HRTF Field: Unifying Measured HRTF Magnitude Representation with Neural Fields

You Zhang, Yuxiang Wang, Zhiyao Duan

Department of Electrical and Computer Engineering,
University of Rochester, New York, USA

Audio Information Research Laboratory

ICASSP 2023
4 - 10 JUNE, RHODES ISLAND, GREECE

ABSTRACT

Head-related transfer functions (HRTFs) are a set of functions of frequency describing the spatial filtering effect of the outer ear (i.e., torso, head, and pinnae) onto sound sources at different azimuth and elevation angles.
Measured HRTFs in existing datasets employ specific spatial sampling schemes, making it difficult to model across datasets with different sampling schemes.

Research question:
- A unified measured HRTF magnitude representation

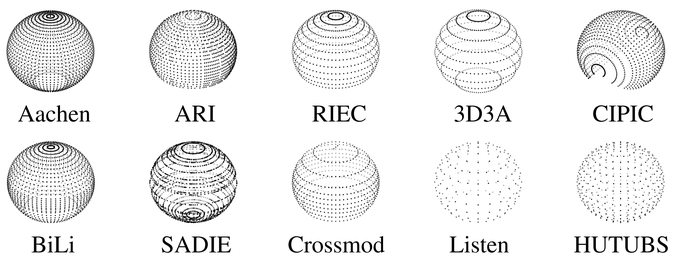
Proposed solution:
- HRTF Representation with **Neural Fields**

Keywords:
head-related transfer function, neural fields, generalized representation across datasets, spatial audio

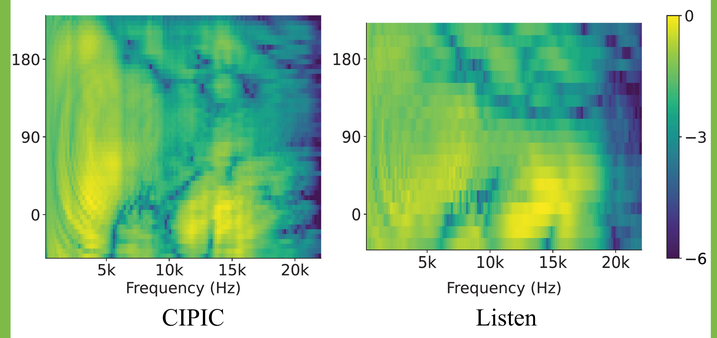
BACKGROUND

Existing measured far-field HRTF databases

Configurations of sound source positions



HRTF magnitude (dB) of the midsagittal planes



METHOD

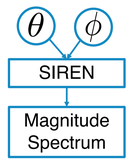
Representing HRTFs of a single subject with a neural field
HRTFs are intrinsically continuous in azimuth and elevation angles.

azimuth angle, elevation angle

$$\mathrm{HRTF}(\theta,\phi) = \frac{\mathbf{p}(\theta,\phi)}{\mathbf{p}_0}$$

$\mathbf{p}(\theta,\phi)$: magnitude spectrum of the sound received
$\mathbf{p}_0$: magnitude spectrum of the source signal

Neural field for one subject
SIREN: a multi-layer perceptron (MLP) with sine activation functions [1]

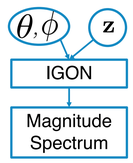


one layer

$$F(\theta,\phi) = \mathbf{W}_n\left(g_{n-1}\circ g_{n-2}\circ\ldots\circ g_0(\theta,\phi)\right) + \mathbf{b}_n,$$
$$\text{where}\quad g_i(\mathbf{h}_i) = \sin\left(\mathbf{W}_i\mathbf{h}_i + \mathbf{b}_i\right).$$

hidden units, weights, bias

Learning HRTF representations across subjects
Implicit Gradient Origin Network (IGON) [2] uses SIREN architecture



Generator: $G(\theta,\phi,\mathbf{z})$ $\quad G:\mathbb{R}^{2+D}\mapsto\mathbb{R}^K$

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

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

EXPERIMENTAL SETUP

Datasets

Name	# Subjects	# Locations	Elevation Range
3D3A [29]	38	648	$[-57^\circ, 75^\circ]$
Aachen [30]	48	2304	$[-66.24^\circ, 90^\circ]$
ARI	97	1550	$[-30^\circ, 80^\circ]$
BiLi [31]	52	1680	$[-50.5^\circ, 85.5^\circ]$
CIPIC [4]	45	1250	$[-50.62^\circ, 90^\circ]$
Crossmod	24	651	$[-40^\circ, 90^\circ]$
HUTUBS [18]	96	440	$[-90^\circ, 90^\circ]$
Listen	50	187	$[-45^\circ, 90^\circ]$
RIEC [32]	105	865	$[-30^\circ, 90^\circ]$
SADIE II [2]	18	2818	$[-90^\circ, 90^\circ]$

Preprocessing
- Map right ear to left, view all ears as left ears
- Normalize by the average energy on the equator

$$\mathrm{H}(\theta,\phi,k) = \frac{\mathrm{HRTF}(\theta,\phi,k)}{\sqrt{\frac{1}{360K}\sum_\theta\sum_k \mathrm{HRTF}(\theta,0,k)^2\Delta\theta}}$$

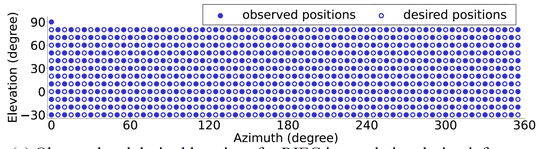
Metrics
Log Spectral Distortion (LSD)

$$LSD(\mathrm{H},\hat{\mathrm{H}}) = \sqrt{\frac{1}{LK}\sum_{\theta,\phi}\sum_k\left(20\log_{10}\left|\frac{\mathrm{H}(\theta,\phi,k)}{\hat{\mathrm{H}}(\theta,\phi,k)}\right|\right)^2}$$

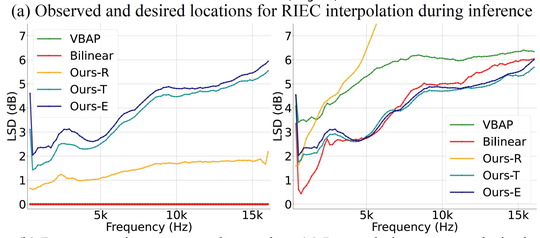
linear-scale magnitude, # spatial locations, frequency index

EXPERIMENTS

HRTF interpolation with the learned HRTF field



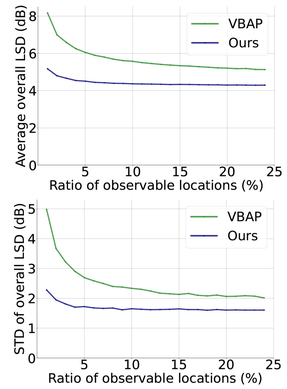
(a) Observed and desired locations for RIEC interpolation during inference



(b) Reconstruction error on observed locations over frequencies

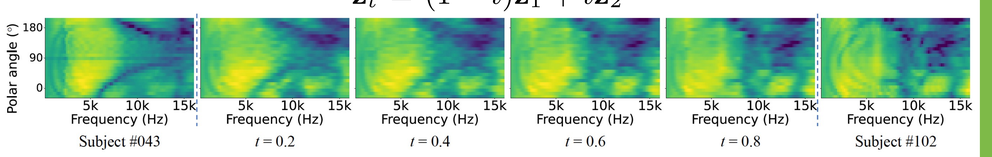
(c) Interpolation error on desired locations over frequencies

Conditional generation from randomly observed locations



Generation by sampling from the latent space

$$\mathbf{z}_t = (1-t)\mathbf{z}_1 + t\mathbf{z}_2$$



CONCLUSIONS

- Neural field is effective in representing HRTFs. It is **agnostic to the specific spatial sampling schemes** adopted by different HRTF datasets.
- A **generative model** learns the manifold information for the neural field representation and can perform well in interpolation and generation.
- We believe that the **proposed HRTF field** will significantly advance **data-driven** research on HRTF personalization.

USEFUL LINKS

Full paper

Code

ACKNOWLEDGMENTS

This work is partially funded by a New York State Center of Excellence in Data Science award and the National Science Foundation (NSF) under grant DGE-1922591.

REFERENCES

[1] Sitzmann, V., Martel, J., Bergman, A., Lindell, D. and Wetzstein, G., "Implicit neural representations with periodic activation functions," in *Proc. NeurIPS*, 2020.
[2] Bond-Taylor, S. and Willcocks, C.G., "Gradient origin networks," in *Proc. ICLR*, 2021.