DOI: 10.61782/fa.2023.1072

forum**acusticum** 2023

CONCEPTS FOR THE EVALUATION OF A PARAMETRIC PINNA MODEL

Felix Perfler* **Piotr Majdak** **Florian Pausch** **Katharina Pollack**
Acoustics Research Institute, Austrian Academy of Sciences, 1040 Vienna, Austria

ABSTRACT

Personalized head-related transfer functions (HRTFs) are essential in systems aiming for realistic binaural sound reproduction, taking the individual anatomy into account. HRTFs are typically acoustically measured, but they can also be numerically calculated provided an accurate 3D geometry of the listener's head and pinnae. A parametric pinna model (PPM) represents a tool to personalize a well-defined 3D mesh of a generic pinna to the actual geometry of the listener's ear. However, the PPM-parameter ranges covering the variety of human ears are yet unclear. In this work, we describe a previously introduced PPM and its key features. We further outline methods for an evaluation of its parameter ranges. The insights gained can be used to create various datasets of pathological and non-pathological ears and train neural networks to automatically parameterize the PPM to describe the individual geometries of human ears.

Keywords: *parametric pinna model, model evaluation, head-related transfer functions, binaural audio*

1. PARAMETRIC PINNA MODELS

The shape of human ears is highly individual and has significant influence on head-related transfer functions (HRTFs) [1–3]. A PPM can be a powerful tool for various applications including the creation of a database of synthetic ears and corresponding HRTFs, which can be used for data-driven HRTF-personalization approaches. The database ideally contains representable examples of ear geometries. The unique shape of the ear is not only relevant for HRTF individualization but can also be utilized for biometric identification, which was the motivation of other pinna models, e.g., [4]. Similar to many face-modeling approaches, their model is a 3D-morphable model (3DMM), however, it is based on 20 ears only [5]. In contrast, other PPMs have been developed with a clear focus on HRTF calculation, [6, 7]. The limitations of these models are unclear, thus, a more general PPM describing the ear shape through a vector of parameters has been proposed [8]. Once the PPM parameters are determined, it is possible to synthesize a pinna geometry using the PPM parameter vector, and numerically calculate HRTFs by application of numerical simulation frameworks such as Mesh2HRTF [9].

This PPM consists of an armature, which is motivated by the biological form of the human ear, as defined in [8]. Using various control points, this structure can be modified, by tweaking the underlying Bézier curves. These changes also effect a mesh which is connected to the bones. The mesh structure can be further altered through parameters controlling the weighting of the mesh. In total, there are 144 parameters which can be changed. Even though the model is motivated by the biological ear form, it is quite similar to blend shape models in face modeling. Various target meshes can be manually registered to the PPM [8]. However, an automatic model registration approach has not yet been introduced. The connection of the PPM-parameter set and the ear geometry is nonlinear and can be described by Equation 1, describing how the ear geometry $\mathscr{T}$ is mapped onto the parameter vector Θ, i.e.

$$f : \mathscr{T}^{NxM} \rightarrow \Theta^{D}, \tag{1}$$

where N and M are the dimension of the geometric ear representation, e.g. images or point clouds, and D is the

**Corresponding author*: *felix.perfler@oeaw.ac.at.*

number of PPM parameters, respectively.

It is not known yet how to automatically register the large variety of human pinnae to that PPM. Deep neural networks (DNNs) represent a promising approach, as DNNs can be trained to map data to non-linear spaces. Such algorithms, however, rely on extensive (labeled) databases. To this end, ear geometries can be synthesized and represented in form of point clouds or rendered images, therefore enabling the creation of large databases for neural network training purposes. In turn, this strategy requires to determine feasible parameter ranges which allow generating human-like ears.

2. MODEL EVALUATION IN OTHER DOMAINS

One of the most groundbreaking face models was built by analyzing laser scans of 200 faces [10]. This model was represented by principal components, therefore the face shape can be defined by

$$S_{\text{model}} = \overline{S} + \sum_{i=1}^{m-1} \alpha_i s_i, \tag{2}$$

where $\overline{S}$ is the average face of the 200 scans and s_i representing the eigenvectors of the orthogonal basis as a result of the principal component analysis (PCA). By changing the weighting factor α_i, new faces can be created. Crucially, algorithms to register unseen faces have been proposed [10], when face representations are available either as 3D scans or images. More recently, the amount of faces forming the basis of the model has been extended drastically for the creation of such 3DMMs. As an example, a model utilizing 10 000 face scans, also including people from a variety of ethnics [11].

Models built on blend shapes are also used in face modeling. These models deform a neutral-base face model through additive deformations by controlling parameters. In [12], a template mesh is deformed to match a target mesh by an artist. While those models are very expressive in the shapes they allow to create, different parameter combinations can result in the same shape. This problem makes automatic registration very challenging. Models combining both PCA and blend shape approaches have also been proposed [13].

3. EVALUATION OF A PARAMETRIC PINNA MODEL

The PPM can be manually registered to a target mesh, which has been obtained by scanning the pinna of a subject. First, the target mesh is loaded into a Blender [1] workspace. In a second step, the PPM in its default parameterization is also loaded in said workspace. The template is the mean ear structure by means of a PCA-based average [14]. Thereafter, the PPM parameters are modified to fit the target mesh. The registration is finished once the PPM mesh and the target mesh coincide as closely as required. The PPM parameters can be manually fit with such precision that the geometric error is below the defined boundary of 1 mm. This error also results in a small enough acoustic error, such that the resulting HRTFs match the individual HRTFs sufficiently enough. Since the manual registration takes a lot of effort and requires the individual target meshes, an automatic algorithm for estimating the PPM parameters would be preferred. Figure 1 shows a supervised approach in which a DNN is trained on a large dataset of PPM instances, as one possible way to tackle the estimation problem.

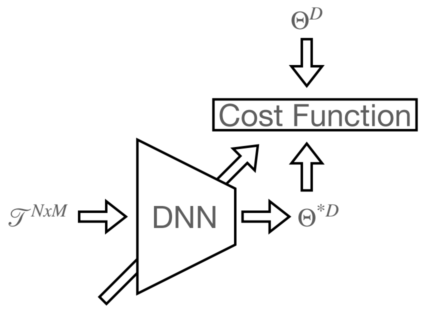

Figure 1: Process of supervised training. During the training, the predicted model output Θ^{*D} from a training sample $\mathscr{T}^{NxM}$ is compared to the actual output Θ^D. The DNN weights are updated according to the cost function.

Such a dataset would consist of the high dimensional pinna geometry $\mathscr{T}$ and the corresponding model parameter vector Θ. It is possible to synthesize a large amount of pinnae from a given PPM-parameter set, but the parameter ranges for representing a diverse set of non-pathological human ears are still unknown. In order to determine feasible parameter ranges more manually fitted PPM-representations could be created. Ideally this should be done by more than one person in order to reduce biases in the usage of the model, since it is possible to create the same shape using different parameter combinations. In the future multimodal distributions of parame-

[1] https://www.blender.org

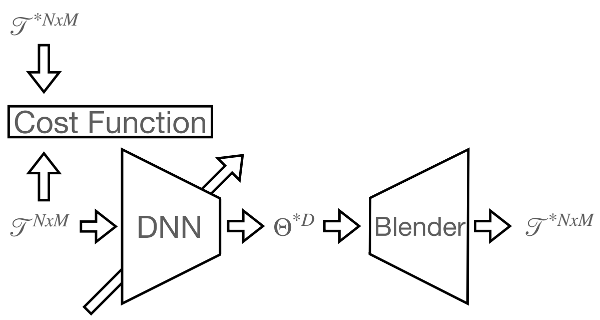

Figure 2: Process of unsupervised training. During the training, the input sample $\mathscr{T}^{NxM}$ is compared to a synthesised version of the input sample $\mathscr{T}^{*NxM}$ created from the predicted output of the network Θ^{*D}.

ters can be derived, from which parameter ranges are set for the generation of a dataset. Alternatively or additionally, pinna experts could be tasked with the creation of pinnae, which are no longer non-pathological, therefore deriving boundaries for the parameters. Neural-network architectures which take unordered point clouds as input have been proposed (see, e.g. [15]). Following such approaches it might make it possible to use an unsupervised method for model evaluation, cf. Figure 2.

The advantage of training a network in an unsupervised way is that the PPM parameters can be unknown, meaning that the training dataset only requires the pinna geometry but not the corresponding PPM-parameter labels. Ideally, if available, real-world databases of pinna geometries could be used for such approach, given that enough data is available. But also the creation of synthetic data would be easier as the ranges of the PPM parameters would not need to be known for data creation, and a PCA-like approach similar to [4] could be used.

4. ACKNOWLEDGMENTS

This study received funding from the European Union's Horizon 2020 research and innovation funding programme (grant no. 101017743, project "SONICOM"–Transforming auditory-based social interaction and communication in AR/VR).

5. REFERENCES

[1] D. J. Kistler and F. L. Wightman, "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," *J Acoust Soc Am*, vol. 91, no. 3, pp. 1637–47, Mar. 1992. [Online]. Available: `http://www.ncbi.nih.gov/entrez/query.fcgi?cmd=retrieve&db=pubmed&dopt=abstract&list_uids=1564200`.

[2] H. Møller, M. F. Sørensen, C. B. Jensen, and D. Hammershøi, "Binaural technique: do we need individual recordings?" *J Audio Eng Soc*, vol. 44, pp. 451–469, May 1996.

[3] J. C. Middlebrooks, "Individual differences in external-ear transfer functions reduced by scaling in frequency," *J Acoust Soc Am*, vol. 106, no. 3 Pt 1, pp. 1480–1492, Sep. 1999. [Online]. Available: `http://www.ncbi.nih.gov/entrez/query.fcgi?cmd=retrieve&db=pubmed&dopt=abstract&list_uids=10489705`.

[4] H. Dai, N. Pears, and W. Smith, "A Data-Augmented 3D Morphable Model of the Ear," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 2018, pp. 404–408.

[5] R. Zolfaghari, N. Epain, C. T. Jin, J. A. Glaunès, and A. I. Tew, "Generating a morphable model of ears," *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1771–1775, 2016.

[6] P. Stitt and B. F. G. Katz, "Sensitivity analysis of pinna morphology on head-related transfer functions simulated via a parametric pinna model," *Journal of the Acoustical Society of America*, vol. 149, no. 4, pp. 2559–2572, Apr. 2021. [Online]. Available: `https://hal.science/hal-03196219`.

[7] D. Fantini, F. Avanzini, S. Ntalampiras, and G. Presti, "HRTF Individualization Based on Anthropometric Measurements Extracted from 3D Head Meshes," in *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, ser. 2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA), IEEE, 2021-9-8, ISBN: 1-66540-999-1.

[8] K. Pollack, F. Pausch, and P. Majdak, "Parametric pinna model for a realistic representation of listener-specific pinna geometry," in *24th Inter-*

national Congress on Acoustics: ICA 2022, Invited paper., Gyeongju, Korea: The Acoustical Society of Korea, Oct. 24, 2022. [Online]. Available: https://ica2022korea.org/data/Proceedings_A21.pdf.

[9] H. Ziegelwanger, W. Kreuzer, and P. Majdak, "Mesh2HRTF: An open-source software package for the numerical calculation of head-related transfer functions," Florence, Italy, 2015, p. 42 583.

[10] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," in *International Conference on Computer Graphics and Interactive Techniques*, 1999.

[11] J. Booth, A. Roussos, S. Zafeiriou, A. Ponniah, and D. Dunaway, "A 3D Morphable Model Learnt From 10,000 Faces," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016.

[12] P. Garrido, L. Valgaert, C. Wu, and C. Theobalt, "Reconstructing Detailed Dynamic Face Geometry from Monocular Video," *ACM Trans. Graph.*, vol. 32, no. 6, Nov. 2013, ISSN: 0730-0301. [Online]. Available: https://doi.org/10.1145/2508363.2508380.

[13] T. Li, T. Bolkart, M. J. Black, H. Li, and J. Romero, "Learning a Model of Facial Shape and Expression from 4D Scans," *ACM Trans. Graph.*, vol. 36, no. 6, Nov. 2017, ISSN: 0730-0301. [Online]. Available: https://doi.org/10.1145/3130800.3130813.

[14] C. Guezenoc and R. Séguier, "A wide dataset of ear shapes and pinna-related transfer functions generated by random ear drawings," *The Journal of the Acoustical Society of America*, vol. 147, no. 6, pp. 4087–4096, Jun. 2020. [Online]. Available: https://doi.org/10.1121%2F10.0001461.

[15] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, *Dynamic Graph CNN for Learning on Point Clouds*, 2019.