EUSIPCO 2024

LISTENER ACOUSTIC PERSONALISATION (LAP) CHALLENGE

Task-1

HRTF Normalization for Merging Different HRTF Datasets

Jiale Zhao, Dingding Yao, Zelin Qiu, Chengzhong Wang, and Junfeng Li

2024-08-29

University of Chinese Academy of Sciences

Institute of Acoustics, Chinese Academy of Sciences

Table of Contents

Introduction

- **Differences between HRTF datasets**
  - Measurement setup
  - Postprocessing
  - ...

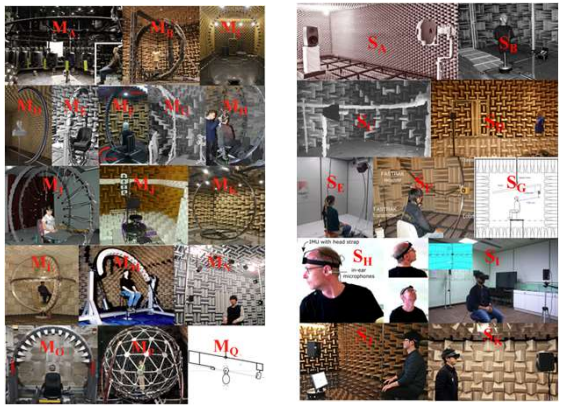


Fig. 8 & Fig. 9 in Li S et al. (2020)

- **Data-driven HRTF modeling**
  - SONICOM, HUTUBS, ...
  - Single dataset → limited performance
  - **Merging datasets**

Task-1

HRTF Normalization for Merging Different HRTF Datasets

- Li S, Peissig J. Measurement of head-related transfer functions: A review[J]. Applied Sciences, 2020, 10(14): 5014.

Introduction

- **Task-1 Constraints**

- **Stage-1**
  - Evaluation with thresholds over different localization metrics computed using an auditory model (*Barumerli et al., 2023*)
  - Differences in localization metrics must be below the threshold for at least 64 of 80 harmonized HRTF datasets (i.e. 80%).

Maintain
Localization Performance

- **Stage-2**
  - The evaluation employs a classifier (*Pauwels and Picinali, 2023*) to test to what extent different collections can be distinguished from each other after harmonization, affirming the removal of measurement setup-induced biases.

Eliminate
Dataset Characteristics

- Barumerli R, Majdak P, Geronazzo M, et al. A Bayesian model for human directional localization of broadband static sound sources[J]. Acta Acustica, 2023, 7: 12.
- Pauwels J, Picinali L. On the relevance of the differences between HRTF measurement setups for machine learning[C]//ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023: 1-5.

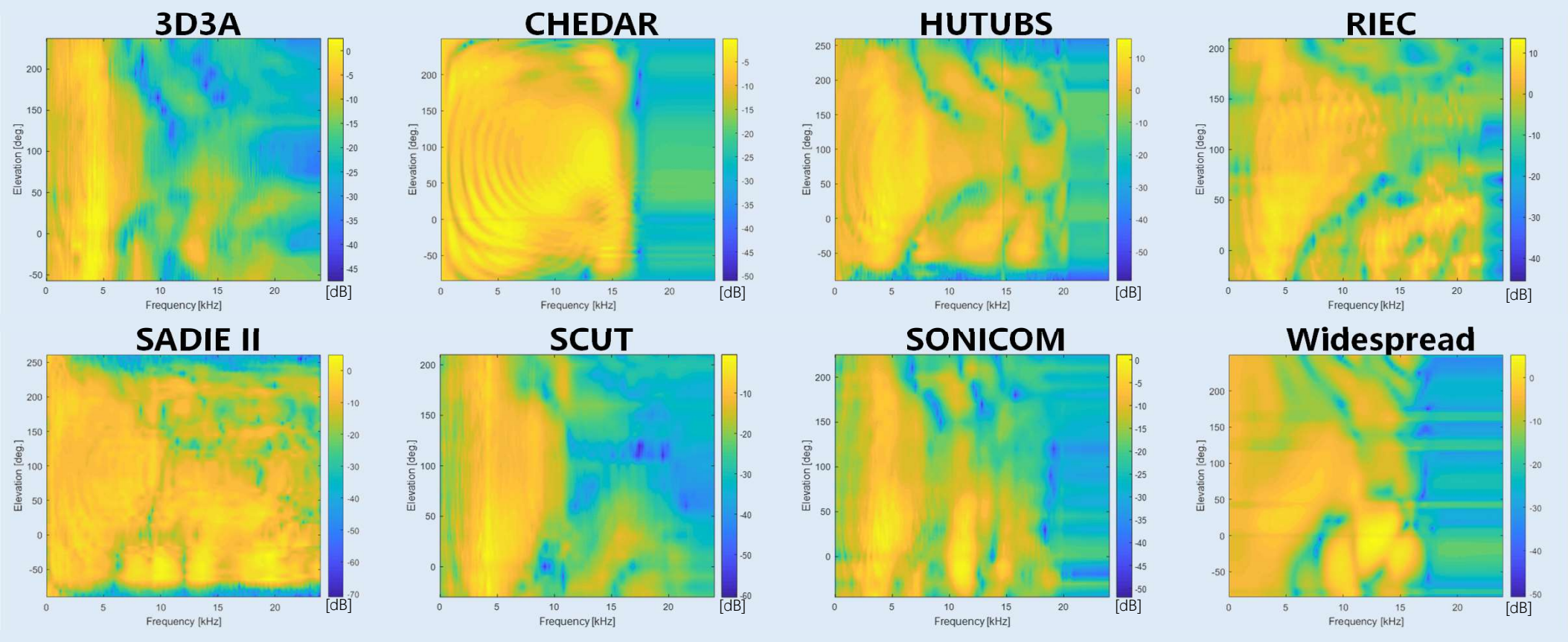
Introduction

- **Dataset Characteristics**

Dataset	Sample Rate	Duration	Radius	Positions	Method	Subjects
3D3A	96 kHz	21.3 ms	0.76 m	648	measured	10
CHEDAR	48 kHz	10 ms	2 m	2522	simulated	10
HUTUBS	44.1 kHz	5.8 ms	1.47 m	440	measured	10
RIEC	48 kHz	10.7 ms	1.5 m	865	measured	10
SADIE II	96 kHz	5.3 ms	1.2 m	2114	measured	10
SCUT	96 kHz	5.3 ms	1 m	864	measured	10
SONICOM	96 kHz	5.3 ms	1.7 m	828	measured	10
Widespread	48 kHz	10 ms	1 m	2522	simulated	10

Introduction

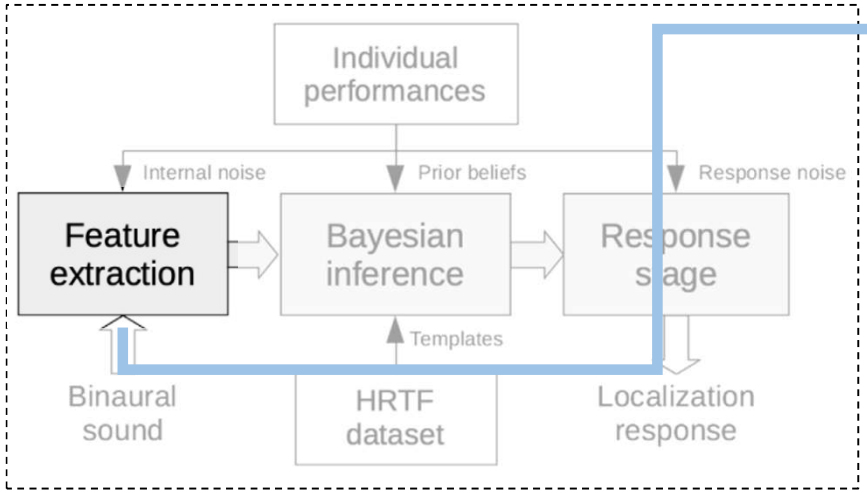
- **Dataset Characteristics**



Contour maps of HRTF spectra on the sagittal plane

Method

- **Maintain Localization Performance**

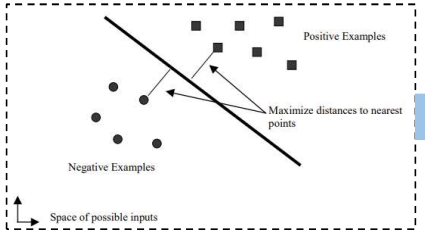


$$\bar{t} = [x_{\mathrm{itd}}, x_{\mathrm{ild}}, \boldsymbol{x}_{\mathrm{L,mon}}, \boldsymbol{x}_{\mathrm{R,mon}}]$$

- **Localization Basis**
  - Interaural Time Difference
  - Interaural Level Difference
  - Binaural gammatone features

Fig. 1 & Eq. 1 in Barumerli et al., 2023

- **Reduce Classification Accuracy**



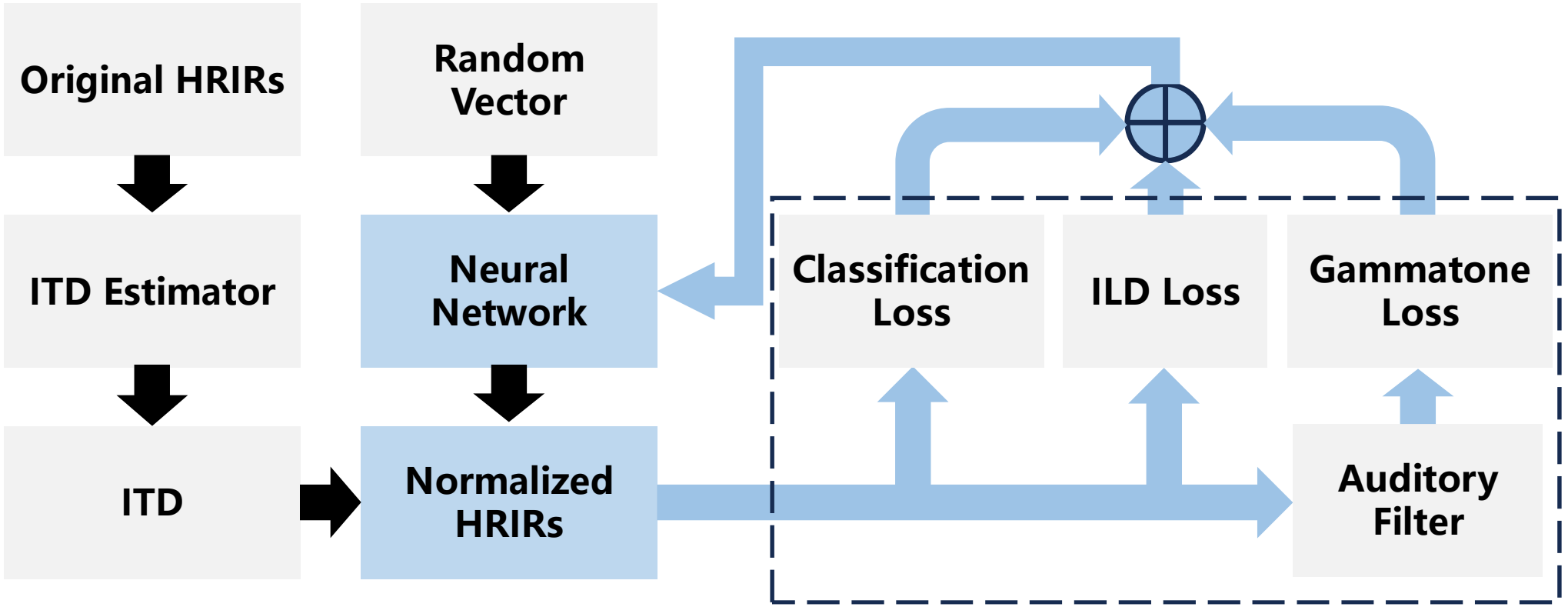
- **Classification Basis**
  - Reduce the distance between classes

Fig. 1 in Platt J, 1998

- Barumerli R, Majdak P, Geronazzo M, et al. A Bayesian model for human directional localization of broadband static sound sources[J]. Acta Acustica, 2023, 7: 12.
- Pauwels J, Picinali L. On the relevance of the differences between HRTF measurement setups for machine learning[C]//ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023: 1-5.
- Platt J. Sequential minimal optimization: A fast algorithm for training support vector machines[J]. 1998.

Method

- Flowchart



Method

- **Localization Loss Function**

$$Loss = Loss_{ILD} + \alpha Loss_{gammatone} + \beta Loss_{class}$$

α is 1 and β is 0.01

$$Loss_{ILD} = \left(\sqrt{\frac{\sum_{i=1}^{N} HRIR_i^{\,2}}{N}} - \sqrt{\frac{\sum_{i=1}^{N} \widehat{HRIR}_i^{\,2}}{N}}\right)^2$$

$HRIR$ is binaural
N is the number of samples

$$Loss_{gammatone} = \left(\mathcal{F}(HRIR) - \mathcal{F}\left(\widehat{HRIR}\right)\right)^2$$

$\mathcal{F}$ → Auditory Model in Stage 1

$$Loss_{class} = \sum_{i=1}^{8} \sum_{j=1}^{i} \left\| c_i - c_j \right\|^2$$

c is class center

Maintain
Localization Performance

Eliminate
Dataset Characteristics

Experiments

- **Preprocessing**
  - **HRIR Resampling:** {44100, 48000, 96000} Hz → 48000 Hz, with 512 samples
  - **ITD estimator:** interaural cross-correlation (IACC)

- **Network Training**
  - **Three blocks:** Fully Connected Layer, Batch Normalization, and ReLU activation in each block
  - **Input & Output:** 512x2 samples
  - **One model for one sampling position:** 126 models in total
  - **Training time:** 600 epochs, Nvidia A40 → 2 hours and 40 minutes

Results

- **Stage-1: Localization Performance**

	Lateral Accuracy [deg.]	Polar Accuracy [deg.]	Lateral RMS error [deg.]	Polar RMS error [deg.]	Quadrant error	Polar gain [1/deg.]
Threshold	5.86	12.67	20.71	5.90	34.56%	0.33
Proposed	**0.64**	**3.45**	**15.13**	**2.60**	**17.54%**	**0.08**

Pass Rate: 76/80=95%

- **Stage-2: Classification Accuracy**

1	2	3	4	5	6	7	8	9	10
28.75%	23.13%	24.38%	25.63%	25.00%	27.50%	25.00%	28.75%	25.63%	25.63%

10 validation runs

Mean: 25.94%

Future Work

- **Perceptual Performance**
  - The perceptual evaluation of normalized HRTFs
  - A perceptually similar localization loss function

Ideal loss function:

1. Input: Target position, Binaural HRTFs — Output: Localization Error
2. Input: Binaural HRTFs — Output: Timbre Error

- **Validation of the Normalized HRTF Dataset**
  - A further investigation whether existing HRTF modeling methods demonstrate improved performance with merged datasets.
  - By comparing the performance across small datasets, large datasets, and merged datasets, the effectiveness of the normalization method could be better confirmed..

EUSIPCO 2024

LISTENER ACOUSTIC PERSONALISATION (LAP) CHALLENGE

Thank you

zhaojiale@hccl.ioa.ac.cn

Code: https://github.com/IOA3Daudio/LAP-Task-1

Jiale Zhao, Dingding Yao, Zelin Qiu, Chengzhong Wang, and Junfeng Li

2024-08-29

中国科学院大学

University of Chinese Academy of Sciences

中国科学院声学研究所

Institute of Acoustics, Chinese Academy of Sciences