TOWARD A NOVEL SET OF PINNA ANTHROPOMETRIC FEATURES FOR INDIVIDUALIZING HEAD-RELATED TRANSFER FUNCTIONS

Davide FANTINI (davide.fantini@unimi.it) (0000-0003-1332-0890),
Stavros NTALAMPIRAS (stavros.ntalampiras@unimi.it) (0000-0003-3482-9215),
Giorgio PRESTI (giorgio.presti@unimi.it) (0000-0001-7643-9915), and
Federico AVANZINI (federico.avanzini@unimi.it) (0000-0002-1257-5878)

Laboratory of Music Informatics (LIM), Department of Computer Science, **University of Milan**, Italy

ABSTRACT

Head-related transfer functions (HRTFs) are essential for spatial audio rendering. However, to provide a proper individual experience, the HRTF should be compliant with the anatomical traits of the listener. Therefore, estimating the individual HRTF based on anthropometric features is a widespread and promising approach. In this work, we propose an extension of the pinna anthropometric parameters commonly used in the related literature. The objective is to provide a more detailed description of the pinna for the purpose of individualizing HRTFs. The entirety of these parameters was extracted using a set of landmarks that were manually annotated on the shape of the pinna. The proposed parameters were compared to the classical ones for selecting the best-match HRTF in the median plane via anthropometry matching. Prior to HRTF selection, we devised a procedure to retain only the relevant anthropometric parameters. The obtained results were evaluated using both objective metrics and auditory model evaluation indicating a significant, though limited, improvement when using the proposed parameters.

1. INTRODUCTION

The direction-dependent effect of human body on sound waves is interpreted by a human listener as spatial cues. Head-related transfer functions (HRTFs) [1] describe this effect as a linear time-invariant (LTI) system. An HRTF set is a collection of transfer functions, one for each source position of interest and for each ear. HRTF sets encode the cues for spatial hearing, such as interaural time and level differences (ITD, ILD) and monaural spectral modifications. Each person has an individual HRTF as these cues depend on the body anatomy. In this respect, the pinna is of particular interest as it is responsible for key monaural spectral cues. The pinna influence on HRTF can be isolated by extracting the pinna-related transfer function (PRTF).

HRTFs find application in multiple fields [2] and are essential in extended reality (XR) frameworks where spatial audio is involved. However, to provide an immersive auditory experience, HRTFs should be compliant with the user's anatomical traits. In end-user applications, a non-individual HRTF (e.g., recorded from a dummy head) is often employed, potentially causing inadequacies such as front-back and up-down confusion, and degradation in elevation localization [3–5]. On the other hand, the HRTF acoustic measurement entails high-cost equipment and time-consuming sessions. For these reasons, the literature includes several works on HRTF individualization, i.e., the estimation of individual HRTF without the need for acoustic measurement. Several approaches for HRTF individualization are based on anthropometric parameters [2, 6, 7]. The CIPIC anthropometric specification [8] is the prevalent one, despite evidence of its incompleteness [9]. Some variations to the CIPIC pinna parameters have been proposed in the literature. However, few attempts have been made to evaluate the relevance of the proposed variations.

In this work, we propose some new pinna parameters in addition to those used in the HUTUBS dataset [10], which are derived from CIPIC. The new parameters are intended to describe in more detail the pinna anatomy for the purpose of individualizing PRTF. In this regard, we especially focused on the concha and fossa triangularis given their influence on HRTF [9, 11], and we considered measurement types different from the classical distances and angles. The effectiveness of new parameters was evaluated for selecting the best-match PRTF via anthropometry matching, in comparison to the sole HUTUBS set. Prior to PRTF selection, an anthropometric selection step was performed to retain only the relevant parameters. The PRTF selection results were evaluated using both spectral distortion (SD) and localization metrics obtained with an auditory model.

The Matlab code to extract the pinna anthropometric parameters described in this paper is publicly available [1]. Also, the data generated in our research as well as a detailed documentation of the pinna parameters are included in supplementary materials [2].

2. BACKGROUND

HRTF individualization is often based on anthropometric parameters [2, 6, 7] due to the relationship between HRTF and body anatomy. These parameters are usually defined

Copyright: © 2024. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

[1] https://github.com/DavideFantini/pinna-anthropometry-extraction
[2] https://doi.org/10.5281/zenodo.10805884

as distances between specific body points, although angles are also used. Several anthropometric specifications have been proposed in the HRTF field, usually defined as two-dimensional sketches [1, Ch. 7]. Early anthropometric specifications include the one used for the KEMAR mannequin's design [12] and the one proposed by Middlebrooks [13]. In 2001, Algazi et al. [8] proposed the CIPIC specification, which includes ten parameters for the pinna and 17 for the head and torso. This specification is still widely used in HRTF datasets and HRTF individualization, although some variations have been proposed. For example, in the HUTUBS dataset, CIPIC parameters, with some variations in their definitions, were reported along with two new pinna parameters for the cavum concha. In the CHEDAR dataset [14], two pinna parameters were proposed for the crus helicis and the parabola effect of the pinna, in addition to five CIPIC parameters. Some of the anthropometric parameters defined in the standard GB/T 2428-1998 [15] were used for HRTF individualization [16] and HRTF datasets [17] in addition to the CIPIC ones. Recently, Iida et al. [18] measured ten distances between the tragus and other points on the pinna, while Stitt and Katz [9] defined a set of control points to measure CIPIC pinna parameters and other new ones. Teng and Zhong [19] proposed five area parameters of different pinna parts. However, the authors did not provide clear definitions for them. Despite the proposed CIPIC variations, currently, there is no complete and inter-independent anthropometric specification for HRTF.

The pinna anatomy is of particular interest in the HRTF field. Several studies have demonstrated its influence on vertical localization and front-back disambiguation [11, 20–23]. The role of pinna cavities has been analyzed with respect to the HRTF magnitude spectrum [20] and the localization in the median plane [23], while some studies focused on anthropometry [24, 25]. Literature findings suggest the influence of concha and triangular fossa on HRTF. In particular, Takemoto et al. [11] demonstrated that the nodes and anti-nodes occurring in these two pinna cavities are related to the notches of the median plane PRTF. Stitt and Katz [9] confirmed these findings by varying anthropometry to deform a 3D pinna model and comparing the corresponding HRTFs obtained via numerical simulation. The authors found a significant impact of concha and fossa triangularis parameters. Such findings suggest the incompleteness of widely used anthropometric specifications, such as CIPIC. For instance, parameters for the fossa triangularis are rare in these specifications.

3. PINNA PARAMETERS

3.1 Landmarks annotation scheme

In the following, we refer to the horizontal and vertical dimensions of depth images as x and y coordinates, respectively, while the z coordinate refers to depth. The proposed pinna parameters are extracted using a set of landmarks that are manually annotated on pinna depth images. Fig. 1 shows the scheme followed to annotate the landmarks, which comprises a set of points placed on the pinna to describe its shape. This scheme is a revised version of the one previously proposed by us [26]. The scheme includes $K = 205$ landmarks, providing a fine resolution of the pinna shape. This scheme divides the pinna shape into four parts: outer helix, concha, fossa triangularis, and inner helix. Fig. 1 shows the starting and ending landmarks indices for each part suggesting the landmarks annotation order. The first 49 landmarks (blue in Fig. 1) describe the outer helix edge starting from the attachment of the helix to the head (superior part of the pinna) and following the helix downwards to the inferior lobulus extremity. Other 68 landmarks (purple in Fig. 1) describe the concha starting from the inferior part of the crus helicis attachment to the cavum conchae, and following the entire concha edge ending under the inner helix edge. Other 18 landmarks (orange in Fig. 1) describe the fossa triangularis from the attachment of the antihelix (crus antihelicis superior) below the helix, to the point where the fossa triangularis edge reaches the border with the cymba conchae. The last 70 landmarks (yellow in Fig. 1) describe the inner helix, starting from the superior edge of the crus helicis attachment to the cymba conchae, and following the helix upward to its superior extremity, then continuing downward.

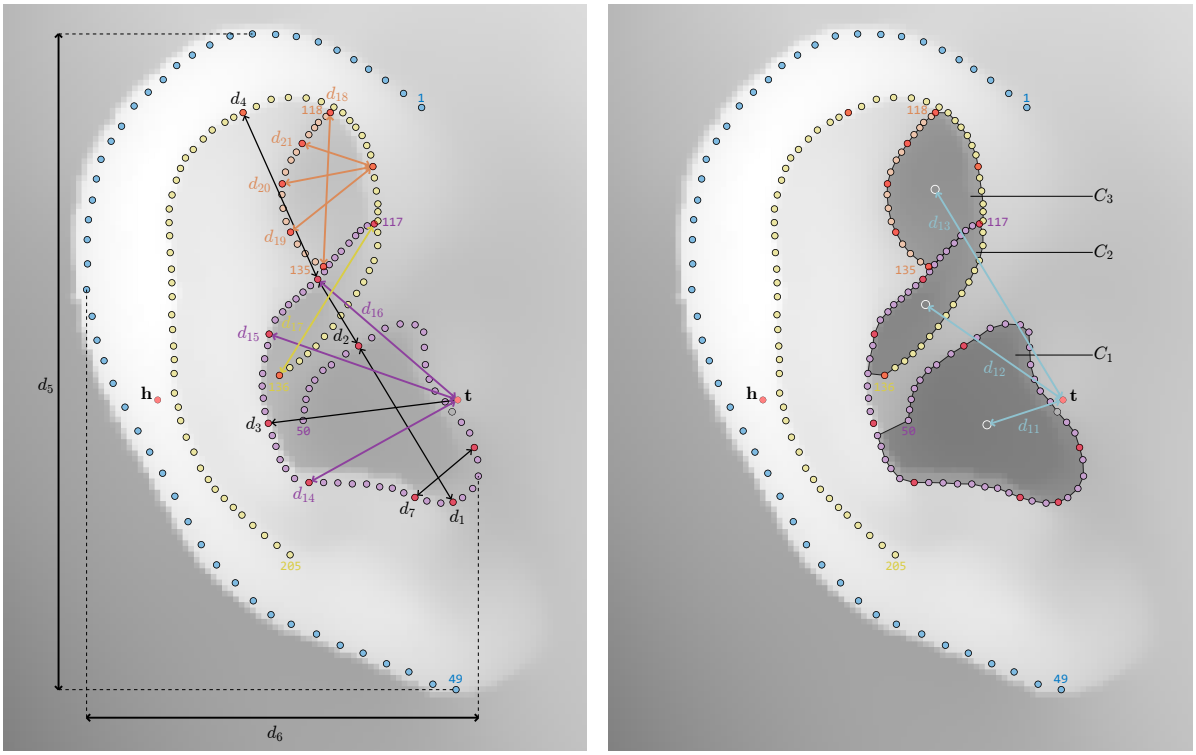
Landmarks are evenly spaced and placed in coherent positions across the pinnae. To ensure this coherence during annotation, we adjusted the landmark of a template shape superimposed on the image to match the current pinna. The landmarks serve to measure anthropometric parameters, therefore only the main contours of the pinna shape involved in the measurements are covered. Landmarks are placed near the border of the pinna contours, but not on the exact edge, to minimize the risk of having a landmark with a significantly incorrect z value. For instance, a slight misplacement of a landmark on the outer helix can result in the landmark being positioned on the head.

3.2 Pinna anthropometry measurement

We developed automated procedures to extract a set $\mathcal{P}$ of 48 pinna anthropometric parameters using the annotated landmarks and pinna depth images. For the sake of conciseness, we provide here a brief, though complete, overview of the parameters. For details on their definitions and measurements, please refer to the anthropometry documentation provided as supplementary material [2].

Most of the parameters are measured using the 2D landmarks (x and y coordinates), while the remaining ones consider the depth information (z coordinate) as well. A subset of $\mathcal{P}$ replicates the pinna parameters defined in the HUTUBS specification. Additionally, we defined new parameters to thoroughly describe the pinna cavities relevant for HRTF, such as the fossa triangularis. We have also explored alternative measurements to the traditional distances and angles, including areas, volumes, and depths. Fig. 1a shows some of the HUTUBS and the proposed pinna distance parameters, while Fig. 1b shows the pinna cavities and their respective distances.

The position $\mathbf{t}$ of the tip of the tragus is used to measure some parameters. Therefore, we estimated $\mathbf{t}$ by selecting the point with the maximum z value in a small area around



(a) (b)

Figure 1. Example of the scheme for manually annotating the pinna shape. Landmarks are color-coded based on the pinna shape part: outer helix edge (blue), concha (purple), fossa triangularis (orange), and inner helix edge (yellow). Numbers on the first and last landmarks of each part suggest the annotation order. The grayscale colormap represents the depth (z) values. (a) Comparison between some of the HUTUBS (black arrows) and the proposed distance parameters represented by arrows (colored as the pinna shape part) between red landmarks. (b) Pinna cavities C_i (shaded areas) along with the distance parameters (light blue arrows) between the tragus $\mathbf{t}$ and the cavities centroids (white empty circles).

a predefined landmark (gray landmark in Fig. 1). Further, we estimated the helix position $\mathbf{h}$ by identifying the point with the maximum z value at the same y coordinate as $\mathbf{t}$.

3.2.1 *HUTUBS distance parameters*

In this paper, we considered the HUTUBS distance parameters d_h with $h = \{1, \ldots, 9\}$ (we neglected d_{10} because we were unable to replicate its measurement). The parameters $d_{1\text{–}4}$, and d_7 are defined as distances between specific pinna points. To measure them, we selected the pairs of landmarks that best approximate the points indicated in the HUTUBS definition (see Fig. 1a) and computed their Euclidean distance in x and y coordinates. The parameters d_5 and d_6, i.e., the pinna height and width, were measured as the landmarks range along the y and x coordinates, respectively. This slightly differs from the definition of d_6 in HUTUBS, but it allowed us to obtain a more robust measurement. An ad hoc procedure was developed for d_8 and d_9. To measure d_9, a horizontal slice of the pinna depth image is taken at the y coordinate of $\mathbf{t}$. Then, d_9 is set as the distance in x and z coordinates between $\mathbf{t}$ and the corner formed by the concha surface and the antihelix wall. The parameter d_8 is measured as the 3D distance between $\mathbf{t}$ and the deepest point in the incisura intertragica, i.e., the region enclosed by tragus and antitragus.

Table 1 shows the absolute difference between the anthropometric values $\mathbf{A}^h$ measured by us and those reported in the HUTUBS dataset. It should be noted that our goal was not to exactly replicate the HUTUBS parameters. However, the obtained errors can be considered sufficient and comparable to those reported in similar studies [26–28].

3.2.2 *Additional distance parameters*

In addition to the HUTUBS parameters, we propose eight new parameters $d_{14\text{–}21}$ (see Fig. 1a) defined as distances in x and y coordinates between specific pairs of landmarks. The parameters $d_{14\text{–}16}$ integrate the HUTUBS parameters d_1, d_3 and d_7 in describing the concha. They are defined as distances between $\mathbf{t}$ and characteristic points on the concha edge. The parameter d_{17} quantifies the length of the cymba conchae and is therefore complementary to d_2. The parameters $d_{18\text{–}21}$ characterize the shape of the fossa triangularis.

3.2.3 *Angle parameters*

In addition to distances, HUTUBS provides two angle parameters: the pinna rotation angle θ_1 and the pinna flare angle θ_2. We measured θ_1 as the angle between the y axis and the line passing through the pinna extremities in y coordinates represented by the same landmarks used to measure d_5. Then, θ_2 was measured as the angle in x and z coordinates between the x axis and the line passing through $\mathbf{t}$ and $\mathbf{h}$. In addition to HUTUBS angles, we

Parameter	Mean ± St. dev. (%)
d_1 [mm]	1.40 ± 1.03 (7.8)
d_2 [mm]	1.65 ± 1.15 (16.4)
d_3 [mm]	1.31 ± 0.99 (7.6)
d_4 [mm]	4.89 ± 2.23 (23.6)
d_5 [mm]	1.83 ± 1.45 (3.0)
d_6 [mm]	3.72 ± 2.59 (12.6)
d_7 [mm]	0.82 ± 0.69 (13.2)
d_8 [mm]	2.89 ± 1.69 (25.4)
d_9 [mm]	2.08 ± 1.38 (17.4)
θ_1 [°]	1.98 ± 1.37 (17.5)
θ_2 [°]	6.15 ± 3.87 (23.9)

Table 1. Mean and standard deviation of the absolute differences between the anthropometric values $\mathbf{A}^h$ measured by us and the values reported in the HUTUBS dataset. The mean error as a percentage of the HUTUBS parameter mean is reported between brackets.

defined the *pinna upper roll* θ_3 as the angle in y and z coordinates between $\mathbf{t}$ and the landmark with the maximum y value. Also, for each distance parameter d_n with $n = \{1, \dots, 21\} \setminus \{5, 6, 8, 9\}$, we defined the corresponding angle r_n as the angle between the y axis and the line passing through the two measurement landmarks of d_n. These angle parameters quantify the relative alignment of each pair of pinna points in addition to their distance d_n.

3.2.4 Pinna cavities parameters

In the pinna shape, we defined the three regions C_1, C_2 and C_3 corresponding to the pinna cavities cavum conchae, cymba conchae and fossa triangularis, respectively. These regions can be easily delimited by selecting specific landmarks as their position on the pinna shape is known. We defined three additional parameters, namely d_{11}, d_{12} and d_{13}, as the distances in x and y coordinates between $\mathbf{t}$ and the centroids of C_1, C_2 and C_3, respectively. These parameters measure the relative position of the cavities in the pinna shape, which is a complementary information to the shape and size considered by other parameters. Fig. 1b shows the regions $C_{11\text{–}13}$ and the parameters $d_{11\text{–}13}$.

Further, the extracted parameters include the area a_i of each region C_i and the volume v_i of the corresponding polyhedron Q_i. To obtain Q_i, we joined the vertices of C_i in 3D coordinates with the depth image pixels enclosed by C_i and computed the 3D convex hull. Also, we measured the depth h_i of each pinna cavity C_i as the range in the z coordinate of the Q_i vertices.

4. PRTF SELECTION

This section describes an HRTF individualization approach based on the anthropometric values $\mathbf{A}$ extracted as discussed previously. This approach includes the preprocessing of $\mathbf{A}$ and HRTFs, the anthropometric parameters selection, and the PRTF selection via anthropometric matching.

4.1 Preprocessing

4.1.1 Pinna parameters

The anthropometric values $\mathbf{A}$ were preprocessed by centering each parameter to have a median of 0, and scaling to have an interquartile range of 1. This normalization technique ensures that the values of distinct parameters fall within a similar range, making their comparison unbiased by their original scales.

4.1.2 HRTF

The preprocessing of HRTFs aims to isolate the influence of the pinna, which is the focus of our work. First, we extracted the PRTF by applying a 1 ms Hann window to the HRIR starting from its onset [22]. The window eliminates the torso and shoulders effects while retaining the pinna one. Next, we extracted the directional transfer function (DTF) magnitude from the PRTFs [29]. The analysis was restricted to the 3–15 kHz frequency range, as the pinna influence above this range is marginal, and below it, the torso and shoulders effects are prevalent [1, 20–22, 25].

Further, the analysis was limited to the median plane, as in related works that focus on the pinna [26, 30–32]. The median plane can be considered a case study for the monaural spectral cues caused by the pinna, which are crucial for elevation perception [11, 21–23]. The findings on the median plane can be extended to other vertical planes by modeling the head and torso with other approaches [33, 34]. In the median plane, we considered the elevations between $-80°$ and $260°$ and discarded the angle $-90°$ (below the subject).

4.2 Anthropometric parameters selection

A selection of the parameters in the set $\mathcal{P}$ is conducted to keep only those relevant to PRTF. Candidates subsets $\mathcal{P}' \subseteq \mathcal{P}$ are compared to find the best set $\mathcal{P}^*$. This comparison is based on spectral distortion (SD). The SD between two PRTFs $\mathbf{H}_1$ and $\mathbf{H}_2$ at elevation φ is so defined [1, Sec. 5.1]:

$$SD^{\varphi}(\mathbf{H}_1, \mathbf{H}_2) = \sqrt{\frac{1}{|F|} \sum_{f \in F} \left(20 \log \frac{|\mathbf{H}_1(f, \varphi)|}{|\mathbf{H}_2(f, \varphi)|} \right)^2} [dB], \tag{1}$$

where F is the set of considered frequency bins. The mean $\bar{SD}$ is obtained by averaging SD^{φ} values of all elevations φ in the considered set Φ. For each pinna t, we computed the distance D between its anthropometry $\mathbf{A}_t$ with the anthropometry $\mathbf{A}_{t'}$ of every other pinna t' in the dataset:

$$D(t, t') = \sqrt{\sum_{p \in \mathcal{P}'} (\mathbf{A}_t^p - \mathbf{A}_{t'}^p)^2}. \tag{2}$$

The pinna t^* minimizing D is selected as the best-match given the candidate set $\mathcal{P}'$. We prevented the right pinna of a subject from being selected as best-match for the left pinna of the same subject, and vice versa. Then we computed the mean of the $\bar{SD}$ values obtained for each pair of PRTFs $\mathbf{H}_t$ and the corresponding best-match $\mathbf{H}_{t^*}$.

The candidate sets of parameters $\mathcal{P}'$ was picked as follows. Starting from $|\mathcal{P}|$ parameters, we iteratively decreased the number m of considered parameters. For each iteration, we evaluate all the combinations of m parameters taken from $\mathcal{P}^*_{m+1}$, i.e., the best set found in the previous iteration. The best set $\mathcal{P}^*_m$ with m parameters is selected as the one minimizing $\bar{SD}$. The search stops if there is no combination of m parameters reducing SD compared to the previous iteration. This procedure permits to evaluate fewer candidates $\mathcal{P}'$ than all possible $2^{|\mathcal{P}|} - 1$ combinations, which is computationally intensive.

4.3 PRTF selection procedure

Selection-based methods for HRTF individualization involve choosing the so-called best-matching HRTF based on the similarity between the test subject and those in a dataset. Similarity can be computed in various ways. In this work, we used a baseline approach commonly found in the literature, i.e., anthropometry matching [16, 28, 33, 35]. For each test pinna t, we computed the Euclidean distance $D(t, t')$ (see Eq. (2)) between its anthropometry $\mathbf{A}_t$ and the anthropometry $\mathbf{A}_{t'}$ of every other pinna t' in the dataset considering the best set of parameters $\mathcal{P}^*$. The PRTF of pinna t^* that minimizes D is selected as the best matched. The selection procedure considers the two pinnae of the subject as separate instances. Therefore, the PRTFs of two different subjects can be selected for the left and right pinnae of the same test subject.

5. RESULTS

5.1 Validation procedure

The PRTF selection approach was evaluated using the HUTUBS dataset, which includes 3D head meshes, anthropometry, and HRTFs measured for 58 subjects. We excluded the FABIAN dummy head and repeated measurements of the same subject, resulting in $N = 55$ subjects for validation. We extracted the pinna depth images from the HUTUBS 3D head meshes. For each pinna, we projected the 3D point cloud from the head lateral view onto a grid representing the depth image. This operation is straightforward as the head meshes are coherently placed in the coordinate system. The extracted depth images have a resolution of 140×160 pixels.

We used leave-one-subject-out cross-validation (LOOCV) to evaluate the PRTF selection procedure. Thus, we performed N splits of the dataset into training and test sets, even if each pinna is a separate instance. In the n-th split, both pinnae of subject n are in the test set, while the remaining subjects are in the training set. This prevented to select the left pinna in the training set for the right pinna of the same subject in the test set, and vice versa. To avoid biases, we performed all statistical measurements of the preprocessing and parameters selection procedures on the training set only. The outcomes were then applied to both the training and test sets.

For each pinna, we obtained PRTFs with $|\Phi| = 33$ elevations and $|F| = 71$ frequency bins by preprocessing the HRTFs from HUTUBS (see Sec. 4.1). In the evaluation, we compared the individual PRTFs $\mathbf{H}$ with two conditions $\hat{\mathbf{H}}^h$ and $\hat{\mathbf{H}}$. The PRTFs $\hat{\mathbf{H}}^h$ are selected using the anthropometry $\mathbf{A}^h$ including the eleven HUTUBS pinna parameters $d_{1\text{–}9}$, θ_1 and θ_2. The PRTFs $\hat{\mathbf{H}}$ are selected using the anthropometry $\mathbf{A}$ including all 48 parameters discussed in this paper. For a fair comparison, we relied exclusively on the anthropometry extracted by us as described in Sec. 3.2, and we ignored the values reported in HUTUBS. Further, we used the left ear PRTF $\mathbf{H}^f$ measured from the FABIAN dummy head as a baseline condition.

5.2 Anthropometric parameters selection

Averaging the LOOCV results of anthropometric parameters selection, for the conditions $\hat{\mathbf{H}}^h$ and $\hat{\mathbf{H}}$ we selected 8.7 and 43.6 parameters with a minimum $\bar{SD}$ of 5.44 and 5.32 dB, respectively. Fig. 2 shows the selection frequency of each parameter for the PRTFs $\hat{\mathbf{H}}$ as a percentage. Several parameters have been selected at every LOOCV iteration, such as d_7, d_{12}, d_{13}, d_{17}, θ_3, a_3, h_1, r_2, r_4, and $r_{19\text{–}21}$. Conversely, d_1 is quite often discarded. Other less frequently selected parameters are a_2, r_1, r_3, and r_{14}. There is no noticeable effect of the type of measurement on the selection frequency.

5.3 Objective evaluation

We computed the SD (see Eq. (1)) with the individual PRTFs $\mathbf{H}$ to objectively evaluate the two conditions $\hat{\mathbf{H}}^h$ and $\hat{\mathbf{H}}$. Averaging the LOOCV results, we obtained the mean SD values 5.56 ± 1.93 dB and 5.34 ± 1.81 dB for $\hat{\mathbf{H}}^h$ and $\hat{\mathbf{H}}$, respectively. As a reference, we also computed the mean SD between $\mathbf{H}$ and the FABIAN PRTF $\mathbf{H}^f$ which is 5.64 ± 2.7 dB. Despite the obtained SD means are close, statistical analysis [3] demonstrated that SD for $\hat{\mathbf{H}}$ is significantly improved compared to $\hat{\mathbf{H}}^h$ and $\mathbf{H}^f$ ($p < 0.001$), while there is no significant difference between $\hat{\mathbf{H}}^h$ and $\mathbf{H}^f$. This result suggests that the proposed parameters provide a significant, though limited, improvement of SD compared to the sole HUTUBS ones. To quantify the extent of this improvement, we computed Cohen's d effect size between $\hat{\mathbf{H}}$ and $\hat{\mathbf{H}}^h$. The obtained value of 0.35 is considered an effect size between small and medium [36].

Fig. 3 shows the elevation-dependent mean SD values for $\hat{\mathbf{H}}^h$ and $\hat{\mathbf{H}}$. The lowest SD values around 3 dB are found above the head, while SD increases up to 9 dB for lower elevations. However, $\hat{\mathbf{H}}$ seems to provide the greatest improvement for these elevations, especially on the back of the head. Table 2 shows the mean and standard deviation values of the SD for $\hat{\mathbf{H}}^h$ and $\hat{\mathbf{H}}$ in different frequency bands. In general, mean SD progressively increases from circa 1.6 dB to more than 6 dB as the frequency increases. Similarly, the standard deviation is larger for higher frequency bands. The mean SD of $\hat{\mathbf{H}}$ is lower for each frequency band with the larger differences found between 8 and 12 kHz, though below 0.5 dB.

[3] Distributions normality hypothesis was rejected by Shapiro-Wilk test. Significant difference between the conditions was found through non-parametric Friedman test. Paired Wilcoxon signed-rank tests were performed to assess the significance of the differences between pairs of conditions. The obtained p-values were adjusted with Holm-Bonferroni correction for the multiple comparisons.

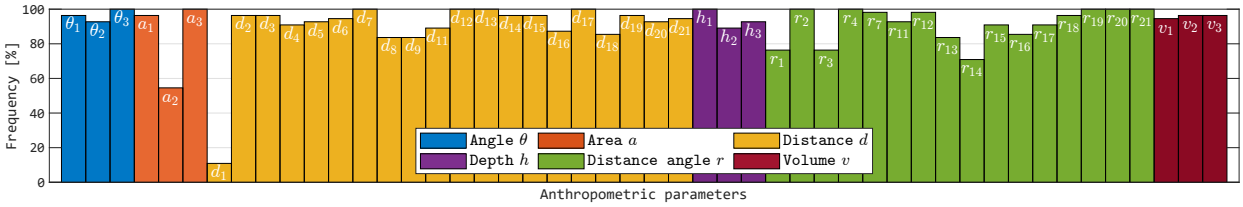


Figure 2. Selection frequency of each parameter obtained in the LOOCV for the condition $\hat{\mathbf{H}}$.

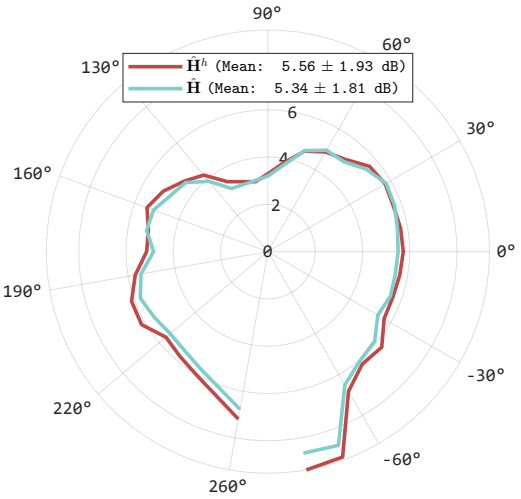


Figure 3. Elevation-dependant mean SD values for the PRTFs $\hat{\mathbf{H}}^h$ selected using HUTUBS anthropometry $\mathbf{A}^h$ (red), and $\hat{\mathbf{H}}$ selected using all parameters $\mathbf{A}$ (light blue).

	Frequency band [kHz]					
	3–4	4–6	6–8	8–10	10–12	12–15
$\hat{\mathbf{H}}^h$	1.64 (1.95)	2.57 (2.36)	4.28 (3.39)	5.53 (3.69)	5.88 (3.47)	6.39 (3.64)
$\hat{\mathbf{H}}$	1.61 (2.09)	2.51 (2.23)	4.09 (3.32)	5.07 (3.51)	5.57 (3.31)	6.21 (3.52)

Table 2. Mean and standard deviation (between brackets) of SD in dB for each frequency band compared for the PRTFs $\hat{\mathbf{H}}^h$ selected using HUTUBS anthropometry $\mathbf{A}^h$ and $\hat{\mathbf{H}}$ selected using all anthropometry $\mathbf{A}$.

5.4 Auditory model evaluation

SD only evaluates the spectral differences between PRTFs and is unsuitable for measuring localization accuracy. Therefore, we compared the selected PRTFs in a virtual localization experiment using the computational auditory model proposed by Baumgartner et al. [37] provided in the auditory modeling toolbox (AMT) [38]. This model estimates the localization performances in the median plane with a template DTF compared to the individual DTF of a subject (target). Several metrics can be computed given actual angles and the subject's responses, and we selected the local polar RMS error (PE) and the quadrant error rate (QE) [39]. PE is the RMS average in degrees of the polar errors with less than 90° in magnitude. QE is the percentage of polar errors higher than 90° in magnitude. We set the frequency range and elevation angles considered by the auditory model according to the ranges defined in the preprocessing step (see Sec. 4.1.2). The sensitivity of the auditory model was set for each subject using the calibration function of Baumgartner's model. We selected the baseline PE and QE values of 25° and 5%, respectively.

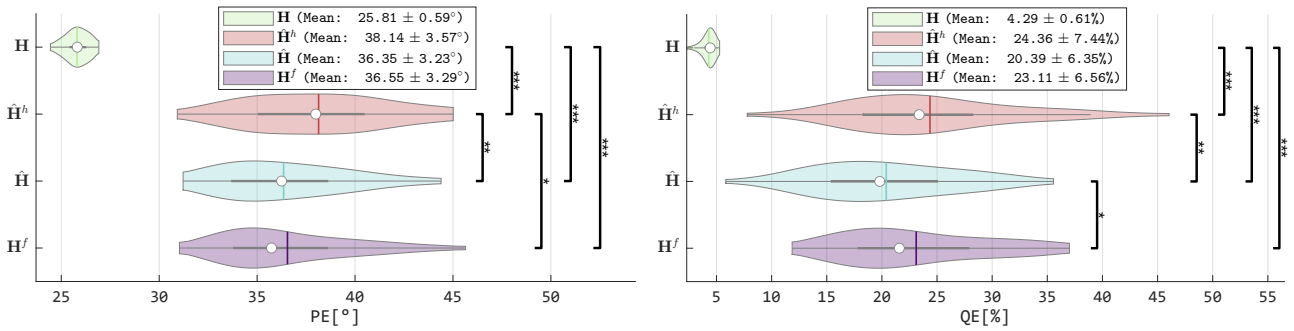
PE and QE values were obtained using the individual PRTF $\mathbf{H}$ as target and other four conditions as templates: the same PRTF $\mathbf{H}$, the selected PRTFs $\hat{\mathbf{H}}^h$ and $\hat{\mathbf{H}}$, and the FABIAN PRTF $\mathbf{H}^f$. The distributions of PE and QE for each condition are shown in Fig. 4. A similar statistical analysis to the aforementioned one for SD was conducted [3]. As expected, the individual PRTF $\mathbf{H}$ performs significantly better than any other condition. Then, the mean PE and QE of $\hat{\mathbf{H}}$ are significantly lower than those of $\hat{\mathbf{H}}^h$ ($p < 0.01$) with an improvement of 1.79° and 3.97%, respectively. We found a medium effect size for these differences in PE and QE distributions corresponding to a Cohen's d of 0.52 and 0.57, respectively. We notice that $\mathbf{H}^f$ has performances similar to $\hat{\mathbf{H}}^h$ and $\hat{\mathbf{H}}$. For PE, $\hat{\mathbf{H}}^h$ is significantly higher than $\mathbf{H}^f$ ($p < 0.05$), while $\hat{\mathbf{H}}$ significantly outperforms $\mathbf{H}^f$ only for QE ($p < 0.05$).

6. CONCLUSIONS

In this paper, we proposed an expansion to the pinna anthropometric parameters commonly used in literature for a more comprehensive description of the pinna in the PRTF individualization field. The parameters were measured using landmarks manually annotated on the pinna shape. We compared the proposed parameters to the HUTUBS ones in a PRTF selection task in the median plane. We found a limited, yet statistically significant, improvement in using the proposed parameters for both SD and localization metrics obtained with an auditory model. Given the limited improvement, further investigation is necessary to assess the relevance of the proposed parameters. This work represents an initial step toward an improved pinna characterization, as the baseline approach used for PRTF selection may not be suitable to exploit the information provided by the proposed parameters. Thus, exploring advanced methods for HRTF individualization, such as those based on machine learning, could be beneficial. In future work, we also plan to automate the landmarks placement on pinna images as done in our previous work [26]. This permits to measure the anthropometry without manual intervention.

Acknowledgments

This work is part of SONICOM, a project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017743.



(a) Local polar RMS error (PE)

(b) Quadrant error rate (QE)

Figure 4. Results of the virtual localization experiment through the auditory model [37] in terms of (a) PE and (b) QE distributions. The compared PRTF conditions are: individual $\mathbf{H}$ (green), $\hat{\mathbf{H}}^h$ selected using the HUTUBS anthropometry $\mathbf{A}^h$ (red), $\hat{\mathbf{H}}$ selected using all anthropometry $\mathbf{A}$ (light blue), and FABIAN $\mathbf{H}^f$ (purple). The white dot in the violin plot represents the median, while the vertical colored line is the mean and the horizontal thick gray line is the interquartile range. The vertical black lines between distributions represent a significant difference, while the stars on their right represent the corresponding p-values (*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$).

7. REFERENCES

[1] B. Xie, *Head-Related Transfer Function and Virtual Auditory Display*, 2nd ed. Florida, US: J. Ross Publishing, 2013.

[2] S. Xu, Z. Li, and G. Salvendy, "Individualization of head-related transfer function for three-dimensional virtual auditory display: A review," in *Virtual Reality*, R. Shumaker, Ed. Springer Berlin Heidelberg, 2007, pp. 397–407.

[3] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization using nonindividualized head-related transfer functions," *The Journal of the Acoustical Society of America*, vol. 94, no. 1, pp. 111–123, 1993.

[4] H. Møller, M. F. Sørensen, C. B. Jensen, and D. Hammershøi, "Binaural technique: Do we need individual recordings?" *Journal of the Audio Engineering Society*, vol. 44, no. 6, pp. 451–469, 1996.

[5] D. R. Begault, E. M. Wenzel, and M. R. Anderson, "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source," *Journal of the Audio Engineering Society*, vol. 49, no. 10, pp. 904–916, october 2001.

[6] R. Bomhardt, "Anthropometric individualization of head-related transfer functions analysis and modeling," Ph.D. dissertation, Institute of Technical Acoustics, Aachen University, Berlin, Germany, 2017.

[7] C. Guezenoc and R. Seguier, "HRTF individualization: A survey," in *Proc. AES Convention*, New York, 2018.

[8] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," in *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No. 01TH8575)*. IEEE, 2001, pp. 99–102.

[9] P. Stitt and B. F. Katz, "Sensitivity analysis of pinna morphology on head-related transfer functions simulated via a parametric pinna model," *The Journal of the Acoustical Society of America*, vol. 149, no. 4, pp. 2559–2572, 2021.

[10] F. Brinkmann, M. Dinakaran, R. Pelzer, P. Grosche, D. Voss, and S. Weinzierl, "A cross-evaluated database of measured and simulated HRTFs including 3D head meshes, anthropometric features, and headphone impulse responses," *Journal of the Audio Engineering Society*, vol. 67, no. 9, pp. 705–718, 2019.

[11] 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," *The Journal of the Acoustical Society of America*, vol. 132, no. 6, pp. 3832–3841, 2012.

[12] M. Burkhard and R. Sachs, "Anthropometric manikin for acoustic research," *The Journal of the Acoustical Society of America*, vol. 58, no. 1, pp. 214–222, 07 1975.

[13] J. C. Middlebrooks, "Individual differences in external-ear transfer functions reduced by scaling in frequency," *The Journal of the Acoustical Society of America*, vol. 106, no. 3, pp. 1480–1492, 09 1999.

[14] S. Ghorbal, X. Bonjour, and R. Séguier, "Computed HRIRs and ears database for acoustic research," in *Audio Engineering Society Convention 148*. Audio Engineering Society, 2020.

[15] N. S. A. o. C. GB/T 2428-1998, "GB/T 2428-1998 standard head-face dimensions of adults," General administration of quality supervision and inspection of China, Tech. Rep., 1998.

[16] Z. Guo, Y. Lu, H. Zhou, Z. Li, Y. Fan, and G. Yu, "Anthropometric-based clustering of pinnae and its application in personalizing hrtfs," *International Journal of Industrial Ergonomics*, vol. 81, p. 103076, 2021.

[17] B. Xie, X. Zhong, D. Rao, and Z. Liang, "Head-related transfer function database and its analyses," *Science in China Series G: Physics, Mechanics and Astronomy*, vol. 50, no. 3, pp. 267–280, 2007.

[18] K. Iida, O. Nishiyama, and T. Aizaki, "Estimation of the category of notch frequency bins of the individual head-related transfer functions using the anthropometry of the listener's pinnae," *Applied Acoustics*, vol. 177, p. 107929, 2021.

[19] Y. Teng and X. Zhong, "An individualized HRTF model based on random forest and anthropometric parameters," in *2023 15th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*. IEEE, 2023, pp. 143–146.

[20] E. A. Shaw and R. Teranishi, "Sound pressure generated in an external-ear replica and real human ears by a nearby point source," *The Journal of the Acoustical Society of America*, vol. 44, no. 1, pp. 240–249, 1968.

[21] V. R. Algazi, C. Avendano, and R. O. Duda, "Elevation localization and head-related transfer function analysis at low frequencies," *The Journal of the Acoustical Society of America*, vol. 109, no. 3, pp. 1110–1122, 2001.

[22] V. C. Raykar, R. Duraiswami, and B. Yegnanarayana, "Extracting the frequencies of the pinna spectral notches in measured head related impulse responses," *The Journal of the Acoustical Society of America*, vol. 118, no. 1, pp. 364–374, 2005.

[23] M. B. Gardner and R. S. Gardner, "Problem of localization in the median plane: effect of pinnae cavity occlusion," *The Journal of the Acoustical Society of America*, vol. 53, no. 2, pp. 400–408, 1973.

[24] J. Fels and M. Vorländer, "Anthropometric parameters influencing head-related transfer functions," *Acta Acustica united with Acustica*, vol. 95, no. 2, pp. 331–342, 2009.

[25] S. Ghorbal, T. Auclair, C. Soladie, and R. Seguier, "Pinna morphological parameters influencing HRTF sets," in *Proceedings of the 20th International Conference on Digital Audio Effects (DAFx-17)*, 2017.

[26] 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)*. IEEE, 2021, pp. 1–10.

[27] M. Dinakaran, P. Grosche, F. Brinkmann, and S. Weinzierl, "Extraction of anthropometric measures from 3D-meshes for the individualization of head-related transfer functions," in *Audio Engineering Society Convention 140*. Audio Engineering Society, 2016.

[28] E. A. Torres-Gallegos, F. Orduna-Bustamante, and F. Arámbula-Cosío, "Personalization of head-related transfer functions (HRTF) based on automatic photo-anthropometry and inference from a database," *Applied Acoustics*, vol. 97, pp. 84–95, 2015.

[29] P. Majdak, M. J. Goupell, and B. Laback, "3-D localization of virtual sound sources: Effects of visual environment, pointing method, and training," *Attention, Perception, & Psychophysics*, vol. 72, no. 2, pp. 454–469, 2010.

[30] K. Iida, H. Shimazaki, and M. Oota, "Generation of the amplitude spectra of the individual head-related transfer functions in the upper median plane based on the anthropometry of the listener's pinnae," *Applied Acoustics*, vol. 155, pp. 280–285, 2019.

[31] X. Liu, W. Huang, H. Zhang, and X. Zhong, "Median-plane head-related transfer function personalization using two-dimensional independent component analysis," in *2022 IEEE 8th International Conference on Computer and Communications (ICCC)*. IEEE, 2022, pp. 2308–2312.

[32] M. G. Onofrei, R. Miccini, R. Unnthorsson, S. Serafin, and S. Spagnol, "3D ear shape as an estimator of HRTF notch frequency," in *Proceedings of the 17th Sound and Music Computing Conference (SMC 2020)*. Axea sas/SMC Network, Jun 2020, pp. 131–137.

[33] D. Zotkin, J. Hwang, R. Duraiswaini, and L. S. Davis, "HRTF personalization using anthropometric measurements," in *2003 IEEE workshop on applications of signal processing to audio and acoustics (IEEE Cat. No. 03TH8684)*. IEEE, 2003, pp. 157–160.

[34] M. Geronazzo, S. Spagnol, and F. Avanzini, "Mixed structural modeling of head-related transfer functions for customized binaural audio delivery," in *2013 18th International Conference on Digital Signal Processing (DSP)*. IEEE, 2013, pp. 1–8.

[35] B. Zhi, D. N. Zotkin, and R. Duraiswami, "Towards fast and convenient end-to-end HRTF personalization," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 441–445.

[36] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*. England, UK: Routledge, 1988.

[37] R. Baumgartner, P. Majdak, and B. Laback, "Modeling sound-source localization in sagittal planes for human listeners," *The Journal of the Acoustical Society of America*, vol. 136, no. 2, pp. 791–802, 2014.

[38] Majdak, Piotr, Hollomey, Clara, and Baumgartner, Robert, "AMT 1.x: A toolbox for reproducible research in auditory modeling," *Acta Acustica*, vol. 6, p. 19, 2022.

[39] J. C. Middlebrooks, "Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency," *The Journal of the Acoustical Society of America*, vol. 106, no. 3, pp. 1493–1510, 1999.