A Landmark-Based Parametric Pinna Model For The Calculation of Head-Related Transfer Functions

Katharina Pollack, Piotr Majdak, Hugo Furtades

A Parametric Pinna Model for the Calculations of Head-Related Transfer Functions

Katharina Pollack[1] **Piotr Majdak**[1] **Hugo Furtades**[2]
[1] Acoustics Research Institute, Austrian Academy of Sciences
[2] Dreamwaves GmbH, Vienna
kpollack@kfs.oeaw.ac.at

ABSTRACT

Personalised head-related transfer functions (HRTFs) are essential for realistic sound reproduction via headphones. They can be acoustically measured, but they also can be numerically calculated from 3D meshes of the listener. While the mesh acquisition is a complex process and involves feature extraction via photogrammetry, it often results in artefacts such as holes and distortions, yielding perceivable artefacts in the calculated HRTFs. We address this problem by proposing a parametric pinna model that modifies a high-resolution template mesh according to anthropometric data of the listener. The model is based on morph target animation and skeletal animation, two methods used for the deformation of 3D geometries in the area of computer animation. We show a systematic evaluation of the technique on a human pinna and discuss its limits in the context of personalised HRTFs. The model aims as a step towards automated HRTF acquisition from photographs, which is an important issue for both the generation of a large number of plausible HRTFs and an easy access to personalised HRTFs for a wide range of audience.

1. INTRODUCTION

Humans are able to localise natural sound sources, i.e., assign direction and distance to a perceived auditory event [1] based on acoustic features of a binaural signal. Binaural signals emerge by filtering of the sound caused by torso, head and outer ears, i.e., pinnae, with the latter being distinctive for individuals. This personalised filtering can be described with head-related transfer functions (HRTFs) [2] which characterise the spatial filtering of the sound of a source relative to the listener's position.

HRTFs are usually acoustically measured, however, because of the rather elaborate acoustical process taking tens of minutes [3] and requiring professional equipment and laboratory conditions, a numerical calculation of HRTFs can be an option. By acquiring a 3D model of the head, the so-called mesh, HRTFs of the listener can be numerically calculated by means of acoustic simulations [4]. The HRTF calculation provides many advantages such as the independence of acoustic laboratory equipment and convenience in data acquisition, e.g., the comfort of not having to sit still for the whole measurement process. However, obtaining an accurate mesh is not trivial and may require much manual post processing.

Algorithms for non-parametric manipulations of pinna meshes have been proposed, e.g., WiDESPREaD [5], and LDDMM [10]. They rely on principal component analysis and diffeomorphism, respectively, and are able to modify meshes within some reasonable ranges. They are not parametric in the sense of providing a simple geometric relation between some pinna landmarks and model parameters.

To this end, we introduce a parametric pinna model (PPM) that can be applied to deform a carefully designed template pinna mesh (see Fig. 1 left) in order to match a potentially noisy and target mesh obtained for an individual listener's pinna (see Fig. 1 right). We show the potential of the PPM by adapting its parameters to a target mesh and evaluating geometric deviations as well as simulation sound-localization predictions.

Note that acquisition of the model parameters is not part of this contribution and can be based on manual retrieval of anthropometric data or automatic generation by means of neural networks.

2. THE MODEL

For the template mesh, we used the high-resolution left-ear pinna mesh of NH5 from [6]. The main idea for the PPM and its deformation was to apply morph target animation and skeletal animation, both methods being state-of-the-art in the field of animation, e.g., for the animation of lip-synchronised movements, i.e., visemes [7]. The deformation type is a rigid registration, i.e., an affine transformation, applied to local regions of the pinna.

Figure 1: Distorted template pinna (left), template pinna (middle), target pinna (right)

Both morph target animation and skeletal animation are implemented in Blender[1] as "bendy bones" and "shape keys". Bendy bones are Beziér-curve representations with control points on each end and construct an armature that

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

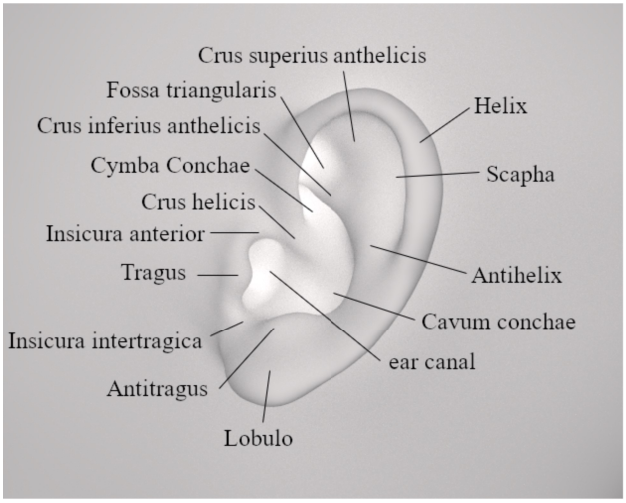

Figure 2: Anatomical structure of a human pinna.

is able to deform vertices around a bendy bone with assignable weights. Bendy bones control the rather coarse deformations of the pinna. Shape keys describe transformations affecting multiple vertices and can deform specific areas in greater detail. In animation, shape keys are used to record the change of vertex positions over time and are a superb tool for modeling organic soft parts, after. After a shape key is defined, one can choose the weight of the shape key itself, not a weight for every vertex. The deformations via bendy bones and via shape keys are affine transformations, thus the process can be represented similar to [8] as

$$\boldsymbol{x} = \boldsymbol{x_0} + \sum_{i=1}^{M} \boldsymbol{V_i}\,\boldsymbol{a_i} + \sum_{j=1}^{N} w_j \boldsymbol{b_j} \qquad (1)$$

where $\boldsymbol{x}$ denotes the deformed target mesh, $\boldsymbol{x_0}$ the base template mesh, $\boldsymbol{a_i}$ the bendy bones, $\boldsymbol{b_j}$ the shape keys, with M and N denoting the amount of bendy bones and shape keys, respectively. The first sum describes the deformation via bendy bones $\boldsymbol{a_i}$ controlled by $\boldsymbol{V_i}$, a matrix modelling the distortion by each control bone in six degrees of freedom, i.e., location and rotation. The second sum describes the deformation via shape keys $\boldsymbol{b_j}$ controlled by w_j, a weight constant ranging from -1 to 1.

Some of the shape keys were assigned to global modifications, i.e., width and length of the pinna; others were used to apply fine tuning in local areas.

As a first step, we allocated pinna vertices to anatomic pinna areas depicted in Fig. 2, e.g., all vertices belonging to the helix were assigned a vertex group called "Helix", all vertices belonging to the Tragus were assigned a vertex group called "Tragus", and so on. The bendy bones were named after the anatomic area they are affecting the most, i.e., Tragus, Antitragus, Antihelix, Crus Inferius Anthelicis, Crus Superius Anthelicis, Lobulo, and the Helix split in three parts (upper Helix, centered Helix, lower Helix, respectively).

However, concave curvatures in the pinna structure, i.e., cavum conchae and cymba conchae, fossa triangularis and to some extent the scapha are crucial areas for HRTFs, meaning that they heavily influence the first two peaks (P1, P2) and first notch (N1) of the HRTFs in the saggital plane [9]. Although the bendy bones heavily influence the prominent inflexions in the pinna, concave surfaces suffer from vertex overlays resulting in unwanted discontinuous sections.

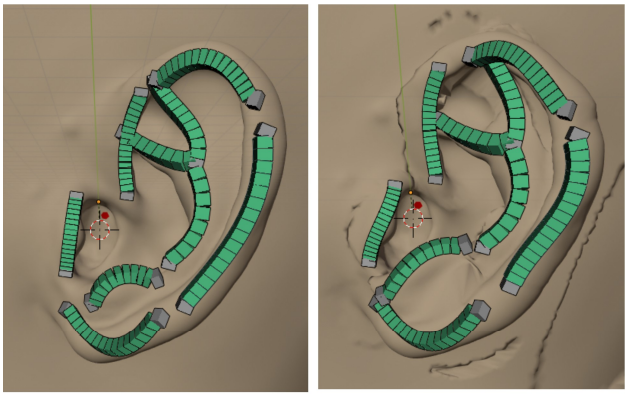

Figure 3: Bendy bones armature (left), distorted to match the target ear (right). Observe the areas of discontinuity due to the weight affecting certain vertices.

Further, we assigned shape keys to said concave surfaces in order to support the influence of the bendy bones in greater detail. They are named after the anatomic area and the dimension they influence, e.g., cavum conchae depth, crus helicis presence, fossa triangularis depth, etc.

In summary, our PPM consists of 17 control bones and 32 shape keys, which result in its overall dimensionality of 134 dimensions. Note that in this representation, the dependency between the shape keys is not considered and we do not consider this dimensionality optimal.

3. EVALUATION

The PPM (with the underlying template mesh) was adapted to the target mesh in several steps, while watching the geometric errors and simulated sound-localization performance in order to study the contribution of various components of the PPM.

The target mesh was that of NH131 from [10]. Note that the target mesh was a clean mesh with the quality similar to the one used for the template mesh, in order to evaluate the PPM's general ability to adapt to a pinna of an other listener.

Methods

For the numerical calculation of the HRTFs, *Mesh2HRTF*[1][4] was used for both distorted template and target mesh. In order to evaluate the geometrical deviation between the adapted mesh and the clean template mesh, Hausdorff distance was applied [11]. It allowed us to determine areas with a large geometrical error, needing more precise adaptation.

For the prediction of sound-localization performance, we used the sagittal-plane localization model [12], implemented in the *Auditory Modelling Toolbox* (AMT)[2]. This model predicts localization performance by means of a quadrant error (QE) and a polar error (PE). The QE rate (in percent, lower is better) describes a listener's rate in confusing the sound-source quadrant by means of localizing the sound source with an error of more than

[1]http://mesh2hrtf.sourceforge.net/

[2]http://amtoolbox.sourceforge.net/

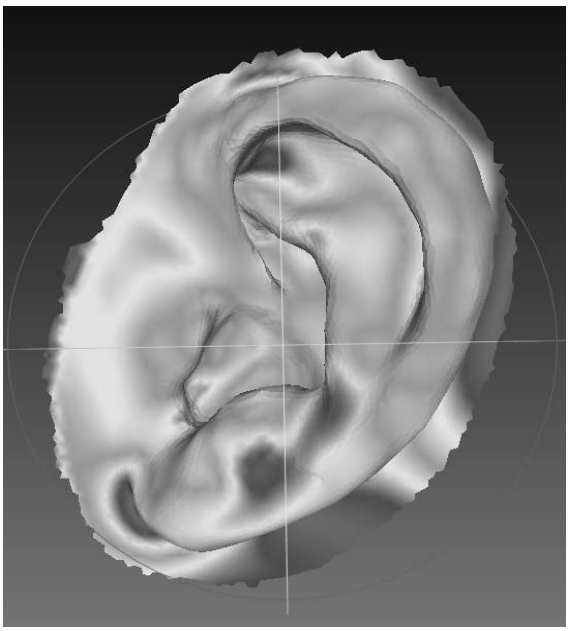

Figure 4: Calculated Hausdorff distance showing areas with geometric error (greyscale colormap, i.e. small errors are decoded in lighter grey, the darker the shade the bigger the error. Note that in order to limit the distance maximum to a minimum, only the area around the pinna has been considered.

90°. The QE rate represents the rate of confusing front from back, top from bottom, etc. The polar error describes the error when localizing the sound source within the correct quadrant. Thus, PE represents the local precision of the localization (in degrees, lower is better).

Adaptation

In a first step, the coarse form of the target ear was manually approximated with the armature (see Fig. 3). This was achieved by a linear transformation of the controls of the bendy bones (depicted as grey cubes in Fig. 3). Additionally, global shape keys were used to deform the pinna width and length.

In a second step, the shape keys were used to control more subtle adaptations as shape key impact smaller vertex groups, and thus enable deformations in greater detail.

After the shape-key modifications, the mesh was cleaned up, i.e., small surface displacements and artefacts from insufficiently precise chosen weights (see Fig. 3) were smoothed out.

Finally, the vertices representing the microphones defined, the distorted mesh was arranged on the mesh representing the head and the composite mesh was arranged for further smoothing. The last step is necessary not only for the local pinna areas that were distorted, but for the whole 3D model of the head with pinnae because any distortion originating from imprecise weight painting may affect other regions such as the back of the head.

Results

Figure 5 shows the QEs and PEs for the various degrees of adaptation:

- "None (NH131 vs NH5)" shows the result of no adaptation, i.e., when NH131 would listen with the template ears of NH5. This result is within the range usually found for localizing sounds with generic HRTFs, [13], and shows that when localizing with NH5's ears, NH131 would have have a poor predicted localization ability.
- "Registration model 1, 2D ref" shows the result of manually adapting in visual inspection process the 3D template according to a 2D reference of a 3D model. After this first registration with a 2D reference, i.e., having both 3D models side-by-side and distort the template via observing the target, it became clear to bring both 3D models within the same space.
- Registration model 1 shows the result of bringing the 3D target mesh into the same space as the template, and adapting the template mesh according to the overlayed 3D reference. The overall performance of NH131 did not improve, thus the Hausdorff distance was applied in order to determine areas with a large geometrical error. In figure 4, darker grey areas describe these areas, one of which is important for saggital plane localization, i.e., the fossa triangularis. With this information, the rotation of the bendy bone Crus Inferius Anthelicis has been adapted and new shape keys have been introduced to cover that area in particular.
- Registration model 2 shows the adapted version with 3D reference. It basically is a better version of the first model, because of the newly introduced shape keys to improve modelling concave surfaces. The adapted model 2 delivers a far better localization result, see figure 5.

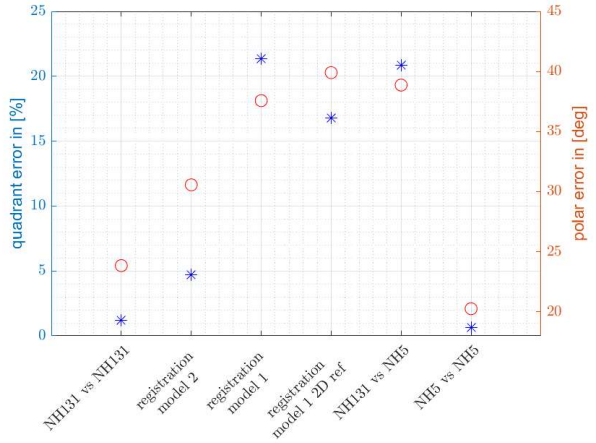

Figure 5: Comparison of quadrant and polar errors of template mesh of NH131 vs several stages of deformation. Far left data depicts the localization errors of the target listener, far right data depicts the localization errors of the template listener with their own HRTF set, respectively. Starting from the second from right, the localization errors of a simulated listener experiment of the target listener with the HRTFs calculated from the manual deformation of the template pinna are shown. The localization errors at different stages of the deformation are plotted on the x-axis from right to left.

For the reference, we also show the predicted listener's performance for the localization with their actual ears, see "NH131 vs NH131" and "NH5 vs NH5". These results are within the usual ranges of sound localization

with listener-specific HRTFs, [13], and show that the predictions are reasonable.

4. CONCLUSIONS

In this study, we propose a PPM consisting of a 3D model of a template pinna and a rig consisting of bendy bones in combination with morph target animation. The PPM is aimed to apply a manual non-rigid registration approach via deformation and affine transformations.

We show that the complex biological structure of a human pinna can be described with a few parameters. However, there is still room for improvements. While we show that the distorted PPM yielded a pinna offering localization performance close to that of listener-specific HRTFs, the accuracy of the parametrization was not good enough in order to obtain *the same* predicted localization performance as that with listener-specific HRTFs.

Still, the proposed PPM can be seen as a starting point for further developments towards an automated process of pinna registration. For example, one may consider preprocessing a template mesh with the proposed PPM before applying it to register target meshes with non-parametric algorithms.

5. REFERENCES

[1] J. Blauert, *Spatial hearing. The Psychophysics of Human Sound Localization*, Revised edition. Cambridge, MA: The MIT Press, 1997.

[2] H. Møller, M. F. Sørensen, D. Hammershøi, and C. B. Jensen, 'Head-related transfer functions of human subjects', *J Audio Eng Soc*, vol. 43, pp. 300–321, May 1995.

[3] P. Majdak, P. Balazs, and B. Laback, 'Multiple exponential sweep method for fast measurement of head-related transfer functions', *J Audio Eng Soc*, vol. 55, pp. 623–637, 2007.

[4] H. Ziegelwanger, W. Kreuzer, and P. Majdak, 'Mesh2HRTF: Open-source software package for the numerical calculation of head-related transfer functions', in *Proceedings of the 22nd International Congress on Sound and Vibration*, Florence, IT, Jul. 2015, pp. 1–8, doi: 10.13140/RG.2.1.1707.1128.

[5] C. Guezenoc and R. Seguier, *A Wide Dataset of Ear Shapes and Pinna-Related Transfer Functions Generated by Random Ear Drawings*. 2020.

[6] H. Ziegelwanger, A. Reichinger, and P. Majdak, 'Calculation of listener-specific head-related transfer functions: Effect of mesh quality', in *Proceedings of Meetings on Acoustics*, Montreal, Canada, 2013, vol. 19, p. 050017, doi: 10.1121/1.4799868.

[7] Fisher Cletus G., 'Confusions Among Visually Perceived Consonants', *J. Speech Hear. Res.*, vol. 11, no. 4, pp. 796–804, Dec. 1968, doi: 10/ghbhss.

[8] V. Barrielle, 'Leveraging Blendshapes for Realtime Physics-Based Facial Animation', Université Bretagne Loire, 2017.

[9] H. Takemoto, P. Mokhtari, H. Kato, R. Nishimura, and K. Iida, 'Mechanism for generating peaks and notches of head-related transfer functions in the median plane', *J Acoust Soc Am*, vol. 132, no. 6, pp. 3832–41, Dec. 2012.

[10] H. Ziegelwanger, P. Majdak, and W. Kreuzer, 'Numerical calculation of listener-specific head-related transfer functions and sound localization: Microphone model and mesh discretization', *J. Acoust. Soc. Am.*, vol. 138, no. 1, pp. 208–222, Jul. 2015, doi: 10.1121/1.4922518.

[11] M. Gromov, *Metric Structures for Riemannian and Non-Riemannian Spaces*. Birkhäuser Basel, 2007.

[12] R. Baumgartner, P. Majdak, and B. Laback, 'Modeling sound-source localization in sagittal planes for human listeners', *J. Acoust. Soc. Am.*, vol. 136, no. 2, pp. 791–802, Aug. 2014, doi: 10.1121/1.4887447.

[13] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, 'Localization using nonindividualized head-related transfer functions', *J Acoust Soc Am*, vol. 94, no. 1, pp. 111–23, Jul. 1993.