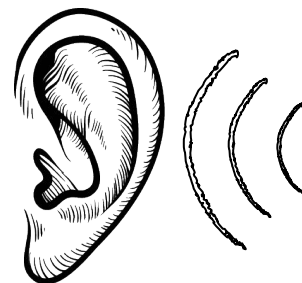


Subjective personalization of HRTFs

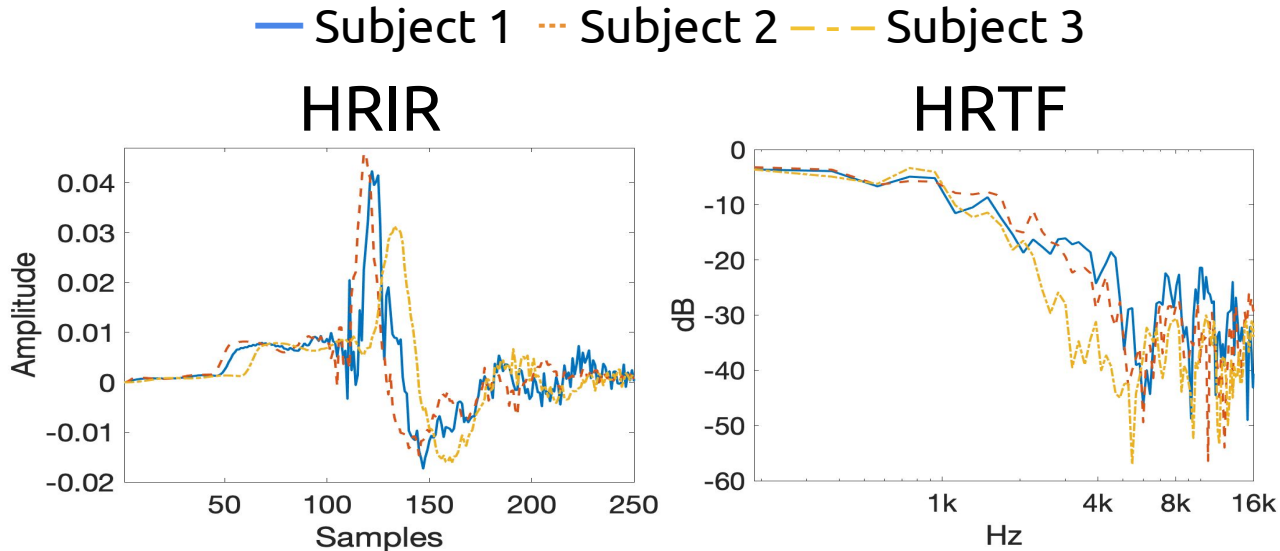


Camilo Arévalo



Individuals HRIRs & HRTFs

Individual's unique ear and head shape significantly influences the way they perceive spatial sound, making the personalization of HRTFs crucial for an accurate and immersive audio experience.

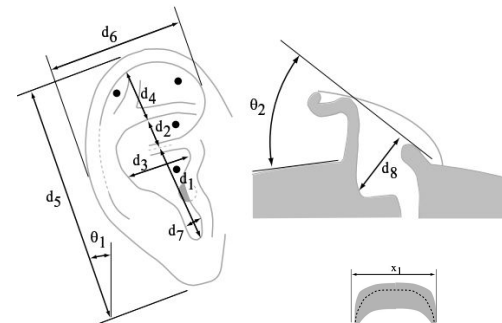
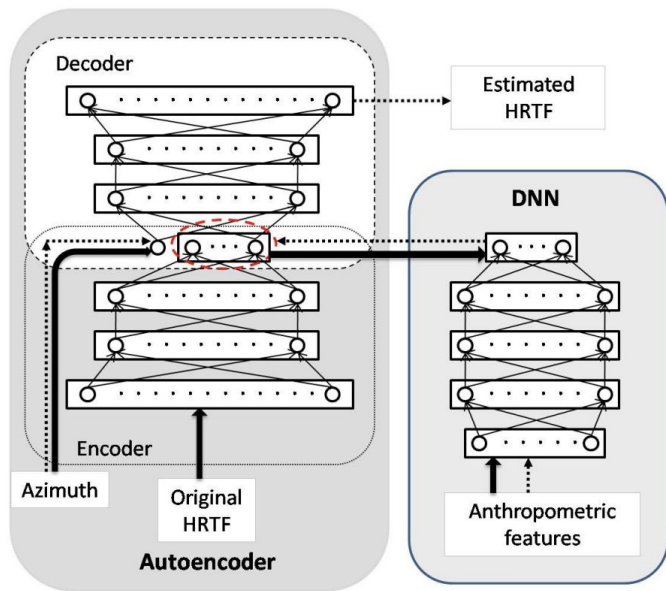


2 Right channel measurements from SADIE database for the front HRTF.

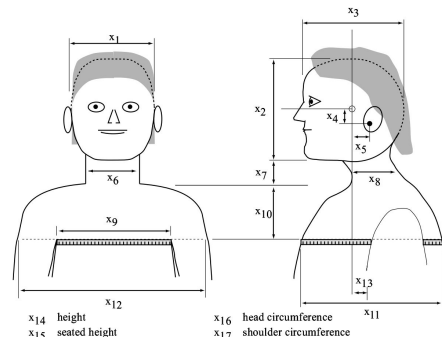


Related work (personalization)

- HRIR personalization using anthropometric features via autoencoders (Chen et al. 2019). SD of 3.2 dB (0.2 -15 kHz) with 20 latent space variables.



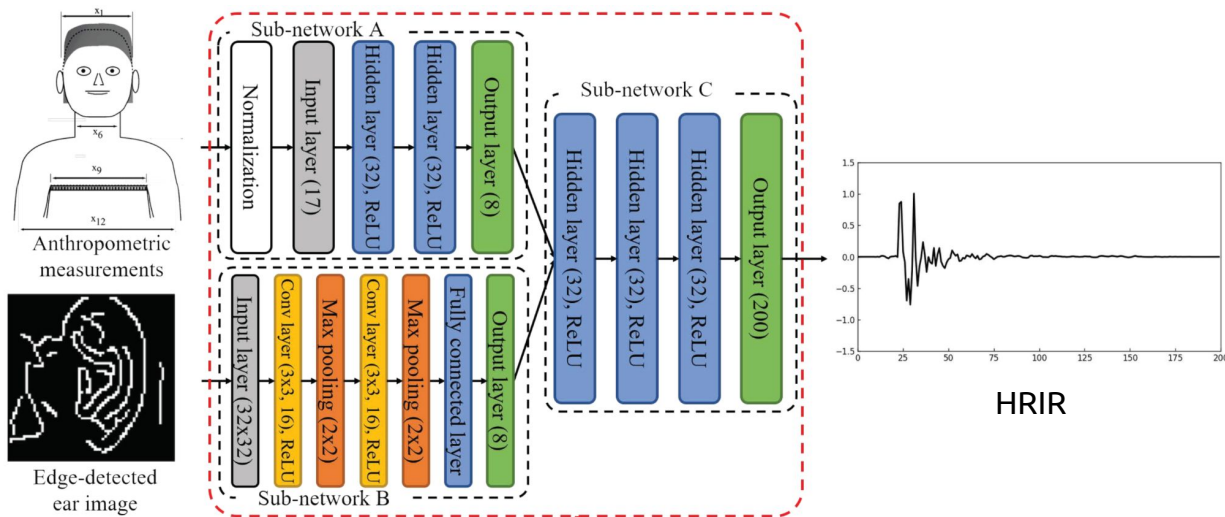
Anthropometric provided in CIPIC database.





Related work (personalization)

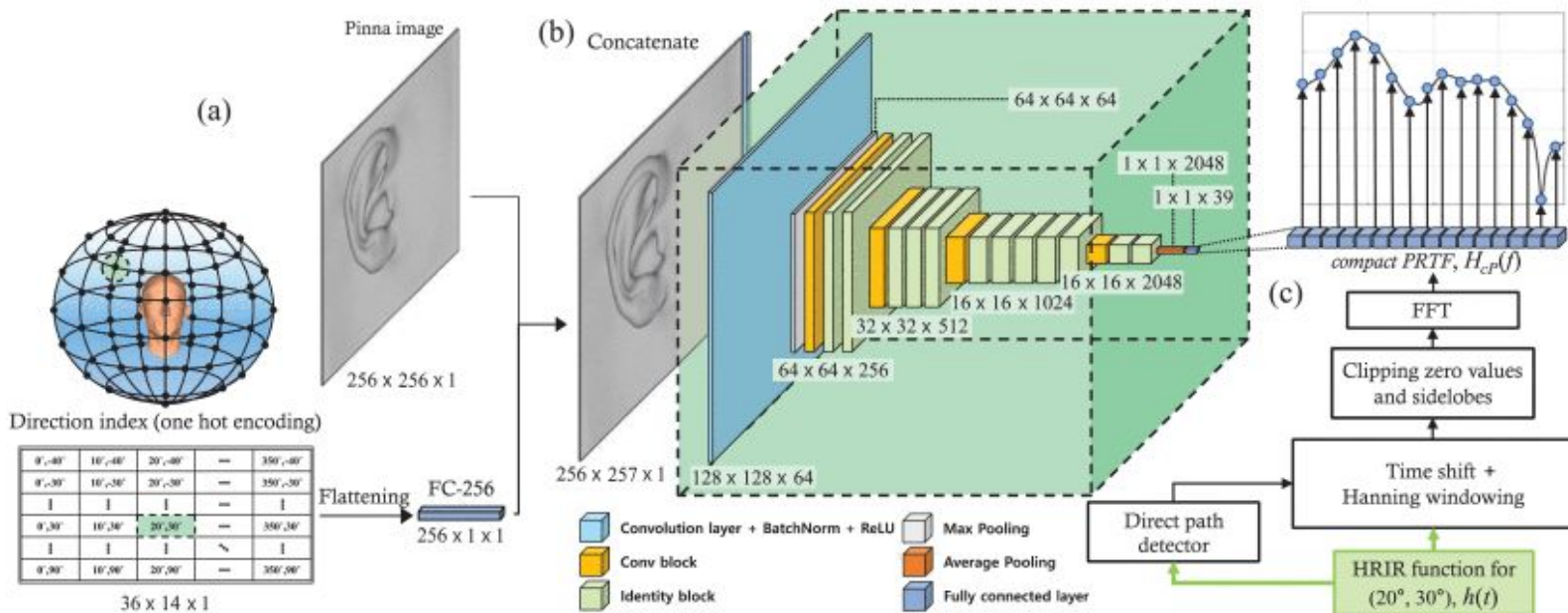
- HRIR personalization using anthropometric features of upper body and pinna pictures via DNN (Lee and Kim. 2018). SD of 4.47 dB.





Related work (personalization)

- Personalization of HRTFs reconstructing first a PRTFs using depth images of pinnae and anthropometric features (Ko et al. 2023). SD of 5.0 dB (4 - 16 kHz).





Related work (personalization)

- HRTF personalization using **anthropometric features** via autoencoders (Chen et al. 2019). SD of 3.2 dB (0.2–15 kHz) with 20 latent space variables.
- HRTF personalization using **anthropometric features** of upper body and pinna pictures via DNN (Lee and Kim. 2018). SD of 4.47 dB.
- Personalization of HRTFs reconstructing first a Pinnae-Related Transfer Functions (PRTFs) using depth images of pinnae and **anthropometric features** (Ko et al. 2023). SD of 5.0 dB (4–16 kHz).



Related work (personalization)

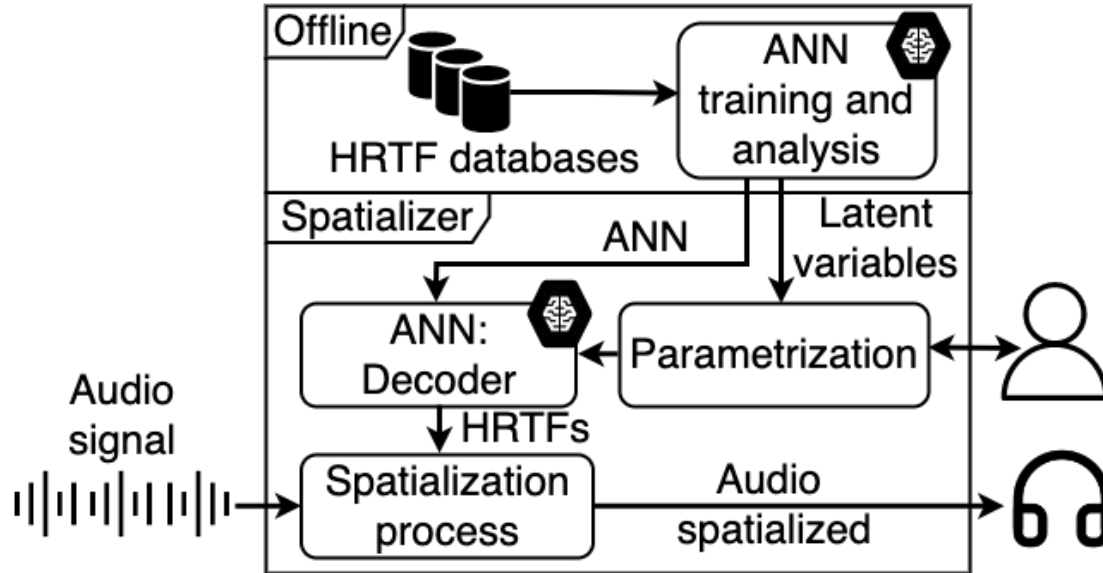
Personalization by the parametrization of a parametric equalizer for the Zenit and interpolate HRTFs for the midsagittal plane HRTFs (Iida and Nakamura 2022).



An experiment performed with two subjects shows that HRTFs provided accurate sound image localization in the midsagittal plane for the front, zenith, and back.

Proposal

We propose the personalization of generic HRIRs based on user feedback. We use autoencoders to reduce the HRIR dimensionality. Autoencoders will create a latent space representation of several databases recorded from different human subjects.





ANN Models explored

- Stacked Auto-Encoders (AE)
- Convolutional Auto-Encoders (CAE)
- Denoising Auto-Encoders (DAE)
- Sparse Auto-Encoders (SAE)
- Multi-Stage Auto-Encoders (MSAE)
- Variational Auto-Encoders (VAE)
- Variational Auto-Encoders - Generative Adversarial Network (VAE-GAN)



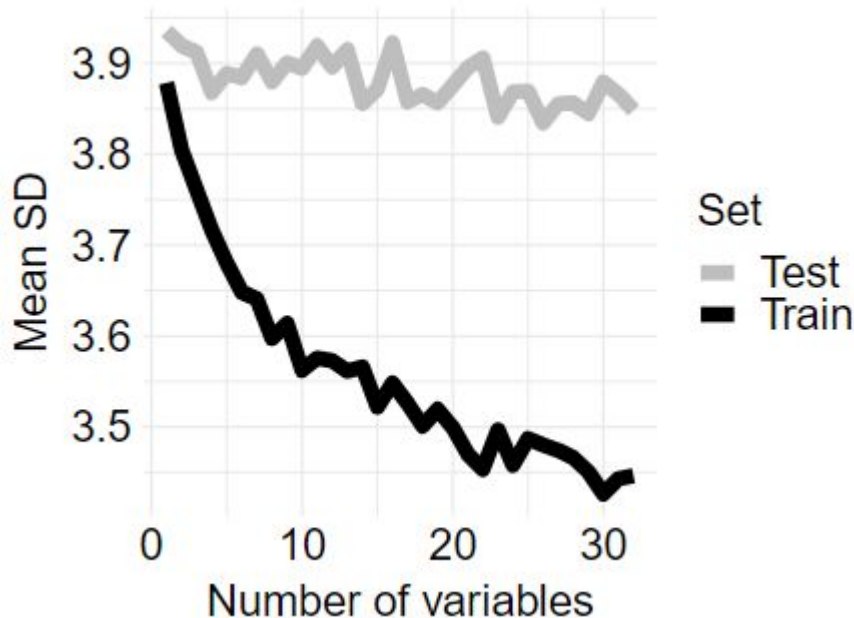
ANN Model

Architecture	HRIRs				HRTFs			
	Training		Test		Training		Test	
	μ	σ	μ	σ	μ	σ	μ	σ
★AE	3.56	1.12	3.82	1.23	4.12	1.45	4.11	1.40
CAE	3.56	1.28	3.88	1.33	6.92	2.96	7.10	3.01
DAE	3.53	1.45	3.82	1.23	4.17	1.46	4.13	1.41
SAE	4.12	1.23	4.12	1.24	4.10	1.40	4.10	1.41
MSAE	3.46	1.12	3.76	1.11	4.05	1.27	4.08	1.42
VAE	4.10	1.22	4.11	1.29	10.76	0.72	10.73	1.48
VAE-GAN	16.41	5.24	16.32	5.28	4.14	1.35	4.16	1.35



Number of latent space variables

Ideally, having a larger number of variables in the latent space could lead to better reconstruction. However, our goal is to minimize the size of the latent space by using as few variables as possible so subjects can memorize and work within a short time.



Considering the Miller's law (Miller 1956), the maximum of variables should be 7 ± 2 . Thus, 5 variables were chosen.



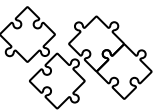
Singular AE

To provide a more organized latent space, we employ singular AE (Pault 2018), a specialized type of AE designed to organize the latent space variables in a manner that prioritizes more meaningful features over less significant ones.

For this implementation, the variables in the latent space created by the SVD are standardized.

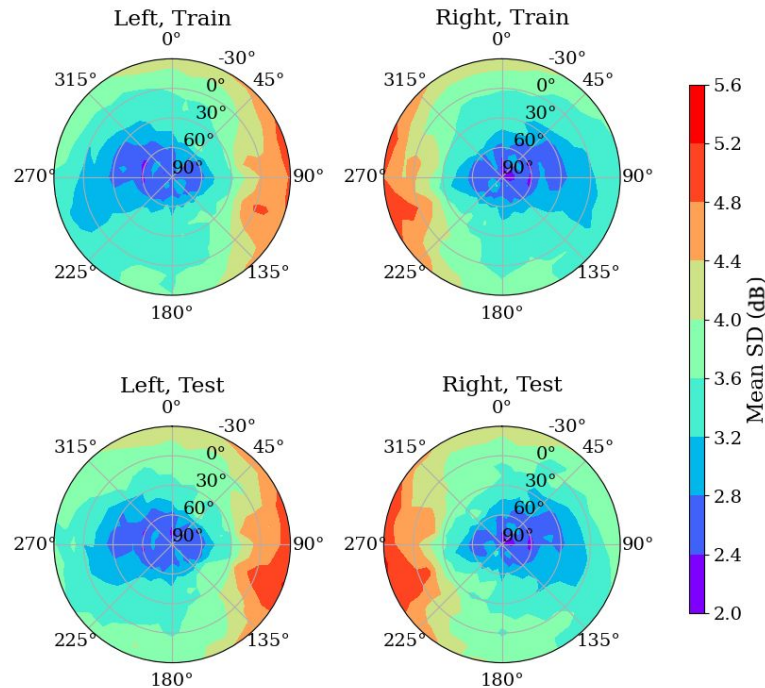
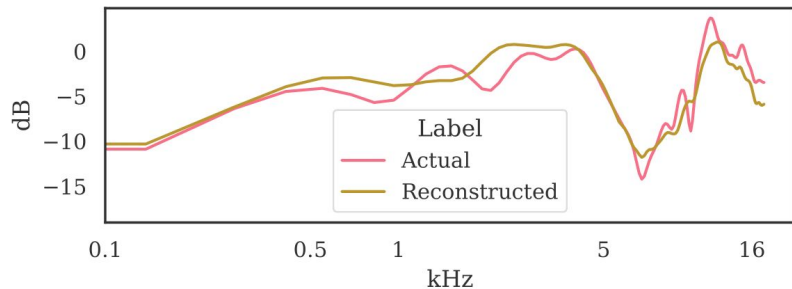
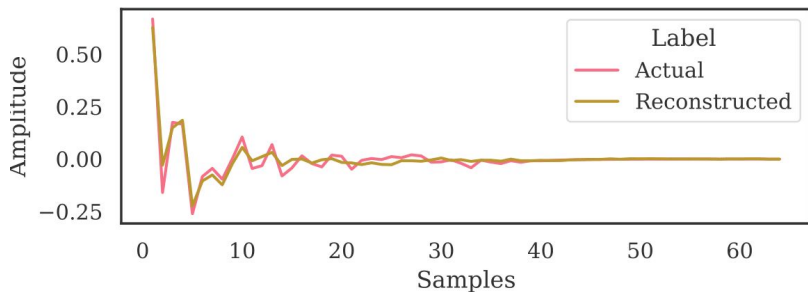
Eigenvalues for the latent space are: 1.59,1.21,0.93,0.75, 0.51.

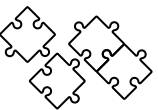
We considered the use of 3 variables to represent the latent space.



Reconstructions of Singular AE

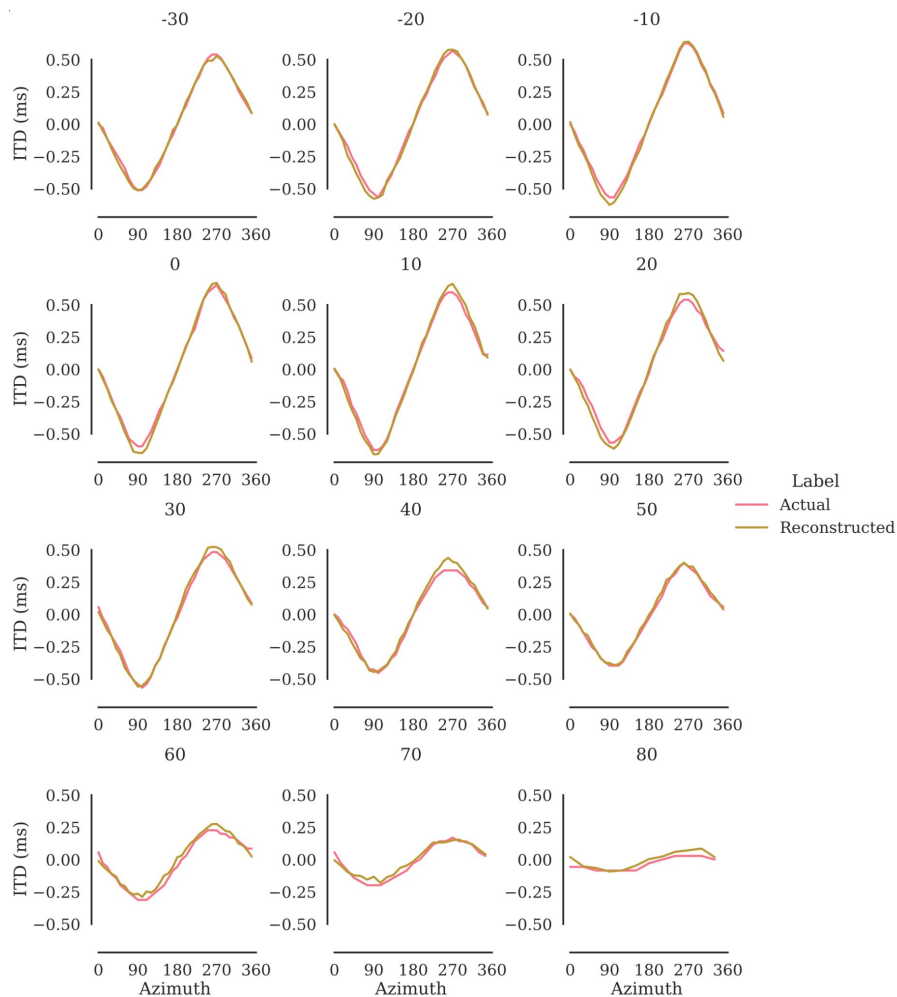
By using 5 latent space variables, model have an average of 3.5 dB ($\sigma=1.1$) for training set and 3.7 dB ($\sigma=1.2$) for the testing set.





Reconstructions (ITDs)

ITDs reconstructions have an average error of 1.4 and 2.6 μs for the training and testing sets, respectively.





HRIRs personalization tool

- Personalization tool expose the three latent space variables normalized.
- Fixed locations and reference elevation angle to evaluate:
 - Reference: 0° .
 - Fixed: -30° , 90° , and 180° (back).
- Memory slots in case subject want to save previous values to compare.
- Illustration of the audio source location.
- Export audio stimuli for the experiment.

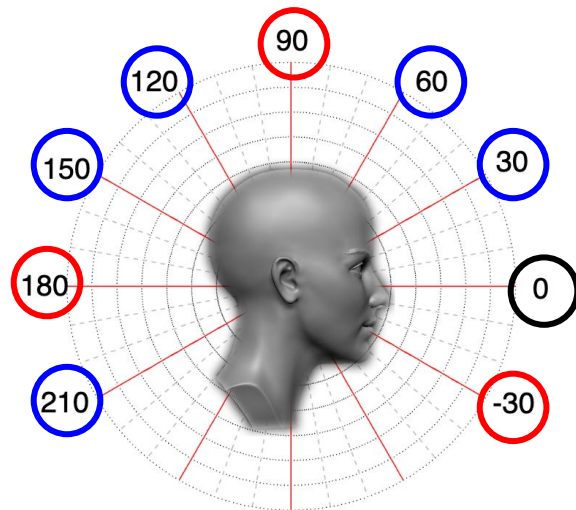
The screenshot displays the HRIRs personalization tool interface, which is divided into several sections:

- Parameters:** Three horizontal sliders labeled 1, 2, and 3, each with a small square handle in the center.
- Location (R):** A dropdown menu showing "[0,-30]" and a radio button labeled "Selected location" which is currently selected. A "Reference" radio button is also present.
- Buttons:** "No loop" (with a toggle switch), "Loop", "Play (P)", "Export", and "Load".
- Azimuth:** A circular plot showing a 3D head model with a blue dot indicating the audio source location. The plot is labeled with angles from 0° to 330° in 30° increments.
- Elevation:** A circular plot showing a 3D head model with a blue dot indicating the audio source location. The plot is labeled with angles from -30° to 90° in 30° increments.
- Memory:** A list of memory slots labeled "memory 1" through "memory 9". A "Save" button is located to the right of the list, and a "Use" button is located below it.



Subjective evaluation

- 14 subjects (one excluded due to hearing issues).
- Stimulus:
 - Two acoustic guitar riffs (~5s, sampled at 35.2 kHz/16 bits) were used. One for practice and personalization and the other one for the main experiment.
 - Angles:



- Reference
- Practice and used for personalization
- Main block

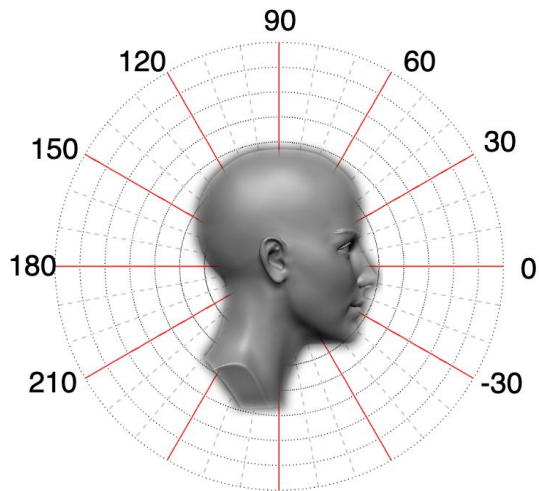


Subjective evaluation

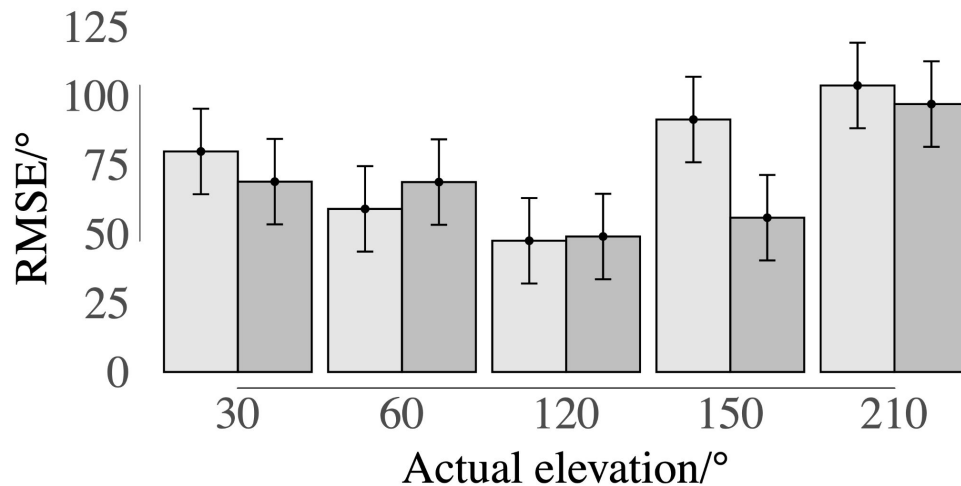
- Experiment is comprised by three blocks: practice and a block for a personalized and non-personalized method. Methods block were permuted between subjects.
- For the non-personalized method, the MIT HRIR database (Gardner and Martin 1995) was used.
- On average, subjects invested 22 minutes and 28 seconds ($\sigma=10$ minutes) in the personalization process.



Results of subjective evaluation



Elevation



□ Non-personalized ■ Personalized

Error bars denote the 95% confidence intervals.



Discussion

- Even though the reconstructions are not perfect, the ANN implemented is comparable to other methods in terms of SD while using only five variables in the latent space (ca. 2.26 dB with 128 variables).
- Personalization method proposed, while using 512 latent space variables ($SD < 2\text{dB}$), could be an alternative to compress HRIRs databases.
- Compared to other methods in the literature, we combined several HRTFs databases to perform the AE training.

MM References

1. T.-Y. Chen, T.-H. Kuo, and T.-S. Chi, "Autoencoding HRTFS for DNN Based HRTF Personalization Using Anthropometric Features," in ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2019, pp. 271–275. doi: 10.1109/ICASSP.2019.8683814.
2. G. W. Lee and H. K. Kim, "Personalized HRTF Modeling Based on Deep Neural Network Using Anthropometric Measurements and Images of the Ear," *Appl. Sci.*, vol. 8, no. 11, p. 2180, Nov. 2018.
3. B.-Y. Ko, G.-T. Lee, H. Nam, and Y.-H. Park, "PRTFNet: HRTF Individualization for Accurate Spectral Cues Using a Compact PRTF," *IEEE Access*, vol. 11, pp. 96 119–96 130, 2023.
4. azuhiro IIDA and Fuka NAKAMURA, "Toolkit For 3d Audio Rendering Using Individualized Head-related Transfer Functions In The Upper Hemisphere," in 27th International Conference on Auditory Display, Jun. 2022. [Online]. Available: http://www.iida-lab.it-chiba.ac.jp/literature/International.Conference.Proceedings/43.ICAD2022Proceedings_lidaNakamura.pdf
5. W. Chen, R. Hu, X. Wang, and D. Li, "HRTF Representation with Convolutional Auto-encoder," in *Proc. of 26th Int. Conf. on MultiMedia Modeling*. Berlin: Springer-Verlag, Jan. 2020, pp. 605–616, DOI:10.1007/978-3-030-37731-1_49.
6. C. Armstrong, L. Thresh, D. Murphy, and G. Kearney, "A Perceptual Evaluation of Individual and Non-Individual HRTFs: A Case Study of the SADIE II Database," *Appl. Sci.*, vol. 8, no. 11, p. 2029, Nov. 2018.
7. H. F. Kaiser, "A NOTE ON GUTTMAN'S LOWER BOUND FOR THE NUMBER OF COMMON FACTORS," *British Journal of Statistical Psychology*, vol. 14, no. 1, pp. 1–2, 1961, doi: 10.1111/j.2044-8317.1961.tb00061.x.

Thank you!

