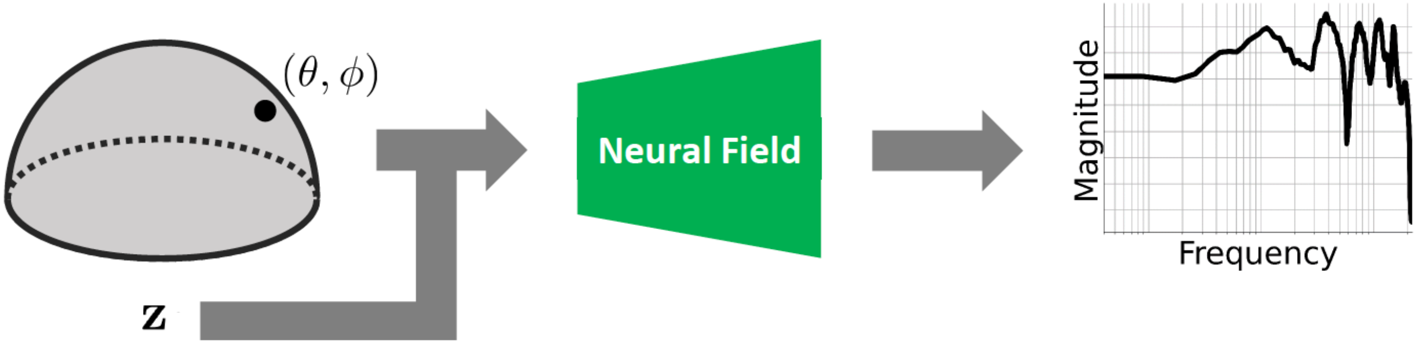


$$f : \mathbb{R}^2 \rightarrow \mathbb{R}^K$$

HRTF Field decouples representation from sampling schemes.

Method: HRTF Field (Cont'd)

Learning HRTF representations across subjects



latent code for a subject

$$\mathbf{z} \in \mathbb{R}^D$$

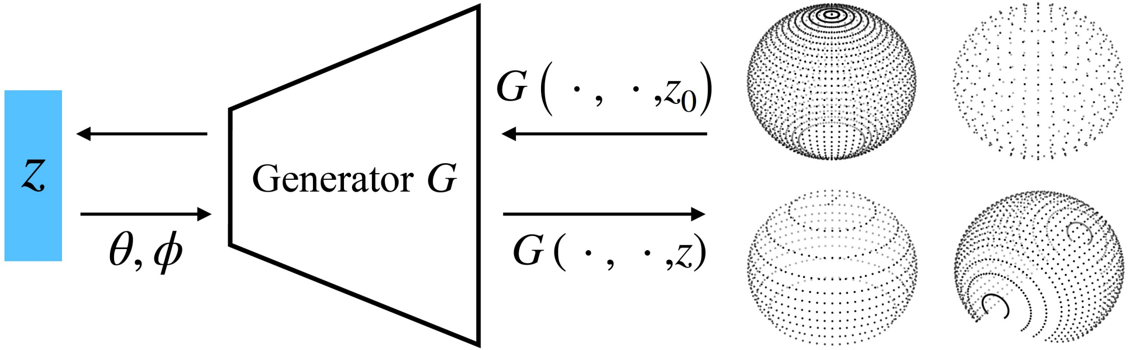
$$G(\theta, \phi, \mathbf{z})$$

frequency bins

$$G : \mathbb{R}^{2+D} \mapsto \mathbb{R}^K$$

HRTF Field Training

$$\mathbf{z} = \mathbf{z}_0 - \nabla_{\mathbf{z}_0} \mathcal{L}_{\mathrm{MSE}} \left(\mathbf{x}, G\left(\cdot, \cdot, \mathbf{z}_0 \right) \right)$$



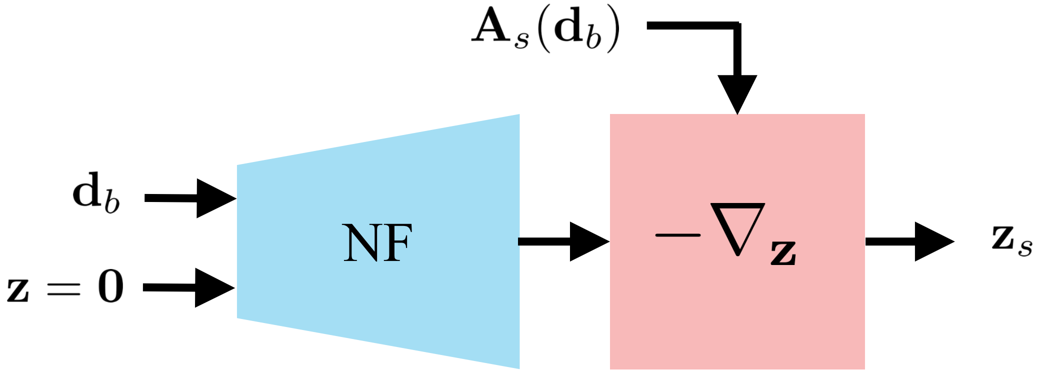
$$\mathcal{L} = \mathcal{L}_{\mathrm{MSE}} \left(\mathbf{x}, G\left(\cdot, \cdot, \mathbf{z} \right) \right)$$

Enables grid-agnostic mix-database training and cross-database evaluation

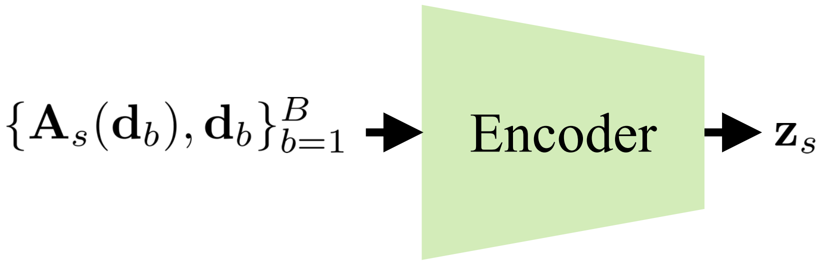
Sam Bond-Taylor, and Chris G. Willcocks. "Gradient origin networks." *ICLR* 2021.

Other Schemes to Train HRTF Field

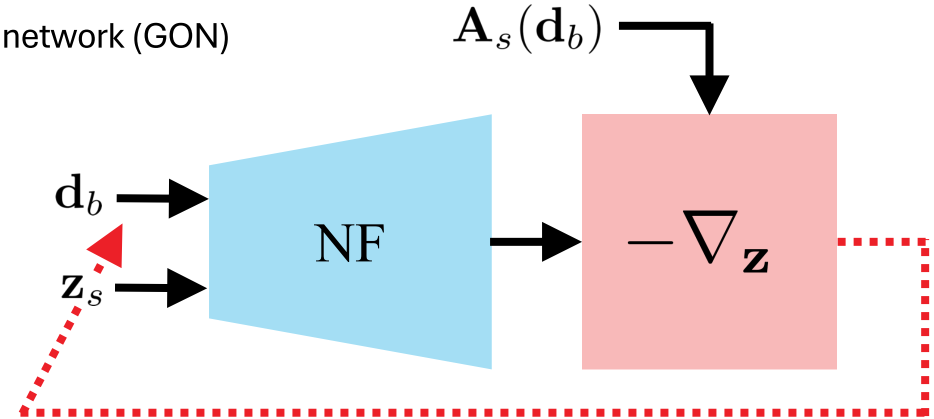
$\mathbf{A}_s(\mathbf{d}_b)$ the **HRTF magnitude (A)** for subject *s* at *b*-th **direction (d)**



Gradient origin network (GON)



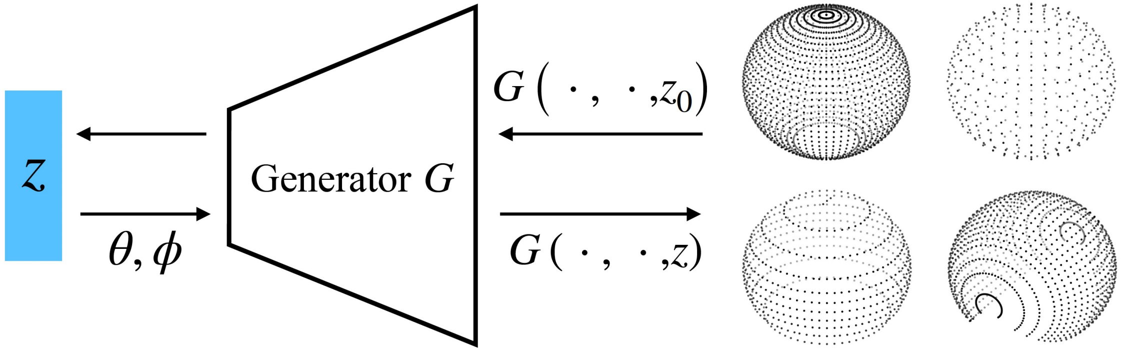
Auto-encoding



Auto-decoding

Summary: Generative Framework for HRTF Representation

- **Goal**: Merge databases with different spatial sampling schemes
- **Approach**: Introduce continuous generative neural fields to model HRTF
- **Contributions**: For the **first time**, it enables mix-database training and cross-database evaluation



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

Hands-on with HRTF Field

https://bit.ly/4qFhGSy

From Neural Fields to Personalized Spatial Audio

Basics

HRTF Modeling with Neural Fields

[HRTF Field, ICASSP’23]

Intermediate

Neural IIR Filter Field

[NIIRF, ICASSP’24]

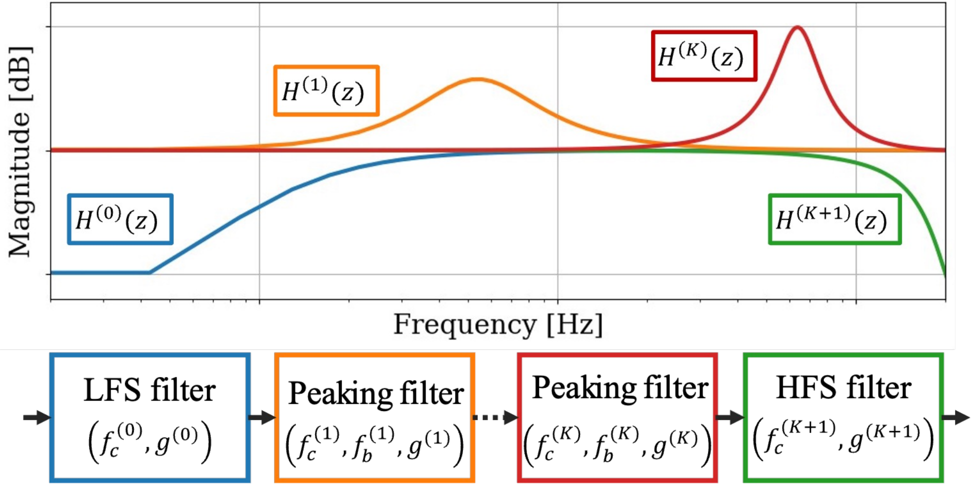
Advanced

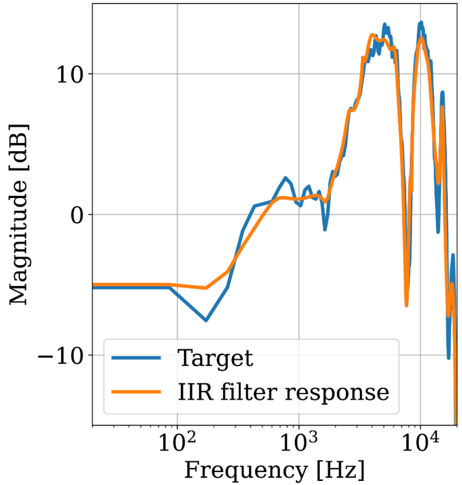
Various Approaches for Improved Modeling

[ICASSP’25] [OJSP’25] [WASPAA’25]

Approximating HRTF through Differentiable DSP (DDSP)

- HRTFs can be well approximated using a cascade of biquad filters [Ramos+2013]
  - Low/High-frequency shelf filters (LFS/HFS) defined by **the cut frequency and gain**
  - Peaking filters defined by **the center frequency, bandwidth, and gain**

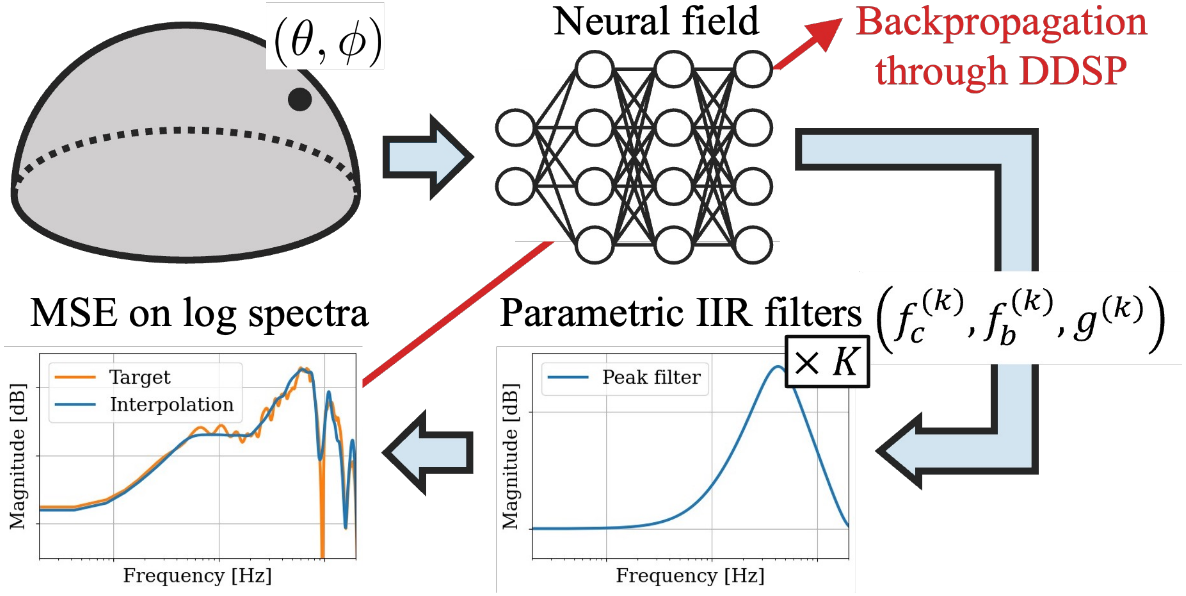




- The filter parameters are interpretable and suitable for interpolation
  - Applying a simple interpolation to the parameters [Ramos+2013]

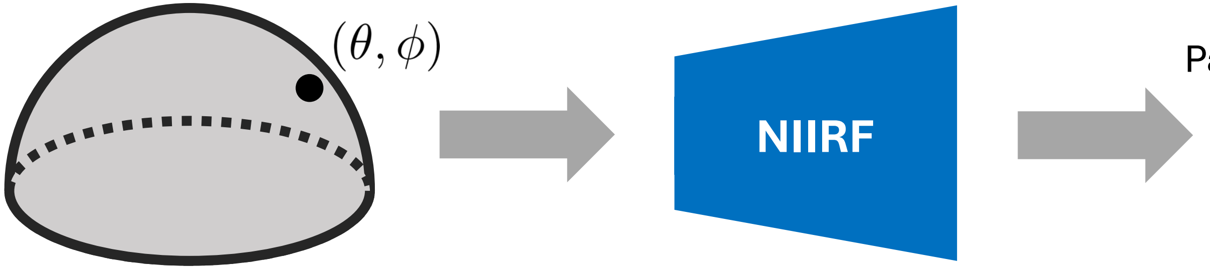
Neural IIR Filter Field (NIIRF) [Masuyama+2024]

- NIIRF incorporates the IIR filter representation into HRTF field
  - Modeling the filter parameters as a continuous function of direction
  - **Training the neural field to match the magnitude response to the true HRTF through DDSP**



Training of NIIRF with DDSP

- NIIRF estimates the parameters of cascaded IIR filters.



Parameters for all IIR filters

$$\left(f_c^{(k)}, f_b^{(k)}, g^{(k)}\right)$$

Filter index

- DDSP allows to train the neural field with a loss function on the filter response
  - Computing the filter coefficients from the parameters and applying frequency sampling

$$H^{(k)}(z) = \frac{b_0^{(k)} z^0 + \cdots + b_N^{(k)} z^{-N}}{1 + \cdots + a_N^{(k)} z^{-N}} \quad \Rightarrow \quad H_{\theta,\phi}[m] = \prod_{k=0}^{K+1} H_{\theta,\phi}^{(k)}(\mathrm{e}^{2\pi \mathrm{j} m/\mathrm{M}})$$

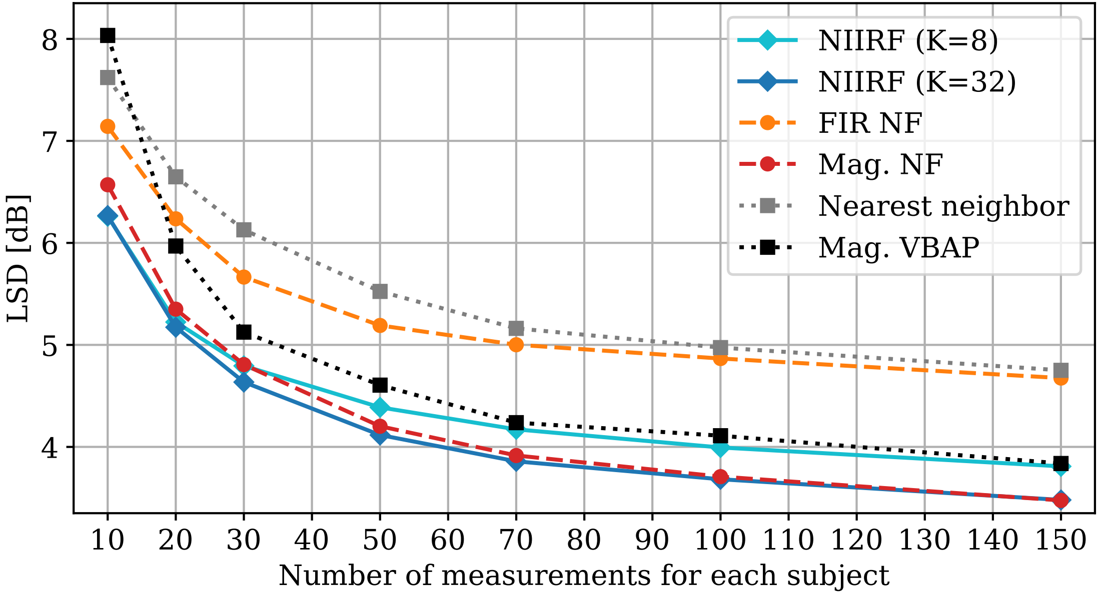
  - Minimizing the mean squared error on the log magnitude response

$$\mathcal{L} = \frac{1}{|\mathcal{D}|M} \sum_{(\theta,\phi)\in\mathcal{D}} \sum_{m=0}^{M-1} \left(20 \log_{10} \left| \frac{H_{\theta,\phi}[m]}{\widetilde{H}_{\theta,\phi}[m]} \right| \right)^2$$

True HRTF

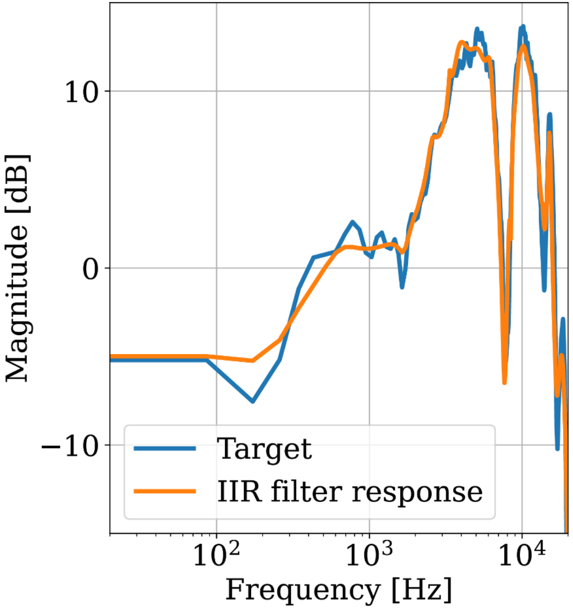
Subject-Wise Training and Evaluation on CIPIC Dataset

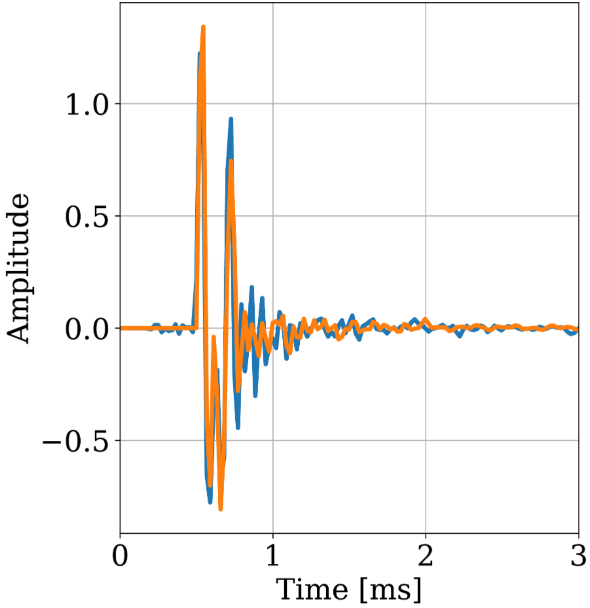
- Neural fields are trained for each of 45 subjects in the CIPIC dataset [Algazi+2001]
  - Training on up to 150 directions, and evaluation on 1000 unseen directions
- NIIRF performs comparably to the vanilla HRTF Field (Mag. NF)



Time-Domain Response of NIIRF Outputs

- NIIRF output matches the time-domain response while using phase-insensitive loss
  - The parametric-IIR-filter representation introduces a good inductive bias





†We manually shift the IIR filter response.

Hands-on with NIIRF

https://bit.ly/4qFhGSy

From Neural Fields to Personalized Spatial Audio

Basics

HRTF Modeling with Neural Fields

[HRTF Field, ICASSP’23]

Intermediate

Neural IIR Filter Field

[NIIRF, ICASSP’24]

Advanced

Various Approaches for Improved Modeling

[ICASSP’25] [OJSP’25] [WASPAA’25]

Various Approaches for Improved Modeling

Advanced

Retrieval-Augmented Neural Field (RANF)
Upsampling and Personalization [ICASSP’25]

Subject- and Dataset-Aware Neural Field (SuDaField)
Harmonization [OJSP’25]

Perception-Informed Neural Field
Personalization [WASPAA’25]

Various Approaches for Improved Modeling

Advanced

Retrieval-Augmented Neural Field (RANF)
Upsampling and Personalization [ICASSP’25]

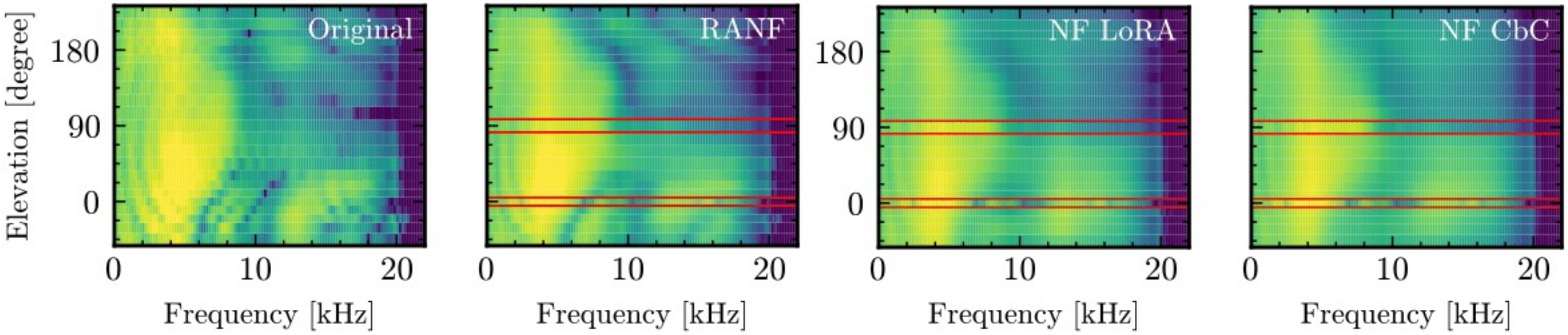
Subject- and Dataset-Aware Neural Field (SuDaField)
Harmonization [OJSP’25]

Perception-Informed Neural Field
Personalization [WASPAA’25]

Qualitative Analysis of Neural-Field Methods on SONICOM Dataset

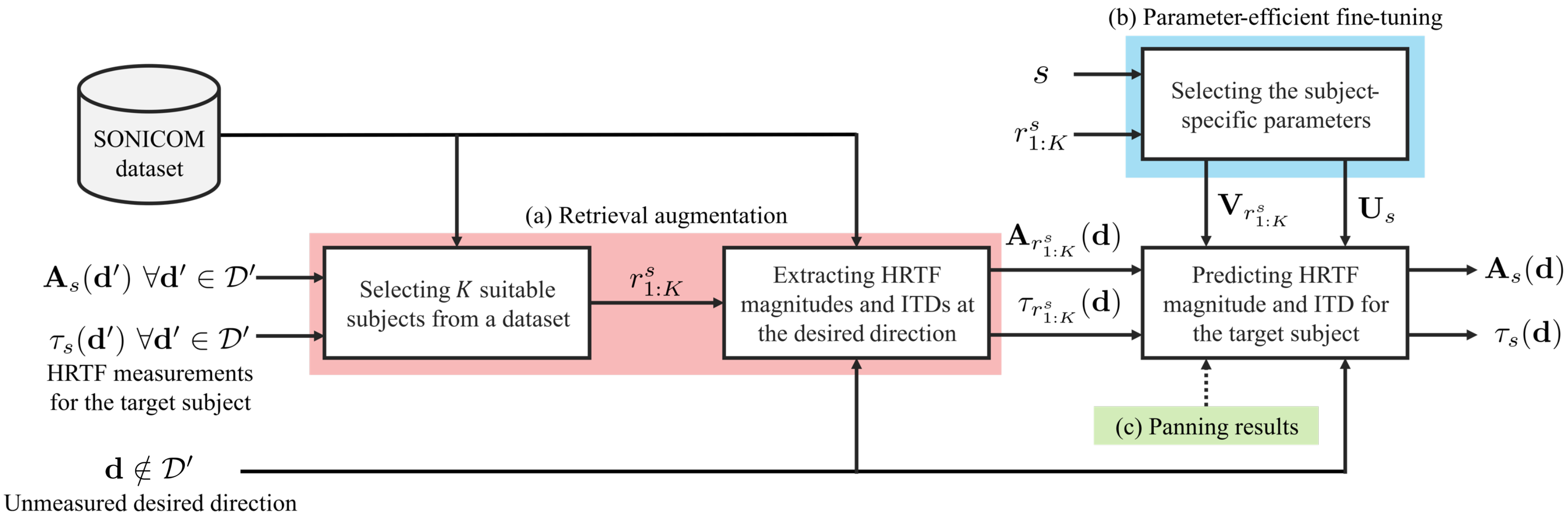
- **Neural-field methods smear the deep spectral peaks and notches**
  - Insufficient cues: Direction and low-dimensional subject-specific parameters

- We explore rich cues to preserve distinctive spectral structures
  - Finding subjects whose HRTFs are close to those of the target subject from a dataset
  - **Leveraging their HRTFs at the desired direction as a direction-dependent high-resolution cue**



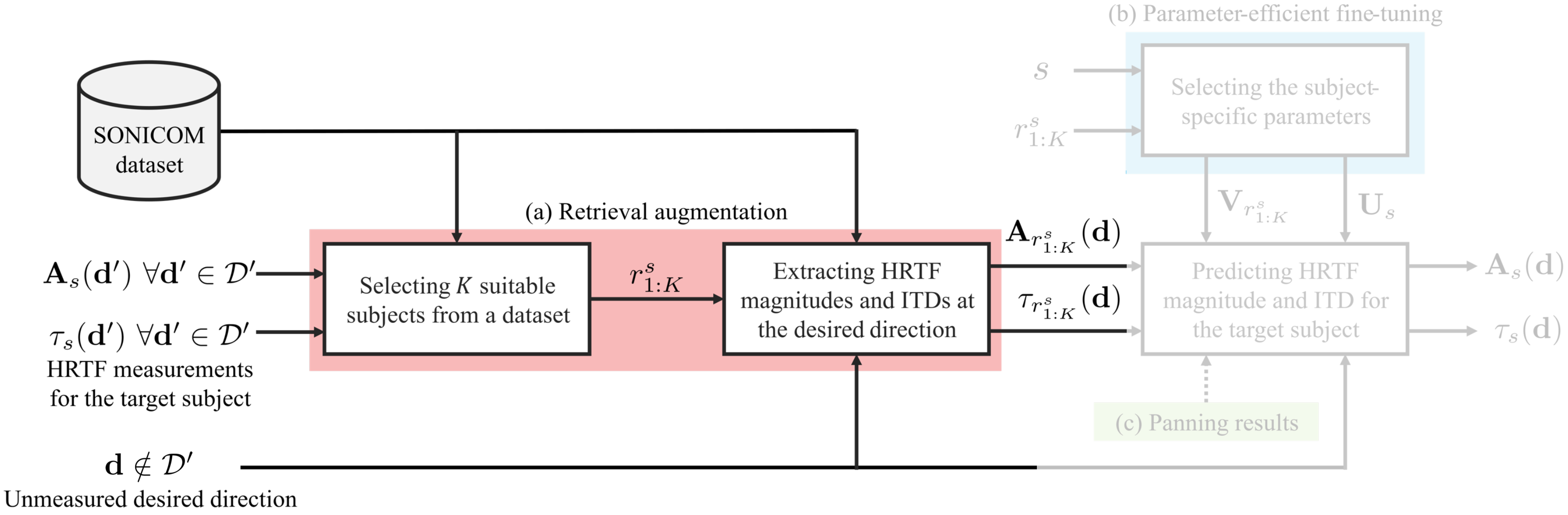
Retrieval Augmented Neural Field (RANF) [Masuyama+2025]

- We retrieve the best-fitting subjects and leverage their dense HRTFs for upsampling
  - Selecting the subjects whose HRTFs are similar to those of the target subject
  - Extracting their HRTF features (magnitude spectra and ITDs) at the desired direction
  - Exploiting the HRTF features to predict the target HRTF, together with subject-specific parameters



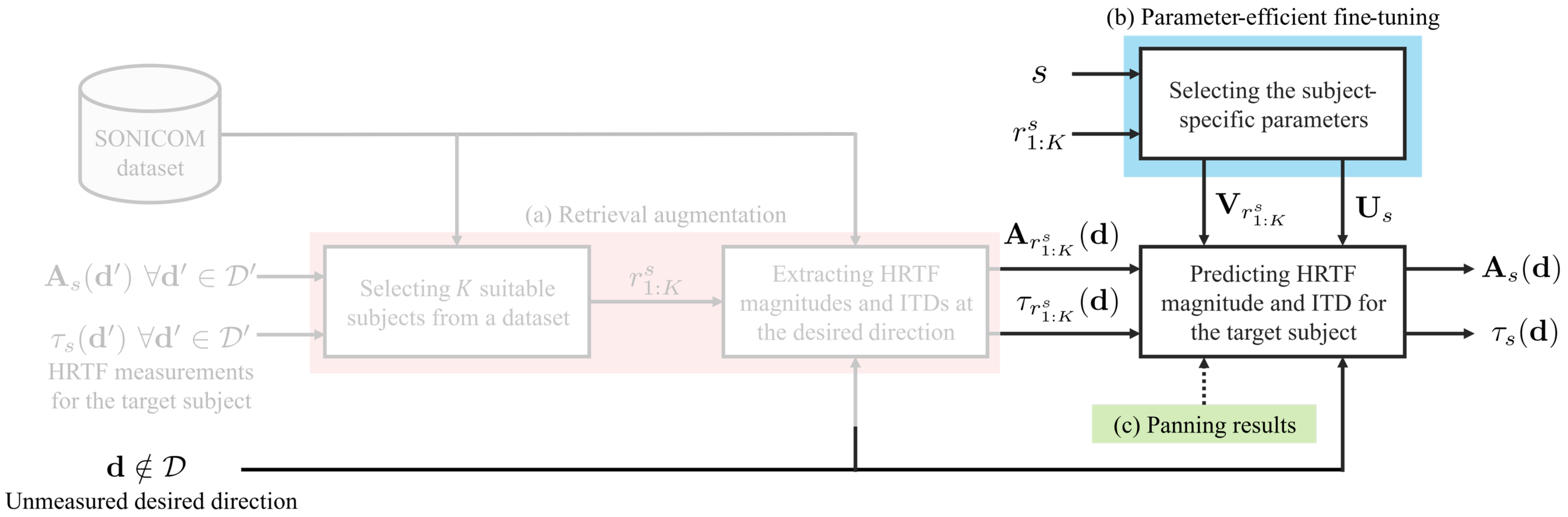
Retrieval Augmented Neural Field (RANF) [Masuyama+2025]

- We retrieve the best-fitting subjects and leverage their dense HRTFs for upsampling
  - **Selecting the subjects whose HRTFs are similar to those of the target subject**
  - Extracting their HRTF features (magnitude spectra and ITDs) at the desired direction
  - Exploiting the HRTF features to predict the target HRTF, together with subject-specific parameters



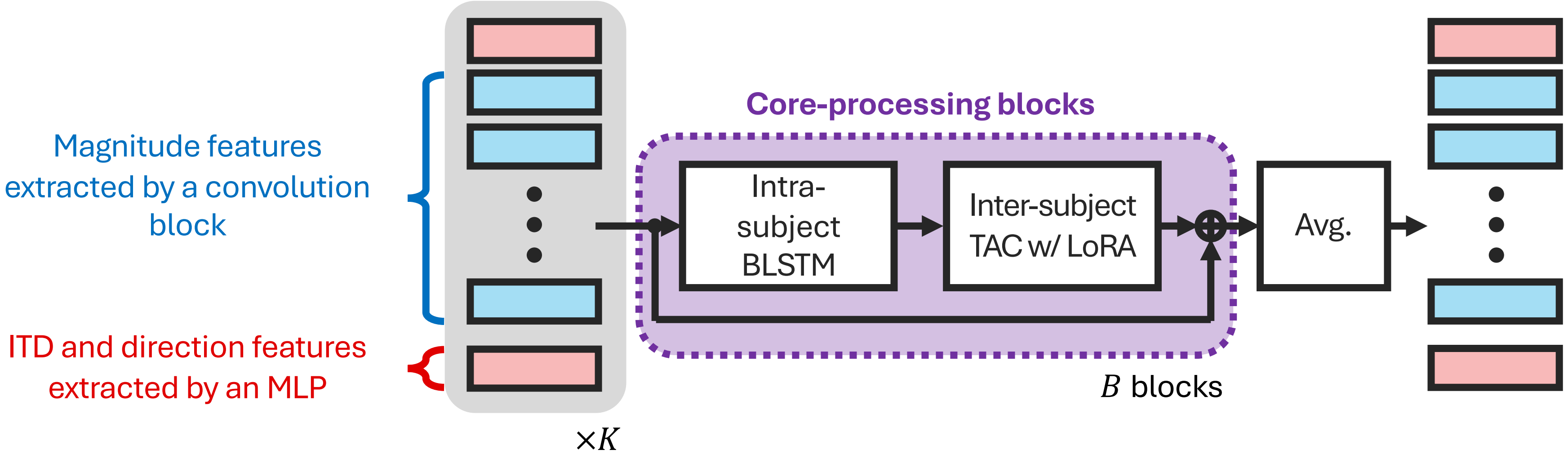
Retrieval Augmented Neural Field (RANF) [Masuyama+2025]

- We retrieve the best-fitting subjects and leverage their dense HRTFs for upsampling
  - Selecting the subjects whose HRTFs are similar to those of the target subject
  - Extracting their HRTF features (magnitude spectra and ITDs) at the desired direction
  - Exploiting the HRTF features to predict the target HRTF, **together with subject-specific parameters**



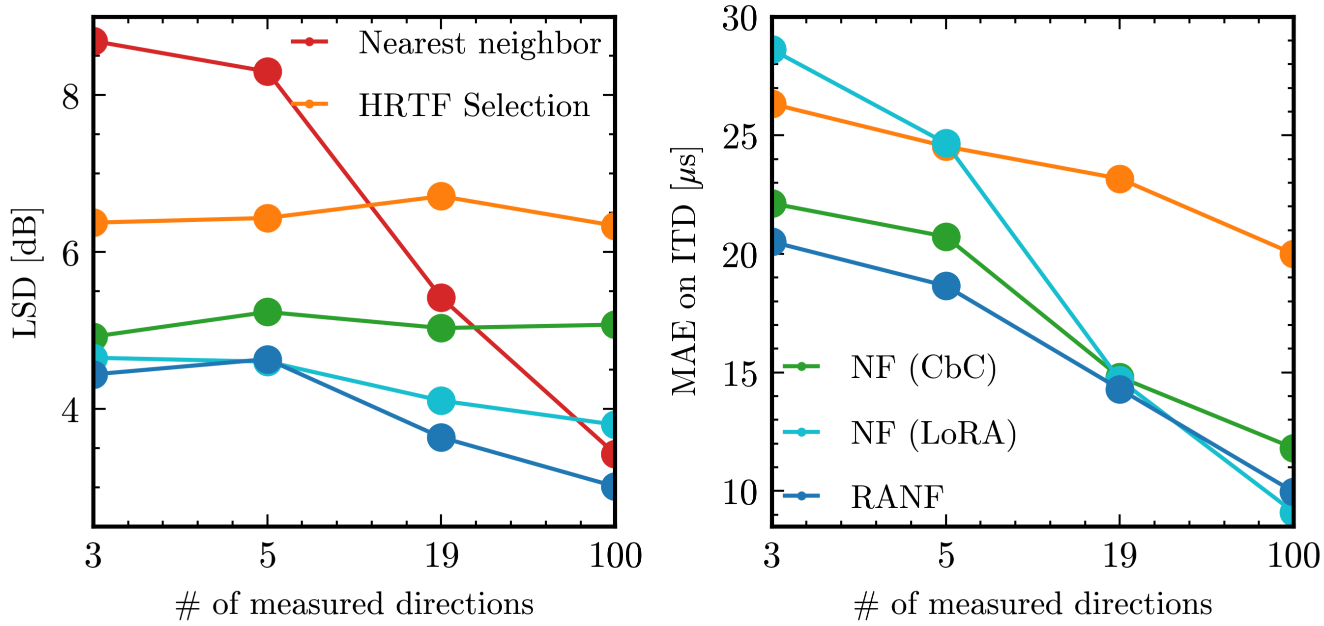
Network Architecture for RANF

- We alternately apply intra- and inter-subject processing
  - BLSTM over the frequency-like dimension
  - Mixing information across the K reference subjects via transform-average-concatenate (TAC)



Quantitative Results on SONICOM Dataset

- **RANF outperforms the selection (retrieving the top-1) and neural-field methods**
  - Learning how to aggregate and refine retrievals
  - Improving the ITD prediction performance under sparse conditions (3 or 5 measurements)



Various Approaches for Improved Modeling

Advanced

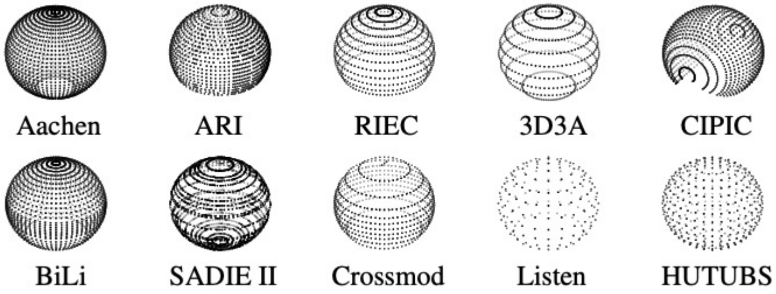
Retrieval-Augmented Neural Field (RANF)
Upsampling and Personalization [ICASSP'25]

Subject- and Dataset-Aware Neural Field (SuDaField)
Harmonization [OJSP'25]

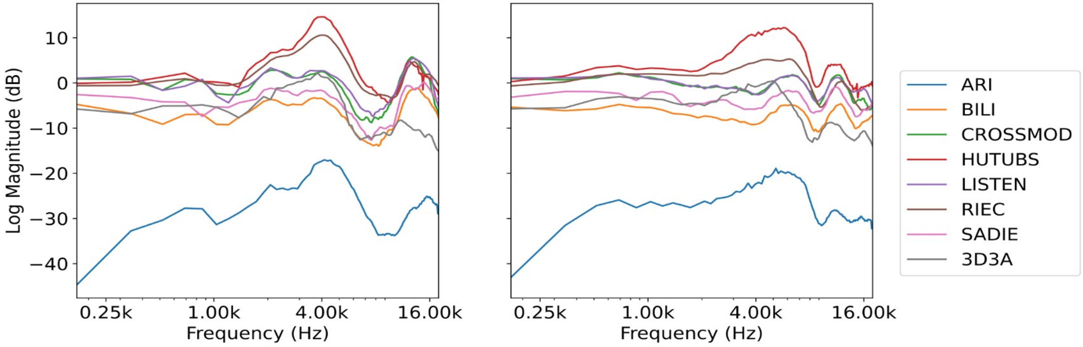
Perception-Informed Neural Field
Personalization [WASPAA'25]

Challenges in Cross-Dataset HRTF Modeling

- **Neural fields can handle HRTFs with different spatial grids** [Zhang+2023]



- Differences in measurement setups and post-processing significantly affect HRTFs



(a) at source position (0, 0)

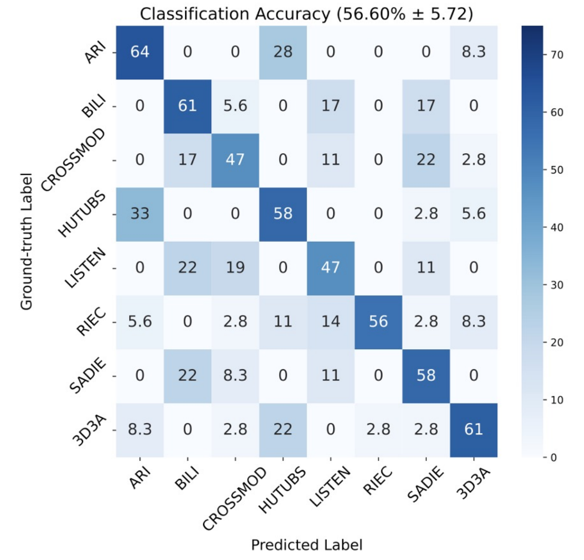
(b) at source position (90, 0)

The average HRTFs from each dataset show a large variation across datasets [Wen+2023]

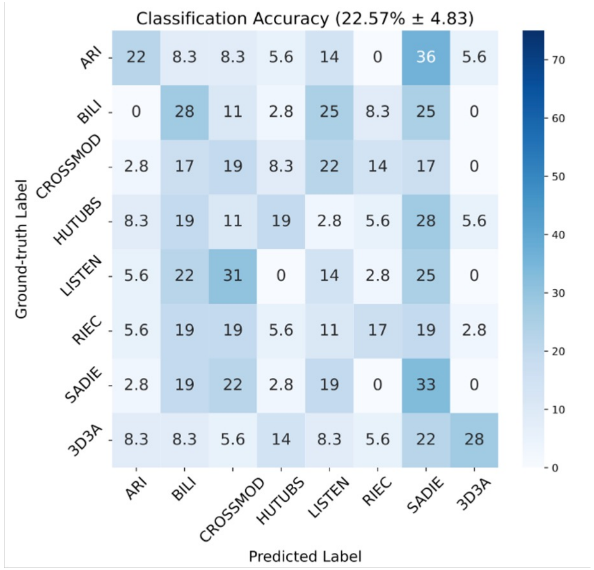
Influence of Measurement Setups

- The original dataset can be easily detected from HRTFs (left)
  - Classifying datasets using an SVM classifier: 18 subjects from each of 8 datasets

- The influence can be mitigated by normalization per dataset and direction [Wen+2023]

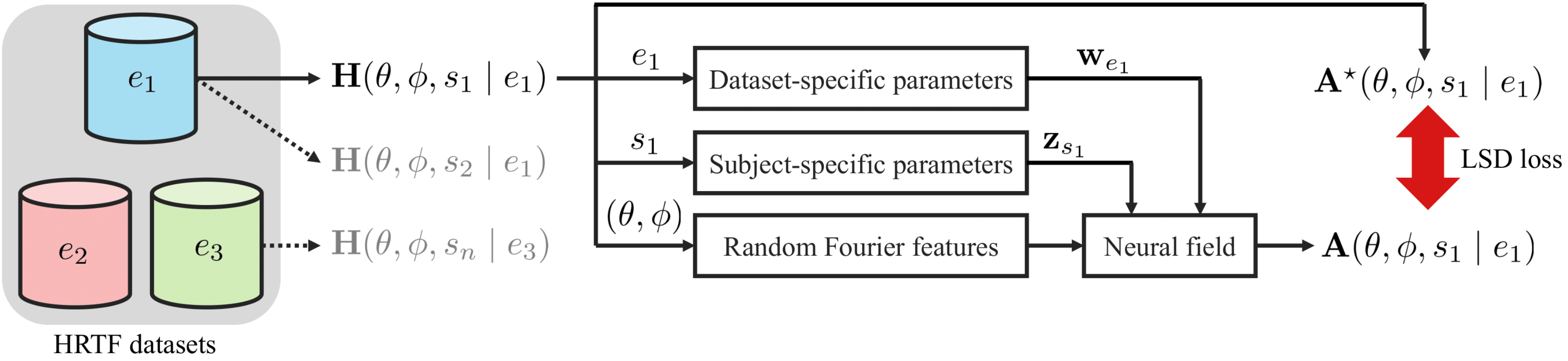


$$\boxed{HRTF_{\text{normalized}}(\theta, \phi) = \frac{Y(\theta, \phi)}{HRTF_{\text{avg}}(\theta, \phi)}}$$



Subject- and Dataset-Aware Neural Field (SuDaField)

- SuDaField adopts dataset-specific parameters instead of the pre-computed average
  - Sharing the same dataset-specific parameters across subjects within the same dataset
  - Capturing the subject-agnostic factor specific to each measurement setup



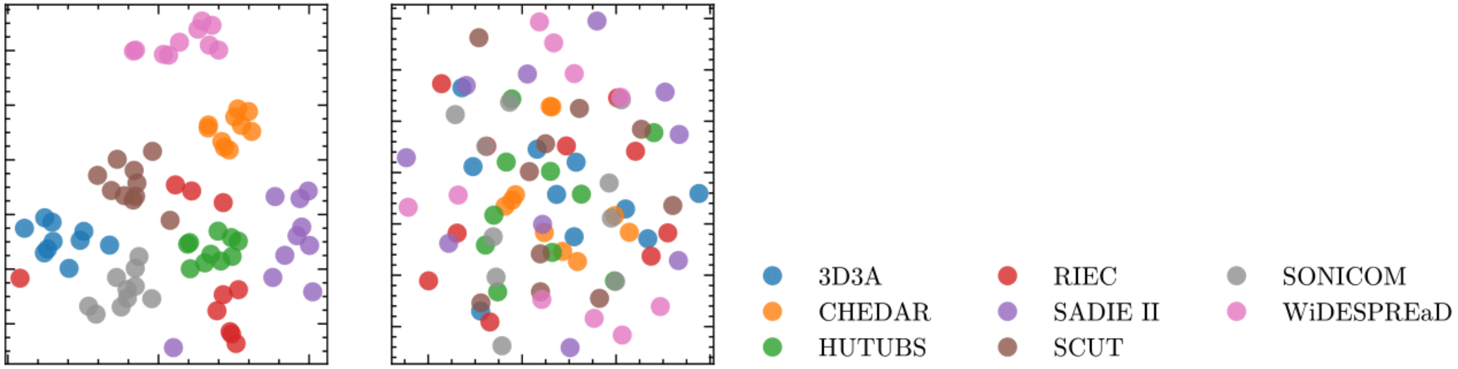
- We exclusively add a subject- or dataset-specific bias to each layer of SuDaField

Qualitative Analysis on LAP Challenge Task 1

- SuDaField disentangles the subject- and dataset-specific factors
  - Training neural fields in the LAP challenge Task 1 setup (10 subjects for each of 8 datasets)
  - Visualizing the subject-specific biases using dimension reduction (t-SNE)

Original
(w/o dataset-specific)

SuDaField
(w/ dataset-specific)



Quantitative Results through HRTF Harmonization on LAP Challenge Task 1

- HRTF harmonization aims to compensate for influence of measurement setups
  - Reducing dataset-specific artifacts while preserving perceptual localization cues

- SuDaField can convert the influence of measurement setups
  - Harmonizing HRTFs by converting them as if they had been measured with a reference setup
  - Replacing the dataset-specific biases with those for the reference dataset during inference
  - **Fooling the dataset classifier the best** while preserving the subject-specific spatial cues

	Localization-cue failure rate↓	Classification accuracy ↓
EqPCA-UoA†	13.8	95.0
CoWiDeq†	10.0	92.3
IOA3D†	5.0	26.9
Magnitude normalization	63.8	53.8
Vanilla neural field	10.0	100.0
SuDaField	12.5	18.1

†Challenge submissions

Various Approaches for Improved Modeling

Advanced

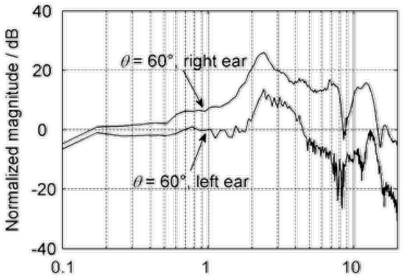
Retrieval-Augmented Neural Field (RANF)
Upsampling and Personalization [ICASSP'25]

Subject- and Dataset-Aware Neural Field (SuDaField)
Harmonization [OJSP'25]

Perception-Informed Neural Field
Personalization [WASPAA'25]

Motivation: Representation not Perceptually Aligned

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



Spectral reconstruction

Perceptually plausible HRTF

If two HRTFs are close in spectral distance, are they close perceptually?

Preliminary: HRTF Perception

Perceptual benefits of individual 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.

Preliminary: Computational Auditory Modeling

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

Externalization: **AEP** [Baumgartner&Majdak2021] — Not differentiable

Localization: **DRMSP** [Barumerli+2023] — Not differentiable

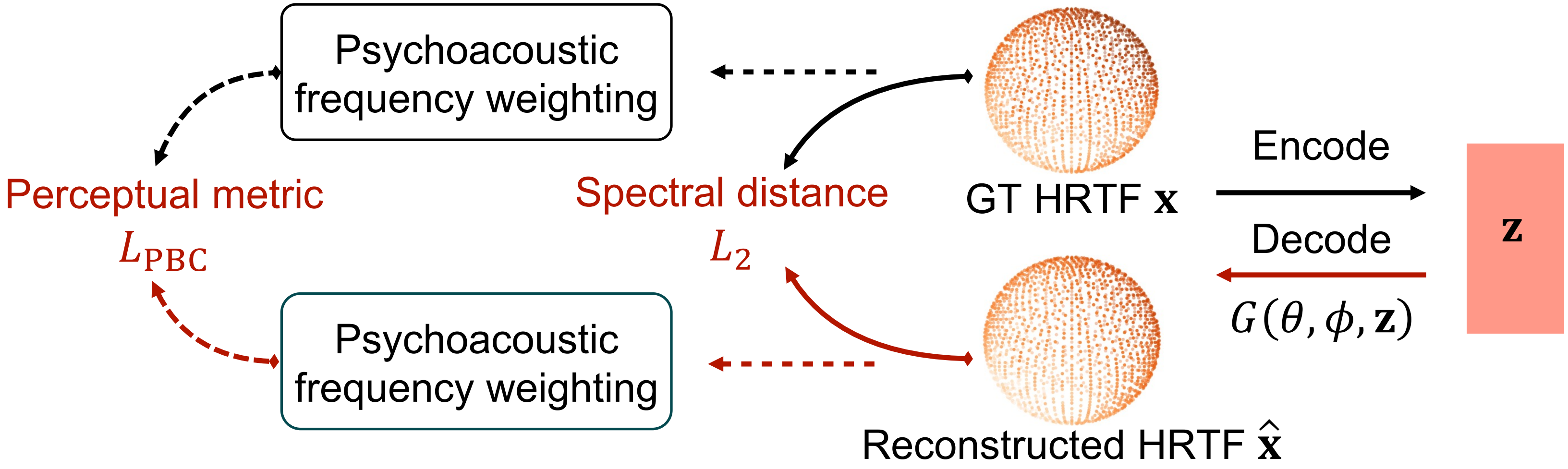
Figure from amtoolbox.org

Thomas McKenzie, Cal Armstrong, Lauren Ward, Damian T. Murphy, and Gavin Kearney. "Predicting the colouration between binaural signals." *Applied Sciences* 2022.
Robert Baumgartner, and Piotr Majdak. "Decision making in auditory externalization perception:model predictions for static conditions." *Acta Acustica* 2021.
Roberto Barumerli, Piotr Majdak, Michele Geronazzo, David Meijer, Federico Avanzini, and Robert Baumgartner. "A Bayesian model for human directional localization of broadband static sound sources." *Acta Acustica* 2023.

Method: Aligning with Perception-Informed Space

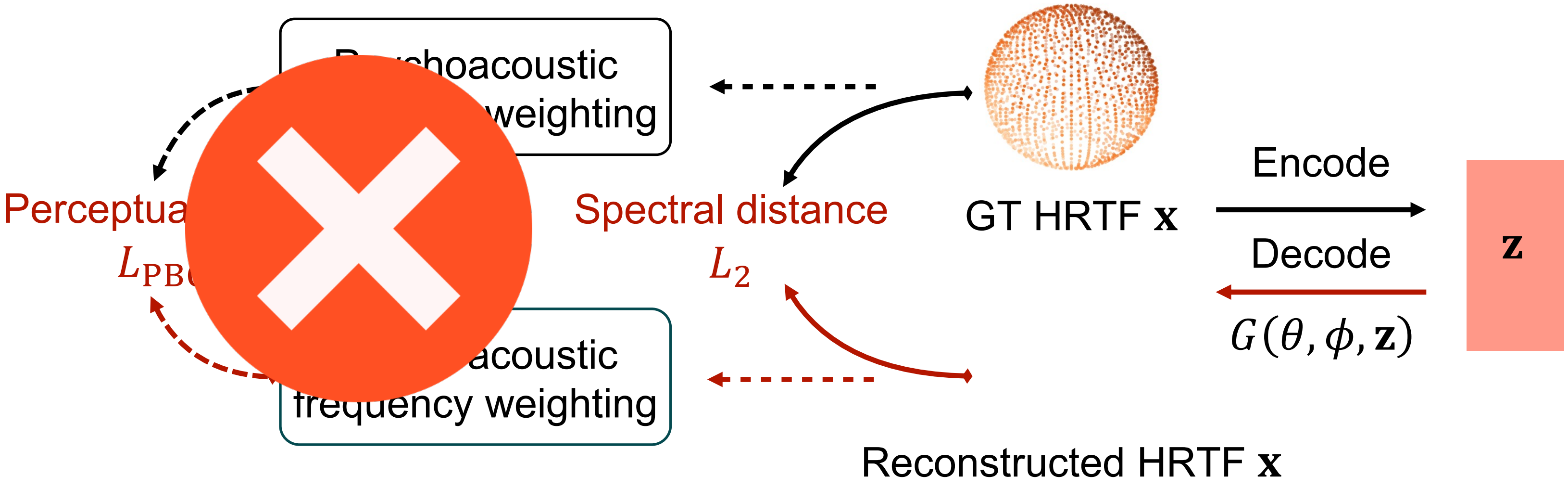
If the perceptual metric is differentiable, a direct perceptual loss can be incorporated.

- This approach applies only to PBC, not to AEP or DRMSP.



Method: Aligning with Perception-Informed Space (Cont'd)

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



Preliminary: Metric multi-dimensional scaling (MMDS)

Optimize a new space that preserves the pairwise distances.

Pairwise distance matrix

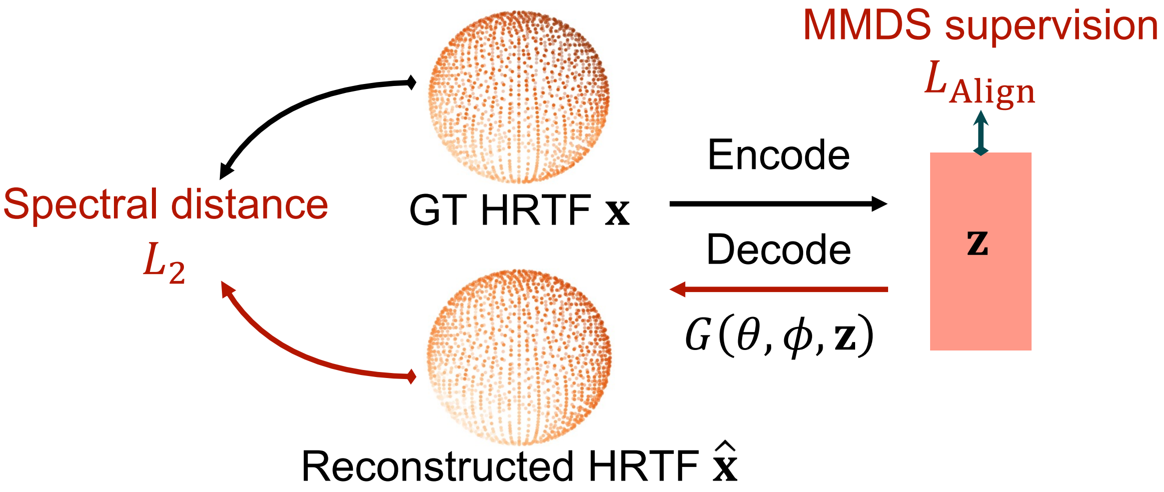
	0	0.4	0.5	0.6	0.9
	0.4	0	0.1	0.1	0.4
	0.5	0.1	0	0.1	0.5
	0.6	0.1	0.1	0	0.6
	0.9	0.4	0.5	0.6	0

New space

Figure adapted from https://youtu.be/VKSJayDi_IQ

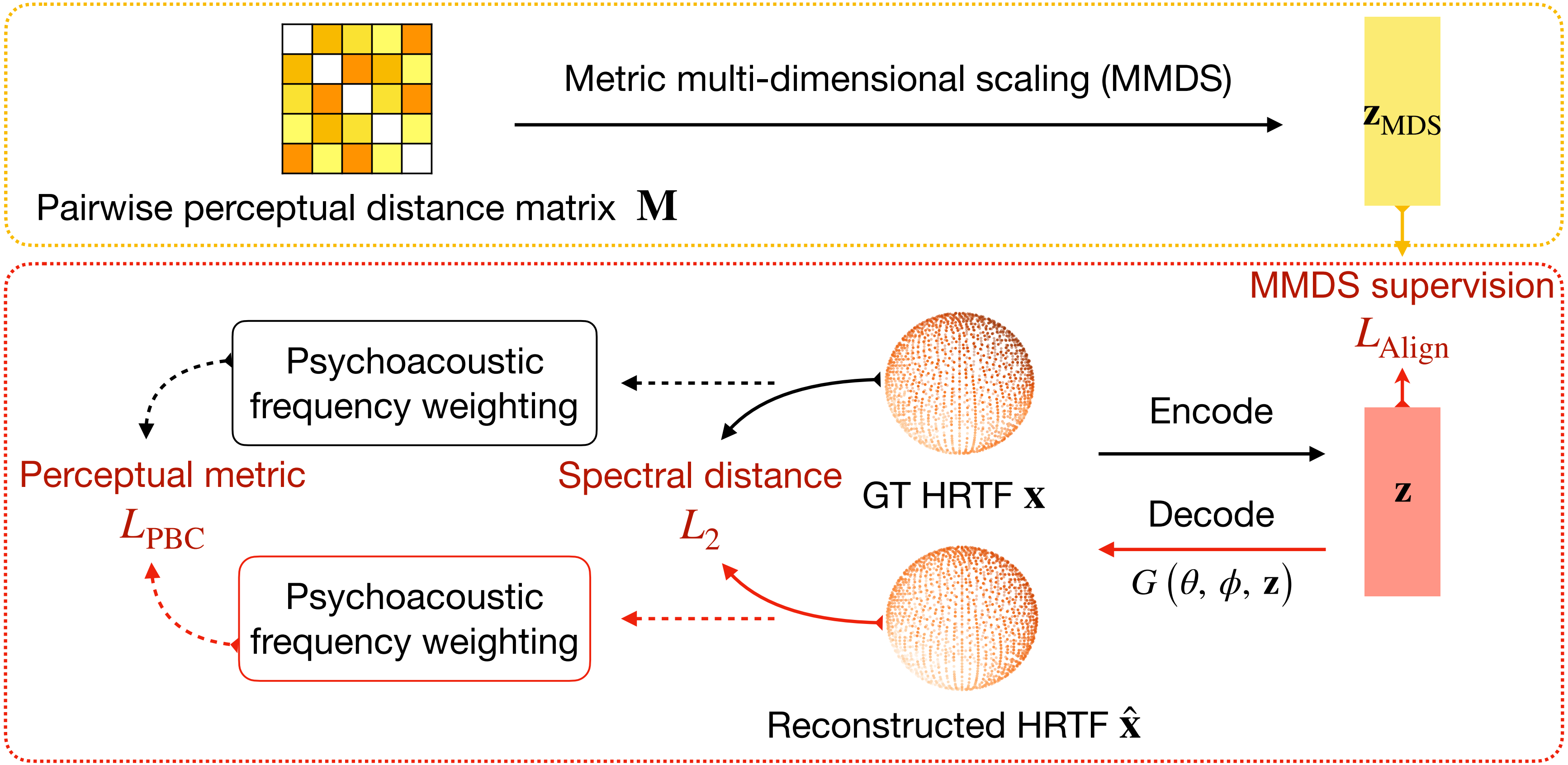
Method: Aligning with Perception-Informed Space (Cont'd)

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



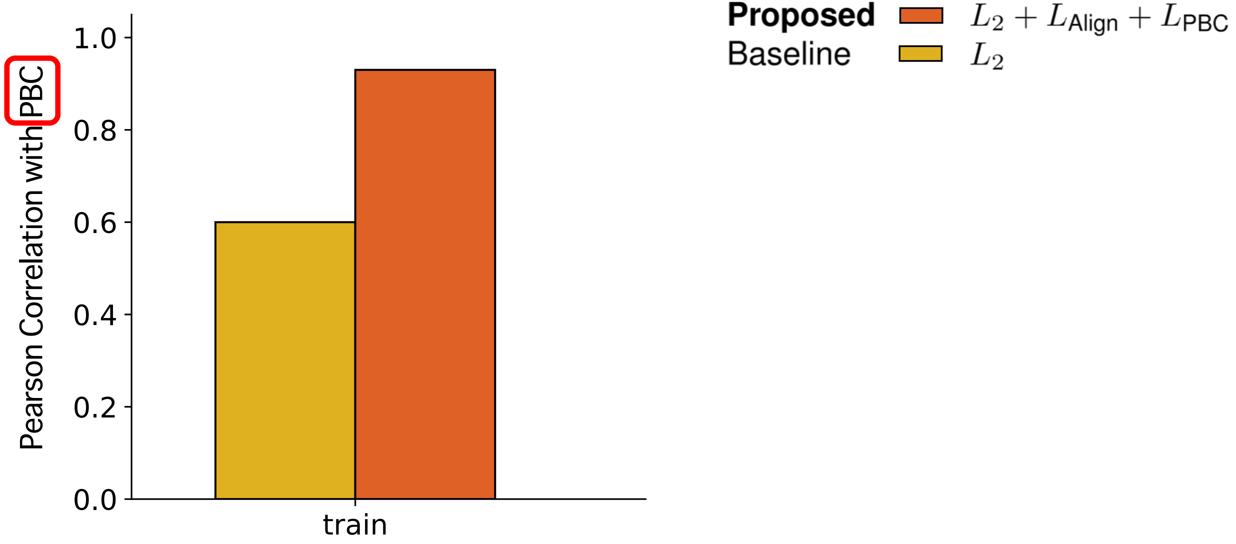
This also applies to **differentiable metrics** (PBC).

Method: Overall Pipeline



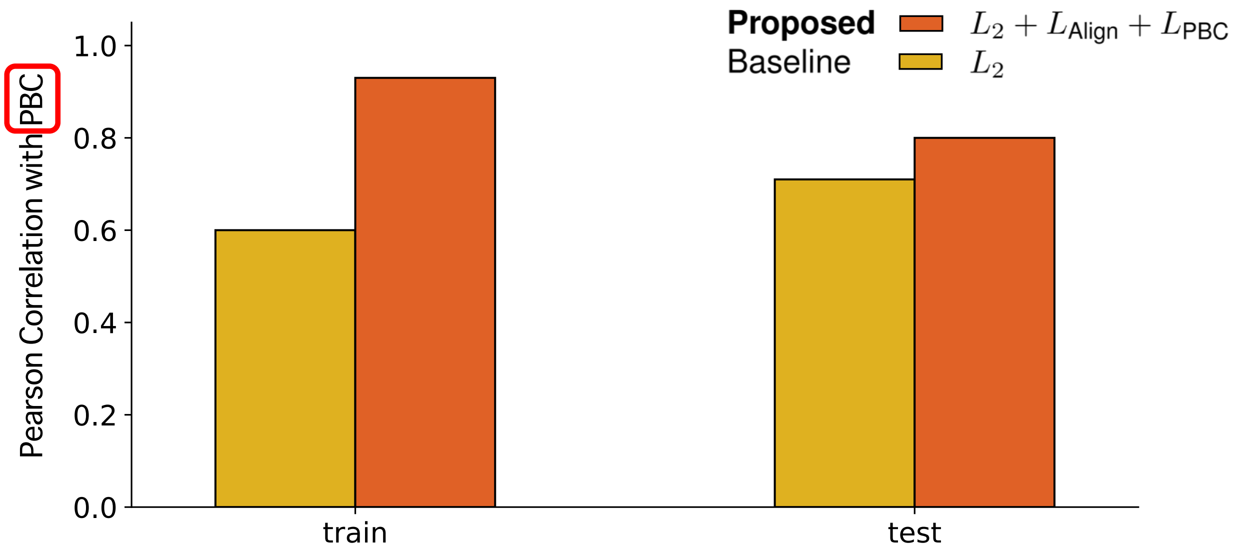
Results: Objective Perceptual Correlation Evaluation on PBC

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



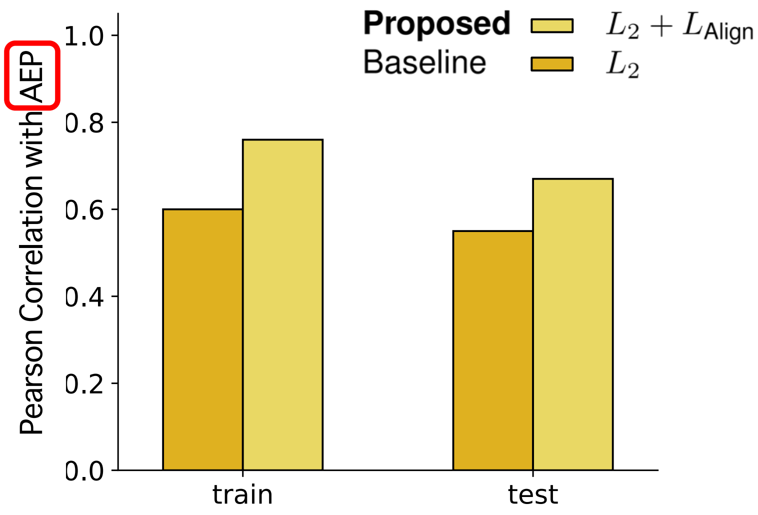
Results: Objective Perceptual Correlation Evaluation on PBC

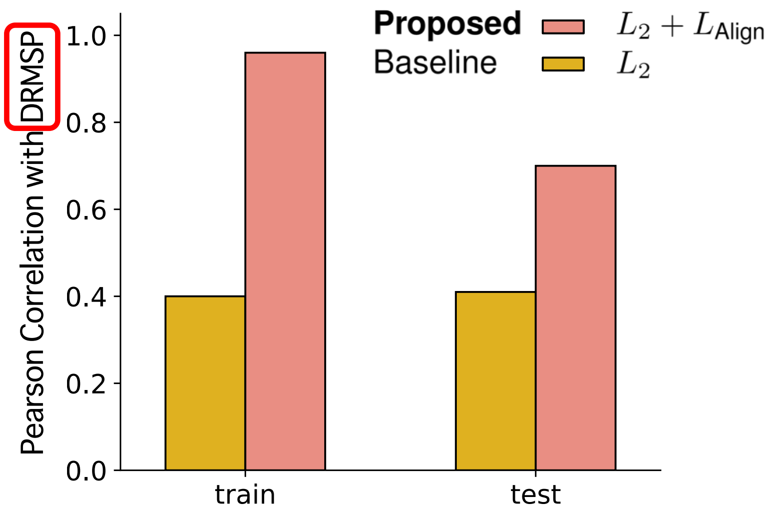
- Our proposed method **achieves better alignment** with perception-informed space.
- The perceptual correlation learned in training transfer to test subjects (unseen).



Generalization to AEP and DRMSP Metrics

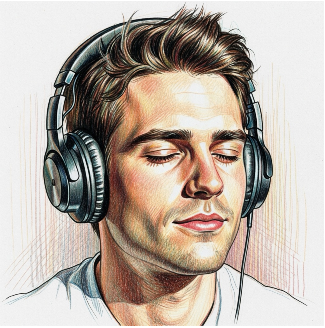
- YES, our proposed correlation improvement method generalizes to externalization and localization.





Results: 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.



	Method	Metric	Spectral Distance (dB)↓
		PBC↓	
Baseline	L_2	1.30	**2.07**
Ours	$L_2 + L_{\text{Align}} + L_{\text{PBC}}$	**1.21**	2.11
		DRMSP↓	
Baseline	L_2	4.21	**2.07**
Ours	$L_2 + L_{\text{Align}}$	**3.20**	2.12

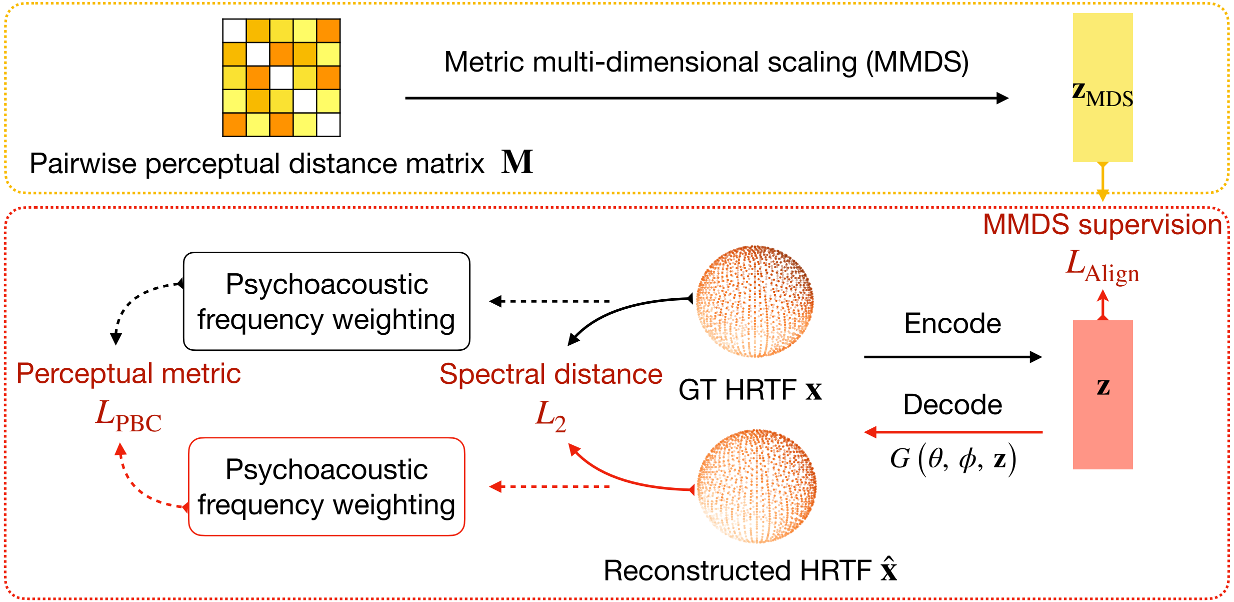
HRTFs selected by our proposed method yield lower perceptual distances.

Perceptually better!

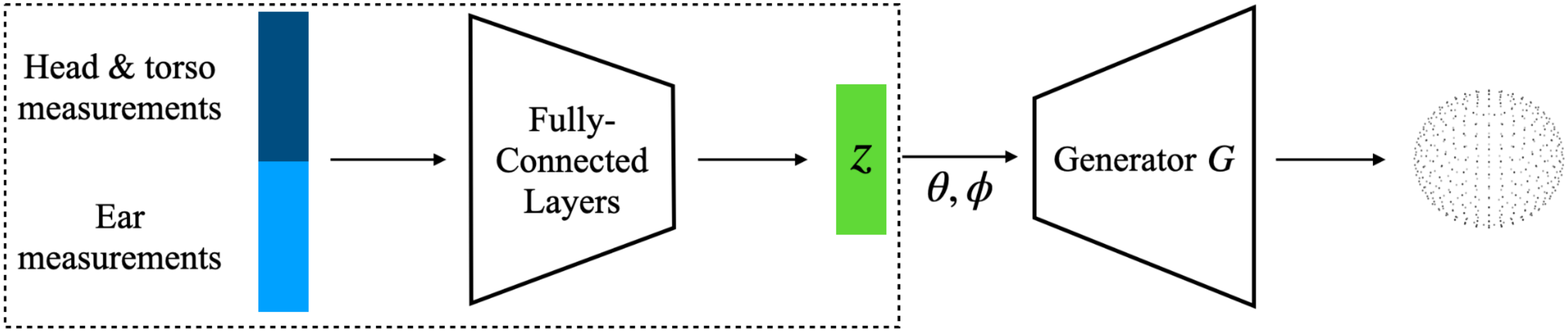
You Zhang, Andrew Francl, Ruohan Gao, Paul Calamia, Zhiyao Duan, and Ishwarya Ananthabhotla. "Towards Perception-Informed Latent HRTF Representations." *WASPAA* 2025.

Summary: Perception-Informed Generative Model for HRTF

- **Goal**: Perceptually grounded HRTF personalization
- **Approach**: Generative neural fields for HRTF modeling with constraints on perceptual alignment using perceptual loss and MMDS
- **Contributions**: Better alignment with human perception in HRTF personalization



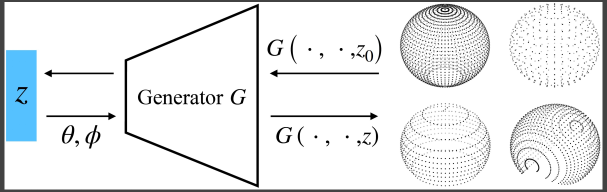
Potential Extension on Personalization



- An encoder to bridge representation learning to personalization
- Subjective validation

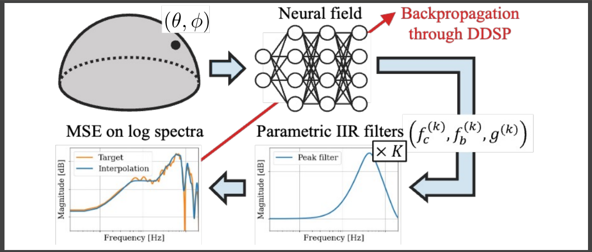
From Neural Fields to Personalized Spatial Audio

Summary



HRTF Modeling with Neural Fields

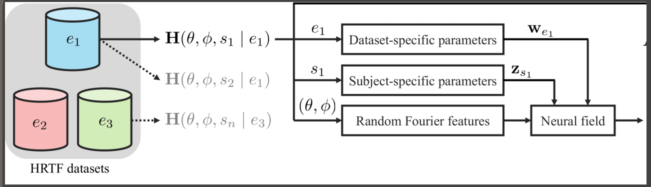
[HRTF Field, ICASSP'23]



Neural IIR Filter Field

[NIIRF, ICASSP'24]

Thank you!
Questions?



Various Approaches for Improved Modeling

[ICASSP'25] [OJSP'25] [WASPAA'25]