Audio Engineering Society

Conference Paper

Presented at the 2022 International Conference on
Audio for Virtual and Augmented Reality
2022 August 15–17, Redmond, WA, USA

HRTF personalization based on ear morphology

Michaela Warnecke[1], Sharon Jamison[1], Sebastian Prepeliță[1], Paul Calamia[1], and Vamsi Krishna Ithapu[1]

[1]*Reality Labs Research at Meta*

Correspondence should be addressed to Michaela Warnecke (mwarnecke@fb.com)

ABSTRACT

On the forefront of realistic spatial audio is the personalization of binaural auditory input. Specifically, personalized head-related transfer functions (HRTFs) have been shown to improve the quality of binaural spatial audio over generic HRTFs. Here, we approached HRTF personalization from a morphological standpoint by calculating the distance between any two three-dimensional models of the ear. Subsequently, a ranking of ears based on the distances provided a similarity estimate. Using measured HRTFs of these ears, we tested how well listeners performed when localizing sounds in virtual space. We show that performance is closest to that of the individual HRTF when listeners used the best-ranked ear's HRTF and furthest when listeners used the worst-ranked ear's HRTF.

1 Introduction

Head-related impulse responses (HRIRs) are vital for enabling realism in virtual sound spaces. They characterize the direction-dependent acoustic filtering needed to transform a given sound signal before entering the eardrum. HRIRs and the corresponding frequency spectra (head-related transfer functions, HRTFs) are unique to each individual, and are defined by the head, ear and torso structure of the listener [1]. It is well known that the differences in the left and right ears' HRTFs create binaural cues that listeners use to determine the perceived direction of a sound source [2]. While an individual can experience a spatial percept using generic (non-personalized) HRTFs, these generic filters create artifacts, such as front-back confusions, angular distortion in elevation perception, and weak externalization [3, 4, 5].

Over the last decade, several groups have proposed approaches to personalizing HRTFs, strongly driven by the interest in building augmented and virtual reality (AR and VR) technologies. However, a robust, scalable solution is still elusive. Measuring individualized HRTFs is a cumbersome process involving expensive acoustic equipment such as anechoic chambers, and the logistics of building a large-scale user study with professionally trained personnel to bring volunteers into controlled lab environments. In addition, such measurement setups are impractical to scale to thousands or more individuals.

An alternative to the tedious HRTF measurement is numerical simulations (e.g., [6, 7]), which require accurate three-dimensional (3D) models of the subject's ears. Despite recent progress [8, 9] and being easier to scale relative to measurements, numerical simulations currently require powerful computing machines while their underlying physical models are not fully and acceptably validated yet [10, 11]. This entails that they cannot yet replace measurements with high fidelity.

Another alternative is presented via statistical modeling and machine learning driven approaches, which have shown promise (e.g., [12]). However, a majority of these methods rely on large, detailed databases of ear shapes (or images) to predict the mapping between mathematical representations of ears and the corresponding spectral shape of HRTFs [13].

A third approach directly utilizes objective or perceptual evaluations to identify and select a *best fit* HRTF from a pre-existing database (e.g., [14, 15, 16]). Such selection-based approaches are implicitly driven by computing distances between two inputs, and range in their approaches from utilizing auditory localization games [15] to extracting anthropometric features of pinna dimensions [17]. Still, it remains unclear what an optimal set of inputs for these prediction models may be, and what metric might define distances in the input space.

Recent efforts have aimed to improve our understanding of the relationship between anthropometric changes of the ear and HRTF spectra and even sound perception [18, 19, 20], but the space of morphological variation, HRTF dimensionality, and perceptual feature evaluation is immense and complex. For example, while much research has studied and utilized anthropometric distances in order to personalize HRTFs, the commonly chosen features - often modeled after the CIPIC database's first suggested measures [21] - are largely ill-defined, likely incomplete, and may fall short of the many intricacies of ear geometry that contribute to changes in the HRTF spectrum [19].

In this work, we present an effort to utilize 3D ear morphology more explicitly. We utilized a database of detailed 3D scanned ear meshes, landmark annotations, as well as corresponding HRTF measurements. In short, we used several landmarks distributed along a detailed 3D scan of the ear to create an aligned representation of that ear, and subsequently calculated and compared the Euclidean distance between the landmarks of an individual's ear to the landmarks of all remaining ears. This approach yielded a ranking of ears for any given individual from closest to farthest in 3D distance. Given that ear shape influences HRTF spectra [22, 23] and that results from physics and mathematical modeling showed that a unique solution (i.e., HRTF) exists for a given geometry (e.g., [24]), we make the assumption that the ear with the shortest distance to any "starter ear" would produce the most similar HRTF set to that of the "starter ear". To evaluate this approach, we employed both objective and perceptual evaluations. First, we compared the HRTF similarity objectively by calculating the spectral difference error (SDE), as well as utilizing a perceptual modeling approach [25]. Second, we evaluated the ranking subjectively by asking listeners to perform a VR localization task.

We hypothesized that localization performance would be least accurate when listeners performed the task with the generic HRTF, most accurate for a given listener's individual HRTF, and increasingly less accurate as the rank of the HRTF increased. We show that HRTF personalization via this method is possible, as early-ranked ear's HRTFs showed improved performance of sound localization over generic HRTFs. Additionally, the last-ranked ear's HRTF showed the worst performance, indicating that Euclidean distance between ears may correspond to changes in perceptual sound localization space.

2 Methods

2.1 Anthropometric and Acoustic Data

3D ear models and their HRTF measurements (see below) were taken from an internally collected database. This database contains detailed 3D scans and landmark annotations of the torso, head and ears of 857 subjects. The scanning pipeline was validated using 3D printed models (similar to [9]) to have a geometrical error below 1 mm for the external ear. The result of the scan was a closed triangulated surface representation called a *mesh*. Additionally, we measured HRTFs for 96 of the subjects with scans using the measurement system described in Cuevas-Rodriguez et al. [26] (section 2). Briefly, these measurements were sampled at 48 kHz and contained a total of 612 far-field locations at 1.53 m from center of the subject's head. The measurement locations cover azimuthal space with 10° resolution from 0° to 360°, and elevation space equally between -66° (bottom) and 85° (top).

2.1.1 Aligned Representation of the Ear

Each head and torso scan in the database was manually annotated with 144 landmark points denoting anatomical features of interest. Of these 144 landmarks, 106 were on the ear surface (53 landmarks per ear). Each

scanned mesh was also wrapped with a template mesh (R3DS Wrap) constructed using the landmarks to guide the wrapping algorithm and achieve anatomical correspondence. The template mesh has a total of 29,742 vertices, 5,985 of which are located on each ear surface. The wrapping process introduced additional geometrical errors relative to the scanned mesh averaging approximately 0.06 mm.

To efficiently compare ears based on pinna shapes alone, without considering location or rotation relative to the head, a local coordinate system (cf. [19]) was defined for each ear (Fig. 1A). This "ear-plane" was defined by a line connecting the bottom of the tragus (landmark point 4) to the highest point on the helix (midpoint between landmark points 37 and 38), and the center of the antihelix (landmark point 20). The origin of the plane was located at the point of inflection between tragus and helix (near landmark 6). Defining a local coordinate system allowed for the alignment of ears based on their morphology.

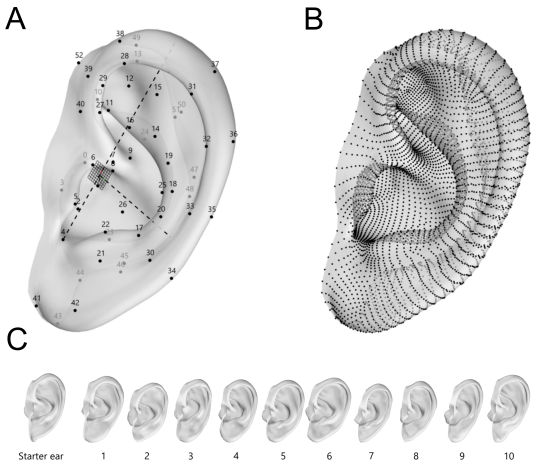

Fig. 1: Visualization of the ear ranking process. (A) Distribution of landmarks along with the defined coordinate system. (B) Distribution of all vertices. (C) Exemplary ranking of the closest ten ears to a starter ear.

Once this plane was defined for all 96 ears, they were aligned by transforming the coordinate system for each ear from its individual plane to a shared target plane (cf. [19]). To achieve a more accurate comparison, the collection of landmarks was expanded to include all the vertices of the ear mesh (Fig. 1B). A similarity ranking was created by selecting a starting ear and measuring the Euclidean distance between each landmark of the starting ear mesh and the corresponding landmark on all other ears. The sum of the distances determined the order of the ranking, with the smallest sum distance indicating the closest ear. Left and right ears were ranked separately, meaning that the distance between the landmarks of the starting ear and those of all other ears was calculated for both left and right ears. Subsequently, we chose the ear with the smallest distance at a specific rank (see below). An example ranking for a selected left "starter ear" with the closest ten ears is illustrated in Figure 1C.

2.2 Objective Testing

As a first step toward evaluating whether the ear ranking might show an effect in HRTF space, we utilized two metrics that take the measured HRTFs of the ranked ears as their input: the spectral difference error (SDE) and Baumgartner et al.'s saggital-plane localization model (SPLM) [25]. While the SDE is an objective measure of the similarity in frequency response between two sets of HRTFs, the SPLM predicts localization errors between two sets of HRTFs.

2.2.1 Spectral Difference Error

SDE was calculated using level-matched HRTFs, as defined in the following equation, for the left and right ear HRTF sets separately. The calculation was band-limited between 200 Hz and 17 kHz. For each frequency bin f, SDE was:

$$\mathrm{SDE}(f) = \frac{1}{n}\sum_{i=1}^{n}\left|20 * \log_{10}\left(\frac{\mathrm{HRTF}_1(\theta_i, f)}{\mathrm{HRTF}_2(\theta_i, f)}\right)\right|$$

where θ_i denotes the discrete angular direction, which here encompassed all azimuth/elevation pairs (and n is total number of such directions). Each frequency-binned SDE value was scaled by a logarithmic function so that each octave was weighted equally. Subsequently, the median SDE across all frequency bins was taken to arrive at a single SDE for each HRTF direction.

2.2.2 Sagittal-Plane Localization Model

The SPLM aims to predict localization performance at different saggital planes by modeling several stages of the auditory pathway in combination with monaural and binaural perceptual factors, and sensorimotor mapping (for details, see Baumgartner et al. [25]). In short, the model uses an "internal spectral representation" of a target sound (i.e. one's own set of HRTFs) and compares it to a "template spectral representation" of that same sound (i.e. another set of HRTFs) at a given saggital plane. We utilized the implementation of the SPLM in the Auditory Modeling Toolbox for Matlab (Mathworks, Natick, MA) (see [27]).

To estimate the impact of ranking of ears on sound localization performance in elevation for the baseline conditions, we input each listener's individual HRTFs as the target HRTF and the generic or their individual HRTF as the template HRTF. For experimental conditions, we input each listener's individual HRTF as the target HRTF and the measured HRTF of the ranked ears for this individual as the template HRTF. Input HRTFs were interpolated using barycentric weights over a Lebedev grid spanning -90° (bottom) to $+90^\circ$ (top) in elevation for a total of $2,131$ directions, excluding directions behind the listener. The model output was a localization probability matrix for a range of target and response polar angles, from which we subsequently extracted the response angle at the maximum localization probability for the target elevation angle that most closely matched the locations which were used in our perceptual evaluation (testing elevations: $[-20^\circ, +45^\circ, 0^\circ, +45^\circ, -20^\circ]$; SPLM elevations extracted: $[-18.7^\circ, +45^\circ, 0^\circ, +45^\circ, -18.7^\circ]$). We then calculated the median localization error across azimuthal locations and listeners, per HRTF condition.

2.3 Perceptual Testing

The VR localization evaluation procedure was the same as that used in the virtual localization study described in [28] (section 3.2). Briefly, participants were asked to localize a 1.5 s duration white noise stimulus, filtered between 200 and 17 kHz, while exposed to a variety of HRTFs (see below). Localization was tested in virtual space using an Oculus Quest2, but the audio stimulus was spatialized via separate over-the-ear headphones (Beyerdynamic DT990 Pro). Tests were carried out at three HRTF directions, where the azimuthal location was randomly mirrored between the left and right hemispheric space on each trial so as to map out variance in the frontal area. Testing locations comprised the following directions: $[\pm 45^\circ, -20^\circ]$, $[0^\circ, 0^\circ]$, $[\pm 20^\circ, +45^\circ]$, where the first coordinate indicates the azimuth in degrees relative to the mid-sagittal plane and the second coordinate indicates the elevation in degrees relative to the horizontal plane through the interaural axis. A total of 7 HRTFs were evaluated for 5 trials at each location, resulting in a total of 105 trials per participant. For each HRTF, we removed the respective initial phase delay (IPD) and all pass delay (APD), and replaced it with the participant's individual IPD and APD to remove the impact of HRTF-specific interaural time differences (ITDs) on localization performance ([29]). ITD personalization based on ear shape alone is likely challenging, but previous work has shown that only a few additional components are necessary to estimate an individual's maximum ITD [22].

The 7 different tested HRTF sets were split into *baseline* and *experimental* conditions. In baseline conditions, listeners were provided with a generic HRTF, which had been selected from the available database, and the individual's own measured HRTF. For experimental conditions, listeners were provided with HRTFs at five ranks of similarity in descending order from the "starter ear" (best, 10th best, 40th best, 80th best, worst). The experimental design was fully randomized. Prior to testing, each listener went through (1) a two-trial familiarization phase, in which they could experience the virtual environment and listen to example sounds, and (2) a 10 to 20 trial training set, which aimed to stabilize the listener's performance using their individual HRTF. The experiment began immediately after training.

To summarize the data, we extracted the listeners' raw indicated locations in azimuth and elevation and calculated the elevation and azimuth error of localization by subtracting the target sound location from the perceived sound location. Subsequently, we averaged the errors first across trials, then across the different locations to arrive at error values per subject and condition.

A total of eleven listeners participated in the VR localization task; 3 subjects were excluded from the analysis, as the localization data with their own HRTF fell outside three median absolute deviations from the median errors. All statistical testing was done in JMP (SAS).

3 Results

In this study, we evaluated both the objective change in HRTF similarity (SDE, SPLM) and the sound localization performance for a collection of HRTFs that were selected based on the similarity of their corresponding ear morphologies. First, we will describe changes in SDE and SPLM results as a function of baseline and experimental HRTF conditions. Subsequently, we will present localization data from listeners who performed a VR localization task with the same conditions.

3.1 Objective Evaluation

Ear morphology is known to influence HRTF spectra [22, 23], which suggests that similarly-shaped ears may produce similar HRTFs for the same stimulus type and spatial location. We thus predicted that the similarity of HRTF spectra might decrease (SDE might increase) as the rank of the matched ear increased. We evaluated this hypothesis objectively using two different measures described below.

3.1.1 SDE

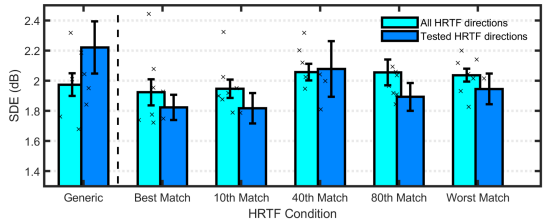

Fig. 2: Log-scaled Spectral Difference Error (SDE; y-axis, ± standard error of the mean) between each individual's set of HRTFs and that of each evaluated condition (x-axis). Crosses indicate variation between subjects.

SDE is a measure between any two input HRTF sets across all available HRTF directions. For this analysis, we compared any individual's own HRTF with the generic and all ranked-ear HRTF sets (1) across all directions (Figure 2, turquoise) and (2) at the directions tested in the perceptual study (Fig. 2, blue). Figure 2 plots average SDE across HRTF conditions. Overall, SDE across all available HRTF directions hovered around 2 dB, showing slight increases from the best-ranked ear's HRTF to the worst-ranked ear's HRTF. Differences could be observed when considering only directions that were part of the perceptual evaluation test. For perceptual testing directions only, the SDE between any individual's HRTF and the generic HRTF was largest, averaging about 2.2 dB, supporting our hypothesis that the generic HRTF might pose an inadequate match for any individual. Further, across experimental conditions, the SDE between any individual's HRTF and that of each ranked ear's HRTF differed for each rank, ranging from 1.8 dB to 2.1 dB. Note that the SDE of the best and 10th ranked ear's HRTF were smallest.

3.1.2 SPLM

Figure 3 shows the predicted localization errors across subjects at elevation target locations and saggital planes tested in our study for both the baseline and experimental conditions. As expected, elevation localization in the baseline conditions showed a larger error for the generic HRTF compared to the individual HRTF. Note that very small errors for the individual HRTF condition are expected, and likely are due to the natural probabilistic uncertainty in human sound localization included in the model. Further, errors were smallest for the HRTF that corresponded to the best-ranked ear, and larger for all other ranks, with the maximum error observed for the HRTF corresponding to the worst-ranked ear. While there was an overall increase from best to worst rank, this increase was not monotonic.

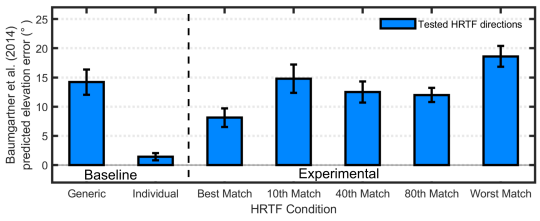

Fig. 3: Predicted elevation error (y-axis, ± standard error of the mean) for each HRTF condition (x-axis) across all HRTF directions tested in the perceptual experiment.

Together, the SDE and SPLM results supported the notion that the best ranked ear may show performance closest to the individual's HRTF. However, the increase in errors to be expected from increasing ranks might not be monotonic, as we hypothesized. The SPLM results further support the hypothesis that the worst

match may show largest performance differences to the individual and best ranked ear's HRTF.

3.2 Perceptual Evaluation

Eight listeners performed a VR localization task while exposed to seven different HRTFs. We were interested in evaluating how localization performance, specifically in elevation, changed as a function of HRTF type. We did not statistically evaluate the change in azimuth localization error, as ITDs were controlled during this experiment to avoid conflict of HRTF-specific ITD information.

In baseline conditions, we presented listeners with two HRTF sets that represent the two ends of the personalization spectrum: a generic HRTF (no personalization) and an individual HRTF (full personalization). We predicted that localization performance, as indicated by errors in elevation localization, would be largest when participants listened with the generic HRTF, as the transfer functions used to render spatialized sounds during the experiment are not personalized to each participant. By contrast, we predicted that localization performance would show the smallest errors when participants listened with the second HRTF in the baseline condition, their individual HRTF.

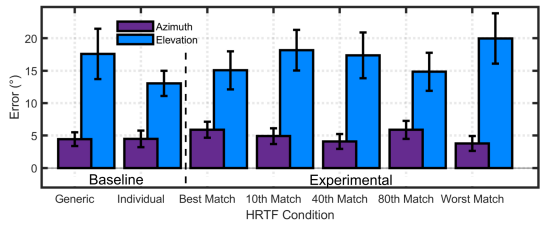

Fig. 4: Error in degrees (y-axis) as a function of baseline and experimental testing condition (x-axis).

Figure 4 shows the average localization error (y-axis; $\pm$ standard error of the mean) across evaluated baseline and experimental conditions. Normality and equal variance of the data were confirmed prior to analysis (Anderson-Darling test, Levene's test). As expected, elevation localization errors were significantly larger when listeners were exposed to the generic HRTF (17.5°) compared to their own (13°; two-sample t-test, $t(13) = 3.14$, $p = 0.0037$). These errors are within the previously-reported range of elevation errors for the same dataset and evaluation method [28].

In the experimental conditions, we presented participants with five HRTF sets that were previously measured for the ears ranked as the best, 10th, 40th, 80th and worst ear for any given "starter ear". We predicted that localization errors would increase as the rank of the ear increased, because the distance in Euclidean space between the ears increased. As HRTFs are intrinsically linked to morphological intricacies of the pinna [22, 23], an increase in difference of the ear morphology suggests an increase in the difference of HRTF spectra, potentially yielding increasing localization errors. We utilized an ANOVA to analyze significant differences between conditions. Figure 4 illustrates a non-significant change in elevation localization error as a function of experimental HRTF condition ($F_{(4,35)} = 0.7$, $p = 0.56$). While localization errors did increase from the best to the 10th-ranked ear's HRTF, they subsequently plateaued and showed a small decrease to 15.1° in the 80th rank condition. Notably, however, a separate ANOVA analysis with posthoc tests revealed that localization errors of the best ranked ear's HRTF were small (15°) and did not differ significantly from the performance with the individual HRTF, while errors were largest for the worst ranked ear's HRTF (20°) and did differ from the performance with the individual HRTF ($F_{(2,21)} = 5.32$, $p = 0.0136$; Tukey HSD: Individual vs. Worst: $p = 0.0106$, Individual vs. Best: $p = 0.14$, Best vs. Worst: $p = 0.42$).

4 Discussion

Our goal in this work was to build an HRTF personalization model that is based on morphological similarity between 3D ear shapes. Using a database of 3D scanned ear meshes and their corresponding HRTF measurements, we created a ranking of ears for any given individual. Recent advances in consumer-available 3D scanning devices and cameras, such as those in the iPhone, have shown promising progress toward providing similar-resolution 3D scans (after processing) as the Artec Spider [30, 31], which was also used for the scans in the present data set. Figure 5A illustrates an example of the ear ranking, showing an individual's left "starter ear" (grey) and the ears at ranks chosen for this study. Each ear is colored to indicate morphological areas that differed in distance from this "starter ear" as a way of indicating change. Additionally, we plot histograms for each ear, outlining the distribution of distances between vertex points of the starter and ranked ear. Finally, each ear's left HRTF set for the

horizontal and median planes is plotted to illustrate changes in the HRTF spectrum for each ranked ear (Fig. 5B).

Previous research showed that ear shape influences HRTF spectra [22, 23], suggesting that an increase in difference of the ear morphology might produce an increase in the difference of HRTF spectra. In an analysis of the differences between HRTF spectra (SDE), we found that this assumption holds true overall (Fig. 2, turquoise), but might be more complex to judge when only a handful of HRTF directions are sampled (Fig. 2, blue).

Evaluating the match of an ear or a corresponding HRTF is a multidimensional problem: multiple acoustic features, which are related to each other in complex ways, can be evaluated (see Simon et al. [5]). Here, we chose to evaluate the *localization* of a virtual sound source, as one of the most common performance metrics that can be reasonably compared across various studies. To better estimate the localization performance we might expect, we utilized the SPLM [25]. The modeling results (Fig. 3) indicated that it was reasonable to assume the best-ranked ear's HRTF could provide the the smallest elevation error for all experimental condition HRTFs, and that the worst-ranked ear's HRTF ought to provide the largest elevation error. For ranks spanning the best and worst ear's HRTFs, however, the SPLM did not suggest that errors might increase with an increase in ranks, as we had initially hypothesized.

Overall, perceptual data from the VR localization test (Fig. 4) matched the relative results of the SPLM: localization errors in the baseline condition were significantly higher for the generic compared to the individual HRTF, which had the overall smallest elevation errors. Further, errors were closest to the individual HRTF performance when listeners were exposed to the best-ranked ear's HRTF. Elevation errors then increased (10th rank) and decreased again to a second low-error performance for the 80th-ranked ear's HRTF. As indicated also in the SPLM, VR localization errors in elevation were overall largest for the worst-ranked ear's HRTF. This validates out hypothesis and certifies the selection procedure we proposed.

However, it is important to note that the 80th-ranked ear's HRTF appeared to be providing the same level of personalization as the best-ranked ear. Closer inspection indicated that on average, subject's performance using the 80th-ranked ear's HRTF was within 1.75° of the best-ranked ear's HRTF performance. This is surprising and further investigation is indeed needed to understand this observation. It may be possible that the ear currently ranked as 80th might change its rank if other factors were included in the current ranking procedure. First, our ranking was based on the sum of the Euclidean distance between all available vertices of two ear meshes (Fig. 1C). Due to the nature of the vertex distribution, there are more vertices in some areas of the ear than others; for example the cavum concha, cymba concha, and fossa triangularis contain more vertices than areas with a larger area such as the lobule. These over-represented areas have previously been proposed to influence HRTF features when modeling HRTFs from anthropometry [32, 33, 34, 35, 36, 19, 37], and a higher density of vertices in these regions introduced implicit weighting. Second, as we aligned the ears to the same plane, we removed their rotation and flare angles on the head as contributing components. Komatsu et al. [38] report that in objective acoustic measurements of median plane HRTFs, pinna rotation angles between 0° to 30° effectively rotated the location of the source. In other words, HRTF magnitudes were similar for equal differences between the pinna rotation angle and the source elevation angle. There was no perceptual evaluation of the effect of rotation angle in this study, but [20] suggest that pinna rotation has a strong importance for localization performance in the median plane. While we did not evaluate localization performance in the median plane (beyond the frontal location $[0°, 0°]$), it is possible that listening to an HRTF obtained from a pinna with a very different rotation angle from one's own might lead to changes in elevation performance at other locations, as well. Rotation angles for our subjects ranged from -2° to 17°. Additionally, Komatsu et al. [38] and Plaskota and Dobrucki [39] investigated the effect of changes in flare angle on the HRTF and show that increasing flare angle from 15° to 45° increased the HRTF amplitude in median plane HRTFs at frequencies between 2 and 7 kHz, and also induced deeper notches. Notch frequency and depth change with increasing elevation [40]; it is conceivable that listening with someone's HRTF that has a different flare angle might alter elevation perception. Flare angles for our subjects ranged from 17° to 38°.

HRTF personalization is a widely researched problem, and has been addressed with different approaches ranging from the utilization of anthropometric measures [41], to using tournament style listening tests [15] and

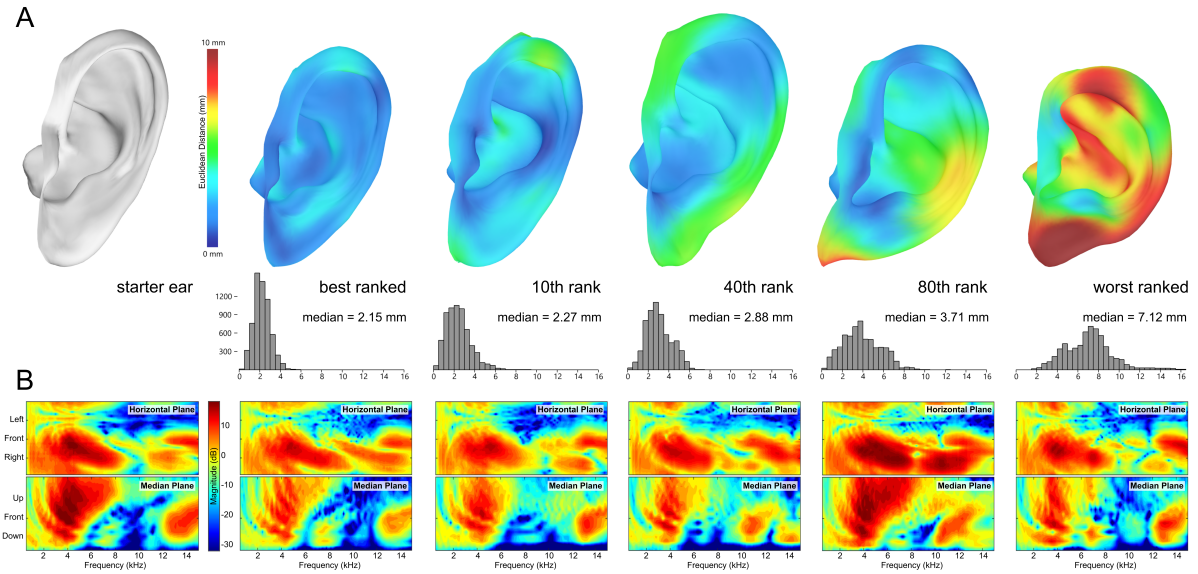

Fig. 5: (A) Exemplary ear shape ranking for one subject, including the binned Euclidean distance from starter ear to each ranked ear. (B) Each ear's set of HRTFs in the horizontal and median planes.

selecting from a database [17], to utilizing numerical simulations based on morphology and anthropometry [7]. In the present approach we provided listeners with an HRTF that, on average, improved their elevation localization performance to within 1.3° of performance with their individual HRTF. To achieve this, we provided listeners with an HRTF by selecting the ear with the smallest Euclidean distance to their own. We also showed that the HRTF of the ear furthest from their own consistently produced largest elevation errors. Further investigation is necessary to understand how changes in ear morphology might need to be weighted in order to map to a similarly increasing distance in HRTF and perceptual space. This is a complex problem, as these three components - ear morphology, HRTF features, and acoustic perception - do not have a clear mapping between them and each vary among several dimensions.

5 Acknowledgment

The authors would like to thank their colleagues at Reality Labs at Meta, Clarissa Munoz and Karisma Kulkarni for their help in collecting the perceptual data, and Zamir Ben-Hur, David Lou Alon and Sam Clapp for the development of an internal toolbox to run the SDE and SPLM model estimations.

References

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

[2] Blauert, J., *Spatial hearing: the psychophysics of human sound localization*, MIT press, 1997.

[3] Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L., "Localization using nonindividualized head-related transfer functions," *The Journal of the Acoustical Society of America*, 94(1), pp. 111–123, 1993.

[4] Møller, H., Sørensen, M. F., Jensen, C. B., and Hammershøi, D., "Binaural technique: Do we need individual recordings?" *Journal of the Audio Engineering Society*, 44(6), pp. 451–469, 1996.

[5] Simon, L. S., Zacharov, N., and Katz, B. F., "Perceptual attributes for the comparison of head-related transfer functions," *The Journal of the Acoustical Society of America*, 140(5), pp. 3623–3632, 2016.

[6] Ziegelwanger, H., Majdak, P., and Kreuzer, W., "Numerical calculation of listener-specific head-related transfer functions and sound localization:

Microphone model and mesh discretization," *The Journal of the Acoustical Society of America*, 138(1), pp. 208–222, 2015.

[7] Katz, B. F., "Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation," *The Journal of the Acoustical Society of America*, 110(5), pp. 2440–2448, 2001.

[8] Prepeliță, S. T., Gómez Bolaños, J., Geronazzo, M., Mehra, R., and Savioja, L., "Pinna-related transfer functions and lossless wave equation using finite-difference methods: Verification and asymptotic solution," *The Journal of the Acoustical Society of America*, 146(5), pp. 3629–3645, 2019, ISSN 0001-4966, doi:10.1121/1.5131245.

[9] Prepeliță, S. T., Bolaños-Gómez, J., Pulkki, V., Savioja, L., and Mehra, R., "Numerical simulations of near-field head-related transfer functions: Magnitude verification and validation with laser spark sources," *The Journal of the Acoustical Society of America*, 148(1), pp. 153–166, 2020, ISSN 0001-4966, doi:10.1121/10.0001409, publisher: Acoustical Society of America.

[10] Brinkmann, F., Dinakaran, M., Pelzer, R., Grosche, P., Voss, D., and Weinzierl, S., "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*, 67(9), pp. 705–718, 2019, doi:10/gf8ww6.

[11] Prepeliță, S. T., Gómez Bolaños, J., Geronazzo, M., Mehra, R., and Savioja, L., "Pinna-related transfer functions and lossless wave equation using finite-difference methods: Validation with measurements," *The Journal of the Acoustical Society of America*, 147(5), pp. 3631–3645, 2020, ISSN 0001-4966, doi:10.1121/10.0001230, publisher: Acoustical Society of America.

[12] Spagnol, S., Miccini, R., Onofrei, M. G., Unnthorsson, R., and Serafin, S., "Estimation of Spectral Notches From Pinna Meshes: Insights From a Simple Computational Model," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29, pp. 2683–2695, 2021.

[13] Guezenoc, C. and Seguier, R., "A wide dataset of ear shapes and pinna-related transfer functions generated by random ear drawings," *The Journal of the Acoustical Society of America*, 147(6), pp. 4087–4096, 2020.

[14] Seeber, B. U. and Fastl, H., "Subjective selection of non-individual head-related transfer functions," Georgia Institute of Technology, 2003.

[15] Iwaya, Y., "Individualization of head-related transfer functions with tournament-style listening test: Listening with other's ears," *Acoustical science and technology*, 27(6), pp. 340–343, 2006.

[16] McMullen, K., Roginska, A., and Wakefield, G. H., "Subjective selection of head-related transfer functions (hrtf) based on spectral coloration and interaural time differences (itd) cues," in *Audio Engineering Society Convention 133*, Audio Engineering Society, 2012.

[17] Iida, K., Ishii, Y., and Nishioka, S., "Personalization of head-related transfer functions in the median plane based on the anthropometry of the listener's pinnae," *The Journal of the Acoustical Society of America*, 136(1), pp. 317–333, 2014.

[18] Mokhtari, P., Takemoto, H., Nishimura, R., and Kato, H., "Acoustic sensitivity to microperturbations of KEMAR's pinna surface geometry," in *Proceedings of the International Congress on Acoustics*, pp. 1–8, International Congress on Acoustics (ICA), Sydney, Australia, 2010, ISBN 978-1-61782-745-7.

[19] Stitt, P. and Katz, B. F., "Sensitivity analysis of pinna morphology on head-related transfer functions simulated via a parametric pinna model," *The Journal of the Acoustical Society of America*, 149(4), pp. 2559–2572, 2021.

[20] Pelzer, R., Dinakaran, M., Brinkmann, F., Lepa, S., Grosche, P., and Weinzierl, S., "Head-related transfer function recommendation based on perceptual similarities and anthropometric features," *The Journal of the Acoustical Society of America*, 148(6), pp. 3809–3817, 2020.

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

[22] Middlebrooks, J. C., "Individual differences in external-ear transfer functions reduced by scaling in frequency," *The Journal of the Acoustical Society of America*, 106(3), pp. 1480–1492, 1999.

[23] Middlebrooks, J. C., Macpherson, E. A., and Onsan, Z. A., "Psychophysical customization of directional transfer functions for virtual sound localization," *The Journal of the Acoustical Society of America*, 108(6), pp. 3088–3091, 2000.

[24] Strauss, W. A., *Partial differential equations: An introduction*, John Wiley & Sons, New York, 2nd edition edition, 2008, ISBN 978-0-470-05456-7, 464 pages.

[25] Baumgartner, R., Majdak, P., and Laback, B., "Modeling sound-source localization in sagittal planes for human listeners," *The Journal of the Acoustical Society of America*, 136(2), pp. 791–802, 2014.

[26] Cuevas-Rodriguez, M., Alon, D. L., Clapp, S., Robinson, P. W., and Mehra, R., *Evaluation of the effect of head-mounted display on individualized head-related transfer functions*, Universitätsbibliothek der RWTH Aachen, 2019.

[27] Søndergaard, P. and Majdak, P., "The Auditory Modeling Toolbox," in J. Blauert, editor, *The Technology of Binaural Listening*, pp. 33–56, Springer, Berlin, Heidelberg, 2013.

[28] Ben-Hur, Z., Alon, D., Robinson, P. W., and Mehra, R., "Localization of virtual sounds in dynamic listening using sparse HRTFs," in *Audio Engineering Society Conference: 2020 AES International Conference on Audio for Virtual and Augmented Reality*, Audio Engineering Society, 2020.

[29] Oppenhein, A. V., Schafer, R. W., and Buck, J. R., "Discrete-time signal processing," *Prince Hall, Sec*, 11, 1999.

[30] Nightingale, R. C., Ross, M. T., Cruz, R. L., Allenby, M. C., Powell, S. K., and Woodruff, M. A., "Frugal 3D scanning using smartphones provides an accessible framework for capturing the external ear," *Journal of Plastic, Reconstructive & Aesthetic Surgery*, 74(11), pp. 3066–3072, 2021.

[31] Ross, M. T., Cruz, R., Brooks-Richards, T. L., Hafner, L. M., Powell, S. K., and Woodruff, M. A., "Comparison of three-dimensional surface scanning techniques for capturing the external ear," *Virtual and Physical Prototyping*, 13(4), pp. 255–265, 2018.

[32] Hu, H., Zhou, L., Zhang, J., Ma, H., and Wu, Z., "Head related transfer function personalization based on multiple regression analysis," in *2006 International conference on computational intelligence and security*, volume 2, pp. 1829–1832, IEEE, 2006.

[33] Hugeng, W. W. and Gunawan, D., "Improved method for individualization of head-related transfer functions on horizontal plane using reduced number of anthropometric measurements," *arXiv preprint arXiv:1005.5137*, 2010.

[34] Zhang, M., Kennedy, R., Abhayapala, T., and Zhang, W., "Statistical method to identify key anthropometric parameters in HRTF individualization," in *2011 Joint Workshop on Hands-free Speech Communication and Microphone Arrays*, pp. 213–218, IEEE, 2011.

[35] Fels, J. and Vorländer, M., "Anthropometric parameters influencing head-related transfer functions," *Acta Acustica united with Acustica*, 95(2), pp. 331–342, 2009.

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

[37] Huang, Q. and Li, L., "Modeling individual HRTF tensor using high-order partial least squares," *EURASIP Journal on Advances in Signal Processing*, 2014(1), pp. 1–14, 2014.

[38] Komatsu, H., Nakano, K., Nakayama, M., and Nishiura, T., "Investigations into the human pinna shapes on head-related transfer functions in the median plane," in *Proceedings of Meetings on Acoustics ICA2013*, volume 19, p. 055042, Acoustical Society of America, 2013.

[39] Plaskota, P. and Dobrucki, A., "The influence of pinna flare angle on Head-Related Transfer Function," in *New Trends in Audio and Video/Signal*

Processing Algorithms, Architectures, Arrangements, and Applications SPA 2008, pp. 43–46, IEEE, 2008.

[40] Takemoto, H., Mokhtari, P., Kato, H., Nishimura, R., and Iida, K., "Mechanism for generating peaks and notches of head-related transfer functions in the median plane," *The Journal of the Acoustical Society of America*, 132(6), pp. 3832–3841, 2012.

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