OPEN ACCESS Freely available online
S. Jain, F. Brinkmann, B. Haubrichs, R. Moeller, and S. Weinzierl,
"On the Relevance of Age, Sex, and Ethnicity for Listener Anthropometry,
HRTFs, and Modeled Auditory Perception,"
J. Audio Eng. Soc., vol. 74, no. 9, pp. 594–611 (2026 Sep.).
https://doi.org/10.17743/jaes.2026.0282.

On the Relevance of Age, Sex, and Ethnicity for Listener Anthropometry, HRTFs, and Modeled Auditory Perception

SANKET JAIN,[1,2] **FABIAN BRINKMANN,**[2] **BJARNE HAUBRICHS,**[1]
(s.jain@campus.tu-berlin.de) (fabian.brinkmann@tu-berlin.de) (bjarne.haubrichs@mercedes-benz.com)

RAINER MOELLER,[1] **AND STEFAN WEINZIERL**[2]*
(rainer.moeller@mercedes-benz.com) (stefan.weinzierl@tu-berlin.de)

[1]*Audio Communication Group, Technische Universität Berlin, Einsteinufer 17c, D-10587, Germany*
[2]*Future Interior, Mercedes-Benz Group, Leibnizstrasse 2, Böblingen,Germany*

Head-related transfer functions (HRTFs) are essential for spatial audio reproduction via headphones, and personalized HRTFs provide more natural audio rendering for a specific listener. Because measuring HRTFs for each individual is usually too costly, most personalization approaches including HRTF prediction via deep learning are based on HRTF databases whose representativeness for a larger population has hardly been investigated. To investigate how biases in HRTF databases might affect personalization, 3D head meshes and anthropometric measures were acquired, and HRTFs of 256 individuals were numerically calculated by equally sampling from 11 age groups, two sexes, and two ethnicities in populations in Sindelfingen, Germany, and Beijing, China. A comparison with existing data showed that the acquired anthropometric features cover a wider range of values, suggesting that the data here are more diverse. Statistical analyses using general linear models confirmed that ear and head size are significantly influenced by age, sex, and ethnicity, in accordance with previous studies. These factors also significantly affect coloration and median and horizontal plane localization errors modeled for pairs of nonindividual HRTFs. Although the effects were small, this suggests that HRTF personalization approaches could benefit from considering age, sex, and ethnicity.

Keywords: Auditory Modeling, Representative Sampling, Spatial Audio

1 INTRODUCTION

Head-related transfer functions (HRTFs) are acoustic filters that characterize how a listener's unique anatomy modifies incoming sound waves in the free field [1]. These filters encode the interaural time (ITDs) and level differences (ILDs) and the spectral cues necessary for 3D sound localization [2, 3]. Spectral cues mainly include peaks and notches in the transfer function that are shaped by pinna geometry [3]. The precise location of these spectral peaks and notches varies across individuals and with the position of the sound source, reflecting interindividual differences in pinna shape [3]. Although ILDs are influenced by both the head and pinna shape, ITDs can be well explained by considering only the head shape [4]. Consequently, understanding how listener anthropometry affects HRTFs is crucial for HRTF individualization, which improves localization accuracy, coloration, and sound externalization [5–7].

Several studies have demonstrated the significant influence of head and pinna anatomy on HRTFs [8–12], with ongoing research focused on identifying key anthropometric features for HRTF individualization [13–15]. However, because of the pinna's complex structure, systematically determining these features remains challenging. Algazi et al. [9] investigated the Center for Image Processing and Integrated Computing (CIPIC) database and found correlations between head size and ITD, and predicted notch frequencies from the pinna rotation, flare angles, and fossa height. They also suggested that additional pinna parameters might be required for an accurate estimation of HRTF features. Iida et al. [8] showed that occluding the scapha, fossa, and conchae resulted in the disappearance of the spectral peak at approximately 4 kHz and the notch at approximately 8–9 kHz, further highlighting the relevance and importance of these anthropometric features on HRTFs.

*To whom correspondence should be addressed, e-mail: stefan.weinzierl@tu-berlin.de

Zhang et al. [10] performed principal component analysis (PCA) on head-related impulse responses for each direction and ear. Multiple linear regression analysis of the resulting PCA components and anthropometric features revealed that head depth significantly affects the weights of the principal components, especially when the sound source is directly above the listener. They also identified seven pinna parameters—cavum conchae width, fossa height, pinna height, pinna width, pinna rotation angle, pinna flare angle, and pinna offset angle—that significantly affect the PCA weights at azimuth angles between -50° and 50°.

Ghorbal et al. [14] employed a parametric model based on blend shapes to generate pinna variations within ± 2 SDs of the population and investigated the effect of these variations on the HRTF magnitude spectrum up to 16 kHz. They found that the cavum conchae width, the fossa height, and the pinna flare angle had the strongest effect on all metrics. Variations in cavum conchae width primarily affect high frequencies above 6 kHz.

Using a similar approach, Stitt and Katz [15] defined 24 control points on an anthropometry-based pinna parametric model, which is a higher number than the ten pinna parameters from the CIPIC database. They then created mesh variations to identify key morphological features for HRTF individualization. Their results indicate that parameters altering the dimensions of the conchae and triangular fossa strongly affect mean spectral distortion and localization errors. Furthermore, they found that relief parameters, such as helix, lobulus, and antihelix, which are less frequently studied, also have a significant impact on HRTFs.

Overall, the most influential pinna parameters for HRTF individualization include cavum conchae height, cavum conchae width, fossa height, pinna height and width, and pinna rotation and flare angles. Additionally, relief features of the helix (root, upper rear, and rear), antihelix, and lobulus are also significant.

Given the great importance of anthropometric features for HRTFs, it is important to understand how sociodemographic factors such as age, sex, and ethnicity affect these features. With this goal in mind, Ban et al. [16] investigated the impact of age on pinna anthropometry in 124 Caucasians. They observed gradual increases in ear height, ear width, cymba conchae height, cavum conchae height, and lobulus height with age. The same study revealed significant sex-based differences among Caucasians, with males exhibiting larger ear height, ear width, and conchae measurements than females, indicating the impact of sex on these anthropometric features.

Furthermore, authors in [17–21] found that the pinna growth differs slightly among certain age groups and even between sexes. Sex-based differences in pinna growth are supported by a large study of $N = 1{,}448$ German individuals from birth to age 92. The authors reported that female ears showed less growth than male ears and found that, after age 20, the growth rate remained constant at 0.19 mm per year for females and 0.26 mm per year for males throughout their lives [22]. Fig. 1 summarizes the available data for the example of pinna height.

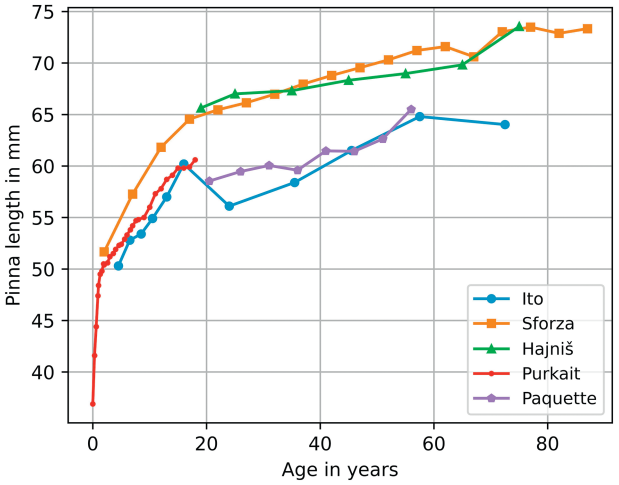


Fig. 1. Pinna height versus age reported in various ear morphology studies. Data are taken from Hajniš [17], Ito [35], Sforza [36], Purkait [37], and Paquette [38], as shown in the legend.

Choi et al. [23] examined the head features of 6,000 U.S. military personnel and found sex-related differences ranging from 5 to 13 mm across four head measurements. Head circumference showed the greatest difference, with males having larger heads than females [23, Fig. 2]. This is also shown in the German standard [24], which additionally shows that head-related features remain constant for adults and do not further grow with age.

Furthermore, Braren and Fels [12] studied the HRTFs of children and demonstrated that the smaller head and pinna geometries of children lead to distinct temporal and spectral differences compared with those of a KEMAR artificial head, with average and maximum ITD differences of 70 µs and 138 µs, respectively, and spectral peaks (second pinna resonance, 5–6 kHz), where KEMAR aligns with the 95th percentile of range of children's data, highlighting the influence of age on head and pinna parameters.

Ethnicity significantly influences head and pinna anthropometry [16, 25–30]. For head morphology, studies reveal distinct differences; for example, Chinese adults ($n = 600$) tend to have rounder heads with shorter lengths and greater widths than Caucasians ($n = 600$) [26], and Hispanic workers exhibit significantly shorter head lengths than Caucasians [31]. Pinna dimensions also vary considerably across ethnic groups. Alexander et al. [27] reported that Indian participants ($n = 56$) had the largest ears (length and width), followed by Caucasians ($n = 288$) and Afro-Caribbeans ($n = 62$). Consistently, Ban et al. [16] found that Koreans ($n = 186$) displayed significantly larger ear height, cavum conchae width, lobulus height, ear depth, and ear rotation angle compared with Caucasians ($n = 124$). These ethnic differences in pinna morphology are further supported by a systematic review highlighting differences among Asian, Western, and African populations [32].

Given the influence of age, sex, and ethnicity on anthropometry, and the relevance of anthropometry for HRTF features, the main objective of this study is to investigate whether age, sex, and ethnicity might be important for HRTF individualization approaches [33, 34], most of

Table 1. Public HRTF databases including 3D meshes.

Name	Subjects	Anthropometry	3D Meshes	Information for subjects
FIU [71]	15	Yes	Pinna only	15 students
RIEC [72]	105	Yes	Head and shoulders	···
SYMARE [73]	61	···	Head and shoulders	48 male, 13 female
ITA [74]	48	Yes	Pinna only	43 Europeans, 3 Asians, 2 South Americans, average age of 29 (5 SD) years
HUTUBS [41]	96	Yes	Head only	93 Students and employees, average age 36 (9 SD) years
SONICOM [75]	372	Yes	Head and shoulders	135 White, 58 Asian/Asian British, 3 Black/African/Caribbean, 10 mixed, 4 other, 1 Arab 133 male, 66 female, average age 30 (10 SD) years
CHEDAR [76], WideSpread [77]	1253	Yes	Head and shoulders	···
3D3A [78]	38	Yes	Head and shoulders	···
SADIE II [79]	20	···	Head only	15 male, 1 female, 2 nonbinary, 2 dummy mannequins, age: 20- 63

CHEDAR = Computed HRIRS and ears database for acoustic research; 3D3A = 3D Audio and Applied Acoustics; FIU = Florida International University; ITA = Institute for Technical Acoustics; RIEC = Research Institute of Electrical Communication; SADIE II = Spatial Audio for Domestic Interactive Entertainment; SYMARE = Sydney-York Morphological and Acoustic Recording of Ears; WideSpread = A Wide Dataset of Ear shapes and Pinna-Related Transfer Functions Generated by Random Ear Drawings.

which rely on HRTF databases acquired from rather homogeneous groups of people (cf. Table 1 for an overview of databases that are often used for HRTF personalization).

To this end, the authors first acquired calculated HRTFs and anthropometric features of 145 European and 111 Asian people systematically stratified by age, sex, and ethnicity. Second, coloration and localization errors were modeled for all combinations of subjects in the database used and investigated in how far errors increase as a result of a mismatch in age, sex, and ethnicity. The results confirm previously reported effects of age, sex, and ethnicity on anthropometry and show that the data here cover a wider range of head and pinna sizes compared with most other databases. Furthermore, a statistically significant influence of age, sex, and ethnicity was observed on all investigated acoustic features. Although effect sizes were generally small, this suggests that these factors influence the perception of nonindividual HRTFs and hence might be important for HRTF individualization.

2 METHODS

2.1 Terminology

Owing to genetic variation, environmental factors, and evolutionary processes, there are statistical differences in ear physiognomy between human populations in widely separated locations around the world that have been repeatedly confirmed empirically [39]. Partly conceptualized as "race," particularly in U.S. studies, ethnicity is understood as a sociohistorical factor and not as a biological category. Age and sex, on the other hand, are biological factors. Although some individuals cannot be clearly assigned to one of the two binary categories of male versus female, this phenomenon has not yet been systematically investigated in existing ear physiognomy research, presumably because of its comparatively infrequent occurrence [40]. Therefore, all available findings, including those of the current study, are based on differences in ear morphology among individuals who clearly identify as male or female.

2.2 Database Acquisition

The METUB-E and METUB-A databases, developed for the current study, comprise 3D head meshes and numerically calculated HRTFs. Both databases are systematically stratified by age and sex, with participants ranging from 18 to 73 years. Specifically, METUB-E includes 145 European head meshes (72 females, 73 males), and METUB-A contains 111 Asian head meshes (55 females, 56 males).

2.2.1 Participants

Because of the presumed differences in head and pinna morphology according to age and sex, 11 age groups between 18 and 73 years old (18–23, 24–28, ..., 69–73) were sampled. Each group spanned 5 years, with the exception of the first group that spanned 6 years. Before data collection, ethical and legal approval was secured from the Ethical Advisory Board, operating under the Research and Development Governance Department of Mercedes-Benz Group. The Advisory Board prohibited the publication of head meshes because they were considered data by which a person can be identified.

For both the databases, the authors aimed at acquiring at least five males and females per group. In case of METUB-E, this criterion could be realized except for the last age group, 69–73 years of age, where it was possible to acquire only three subjects per sex. The three to eight subjects in each group resulted in a sample size of 145 individuals. For METUB-A, five to six subjects were acquired in each group, resulting in a sample size of 111 individuals. Because of legal constraints and practical considerations, it was decided not to include individuals under 18 years of age in the sample.

The ethnicity criterion was established by including only participants who self-reported that they and their parents were born in Europe in case of METUB-E and Asia for METUB-A. In case of METUB-E, because the sample was acquired in Sindelfingen, most participants can be assumed to be from Southwestern Germany. For METUB-A, the sample was acquired by Kantar China, in Beijing City of China, and the participants were randomly sampled from the city. All participants were paid for their participation.

2.2.2 3D Head Meshes

For the acquisition of 3D head meshes, the subjects were seated in an immobile wooden chair and instructed to sit upright with their heads in a natural position. Numbered orientation marks at different heights on the wall in front of them help them maintain their head position and served as fixation points during data acquisition. The chair was immobilized by placing it on a wooden plate with holes for its legs. Two scanners were used to acquire the 3D head meshes: an Artec Eva (resolution: 0.2 mm; working distance: 0.4–1 m) for scanning the heads, and an Artec Spider (resolution: 0.1 mm; working distance: 0.2–0.3 m) for scanning the ears. Subjects wore a swim cap (approximately 1 mm thick) to capture the actual shape of the head and minimize the influence of hair on the scans. Those with long hair were asked to tie it in a ponytail. During postprocessing, ponytails were manually removed using the rectangular and lasso selection eraser tools within Artec Studio software, ensuring no impact on other head regions.

First, a complete head scan was generated by slowly moving the scanner around the subject. To obtain detailed pinna scans, approximately five to six scans were taken of each pinna from different positions to capture detailed features and folds that would be inaccessible with a long-range scanner. For the right pinna scan, the subjects remained seated while the chair was rotated to the remaining four holes to acquire the left pinna scan. Scanning one subject took approximately 15–20 min.

During postprocessing, the meshes from the different scans were fused to create a mesh of the head and upper torso with highly detailed pinna regions. Using the Artec Studio software, all holes were closed to obtain watertight meshes (see [41] for details). The meshes were then cleaned and smoothed further in Blender version 4.3 to yield the final head-and-torso meshes. Additional head-only meshes were obtained by making a cut in Blender using a cutting plane that passed through automatically detected points on the neck (cf. [42, lower points in Fig. 6]). The resulting hole in the neck was manually filled, and the edges were smoothed.

The high-resolution pinna scans contained the ear canal entry, which was used to align the interaural axis (passing through the center of the ear canal entries) with the y-axis of the global coordinate system. Next, the interaural center, defined as the midpoint between the two ear canal entries, was moved to the origin of the coordinate system. Because the detailed geometry of the pinna provides a good estimate of the ear canal entry position, this information was used to manually block the ear canal entries with Blender's sculpting tools. This ensured that the conchae region of the pinna remained undisturbed and only the ear canal entries were blocked. All heads are viewed in the positive x-direction, which is consistent with the widely used spatially oriented format for acoustics (SOFA) file format [43]. Fig. 2 shows an example head and a pinna mesh acquired in the database.

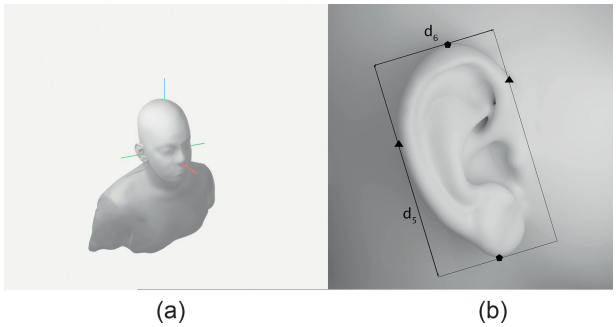


Fig. 2. (a) Example head mesh with red, green, and blue lines indicating the x, y, and z-axis. (b) Right pinna mesh of the same head showing the definition of, ear height (d_5) and ear width (d_6).

2.2.3 Numerical HRTF Calculation

The HRTFs were numerically calculated using the multilevel fast multipole method in combination with the Boundary element method (BEM), as implemented in Mesh2HRTF v1.2.3 [44, 45]. The authors calculated 128 linearly spaced frequencies of the single-sided HRTF spectrum up to 22.05 kHz in steps of 172.27 Hz (excluding 0 Hz). The HRTFs were calculated using a Lebedev grid with a radius of 1.0 m and a spherical harmonic order of 44, resulting in an evaluation grid of 2,702 source positions.

To speed up the calculations, acoustic reciprocity was exploited by separately calculating the left and right ear HRTFs using volume velocity sources placed at the center of the blocked ear canals. To further reduce the calculation time and memory requirements, the high-resolution head meshes were automatically remeshed to a target edge length of 0.8 mm at the ear of interest (sufficient for obtaining perceptually valid HRTFs, as described in [46]), and 10-mm target edge length at the contralateral ear. The remeshing was realized using curvature-based hybrid mesh grading [47], resulting in 20,000–30,000 faces per mesh.

All calculations were performed on the high-performance cluster of Technische Universität Berlin. Calculating one HRTF set (left and right ear) took approximately 2 h using one CPU node of the cluster. In total, 145 HRTFs were calculated in less than a day. During postprocessing, the HRTFs of the left and right pinna were merged, the 0 Hz bin was set to 1 (0 dB), and the single-sided spectrum was mirrored under consideration of the symmetry properties of the Fourier transform of real-valued time signals to obtain a both-sided spectrum. In a last step, head-related impulse responses with a sampling rate of 44.1 kHz length of 256 samples were obtained via an inverse Fourier transform (see [44, SEC. 3] for details) and saved as SOFA files [48].

2.3 Anthropometric Features

Head width and head depth were automatically extracted based on projected 2D outlines of the head meshes [9, 42], and pinna height and width were manually measured on the 3D meshes in Blender using its measurement tool. The authors followed the Anthropometric Survey (ANSUR) II definition of pinna height as the Euclidean distance between the lowest and highest points along the long axis of the ear and pinna width as the maximum breadth of the ear perpendicular to its long axis [38] [see Fig. 2(b)].

2.3.1 Statistical Analyses

The anthropometric features were analyzed by applying a general linear model (GLM) to the ANSUR II and METUB databases using *age*, *sex*, *ethnicity* and their interaction *age* × *sex*, *age* × *ethnicity*, and *sex* × *ethnicity* as independent variables and the anthropometric features as dependent variables. The analysis was done in R using the jamovi graphical user interface (v2.6.19) [49] and the gamlj package (v2.6.6) [50]. GLMs require normally distributed residuals, which were visually confirmed using histograms and quantile-quantile plots.

Because the ANSUR II database does not contain Europeans and does not detail the criteria used to define ethnicity, it was assumed that the *White* subsample was the closest match for a comparison with database used for this study. For all statistical analyses, the data of the right pinna were used for two reasons. First, the ANSUR II database provides only right ear data [51], and second, previous studies report strong bilateral ear symmetry with large bivariate Pearson correlation coefficients for ear height and width ranging from 0.74 to 0.96 in males ($n = 201$) and females ($n = 213$), across Caucasian ($n = 288$), Afro-Caribbean ($n = 62$), and Indian subcontinent ($n = 56$) populations [27]. Additionally, a study by Bozkir et al. [52] on a Turkish population ($n = 191$ males, $n = 150$ females) indicated that bilateral asymmetry in total ear height and width was less than 1 mm.

The authors only compared METUB to ANSUR II because the public databases listed in Table 1 are not stratified by age, sex, or ethnicity.

2.4 Acoustic Features

To investigate to what extent differences in age, sex, and ethnicity can explain differences in acoustic features Φ, median plane localization, differences in binaural localization cues (ITDs and ILDs), and the auditory basic spectral differences (BSDs) were modeled as a predictor for coloration. Each feature was computed for all combinations of subjects from the European and Asian database, resulting in a 256 × 256 matrix of 65,536 values per feature. In all cases, the values on the diagonal were discarded because they denote zero differences between HRTFs of the same subject. Only the upper triangular matrix was computed for symmetric features $\Phi(i, j) = \Phi(j, i)$, where i and j denote the subject number.

Median plane localization was modeled using the probabilistic model of Baumgartner et al. [53]. The model expects a template and a target set of directional transfer functions (DTFs) and estimates localization performance by comparing the positive spectral gradients extracted from the DTFs. As noted above, template and target DTFs from different subjects were used, thereby modeling localization performance for nonindividual HRTFs. The sensitivity of the model was set to 0.21, thereby simulating localization performance of the best (most sensitive) listener reported by Baumgartner et al. [53]. This was done to emphasize possible effects of age, sex, and ethnicity and can be expected to result in relatively large localization errors resulting from modeling localization performance with nonindividual HRTFs.

To compute localization errors in the median plane, the HRTFs were interpolated from the original grid of 2,702 sampling points to a grid of 360 points in the sagittal plane using spherical harmonic interpolation of order 44 [54] and the authors converted them to DTFs using *SOFAhrtf2dtf* from the SOFA toolbox. For each combination of subjects, polar errors (PEs) and quadrant errors (QEs) were computed. QE describes the percentage of localization errors greater than 90°, and PE gives the root mean square error of responses inside the hemifield (i.e., for errors smaller than 90° [6, 55]). This feature is asymmetric and was computed for the entire 256 × 256 feature matrix excluding the diagonal.

Human sound localization in the horizontal direction relies on ITDs and ILDs, both of which are binaural cues [1]. To compute ITD errors, the authors first estimated ITDs for each subject and source position using the onset threshold detection method [56] from the *itdestimator* contained in the auditory modeling toolbox [57]. Second, absolute ITD differences between each pair of subjects were computed and converted to just noticeable difference (JND) units. To convert the differences to JND units, the authors estimated the JND for a given lateral angle using linear interpolation between the JND values of 20 µs for 0° and 100 µs for 90° based on the findings of Mossop and Culling [58] and divided the differences by the respective JND.

Broadband ILDs were estimated as energetic differences between the left and right ear HRTFs; absolute differences were computed for all combination of subjects, and the JND was assumed to be 1 dB independent of the source position [1].

In a last step, ITD and ILD differences were averaged across source positions using the quadrature weights of the Lebedev sampling grid. ITD and ILD are symmetric features owing to taking absolute differences between subjects and were computed only for the upper triangular 256 × 256 feature matrix.

Lastly, the auditory BSD_A [59, Eq. (4)] was used as a simple predictor for coloration, which computes energetic differences between sets of HRTFs in auditory bands separately for the left and right ear. The results were analyzed in five frequency bands (0.1–1 kHz, 1–5 kHz, 5–10 kHz, 10–15 kHz, and 15–20 kHz) by taking the unweighted absolute BSD_A average across source positions and ears and within each of the five frequency bands. To speed up calculation, the original grid was interpolated to a Fliege-Maier

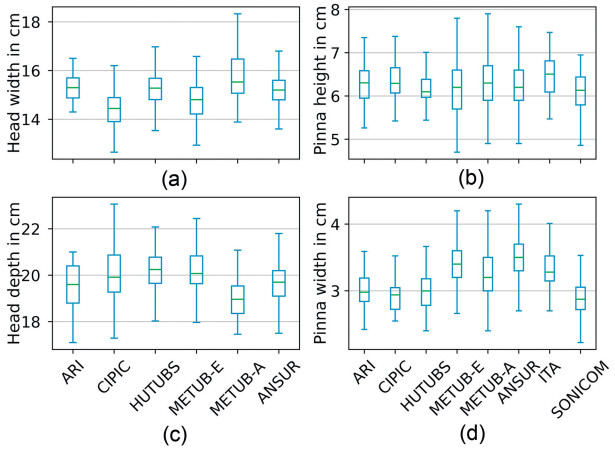


Fig. 3. Distribution of head and pinna parameters in different databases. Head width (a), head depth (c), pinna height (b), and pinna width (d). Horizontal lines indicate the median value; boxes indicate the interquartile range; and whiskers indicate the minimum and maximum values.

spherical sampling grid [60] of 324 points using spherical harmonic interpolation of order 44. The BSD_A is a symmetric feature and was computed only for the upper triangular 256×256 feature matrix.

2.4.1 Statistical Analysis

To investigate the extent to which differences in the acoustic features can be explained by differences in age, sex, and ethnicity, a separate GLM analysis was performed using each acoustic feature Φ as dependent variable, and *age difference*, *sex difference*, *ethnicity difference*, and their interactions *age difference* $\times$ *sex difference*, *age difference* $\times$ *ethnicity difference*, and *sex difference* $\times$ *ethnicity difference* as independent variables. Sex and ethnicity differences between subjects were dummy coded by assigning a value of 0 to subjects of the same category and a value of 1 otherwise. Normality of distributed residuals was again visually confirmed using histograms and quantile-quantile plots.

The extent to which differences in acoustic features can be explained by differences in anthropometric features was also studied by employing a GLM for each feature (Φ) as the dependent variable and absolute differences in *pinna length*, *pinna width*, *head width*, and *head depth* as the independent variables. This was considered a baseline for providing the maximum amount of variance that could be explained by differences in age, sex, and ethnicity (if anthropometric and acoustic features were fully correlated, which is not to be expected).

3 RESULTS

3.1 Distribution of Anthropometric Features

Fig. 3 shows the distributions of anthropometric features from METUB-E, METUB-A, and five publicly available databases. For pinna length, the databases cover the largest range of values. For pinna width, the range is comparable to ANSUR II and larger than the that of the remaining databases. This indicates that the METUB-E and METUB-A data are more diverse for these cases. For the head width and depth, the range covered by the database used here is comparable to that of other databases, with METUB-A having the largest range of values in head width. In case of head depth, the CIPIC database features a larger range of values. The range of values for the CIPIC database is comparable to most other databases for the head width but larger compared with all other databases for the head depth. Thus, the large range of head depths may be a measurement artifact. However, this can only be hypothesized because the CIPIC database contains no information about how the anthropometric parameters were measured.

Table 2. Explained variance (partial η_p^2) in percentage and significance values for the models predicting anthropometric measures (pinna height and width, and head width and depth) from age, sex, and ethnicity for the ANSUR II (A) and METUB (M) databases.

	Model		
Variable	Database	η_p^2	p
Pinna height	A	30.3	<**0.01**
	M	48.4	<**0.01**
Pinna width	A	20.9	<**0.01**
	M	28.9	<**0.01**
Head width	A	26.2	<**0.01**
	M	52.3	<**0.01**
Head depth	A	36.6	<**0.01**
	M	52.3	<**0.01**

Note. Bold numbers denote statistically significant effects.

3.2 Effect of Age, Sex, and Ethnicity on Anthropometric Features

Tables 2, 3 and Fig. 4 give the results of the GLMs, analyzing how far anthropometric features depend on sociodemographic factors in ANSUR and METUB. Across the investigated measures, the majority of the variance was explained by age and sex, followed by ethnicity. Age consistently emerges as a significant factor influencing pinna measures, particularly for METUB. For pinna height, age explains approximately 33% of the variance in METUB and 2.3% in ANSUR. Similarly, for pinna width, age accounts for 16% variance in METUB and 0.3% in ANSUR. These effects show that pinnae grow with age, with an average increase of 0.2 mm/year in METUB and 0.18 mm/year in ANSUR for pinna height and 0.08 mm/year in METUB and 0.05 mm/year in ANSUR for pinna width. The fact that age explains less variance in ANSUR might indicate larger errors in the anthropometric measurements that were done manually and probably by multiple experimenters.

Age was not a significant predictor for head width in either database. However, age significantly influenced head depth in both the ANSUR and METUB databases, with a larger effect of 2.6% explained variance in METUB compared with approximately 0.1% in ANSUR, suggesting a slight dependency on age, with a decrease of 0.07 mm/year

Table 3. Statistical analysis of ANSUR II and METUB databases based on GLM analysis.

	ANSUR II							METUB						
Predictor	*B*	95% CI	SE	β	*t*	*p*	η_p^2	*B*	95% CI	SE	β	*t*	*p*	η_p^2
							Pinna height							
Intercept	56.49	[55.35, 57.63]	0.58	0.00	97.21	<**0.01**	···	52.88	[51.00, 54.60]	0.90	0.00	58.18	<**0.01**	···
Age	0.18	[0.14, 0.22]	0.01	0.33	9.72	<**0.01**	2.3	0.20	[0.17, 0.24]	0.01	0.50	11.04	<**0.01**	32.9
Sex	3.00	[1.91, 4.20]	0.58	1.03	5.21	<**0.01**	0.7	9.24	[5.68, 12.80]	1.81	0.90	5.10	<**0.01**	9.5
Ethnicity	–1.80	[–4.10, 0.36]	1.10	–0.02	–1.65	0.10	0.1	4.29	[0.70, 7.80]	1.81	0.14	2.36	**0.02**	2.2
Age × sex	0.06	[0.02, 0.09]	0.01	0.11	3.68	<**0.01**	0.3	–0.07	[–0.14, 0.00]	0.03	–0.17	–1.96	0.05	1.5
Age × ethnicity	0.05	[–0.01, 0.13]	0.03	0.10	1.58	0.11	0.1	–0.07	[–0.14, 0.00]	0.03	–0.18	–1.97	<**0.05**	1.5
Sex × ethnicity	0.70	[–0.50, 1.91]	0.61	0.14	1.14	0.25	0	–0.90	[–3.27, 1.36]	1.17	–0.14	–0.81	0.42	0.3
							Pinna width							
Intercept	32.66	[31.85, 33.47]	0.41	0.00	78.98	<**0.01**	···	29.10	[27.80, 30.30]	0.62	0.00	46.59	<**0.01**	···
Age	0.05	[0.02, 0.07]	0.01	0.13	3.65	<**0.01**	0.3	0.08	[0.06, 0.11]	0.01	0.30	6.88	<**0.01**	15.9
Sex	2.64	[1.82, 3.46]	0.41	0.90	6.32	<**0.01**	1	1.38	[–1.06, 3.84]	1.24	0.50	1.12	0.27	0.5
Ethnicity	–0.25	[–1.84, 1.34]	0.81	–0.30	–0.30	0.76	0	–1.54	[–4.00, 0.90]	1.24	–0.50	–1.23	0.22	0.6
Age × sex	0.01	[–0.01, 0.03]	0.01	0.02	0.67	0.51	0	0.01	[–0.03, 0.06]	0.02	0.07	0.67	0.50	0.2
Age × ethnicity	–0.02	[–0.07, 0.02]	0.02	–0.06	–0.92	0.36	0	–0.01	[–0.06, 0.04]	0.02	–0.04	–0.40	0.68	0.1
Sex × ethnicity	–0.06	[–0.90, 0.70]	0.40	–0.02	–0.16	0.88	0	–0.50	[–2.10, 1.07]	0.80	–0.13	–0.64	0.52	0.2
							Head width							
Intercept	151.87	[150.37, 153.37]	0.76	0.00	198.68	<**0.01**	···	153.08	[150.44, 155.73]	1.34	0.00	114.05	<**0.01**	···
Age	0.03	[–0.01, 0.08]	0.02	0.04	1.21	0.22	0	–0.01	[–0.06, 0.04]	0.02	–0.01	–0.20	0.84	0
Sex	6.06	[4.55, 7.58]	0.77	1.07	7.85	<**0.01**	1.5	12.11	[6.84, 17.38]	2.67	1.09	4.53	<**0.01**	7.6
Ethnicity	1.79	[–1.16, 4.74]	1.50	0.74	1.18	0.23	0	17.59	[12.30, 22.88]	2.68	0.92	6.55	<**0.01**	14.7
Age × sex	0.01	[–0.02, 0.06]	0.02	0.02	0.75	0.45	0	–0.02	[–0.13, 0.07]	0.05	–0.04	–0.54	0.59	0.1
Age × ethnicity	0.09	[–0.00, 0.19]	0.04	0.13	1.86	0.06	0.1	–0.18	[–0.29, –0.07]	0.05	–0.29	–3.34	<**0.01**	4.3
Sex × ethnicity	–0.72	[–2.30, 0.87]	0.80	–0.11	–0.88	0.37	0	1.88	[–1.53, 5.30]	1.73	0.19	1.08	0.28	0.5
							Head depth							
Intercept	193.36	[191.43, 195.28]	0.98	0.00	196.78	<**0.01**	···	199.29	[196.52, 202.06]	1.40	0.00	141.71	<**0.01**	···
Age	–0.07	[–0.13, –0.01]	0.03	–0.07	–2.33	**0.02**	0.1	–0.07	[–0.13, –0.02]	0.02	–0.11	–2.56	**0.01**	2.6
Sex	7.11	[5.16, 9.05]	0.99	1.17	7.16	<**0.01**	1.3	11.40	[5.80, 16.90]	2.80	0.80	4.06	<**0.01**	6.2
Ethnicity	–3.16	[–6.96, 0.63]	1.93	–0.75	–1.63	0.10	0.1	–12.09	[–17.60, –6.50]	2.81	–1.15	–4.29	<**0.01**	6.9
Age × sex	0.09	[0.03, 0.15]	0.02	0.09	3.25	<**0.01**	0.3	–0.06	[–0.18, 0.04]	0.05	–0.10	–1.15	0.25	0.5
Age × ethnicity	–0.10	[–0.23, 0.01]	0.06	–0.11	–1.71	0.08	0.1	0.00	[–0.11, 0.11]	0.05	0.00	0.05	0.96	0
Sex × ethnicity	–2.41	[–4.40, –0.36]	1.04	–0.28	–2.31	**0.02**	0.1	–1.72	[–5.28, 1.88]	1.81	–0.17	–0.93	0.35	0.4

Note: Predictors include age, sex (male and female), and ethnicity (Asian and White). Statistical significance was set at $p < 0.05$. Bold numbers denote statistically significant effects. All p-values are rounded to two decimal places. The models are specified by the intercept in millimeters (obtained by evaluating the regression line for an age of 0 years), the mean difference for sex in millimeters, the growth in millimeters per year (age), and the mean difference for ethnicity in millimeters. Bold numbers denote statistically significant effects.
B = unstandardized coefficient;β = standardized coefficient; t = t-statistic; η_p^2 = partial eta squared as explained variance in percentage.

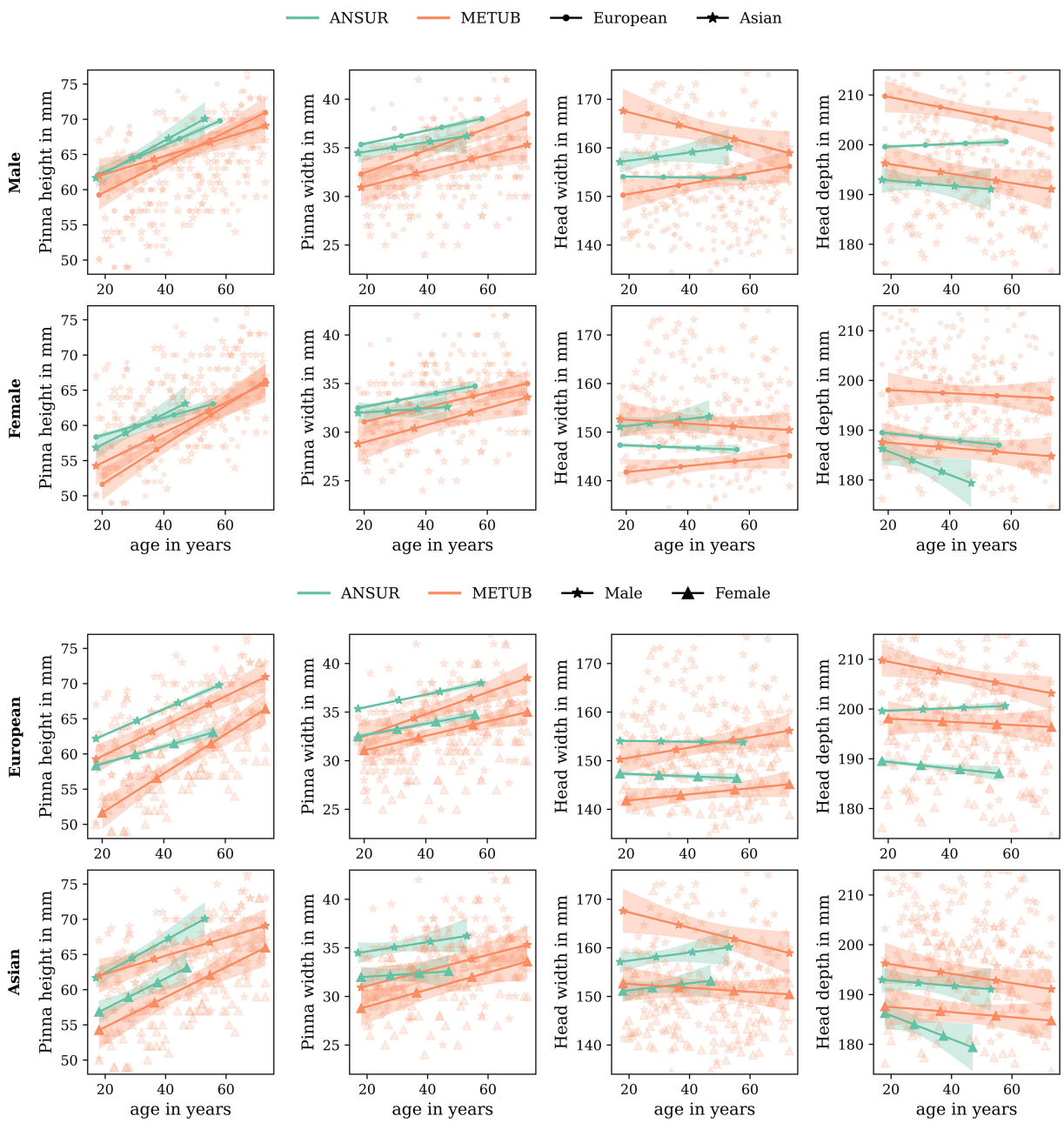


Fig. 4. Regression lines for ANSUR (green) and METUB (orange) are shown, overlaid on a scatter plot of the METUB database. The shaded areas represent the 95% CIs around the mean values. The first two rows present data separated by sex (male and female), and the last two rows represent data separated by ethnicity (European and Asian). The x and y-axis show the range of values covered by the METUB databases.

in both databases. It was also observed that there was a significant age $\times$ ethnicity interaction for head width, explaining approximately 4.3% variance in METUB. In this case head width increases with age for the European sample but decreases for the Asian sample.

In both the databases, sex was a significant predictor for head and pinna measures, except for pinna width in METUB. For pinna height, sex explained 9.5% of the variance in METUB, which was higher than 0.7% in ANSUR. In case of pinna width, sex was a significant factor in ANSUR, accounting for 1% variance, but not in METUB, where the explained variance was only 0.5%. On average, males have 9.2 mm larger and 1.3 mm wider ears than females (in the case of METUB) and 3 mm and 2.64 mm for ANSUR.

Similarly, for head measures, METUB showed larger effects than ANSUR. For head width, sex accounted for 7.6% variance in METUB versus 1.5% in ANSUR. For head depth, sex was a significant predictor, explaining approximately 6.2% variance in METUB compared with 0.1% in ANSUR. Males demonstrated larger head widths than females: 12.1 mm in METUB and 6 mm in ANSUR. Head depths were also greater in males: 11.4 mm in METUB and 7 mm in ANSUR. Overall, sex-related effects were larger in METUB than in ANSUR, and it is not clear what caused these differences.

The effect of ethnicity was less pronounced on pinna features than on head features, yet it significantly predicted all features in the METUB database except pinna width. Ethnicity was not a significant predictor for any feature in the ANSUR database. For pinna height, ethnicity contributed 2.2% of variance in METUB. In this case, Asians had 4.3 mm larger pinna heights than Europeans, whereas Europeans had 1.54 mm larger pinna width than Asians. Moreover, ethnicity accounted for 14.7% of variance in head width and 7% in head depth, with Asians exhibiting 17.6 mm wider heads and Europeans having 12.1 mm deeper heads.

3.3 Effect of Age, Sex, and Ethnicity on Acoustic Features

Fig. 5 shows the calculated left-ear HRTFs from both databases grouped by sex and age for an ipsilateral source on the horizontal plane at an azimuth angle of 90° and an elevation of 0°. For the selected source position, HRTFs averaged across age, sex, or ethnicity, are similar up to approximately 2 kHz. In this frequency range, the HRTF is dominated by acoustic cues evoked by the head, whereas it is dominated by cues generated by the pinna above approximately 3 kHz [61]. Above this frequency, three distinct peaks at approximately 6, 12, and 19 kHz and two dips at approximately 10 and 15 kHz can be seen in the averaged HRTFs.

Fig. 5 (first row) shows that the frequencies of these peaks and dips are on average smaller for the European sample compared with the Asian sample and for male compared with female. In both cases, this can be related to the effect of these factors on pinna height and width (cf. SEC. 3.2). For age, this effect is less clear, but there is a tendency for the gains of the first peak to increase with age in both samples.

3.3.1 Median Plane Localization

The GLM analysis revealed small yet significant effects of age, sex, and ethnicity differences on median plane localization errors, and the results are summarized in Tables 4 and 5. Overall, sex and ethnicity differences were the most significant factors, followed by age differences. These systematic, albeit small, differences in PEs and QEs are also illustrated in the first row of Fig. 6. For PEs, the model explains 3.1% of the variance, with sex differences accounting for 1% and ethnicity differences counting for 0.7% of variance. PEs are on average 1.03° larger if subjects are not of same sex and 0.96° larger when subjects do not share the same ethnicity. Moreover, they increase by 0.01° per year of age difference.

Similarly for QEs, the model explains 2.4% of the variance in total, sex differences explain 1.1%, and ethnicity differences explain 0.3%. QEs were on average 2.1% larger for subjects of different sexes and increased by 1.21% with differing ethnicity. Age also had a minor effect, increasing QEs by 0.01% per year.

Fig. 6 shows that these effects are approximately constant for age differences up to 35 years. Beyond this, the results get more noisy, as indicated by the increasing confidence intervals. This is presumably caused by the fact that there are less cases of subjects in the database with such large age differences, and this might have caused the significant interactions with age that, however, explain only 0.1% of the variance.

Table 4. Variance (partial η_p^2) in percentage and significance values for the models predicting differences between acoustic features using demographic factors (D; difference in age, sex, and ethnicity) and anthropometric measures (A; differences in pinna height and width and head width and depth).

	Model		
Acoustic Feature	Predictors	η_p^2	p
PE	D	3.1	<**0.01**
	A	11.8	<**0.01**
QE	D	2.4	<**0.01**
	A	11.0	<**0.01**
ITD	D	14.6	<**0.01**
	A	51.0	<**0.01**
ILD	D	0.1	<**0.01**
	A	0.6	<**0.01**
Coloration (by frequency band)			
Band 1: 0.1–1 kHz	D	10.3	<**0.01**
	A	21.3	<**0.01**
Band 2: 1–5 kHz	D	3.5	<**0.01**
	A	11.0	<**0.01**
Band 3: 5–10 kHz	D	2.0	<**0.01**
	A	10.4	<**0.01**
Band 4: 10–15 kHz	D	1.5	<**0.01**
	A	2.4	<**0.01**
Band 5: 15–20 kHz	D	6.0	<**0.01**
	A	4.0	<**0.01**

Note. Bold numbers denote statistically significant effects.

Using differences in head and pinna features as independent variables directly, larger and significant effects on differences in localization errors with the models for PE and QE were observed, accounting for 11.8% and 11% of the variance, respectively.

Pinna length differences were the primary drivers, explaining 5.4% of the variance in PE and 5.9% of the variance in QE, whereas pinna width differences explained between 1.8% to 2.7% of the variance. On average, PEs increase by 1.3 degrees per centimeter of pinna length difference and by 1.5 degrees per centimeter of width difference. QEs increase by approximately 2.6% and 2.3% per centimeter of pinna length and width difference, respectively. Differences in head depths also showed a small but significant effect, explaining approximately 2.6% of the variance for both PE and QE. PEs increase by 0.5 degrees and QEs by 0.8% per centimeter of head depth difference. Differences in head width explain no variance in PE and approximately 0.7% of the variance in QE.

3.3.2 Horizontal Plane Localization

For the ITD, the model accounts for 14.6% of the variance, with sex differences explaining the largest portion of approximately 6%. Ethnicity differences contribute a smaller 0.2% of the variance, whereas age differences did not explain any variance in the ITD. These effects are summarized in Tables 4 and 5 and are clearly seen in the sec-

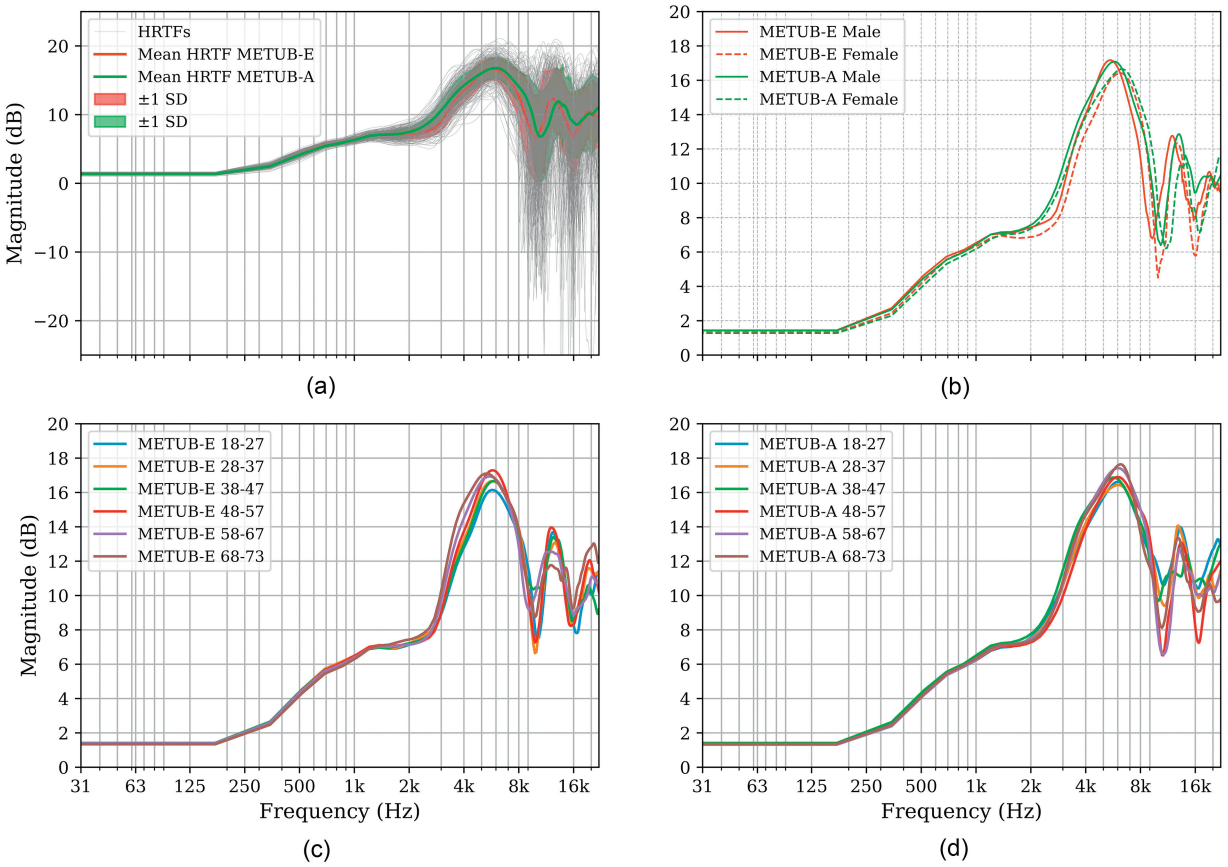


Fig. 5. Calculated METUB-E and METUB-A left-ear HRTFs for a sound source positioned at $(\theta, \phi) = (90,0)$. (a) HRTFs for all 145 and 111 subjects from both databases, along with ±1 SD around the mean. (b) Average male and female HRTFs across both databases. (c) Average HRTFs for six age groups from METUB-E. (d) Average HRTFs for six age groups from METUB-A.

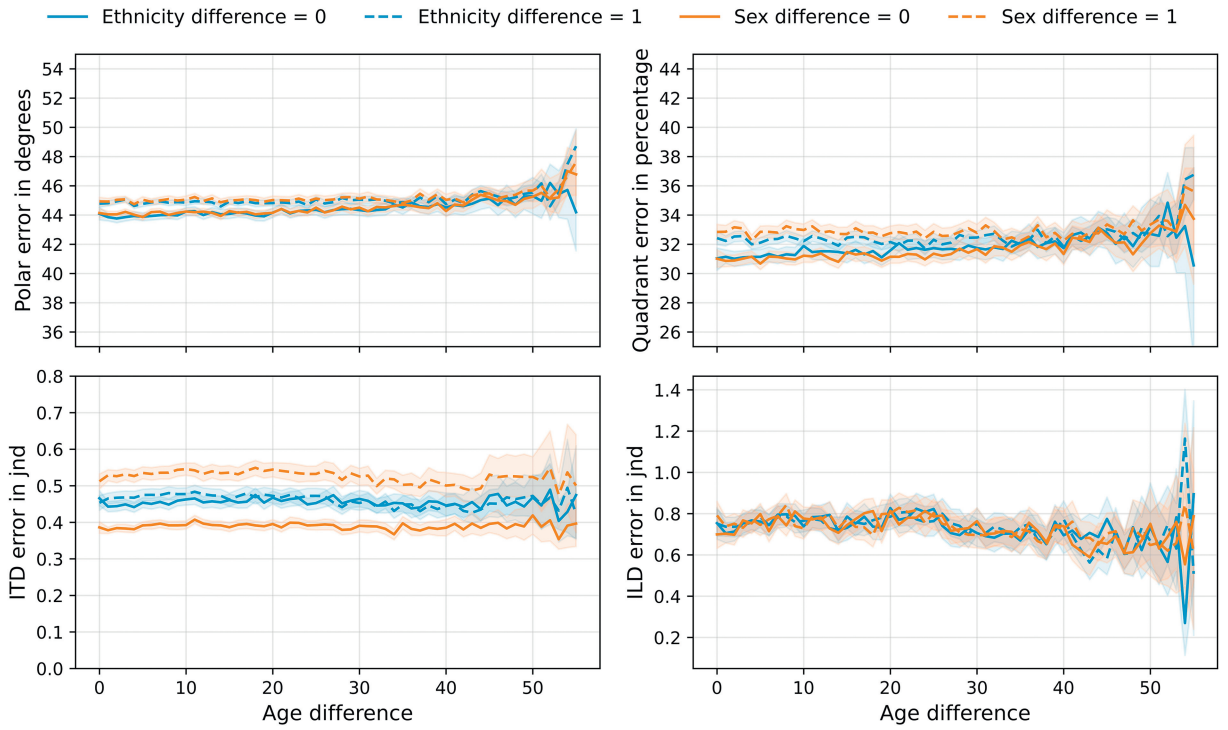


Fig. 6. Localization errors (PE, QE, ITD, and ILD) are shown as a function of age, sex and ethnicity differences. Lines represent mean errors. The shaded areas represent the 95% CIs around the mean values. The full range of observed PE was 31° to 55.4°, for QE it was 7.46% to 56.2%, for ITD error it was 0.07 to 1.85 JND, and for ILD error it was 0 to 5.47 JND. Variables and differences are distinguished by colors and line styles (see legend). The *x*-axis indicates the age difference in years, and *y*-axis indicates the mean errors.

Table 5. Statistical analysis of acoustic features of METUB based on GLM analysis.

	Localization error							Coloration error						
Predictor	B	95% CI	SE	β	t	p	η_p^2	B	95% CI	SE	β	t	p	η_p^2
	PE							Band 1: 0.1−1 kHz						
Intercept	44.28	[44.24, 44.33]	0.02	0.00	2037.11	<**0.01**		0.19	[0.19, 0.192]	0	0	327.15	<**0.01**	
Δ_A	0.015	[0.012, 0.016]	0	0.05	14.71	<**0.01**	0.3	0	[0,0]	0	0	0.08	0.93	0
Δ_S	1.03	[0.94, 1.11]	0.04	0.2	23.75	<**0.01**	0.9	0.04	[0.03, 0.04]	0	0.6	35.69	<**0.01**	3.8
Δ_E	0.96	[0.88, 1.05]	0.04	0.21	22.20	<**0.01**	0.7	0.01	[0.010, 0.014]	0	0.18	10.85	<**0.01**	0.4
$\Delta_A \times \Delta_S$	−0.01	[−0.01, −0.009]	0.001	−0.05	−6.5	<**0.01**	0.1	0	[0,0]	0	−0.02	−2.43	**0.01**	0
$\Delta_A \times \Delta_E$	−0.01	[−0.018, 0.010]	0.001	−0.05	−7.23	<**0.01**	0.1	0	[0,0]	0	0	−0.5	0.57	0
$\Delta_S \times \Delta_E$	−0.11	[−0.2, −0.01]	0.05	−0.03	−2.31	**0.02**	0	0	[-0.003, 0.001]	0.001	−0.01	−0.5	0.58	0
	QE							Band 2: 1−5 kHz						
Intercept	31.65	[31.57, 31.73]	0.004	0	774	<**0.01**		1.14	[1.14, 1.15]	0	0	366.67	<**0.01**	
Δ_A	0.01	[0.012, 0.019]	0.001	0.03	8.71	<**0.01**	0.1	0	[0,0]	0	0.04	7.95	<**0.01**	0.2
Δ_S	2.1	[2,2.3]	0.08	0.2	26.36	<**0.01**	1.1	0.11	[0.09, 0.12]	0.006	0.31	17.69	<**0.01**	1.0
Δ_E	1.21	[1,1.3]	0.08	0.12	14.80	<**0.01**	0.3	0.07	[0.06, 0.09]	0.006	0.18	12.47	<**0.01**	0.5
$\Delta_A \times \Delta_S$	−0.03	[-0.03, -0.02]	0.003	−0.06	−8.57	<**0.01**	0.1	0	[0,0]	0	−0.01	−1.24	0.21	0
$\Delta_A \times \Delta_E$	−0.02	[-0.03, -0.01]	0.003	−0.05	−6.82	<**0.01**	0.1	0	[0,0]	0	−0.03	−3.3	<**0.01**	0
$\Delta_S \times \Delta_E$	−0.13	[-0.32, 0.046]	0.09	−0.02	−1.46	0.14	0	−0.001	[-0.01, 0.01]	0.007	−0.004	−0.21	0.82	0
	ITD							Band 3: 5−10 kHz						
Intercept	0.46	[0.461, 0.468]	0.001	0	282.48	<**0.01**		2.81	[2.79, 2.82]	0.005	0	476.33	<**0.01**	
Δ_A	0	[0,0]	0	−0.01	−3.6	<**0.01**	0	0.002	[0.002, 0.003]	0	0.05	9.61	<**0.01**	0.3
Δ_S	0.14	[0.143, 0.15]	0.003	0.75	45.41	<**0.01**	6	0.13	[0.11, 0.15]	0.13	0.19	11.28	<**0.01**	0.4
Δ_E	0.025	[0.01, 0.03]	0	0.06	7.8	<**0.01**	0.2	0.11	[0.08, 0.13]	0.01	0.17	9.50	<**0.01**	0.3
$\Delta_A \times \Delta_S$	0	[0,0]	0	−0.03	−3.4	<**0.01**	0	0	[0,0]	0	−0.01	−1.25	0.21	0
$\Delta_A \times \Delta_E$	0	[0,0]	0	−0.05	−5.3	<**0.01**	0.1	0	[0,0]	0	−0.006	−0.59	0.55	0
$\Delta_S \times \Delta_E$	0.002	[-0.004, 0.01]	0	0.01	0.68	0.49	0	−0.004	[−0.03, 0.02]	0.01	−0.007	−0.36	0.71	0

Table 5. *Continued*

Predictor	Localization error							Coloration error						
	B	95% CI	SE	β	*t*	*p*	η_p^2	*B*	95% CI	SE	β	*t*	*p*	η_p^2
	ILD							Band 4: 10−15 kHz						
Intercept	0.77	[0.76, 0.78]	0.005	0	133.80	<**0.01**		3.68	[3.67, 3.70]	0.007	0	517.15	<**0.01**	
Δ_A	0	[0, 0]	−0.001	−0.03	−5.84	<**0.01**	0.1	0.002	[0.002, 0.003]	0	0.04	8.22	<**0.01**	0.2
Δ_S	0.01	[−0.01, 0.03]	0.011	0.007	0.87	0.38	0	0.05	[0.02, 0.08]	0.01	0.05	3.89	<**0.01**	0
Δ_E	0	[−0.02, 0.02]	0.011	0	0.02	0.98	0	0.20	[0.18, 0.23]	0.01	0.21	14.67	<**0.01**	0.7
$\Delta_A \times \Delta_S$	0	[0,0]	0	−0.006	−0.6	0.54	0	0	[0,0]	0	−0.01	−1.35	0.18	0
$\Delta_A \times \Delta_E$	0	[0, 0]	0	−0.003	−0.31	0.75	0	−0.002	[−0.004, −0.001]	0	−0.04	−4.4	<**0.01**	0.1
$\Delta_S \times \Delta_E$	0	[−0.02, 0.02]	0.01	0	−0.03	0.97	0	−0.04	[−0.07, −0.009]	0.01	−0.05	−2.53	**0.01**	0
								Band 5: 15−20 kHz						
Intercept								3.77	[3.76, 3.78]	0.006	0	559.6	<**0.01**	
Δ_A								0.0036	[0.003, 0.004]	0	0.06	11.84	<**0.01**	0.4
Δ_S								0.10	[0.07, 0.12]	0.01	0.11	7.50	<**0.01**	0.2
Δ_E								0.38	[0.35, 0.40]	0.01	0.45	28.17	<**0.01**	2.4
$\Delta_A \times \Delta_S$								0	[0,0]	0	−0.01	−1.82	0.07	0
$\Delta_A \times \Delta_E$								−0.003	[−0.004, -0.001]	0	−0.05	−4.9	<**0.01**	0.1
$\Delta_S \times \Delta_E$								−0.001	[−0.03, 0.02]	0.01	−0.001	−0.07	0.93	0

* Note. Predictors include ΔA as age difference, ΔS as sex difference (different sex–same sex), and ΔE as ethnicity difference (different ethnicity–same ethnicity). Statistical significance was set at $p < 0.05$. Bold numbers denote statistically significant effects. All p-values are rounded to two decimal places. Bold numbers denote statistically significant effects.
B, unstandardized coefficient; β = standardized coefficient; *t* = t statistic; η_p^2 = partial eta squared as explained variance in percentage.

ond row of Fig. 6. All interaction terms are significant in predicting ITD, except for the sex × ethnicity interaction. Differences in ITD because of sex and ethnicity differences are on average only 0.14 and 0.025 JND units and hence potentially inaudible. Although the model predicting ILD differences is statistically significant, it explains only 0.1% of the variance, which was considered to be negligible.

A second GLM analysis using head and pinna differences as independent variables revealed significant effects on ITD explaining between 0.3% to 44.7% of the variance. Overall, this model accounted for 51% of the variance, with differences in head width accounting for 44.7% of the variance and ITD errors increasing by 0.14 JND units/cm difference. Pinna height differences accounted for 10.4% of the variance and resulted in an ITD error increase of 0.08 JND units/cm difference. The remaining anthropometric parameter differences were significant as well but contributed changes of less than 0.03 JND units/cm difference. For the ILD, only pinna length and width differences are significant predictors, however, both explaining less than 0.5% of the variance.

3.3.3 Coloration

Fig. 7 shows BSD_A values ranging from 0.03 to 9.27 dB across selected frequency bands grouped by sex and ethnicity differences. Overall, the models explain between 1.4 and 10.3% of the variance in coloration. Sex and ethnicity differences were more influential than age differences. Sex differences account for up to 3.8% of the variance, observed in the 0.1–1 kHz frequency band. Ethnicity differences explain up to 2.4% of the variance in case of the 15–20 kHz frequency band. Age differences had the smallest effect, contributing up to 0.4% of the variance. Although some interaction terms were significant, their effect sizes were considered negligible (see Table 5). However, the influence on coloration was small in all cases, with differences in sex and ethnicity resulting in differences in coloration of less then 0.4 dB on average and the effect of age being close to 0 dB.

A second GLM analysis revealed significant effects of anthropometric feature differences on coloration, with the models explaining between 2.4% (band 4) and 21.3% (band 1) of the variance. Differences in in head parameters explain up to 13.2% of the variance (head depth, band 1), and differences in pinna parameters up to 4.8% (pinna height, band 2). But in all cases, the influence on coloration is less than 0.33 dB/cm of anthropometric feature difference.

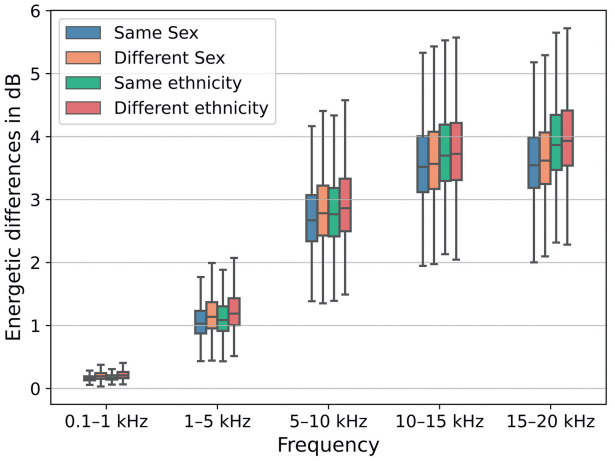


Fig. 7. Box plot showing differences in coloration for all the frequency bands, with the horizontal line inside the box indicating the median value. Whiskers indicate the minimum and maximum values unless there are outliers. The *x*-axis indicates five frequency bands, and the *y*-axis shows energetic differences in decibels.

4 DISCUSSION

In this paper, the acquisition and analysis of an HRTF database stratified by age, sex, and ethnicity is presented. The database contains 3D head meshes, numerically calculated HRTFs, and anthropometric features from 145 European and 111 Asian individuals.

Compared with existing databases, the anthropometric features cover a wider range of pinna sizes and a comparable range of head sizes. The CIPIC database is an outlier, showing larger head depths than all other databases. Because it is not detailed how the CIPIC anthropometric parameters were measured, this difference may be a measurement artifact. Consistent with previous studies, it was found that pinna height and width increase with age and that males have larger heads and pinnae than females. The METUB database shows a 2-mm/decade increase in pinna height and an 0.8-mm/decade increase in pinna width. On average, males have head widths that are 12.11 mm larger and head depths that are 11.4 mm larger than females. Similarly, male pinnae were 9.2 mm larger and 1.3 mm wider than those of females. Asians showed head widths that are 17.6 mm larger, head depths that are 12.09 mm smaller, and pinnae that are 4.3 mm larger than Europeans.

In the sample used here, head depth decreases with age, and this is effect is more pronounced in males. Because previous studies showed no systematic decrease of head depth for adults [24], the authors hypothesize that this could be related to measurement artifacts. First, smoothing of the swim caps during postprocessing of the meshes could have slightly distorted the head measures. Second, age-related hair loss, which is more severe in males [62], may contribute to a decrease in the effective head depth with age when wearing a swim cap.

Overall, the effects of age, sex, and ethnicity are more pronounced and explain more variance in the case of METUB than in the case of ANSUR II. Because of the large size of ANSUR II, the manual anthropometric measurements may have been taken by multiple experimenters, making them less consistent and thus leading to less pronounced effects. Because ethnicity is a sociohistoric factor, sampling from two distinct locations (Germany and China in case of METUB) might yield larger effects than sampling only in the United States for ANSUR II.

Statistically significant effects of age, sex, and ethnicity differences were found on acoustic features in most cases. However, the effects were generally small, and the authors consider them to be hardly relevant for most audio applica-

tions. The modeled median plane PE increases by 1° as a result of a mismatch in sex and ethnicity and by 0.01°/year of age difference. This is small compared with the median plane localization blur for familiar speech, which is between approximately 10° and 20°, depending on the source position [1, SEC. 2.1].

For ITD errors, a systematic increase of 0.14 JND units was observed in the case of sex mismatch, and an increase of 0.02 JND units owing to a mismatch of ethnicity was observed, which was also considered to be negligible. The authors observed no significant effects for age difference between subjects on ITD and no mentionable significant effects for the ILD. Because ITD and ILD are considered the dominant cues for source localization in the lateral (left/right) direction, these findings suggest that localization in the lateral direction is not systematically affected by the investigated factors.

The amount of coloration that could be explained by differences in age, sex, and ethnicity is far below the assumed JND of 1 dB for all cases, potentially rendering the influence of age, sex, and ethnicity irrelevant in many practical cases.

To put the results of the models using differences in age, sex, and ethnicity to predict differences in acoustic features into perspective, differences in anthropometric features were also used as independent variables. Despite the limited set of only four predictors (pinna height/width, head width/depth), these models explained considerably more variance and showed larger effects than models using demographic factors for the prediction. This shows, that the selected acoustic features are influenced by the used anthropometric features.

A limited set of acoustic features were used to investigate the influence of age, sex, and ethnicity on the modeled perception of nonindividual HRTFs. This investigation could be extended for models for localization in the lateral direction [63–65], models for other perceptual qualities [66], or perceptual evaluations. Moreover, effects on localization and coloration in isolation were investigated. Even if the systematic errors caused by differences in demographic factors are below the JND of each single quality, they may become audible when considered in combination (cf. [67]). Moreover, a stratified sampling might be required for other applications, for example, when generating an unbiased statistical morphable pinna model [68]. Future work will involve building such a model, which can be used to generate an arbitrary amount of head and pinna shapes, computer rendered images, and numerically calculated HRTFs. Such a dataset might be used to improve deep learning-based HRTF individualization or spatial upsampling.

HRTFs were computed based on head models without torso, and it is known that the torso adds a reflection and shadowing effects to the HRTF. The reflection is strongest when the ear, shoulder, and source are approximately aligned to each other, and shadowing is largest for sources at very low elevations when the shoulders or other parts of the torso are blocking the direct path from the source to the ear [61]. These effects were shown to be audible and slightly affect source localization [61, 69, 70].

Additionally, the position of the torso relative to the head differs across subjects, which affects the timing of the torso reflection. As a consequence, the peak frequencies and depth of the comb filter that is induced by the torso reflection differ across subjects [70]. Differences in torso geometry might hence also systematically contribute to differences between HRTFs, and to the best of the authors' knowledge, no studies have yet investigated the acoustic effect of torso anthropometric features (e.g., the abdominal and shoulder circumference) in relation to age, sex, and ethnicity, making this a potential direction for future research.

5 CONCLUSION

The present study acquired 3D head meshes, anthropometric measurements, and numerically calculated HRTFs from a sample of 256 individuals, equally sampled from 11 age groups, two sexes (male and female), and two ethnicities (European and Asian). Analyzing the influence on anthropometric features using GLMs revealed that pinna height, pinna width, and head depth significantly change with age, whereas pinna height, head width, and head depth significantly change with sex and ethnicity.

Analyzing the influence of age, sex, and ethnicity on acoustic performance, as modeled by PEs, QEs in the median plane, ITD and ILD differences in the horizontal plane, and auditory BSD as a predictor of timbre differences also showed significant effects on almost each of these measures. The size of these effects, however, calculated for all 256×256 pairs of participants in the sample, was always smaller than the JND of the individual measures. Considering the possible combined effect on the different performance metrics, HRTF personalization approaches will improve significantly and possibly also audibly if the underlying HRTF databases were more representative in terms of age, sex, and ethnicity than most databases currently used.

6 ACKNOWLEDGMENTS

The authors would like to thank Steffen Lepa for advice on sample design and statistical analysis and Daniel Fuhrmann and other colleagues at Mercedes-Benz group for their assistance with the preparation of the data acquisition.

7 DATA AVAILABILITY

The anthropometric data (head width/depth, pinna height/width), are publicly available and documented in [80]. The algorithm for estimating the head width and depth from 3D head meshes is available from https://github.com/sanketjain1101/Anthropometric_feature_extraction-for-Head-Related-Transfer-Functions.

8 REFERENCES

[1] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, MA,, USA 1996), https://doi.org/10.7551/mitpress/6391.001.0001.

[2] B. Xie, *Head-Related Transfer Function and Virtual Auditory Display* (J. Ross Publishing, Plantation, FL, USA, 2013), 2nd ed.

[3] K. Iida, *Head-Related Transfer Function and Acoustic Virtual Reality* (Springer Nature Singapore, Singapore, 2019), https://doi.org/10.1007/978-981-13-9745-5.

[4] V. R. Algazi, C. Avendano, and R. O. Duda, "Estimation of a Spherical-Head Model from Anthropometry," *J. Audio Eng. Soc.*, vol. 49, no. 6, pp. 472–479 (2001 Jun.). https://aes.org/publications/elibrary-page/?id=10188.

[5] 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. 111–123 (1993 Jul.), https://doi.org/10.1121/1.407089.

[6] J. C. Middlebrooks, "Virtual Localization Improved By Scaling Nonindividualized External-Ear Transfer Functions in Frequency," *J. Acoust. Soc. Am.*, vol. 106, no. 3, pp. 1493–1510 (1999 Sep.), https://doi.org/10.1121/1.427147.

[7] W. M. Hartmann and A. Wittenberg, "On the Externalization of Sound Images," *J. Acoust. Soc. Am.*, vol. 99, no. 6, pp. 3678–3688 (1996 Jun.), https://doi.org/10.1121/1.414965.

[8] K. Iida, M. Yairi, and M. Morimoto, "Role of Pinna Cavities in Median Plane Localization," in Proc. *16th Int. Cong. on Acoustics*, pp. 845–846, Seattle, WA, USA (1998 Jun.).

[9] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF Database," in Proc. *IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No. 01TH8575),* pp. 99–102, New Platz, NY, USA (2001 Oct.), https://doi.org/10.1109/ASPAA.2001.969552.

[10] M. Zhang, R. Kennedy, T. Abhayapala, and W. Zhang, "Statistical Method to Identify Key Anthropometric Parameters in HRTF Individualization," in Proc. *Joint Workshop on Hands-free Speech Communication and Microphone Arrays*, pp. 213–218, Edinburgh, UK (2011 Jun.), https://doi.org/10.1109/HSCMA.2011.5942401.

[11] S. Spagnol, M. Geronazzo, and F. Avanzini, "On the Relation Between Pinna Reflection Patterns and Head-Related Transfer Function Features," *IEEE Trans. Audio Speech Lang. Process.*, vol. 21, no. 3, pp. 508–519 (2012 Mar.), https://doi.org/10.1109/TASL.2012.2227730.

[12] H. S. Braren and J. Fels, "Towards Child-Appropriate Virtual Acoustic Environments: A Database of High-Resolution HRTF Measurements and 3D-Scans of Children," *Int. J. Environ. Res. Public Health*, vol. 19, no. 1, paper 324:1–17 (2021 Dec.), https://doi.org/10.3390/ijerph19010324.

[13] S. Spagnol and F. Avanzini, "Frequency Estimation of the First Pinna Notch in Head-Related Transfer Functions With a Linear Anthropometric Model," in Proc. *18th Int. Conf. Digital Audio Effects (DAFx)*, pp. 231–236, Trondheim, Norway (2015 Nov.-Dec.), https://hdl.handle.net/11577/3232314.

[14] S. Ghorbal, T. Auclair, C. Soladie, and R. Seguier, "Pinna Morphological Parameters Influencing HRTF Sets," in Proc. *20th Int. Conf. Digital Audio Effects (DAFx)*, pp. 353–359, Edinburgh, UK (2017 Sep.).

[15] P. Stitt and B. F. Katz, "Sensitivity Analysis of Pinna Morphology on Head-Related Transfer Functions Simulated via a Parametric Pinna Model," *J. Acoust. Soc. Am.*, vol. 149, no. 4, pp. 2559–2572 (2021 Apr.), https://doi.org/10.1121/10.0004128.

[16] K. Ban and E. S. Jung, "Ear Shape Categorization for Ergonomic Product Design," *Int. J. Ind. Erg.*, vol. 80, paper 102962 (2020 Nov.), https://doi.org/10.1016/j.ergon.2020.102962.

[17] K. Hajniš, "Die Veränderungen der Ohrmuscheln Beim Erwachsenen," *Z. Morphol. Anthropol.*, vol. 61, no. 1. pp. 42–56 (1969 May).

[18] D. Pellnitz, "Über das Wachstum der Menschlichen Ohrmuschel: Mit 6 Textabbildungen," *Arch. Ohren. Nasen. Kehlkopfheilkd.*, vol. 171, no. 2, pp. 334–340 (1958 Jan.), https://doi.org/10.1007/BF02127295.

[19] G. F. Schwalbe, *Beiträge zur Anthropologie des Ohres* (Verlag Nicht Ermittelbar, 1891).

[20] C. Susanne, "Individual Age Changes of the Morphological Characteristics," *J. Hum. Evol.*, vol. 6, no. 2, pp. 181–189 (1977 Feb.), https://doi.org/10.1016/S0047-2484(77)80126-7.

[21] E. Gualdi, "Longitudinal Study of Anthropometric Changes With Aging in an Urban Italian Population," *Homo.*, vol. 49, no. 3, pp. 241–259 (1998), https://hdl.handle.net/11392/1203174.

[22] C. Niemitz, M. Nibbrig, and V. Zacher, "Human Ears Grow Throughout the Entire Lifetime According to Complicated and Sexually Dimorphic Patterns—Conclusions From a Cross-Sectional Analysis," *Anthropol. Anz.*, vol 65, no. 4, pp. 391–413 (2007 Dec.), https://www.jstor.org/stable/29542890.

[23] H. J. Choi, T. N. Garlie, and J. L. Parham, "Anthropometric Analyses of Head and Face Shape to Design Protective Headgear for US Army Personnel," in Proc. *Int. Conf. Applied Human Factors and Ergonomics*, pp. 533–545, Washington, DC, USA (2019 Jul.), https://doi.org/10.1007/978-3-030-20216-3_50.

[24] DIN, *33402-2: Ergonomics – Human Body Dimensions. Part 2: Values* (2005).

[25] L. G. Farkas, M. J. Katic, and C. R. Forrest, "International Anthropometric Study of Facial Morphology in Various Ethnic Groups/Races," *J. Craniofac. Surg.*, vol. 16, no. 4, pp. 615–646 (2005 Jul.), https://doi.org/10.1097/01.scs.0000171847.58031.9e.

[26] R. Ball, C. Shu, P. Xi, M. Rioux, Y. Luximon, and J. Molenbroek, "A Comparison Between Chinese and Caucasian Head Shapes," *Appl. Ergon.*, vol. 41, no. 6, pp. 832–839 (2010 Oct.), https://doi.org/10.1016/j.apergo.2010.02.002.

[27] K. S. Alexander, D. J. Stott, B. Sivakumar, and N. Kang, "A Morphometric Study of the Human Ear," *J. Plast. Reconstr. Aesthet. Surg.*, vol. 64, no. 1, pp. 41–47 (2011 Jan.), https://doi.org/10.1016/j.bjps.2010.04.005.

[28] O. Oludiran and D. R. Omotoso, "A Morphometric Study of the External Ears at Benin City," *Niger. J. Plast. Surg.*, vol. 8, no. 1, pp. 1–5 (2012 Jun.).

[29] A. Mumin, B. Olabu, K. Ongeti, and H. Saidi, "Ethnic Differences in the Morphology of the Pinna," *Anat. J. Afr.*, vol. 7, no. 1, pp. 1097–1102 (2018 Apr.).

[30] S. Kaneko, T. Suenaga, H. Okumura, and S. Kim, "Automatic Statistical Analysis of Ear Shape Anthropometrics of an International Ear Shape Database," in Proc. *AES Int. Conf. Immersive and Interactive Audio* , paper 8:1–7, York, UK (2019 Mar.).

[31] Z. Zhuang, D. Landsittel, S. Benson, R. Roberge, and R. Shaffer, "Facial Anthropometric Differences Among Gender, Ethnicity, and Age Groups," *Ann. Occup. Hyg.*, vol. 54, no. 4, pp. 391–402 (2010 Jun.), https://doi.org/10.1093/annhyg/meq007.

[32] F. Fu and Y. Luximon, "A Systematic Review on Ear Anthropometry and Its Industrial Design Applications," *Hum. Factors Ergon. Manuf.*, vol. 30, no. 3, pp. 176–194 (2020 May), https://doi.org/10.1002/hfm.20832.

[33] L. Picinali and B. F. G. Katz, "System-to-User and User-to-System Adaptations in Binaural Audio," in M. Geronazzo and S. Serafin (Eds.), *Sonic Interactions in Virtual Environments*, Human-Computer Interaction Series, pp. 115–143 (Springer, Cham, Switzerland, 2022 Oct.), https://doi.org/10.1007/978-3-031-04021-4.

[34] D. Fantini, M. Geronazzo, F. Avanzini, and S. Ntalampiras, "A Survey on Machine Learning Techniques for Head-Related Transfer Function Individualization," *IEEE Open J. Signal Process.*, vol. 6, pp. 30–56 (2025 Jan.), https://doi.org/10.1109/OJSP.2025.3528330.

[35] I. Ito, M. Imada, M. Ikeda, K. Sueno, T. Arikuni, and A. Kida, "A Morphological Study of Age Changes in Adult Human Auricular Cartilage With Special Emphasis on Elastic Fibers," *Laryngoscope*, vol. 111, no. 5, pp. 881–886 (2001 May), https://doi.org/10.1097/00005537-200105000-00023.

[36] C. Sforza, G. Grandi, M. Binelli, D. G. Tommasi, R. Rosati, and V. F. Ferrario, "Age- and Sex-Related Changes in the Normal Human Ear," *Forensic Sci. Int.*, vol. 187, no. 1–3, pp. 110.e1–110.e7 (2009 May), https://doi.org/10.1016/j.forsciint.2009.02.019.

[37] R. Purkait, "Progression of Growth in the External Ear From Birth to Maturity: A 2-Year Follow-Up Study in India," *Aesthet. Plast. Surg.*, vol. 37, pp. 605–616 (2013 Jun.), https://doi.org/10.1007/s00266-013-0097-1.

[38] S. Paquette, *Anthropometric Survey (ANSUR) II Pilot Study: Methods and Summary Statistics* (Anthrotch, US Army Natick Soldier Research, Development and Engineering Center, Natick, MA, USA, 2009).

[39] P. Verma, H. K. Sandhu, K. G. Verma, S. Goyal, M. Sudan, and A. Ladgotra, "Morphological Variations and Biometrics of Ear: An Aid to Personal Identification," *J. Clin. Diagn. Res.*, vol. 10, no. 5, pp. ZC138–ZC142 (2016 May), https://doi.org/10.7860/JCDR/2016/18265.7876.

[40] L. Sax, "How Common Is lntersex? A Response to Anne Fausto-Sterling," *J. Sex Res.*, vol. 39, no. 3, pp. 174–178 (2002 Jan.), https://doi.org/10.1080/00224490209552139.

[41] 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," *J. Audio Eng. Soc.*, vol. 67, no. 9, pp. 705–718 (2019 Sep.), https://doi.org/10.17743/jaes.2019.0024.

[42] 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 Proc. *140th Conv. Audio Eng. Soc.*, paper 9579:1–9, Paris, France (2016 Jun.).

[43] AES, *AES69-2022: AES Standard for File Exchange – Spatial Acoustic Data File Format* (2022).

[44] F. Brinkmann, W. Kreuzer, J. Thomsen, et al., "Recent Advances in an Open Software for Numerical HRTF Calculation," *J. Audio Eng. Soc.*, vol. 71, no. 7/8, pp. 502–514 (2023 Jul./Aug.), https://doi.org/10.17743/jaes.2022.0078.

[45] W. Kreuzer, K. Pollack, F. Brinkmann, and P. Majdak, "NumCalc: An Open-Source BEM Code for Solving Acoustic Scattering Problems," *Eng. Anal. Bound. Elem.*, vol. 161, pp. 157–178 (2024 Apr.), https://doi.org/10.1016/j.enganabound.2024.01.008.

[46] 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 (2015 Jul.), https://doi.org/10.1121/1.4922518.

[47] T. Palm, S. Koch, F. Brinkmann, and M. Alexa, "Curvature-Adaptive Mesh Grading for Numerical Approximation of Head-Related Transfer Functions," in Proc. *Fortschritte der Akustik (DAGA)*, pp. 1111–1114, Vienna, Austria (2021 Aug.).

[48] P. Majdak, F. Zotter, F. Brinkmann, J. De Muynke, M. Mihocic, and M. Noisternig, "Spatially Oriented Format for Acoustics 2.1: Introduction and Recent Advances," *J. Audio Eng. Soc.*, vol. 70, no. 7/8, pp. 565–584 (2022 Jul./Aug.), https://doi.org/10.17743/jaes.2022.0026.

[49] J. Love, D. Dropman, and R. Selker, "The Jamovi Project (Version 2.6.19)," https://www.jamovi.org (accessed 7 Feb. 2026).

[50] M. Gallucci and C. Simpson, "GAMLj: General Analysis for Linear Models," https://github.com/gamlj/gamlj (accessed 7 Feb. 2026).

[51] J. Hotzman, C. C. Gordon, B. Bradtmiller, et al., "2012 Measure's Handbook: US Army and Marine Corps Antropometric Surveys, 2010-2011," Tech. Rep. NATICK/TR-11/017 (2011 Aug.). https://apps.dtic.mil/sti/tr/pdf/ADA548497.pdf.

[52] M. G. Bozkır, P. Karakaş, M. Yavuz, and F. Dere, "Morphometry of the External Ear in Our Adult Population," *Aesthet. Plast. Surg.*, vol. 30, no. 1, pp. 81–85 (2006 Feb.), https://doi.org/10.1007/s00266-005-6095-1.

[53] 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 (2014 Aug.), https://doi.org/10.1121/1.4887447.

[54] B. Rafaely, *Fundamentals of Spherical Array Processing* (Springer, Berlin, Germany, 2019), 2nd ed. https://doi.org/10.1007/978-3-319-99561-8.

[55] J. C. Middlebrooks, "Individual Differences in External-Ear Transfer Functions Reduced By Scaling in Frequency," *J. Acoust. Soc. Am.*, vol. 106, no. 3, pp. 1480–1492 (1999 Sep.), https://doi.org/10.1121/1.427176.

[56] A. Andreopoulou and B. F. Katz, "Identification of Perceptually Relevant Methods of Inter-Aural Time Difference Estimation," *J. Acoust. Soc. Am.*, vol. 142, no. 2, pp. 588–598 (2017 Aug.), https://doi.org/10.1121/1.4996457.

[57] P. Majdak, C. Hollomey, and R. Baumgartner, "AMT 1.x: A toolbox for Reproducible Research in Auditory Modeling," *Acta Acust.*, vol. 6, paper 19:1–17 (2022 May), https://doi.org/10.1051/aacus/2022011.

[58] J. E. Mossop and J. F. Culling, "Lateralization of Large Interaural Delays," *J. Acoust. Soc. Am.*, vol. 104, no. 3, pp. 1574–1579 (1998 Sep.), https://doi.org/10.1121/1.424369.

[59] T. McKenzie and F. Brinkmann, "Toward an Improved Auditory Model for Predicting Binaural Coloration," *J. Audio Eng. Soc.*, vol. 73, no. 3, pp. 115–126 (2025 Mar.), https://doi.org/10.17743/jaes.2022.0192.

[60] J. Fliege and U. Maier, "The Distribution of Points on the Sphere and Corresponding Cubature Formulae," *IMA J. Numer. Anal.*, vol. 19, no. 2, pp. 317–334 (1999 Apr.), https://doi.org/10.1093/imanum/19.2.317.

[61] V. R. Algazi, C. Avendano, and R. O. Duda, "Elevation Localization and Head-Related Transfer Function Analysis at Low Frequencies," *J. Acoust. Soc. Am.*, vol. 109, no. 3, pp. 1110–1122 (2001 Mar.), https://doi.org/10.1121/1.1349185.

[62] L. Iyengar and J. Li, "Male and Female Pattern Hair Loss," *Aust. Prescr.*, vol. 48, no. 3, pp. 93–97 (2025 Jun.), https://doi.org/10.18773/austprescr.2025.020.

[63] M. Dietz, S. D. Ewert, and V. Hohmann, "Auditory Model Based Direction Estimation of Concurrent Speakers from Binaural Signals," *Speech Commun.*, vol. 53, no. 5, pp. 592–605 (2011 May/Jun.), https://doi.org/10.1016/j.specom.2010.05.006.

[64] R. Barumerli, P. Majdak, M. Geronazzo, D. Meijer, F. Avanzini, and R. Baumgartner, "A Bayesian Model for Human Directional Localization of Broadband Static Sound Sources," *Acta Acust.*, vol. 7, paper 12:1–13 (2023 May), https://doi.org/10.1051/aacus/2023006.

[65] A. Souchaud, P. Lladó, A. Neidhardt, Z. Cvetković, and E. De Sena, "White-Box Differentiable Model of Perceived Localisation," in Proc. *IEEE 27th Int. Workshop Multimedia Signal Processing (MMSP)*, pp. 264–268, Beijing, China (2025 Sep.).

[66] P. Lladó, A. Neidhardt, F. Brinkmann, and E. De Sena, "Spatial Audio Models Inventory to Cover The Attributes from the Spatial Audio Quality Inventory," in Proc. *Forum Acusticum | Euronoise*, pp. 1–9, Malaga, Spain (2025 Jun.).

[67] N. Meyer-Kahlen, S. V. A. Garí, I. Ananthabhotla, and P. Calamia, "A Two-Dimensional Threshold Test for Reverberation Time and Direct-to-Reverberant Ratio," in Proc. *Mmersive and 3D Audio: From Architecture to Automotive (I3DA)*, pp. 1–8, Bologna, Italy (2023 Sep.), https://doi.org/10.1109/I3DA57090.2023.10289365.

[68] B. Egger, W. A. P. Smith, A. Tewari, et al., "3D Morphable Face Models—Past, Present, and Future," *ACM Trans. Graph.*, vol. 39, no. 5, paper 157:1–38 (2020 Jun.), https://doi.org/10.1145/3395208.

[69] F. Brinkmann, R. Roden, A. Lindau, and S. Weinzierl, "Audibility and Interpolation of Head-Above-Torso Orientation in Binaural Technology," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 5, pp. 931–942 (2015 Aug.), https://doi.org/10.1109/jstsp.2015.2414905.

[70] J. Johansson, A. Mäkivirta, and M. Malinen, "Effects of Torso Location and Rotation to HRTF," *J. Audio Eng. Soc.*, vol. 72, no. 7/8, pp. 493–503 (2024 Jul.), https://doi.org/10.17743/jaes.2022.0152.

[71] N. Gupta, A. Barreto, M. Joshi, and J. C. Agudelo, "HRTF Database at FIU DSP Lab," in Proc. *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, pp. 169–172, Dallas, TX, USA (2010 Mar.), https://doi.org/10.1109/ICASSP.2010.5496084.

[72] K. Watanabe, Y. Iwaya, Y. Suzuki, S. Takane, and S. Sato, "Dataset of Head-Related Transfer Functions Measured With a Circular Loudspeaker Array," *Acoust. Sci. Technol.*, vol. 35, no. 3, pp. 159–165 (2014 May), https://doi.org/10.1250/ast.35.159.

[73] C. T. Jin, P. Guillon, N. Epain, R. Zolfaghari, A. van Schaik, and A. I. Tew, "Creating the Sydney York Morphological and Acoustic Recordings of Ears Database," *IEEE Trans. Multimed.*, vol. 16, no. 1, pp. 37–46 (2014 Jan.), https://doi.org/10.1109/TMM.2013.2282134.

[74] R. Bomhardt, M. de la Fuente Klein, and J. Fels, "A High-Resolution Head-Related Transfer Function and Three-Dimensional Ear Model Database," Proc. *Mtgs. Acoust.*, vol. 29, no. 1, paper 050002:1–11 (2016 Nov.), https://doi.org/10.1121/2.0000467.

[75] I. Engel, R. Daugintis, T. Vicente, et al., "The SONICOM HRTF Dataset," *J. Audio Eng. Soc.*, vol. 71, no. 5, pp. 241–253 (2023 May), https://doi.org/10.17743/jaes.2022.0066.

[76] S. Ghorbal, X. Bonjour, and R. Séguier, "Computed HRIRs and Ears Database for Acoustic Research," in Proc. *148th Conv. Audio Eng. Soc.*, paper 13061:1–1, Virtual (2020 Jun.), https://doi.org/10.13140/RG.2.2.13880.57608.

[77] C. Guezenoc and R. Seguier, "A Wide Dataset of Ear Shapes and Pinna-Related Transfer Functions Generated by Random Ear Drawings," *J. Acoust. Soc. Am.*, vol. 147, no. 6, pp. 4087–4096 (2020 Jun.), https://doi.org/10.1121/10.0001461.

[78] J. G. Tylka R. Sridhar, and E. Y. Choueiri, "A Database of Head-Related Transfer Function and Morphological Measurements," in Proc. *143rd Conv. Audio Eng. Soc.* , paper 357:1–5, New York, NY, USA (2017 Oct.).

[79] C. Armstrong, L. Thresh, D. Murphy, and G. Kearney, "A Perceptual Evaluation of Individual and Non-Individual HRTFs: A Case Study of the SADIE II Database," *Appl. Sci.*, vol. 8, no. 11, paper 2029:1–21 (2018 Oct.), https://doi.org/10.3390/app8112029.

[80] S. Jain, F. Brinkmann, R. Moeller, B. Haubrichs, and S. Weinzierl, "Anthropometric Measures of METUB Database," (2026), https://doi.org/10.14279/depositonce-25270 (accessed 18 Feb. 2026).

THE AUTHORS


Sanket Jain


Bjarne Haubrichs


Fabian Brinkmann


Rainer Moeller


Stefan Weinzierl

Sanket Jain is currently pursuing his Ph.D. at Technische Universität Berlin, where his research is dedicated to the individualization of head-related transfer functions (HRTFs) using anthropometric and machine learning–based approaches for spatial audio applications. He earned his Bachelor's degree in electronics and telecommunication from Vidyalankar Institute of Technology (VIT) Mumbai, followed by a Master's degree in communication and media engineering from Hochschule (HS) Offenburg in 2018 and 2022, respectively. During his master's studies, he completed an internship and master's thesis at Mercedes-Benz Group, focusing on crosstalk cancellation systems required for spatial audio systems.

•

Bjarne Haubrichs currently works as a software engineer at Schaerer AG. He previously studied medical engineering at Offenburg University of Applied Sciences. As part of his studies, he completed an internship and a master's thesis at Mercedes-Benz AG in the field of audio and signal processing research.

•

Fabian Brinkmann is head of the binaural processing and perception team in the Audio Communication Group at Technische Universität Berlin, where he earned his Ph.D. in 2019. His research focuses on signal processing and evaluation approaches for spatial audio, including perception and model-based evaluations and the development of algorithms for room acoustic simulation and binaural rendering. He is also one of the developers of the Python packages for acoustics research (pyfar).

•

Rainer Moeller leads the Future Audio Laboratory at Mercedes Benz Research. With a background in precision engineering—covering electronics, optics, and mechanics—he brings an interdisciplinary perspective to audio technology. His work centers on advanced in car audio, perceptual evaluation, and spatial sound concepts. Over the past seven years, he has supervised numerous undergraduate, graduate, and doctoral research projects and collaborates closely with academic partners to advance audio and acoustic research in the automotive domain.

•

Stefan Weinzierl is head of the Audio Communication Group at the Technische Universität (TU) Berlin. His research is focused on audio technology, virtual acoustics, room acoustics, and musical acoustics. With a diploma in physics and sound engineering, he received his Ph.D. in musical acoustics. He is coordinating a master program in audio communication and technology at TU Berlin and has coordinated several international research consortia in the field of virtual acoustics and music information retrieval.