Fabian Brinkmann, Manoj Dinakaran, Robert Pelzer, Peter Grosche, Daniel Voss, Stefan Weinzierl

A Cross-Evaluated Database of Measured and Simulated HRTFs Including 3D Head Meshes, Anthropometric Features, and Headphone Impulse Responses

Citation details
Brinkmann, Fabian; Dinakaran, Manoj; Pelzer, Robert; Grosche, Peter; Voss, Daniel; Weinzierl, Stefan (2019). 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. https://doi.org/10.17743/jaes.2019.0024.

A Cross-Evaluated Database of Measured and Simulated HRTFs Including 3D Head Meshes, Anthropometric Features, and Headphone Impulse Responses

FABIAN BRINKMANN[1], MANOJ DINAKARAN[1,2], ROBERT PELZER[1],
(fabian.brinkmann@tu-berlin.de) (manoj.dinakaran@huawei.com) (r.pelzer@posteo.de)

PETER GROSCHE[2], DANIEL VOSS[3], AND STEFAN WEINZIERL[1]
(peter.grosche@huawei.com) (daniel.voss@sennheiser.com) (stefan.weinzierl@tu-berlin.de)

[1]Audio Communication Group, Technical University of Berlin, Einsteinufer 17c, D-10587, Germany
[2]Huawei Technologies, Munich Research Centre, Riesstrasse 25, D-80992 Munich, Germany
[3]Sennheiser electronic GmbH & Co. KG, Am Labor 1, D-30900 Wedemark, Germany

The individualization of head related transfer functions (HRTFs) can make an important contribution to improving the quality of binaural technology applications. One approach to individualization is to exploit relations between the shape of HRTFs on the one hand and anthropometric features of the ears, head, and torso of the corresponding listeners on the other hand. To identify statistically significant relations between the two sets of variables, a relatively large database is required. For this purpose, full-spherical HRTFs of 96 subjects were acoustically measured and numerically simulated. A detailed cross-evaluation showed a good agreement to previous data between repeated measurements and between measured and simulated data. In addition to 96 HRTFs, the database includes high resolution head-meshes, a list of 25 anthropometric features per subject, and headphone transfer functions for two headphone models. It is publicly available under a free culture license from https://doi.org/10.14279/depositonce-8487.

0 INTRODUCTION

Individualizing head-related transfer functions (HRTFs) provides an approach to improve the spatial audio quality in binaural technology applications, e.g., for achieving a localization accuracy that is comparable to the performance of listeners in the corresponding real sound field [1]. The most precise approaches to obtain individual HRTFs are acoustic measurements or numerical simulations [1, 2]. Both require specialized soft- and hardware making it impractical for consumers to determine their individual HRTFs in this way. Alternatively, HRTFs can be individualized by exploiting relations between the listener's anthropometry and acoustic features of HRTFs. This seems a reasonable approach as the salient auditory features in HRTFs originate from the impact of the listener's torso, head, and outer ears (pinnae) on the incident sound field [3]. This, however, requires databases comprising individual HRTFs with high spatial resolution and accurate anthropometric measurements.

Several of such databases were published within the last two decades, as can be seen from the overview in Table 1 (cf., Bomhardt et al. [10] for a more detailed summary). All datasets were measured at a distance of 1 m or more to avoid proximity effects of the binaural cues and acoustic parallax [11] but differ in the number of included HRTF sets and the spatial sampling grid. For the latter, different sampling strategies and mechanical restrictions of the measurement systems have led to different spatial resolutions and missing points below certain elevations.

While for early databases HRTFs were measured sequentially, i.e., one sampling point after the other, for some of the later databases the measurement was accelerated by interleaving the measurement signals using the (optimized) multiple exponential sweep method (ARI, ITA) [12–14]. Although methods for a continuous rotation of the subjects would have been available to further reduce the measurement time [15], all previous databases were measured at discrete angles by stepwise rotation of the loudspeakers or subjects, possibly to avoid noise from the rotation device that would reduce the signal to noise ratio of the measured HRTFs [16].

Anthropometric measures of the head and torso were taken directly from the subjects using tape measures, while pinnae features were either estimated from photographs

Table 1. Public HRTF databases that include anthropometry in chronological order of publication. $\Delta\phi$ and $\Delta\theta$ specify the spatial resolution in azimuth and elevation and r gives the radius of the spherical sampling grid. (1: great circle distance (GCD) between neighboring points of the same elevation; 2: Simulated HRTFs have a higher spatial resolution)

Name	Subjects	Directions	Spatial sampling	Room	Anthropometry	3D meshes	Simulated HRTFs	Spatially continuous
CIPIC [4]	45	1250	$\Delta\phi = 5°$; $\Delta\theta = 5.6°$ $-45° \leq \theta \leq 80°$; $r = 1$ m	room with absorbers	43 subjects	–	–	–
LISTEN [5]	51	187	$\Delta\phi = 15°$; $\Delta\theta = 15°$ $-45° \leq \theta \leq 90°$; $r = 1.95$ m	anechoic	yes	–	–	–
FIU [6]	15	72	$\Delta\phi = 30°$; $\Delta\theta = 18°$ $-36° \leq \theta \leq 54°$	room with absorbers	yes	pinnae only	–	–
ARI [7]	150	1550	$\Delta\phi = 2.5°, 5°$; $\Delta\theta = 5°$ $-30° \leq \theta \leq 80°$; $r = 1.2$ m	hemi-anechoic	60 subjects	–	–	–
RIEC [8]	105	865	$\Delta\phi = 5°$; $\Delta\theta = 10°$ $-30° \leq \theta \leq 90°$; $r = 1.5$ m	anechoic	39 subjects	head and shoulder	–	–
SYMARE [9]	10	393	$\Delta\phi \approx 10°$ GCD[1]; $\Delta\theta = 10°$ $-45° \leq \theta \leq 90°$; $r = 1$ m	anechoic	yes	head and shoulder	up to 16 kHz	–
ITA [10]	48	2304	$\Delta\phi = 5°$; $\Delta\theta = 5°$ $-66° \leq \theta \leq 90°$; $r = 1.2$ m	hemi-anechoic	yes	pinnae only	–	–
HUTUBS[2] (this work)	96	440	$\Delta\phi \approx 10°$ GCD[1]; $\Delta\theta = 10°$ $-90° \leq \theta \leq 90°$; $r = 1.47$ m	anechoic	yes	head without shoulder	up to 22 kHz	yes

(CIPIC, LISTEN, ARI) or from 3D surface meshes of the subjects ears. The latter appears to be preferable, because picture based anthropometry requires a well calibrated experimental environment regarding the illumination, camera optics, exposure settings, position, posture, and distance to the subject [17]. As this was not given in the respective studies, biases related to wrong projections of the 3D-structure onto the 2D-image and uncertainties in the scale of the image might have been introduced. In all cases, the anthropometric measures were extracted manually, which can introduce a bias and uncertainty due to the experimenter.

A critical aspect of HRTF measurement systems is their evaluation. In lack of an analytical or other reliable reference for HRTFs, this can only be done indirectly by means of a cross-evaluation against natural listening, measurements from other systems or numerically simulated HRTFs. The latter was, for example, done during the acquisition of the SYMARE database [9].

As detailed in Sec. 1, the current database contains acoustically measured and numerically simulated HRTFs of 96 subjects, including two repeated measurements for cross-evaluation. The simulated HRTFs were calculated based on high resolution head meshes using the boundary element method. In contrast to earlier studies, we used a continuous rotation of the subjects to accelerate the acoustic HRTF measurements and acquired data on a full spherical sampling grid optimized to allow for a spatially continuous HRTF representation by means of spherical harmonics. Most anthropometric pinna measures were acquired automatically or semi-automatically to reduce the bias and uncertainty from the experimenter. In addition, headphone transfer functions (HpTFs) for two headphone models are provided for auralization. The availability of the data and the data format are addressed in Sec. 2. It is followed by a cross-evaluation of (a) the current database to previous HRTF measurements, (b) repeated measurements within the database, and (c) measured and simulated HRTFs in Sec. 3.

1 DATABASE ACQUISITION

1.1 3D Head Meshes

The 3D head meshes were acquired by a hybrid method using a Kinect 3D scanner for the head and a high resolution Artec Space Spider scanner for the ears. An evaluation of this hybrid scanning method against scans by a reference system (GOM ATOS I scanner) showed geometric deviations between the meshes of only 0.14 mm on average (SD 0.24 mm) [18].

The Kinect 3D scanner with the Kinect fusion developer toolkit browser v1.8.0 [19] was set up at eye level at a distance of about 1 m from the subjects who were sitting on a swivel chair with their natural head position. This was shown to be a reproducible position obtained when the subject is in a relaxed position, sitting or standing, and looking at the horizon or an external reference point at eye level [20]. Subjects wore a swim cap (approx. 1 mm thick) to acquire the actual head shape, i.e., to reduce the influence of hair on the scans, and were asked to make a pigtail if they had long hair. The pigtails were manually removed in post-processing, and might occasionally have caused slight differences between the meshes and the natural head shape. The Kinect scans were taken in a two-step procedure with the resolution set to the maximum of 768 voxels per meter. First, a complete mesh was generated by slowly rotating the subject 360°. In this case, the rotation caused a slight spatial smoothing in the mesh. To obtain non-smoothed meshes, separate scans of the left and right side of the face and the ears were taken while carefully rotating the subjects back and forth a couple of degrees until a closed and fine structured mesh was obtained. In post-processing

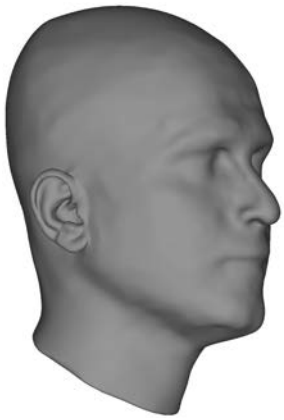

Fig. 1. Mesh of subject 2 after post-processing.

unwanted parts in the mesh were removed (e.g., surroundings and irregular parts close to holes), holes were filled, and the separate scans were merged using Meshlab [21] and Geomagic's point based glue tool. In addition the shoulders and torso were removed by cutting the mesh at the bottom of the neck.

Then, the hand-held blue structured light scanner Artec Space Spider with 0.05 mm point spacing resolution from a working distance of 0.2 m to 0.3 m was used to obtain high quality surface scans of the left and right pinna. About 10 to 20 scans for each pinna were taken at different angles in order to acquire the shape of the pinna accurately. To obtain complete pinna meshes, the separate scans were aligned and fused using the Artec Studio Professional 12 software. Despite the small size of the scanners field of view, small holes remained in the meshes occasionally at spots behind the ear, inside the ear canal, or in some cases where the crus anthelicis lead below the helix (cf., Fig. 5(d) for the anthropometry of the ear). These holes were automatically closed under consideration of the neighboring elements using Artec Studio's *watertight* option during the fusion process. In a final step, the ear canal entrances were closed using Meshlab to be flush with the bottom of the cavum concha, and the Artec scans were used to replace the ears of the Kinect scans using the Geomagic point based glue tool. The scanning of one subject took about 15 to 20 minutes.

The final meshes were then aligned to the global coordinate system based on the interaural axis, defined as the axis connecting the centers of the entrances to the ear canals. The alignment was done semi-automatically using a Python script for the open-source software Blender that required the selection of three points in the mesh (center of the left/right ear canal and a point on the nose). The ear canal points were used to move the midpoint of the interaural axis to the origin of coordinates, and to rotate the mesh about the vertical axis (z-axis) until the interaural axis fell onto the y-axis. The arbitrary point on the nose was only used to assure that the head is viewing in positive x-direction. Because the natural head position was already established during the scans, the script did not rotate the head around the interaural axis. An example for a final mesh after post-processing is shown in Fig. 1.

1.2 Numerical HRTF Simulation

HRTFs were simulated for frequencies between 100 Hz and 22 kHz in steps of 100 Hz using the boundary element method as implemented in MESH2HRTF [22]. A previous study showed that spectral differences between HRTFs calculated with such meshes and HRTFs calculated from high resolution reference meshes were below 0.5 dB on average [18]. The complex pressure was calculated on a $Q = 1730$ point Lebedev grid [23] with a radius of 1.47 m by assuming reciprocity, i.e., interchanging the positions of loudspeakers and microphones. This reduces the computational cost and was realized by assigning a volume velocity to a single mesh element in the center of the blocked ear canal [24]. To further reduce the processing cost, HRTFs were simulated separately for the left and right ear. For this purpose, the edge length of the meshes were gradually increased from 1 mm at the simulated ear to 10 mm at the opposite ear using the OpenFlipper plug-in contained in MESH2HRTF [25, 22], which resulted in 14,000 to 20,000 elements per mesh. HRTFs simulated from such meshes showed only negligible spectral distortion and deterioration in localization performance in comparison to HRTFs simulated from high resolution reference meshes [26]. MESH2HRTF requires that surface materials of the head mesh are named *Skin*, *Left ear*, and *Right ear*, which was done automatically with a python script for Blender. Calculating one HRTF set (left and right ear) took approximately 13 hours using 4 cores of an Intel i7 4 GHz CPU and 32 GB RAM.

In post-processing, the complex HRTF spectra were referenced to a point source in the coordinate origin by spectral division and normalized with respect to the surface area of the sound emitting mesh element at the blocked ear canal and the assigned volume velocity[1]. The referencing and normalization agrees with the definition of the HRTF that is given by the pressure at the ear canal divided by the pressure in the center of the head with the head absent [3]. Afterwards, the 0 Hz bin was set to 1 (0 dB), the single sided spectra were mirrored using the complex conjugate, and head-related impulse responses (HRIRs) were obtained by inverse Fourier transform. Due to the point source referencing, some HRIRs showed negative onset times. This was corrected by a circular shift of 60 samples that was applied to all HRIRs. Finally, HRIRs were shortened to 256 samples at a sampling rate of 44.1 kHz, by applying squared sine fade-ins of 10 samples and fade-outs of 20 samples.

To arrive at a spatially continuous HRTF representation, the complex spectra were subjected to a spherical harmonics (SH) transform of order 35 [27]

$$f_{nm} \approx \sum_{q=1}^{Q} \alpha_q \, f_q \, Y^*_{nm,q} \,, \tag{1}$$

[1] The Matlab scripts for referencing and normalization were added to MESH2HRTF as a result of this work along with the Python scripts for mesh alignment and material assignment and Matlab code for adding custom spatial sampling grids. The new features are available from version 0.2 at http://mesh2hrtf.sourceforge.net.

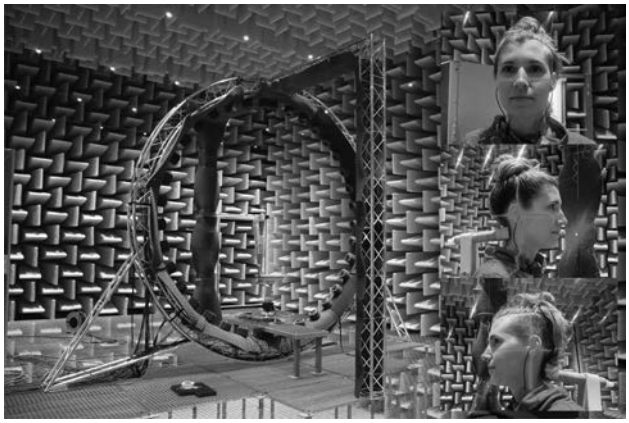

Fig. 2. HRTF measurement system and close ups to illustrate the subject positioning procedure using two cross line lasers.

with f_{nm} being the SH coefficients of order n and degree m, and α_q and f_q the sampling weights and HRTF frequency bins at the Q sampling points. Y^*_{nm} denotes the complex conjugate of the SH basis functions at the angle Ω

$$Y^*_{nm} = (-1)^m Y_{n,-m} \tag{2}$$

$$Y_{nm}(\Omega) \equiv \sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}} P_{nm}(\sin\theta) \mathrm{e}^{jm\phi}, \tag{3}$$

where $\Omega \equiv (\phi, \theta)$ gives the azimuth $\phi = [0^\circ, 360^\circ)$ (measured counter clockwise in the xy-plane, starting at positive x) and elevation $\theta = [-90^\circ, 90^\circ]$ (90° at positive z). P_{nm} denotes the associated Legendre function and $j = \sqrt{-1}$ the imaginary unit. From this representation, the HRTF can be interpolated to arbitrary Ω by means of the inverse SH transform

$$f(\Omega) = \sum_{n=0}^{N} \sum_{m=-n}^{n} f_{nm} Y_{nm}(\Omega). \tag{4}$$

1.3 Acoustic HRTF Measurements

HRTFs were measured in the anechoic chamber of the Technical University Berlin with a sampling rate of 44.1 kHz (cf., Fig. 2). The temperature and humidity were not tracked and assumed to be constant since the room was neither heated nor ventilated.

The measurement system consists of 37 Peerless NE65-04 2-inch drivers in custom made cylindrical closed boxes with an operating range between 200 Hz and 20 kHz. The loudspeaker signals were fed to five Cloud CXA850 8 channel amplifiers (50 W/ch. RMS @ 4 Ω) and converted by a cascade of an RME HDSPe AIO and an RME HDSPe RayDat sound card. The speakers were mounted to the inner ring of the measurement system, which was attached to the outer construction by eight adjustable lashing straps. The distance from the center of the array to the membrane of the speakers was 1.47 m. The speakers are arranged with distances of 10°, and a resolution of 5° is obtained by combining the front and back semicircle of the inner construction [28]. The chair for the subjects was mounted on a custom built belt driven turntable equipped with a brushless motor from Ott GmbH. The turntable stood on a supporting structure that bridged the bottom of the loudspeaker ring. The sound emitted from the turntable was below the level of the environmental noise measured at the position of subject's head with an NTI XL2 analyzer equipped with a class 1 half inch measuring microphone (NTI MC230 with NTI MA220). To minimize reflections from the measurement system, the outer and inner constructions, as well as the sides of the loudspeakers were wrapped in absorbers. During the measurements, the floor of the anechoic chamber and the supporting structure of the chair were covered as well (not shown in Fig. 2).

Before being seated, the subjects inserted custom made in-ear microphones that showed a low positioning variability in a previous study [29] and were connected to the analogue inputs of the RME HDSPe AIO interface via a Lake People C360 microphone pre-amplifier. In a next step, the fit of the microphones was checked by the experimenter and adjusted to be flush with the entrance of the ear canal if necessary, after which the microphone cables were fixed using medical tape. Afterwards, the height and depth of the chair, as well as the depth of the chair's neck support were adjusted for a comfortable and natural sitting position. The subjects' ear canals were aligned with two Bosh Quigo cross-line lasers that marked the center of the array (cf., Fig. 2). Once the seating was finished, the subjects directly faced the speaker at 0° elevation and the position of the turntable was locked. In addition to the optical positioning procedure, a single HRTF for frontal sound incidence was measured to inspect if the left and right ear signals had the same level and time of arrival.

HRIRs were then measured under continuous rotation of the subject using normalized least mean squares (NLMS) adaptive filters [15]. Simulated measurements showed that a rotation speed of 60 s/360° and a step-size of $\mu = 0.5$, which controls the adaption speed and noise suppression of the NLMS filter, are sufficient to obtain high quality HRIRs with a length of 256 samples [30]. For the measurements, the initial length of the HRIRs was set to 1024 samples to account for reflections from opposing speaker membranes and the door of the anechoic chamber [28]. The rotation speed was thus adjusted accordingly to 240 s/360°. For measuring the HRIRs, the subjects were positioned as described above and rotated for 285 s while running the NLMS system identification. The first 45 s were used to allow the filter to adapt and to warm up the speakers. The position of the turntable was tracked by measuring the resistance of a high precision endless potentiometer built into the axis of rotation and by linear interpolation from a look-up table with values stored in 1° resolution. The tracking procedure had a precision of approximately 0.1° due to slight non-linearities of the potentiometer.

An initial inspection of HRIRs measured with speakers from opposing sides of the inner construction showed slight level differences and misalignments (i.e., differences in times of arrival) for the majority of subjects. It was assumed that this was caused by a slight subject dependent tilt of the hydraulic chair's vertical axis of rotation. All data for odd elevations, that were measured with the speakers

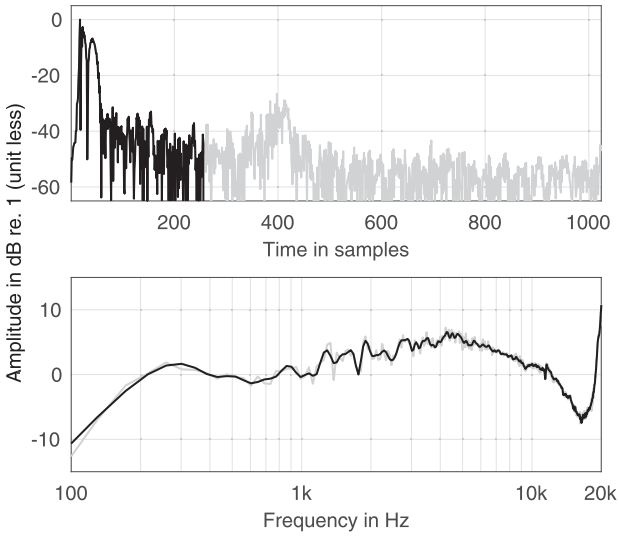

Fig. 3. Impulse response (top) and magnitude spectrum (bottom) of the reference measurement for the speaker at the equator. Gray curves show the initial filter length that was used for measuring the data; black curves show the final filter length. The frequency response of the microphone (DPA 4060) was removed.

of the rear semicircle of the inner construction, were thus discarded, and impulse responses were saved with a resolution of 10° in elevation. The azimuthal resolution can be freely chosen because the NLMS outputs quasi continuous HRIRs [15]. The chosen resolution in azimuth slowly increases with elevation to yield an almost constant great circle distance between neighboring points of the same elevation ϑ ($10^\circ @ |\vartheta| \leq = 30^\circ$; $12^\circ @ |\vartheta| = 40^\circ$; $15^\circ @ |\vartheta| = 50^\circ$; $20^\circ @ |\vartheta| = 60^\circ$; $24^\circ @ |\vartheta| = 70^\circ$; $60^\circ @ |\vartheta| = 80^\circ$; $360^\circ @ |\vartheta| = 90^\circ$). This resulted in a full-spherical sampling grid with $Q = 440$ points.

HRTFs were then obtained by spectral division of the pressure at the blocked ear canal by the pressure in the center of the head with the head absent [3]. The latter was obtained in reference measurements carried out separately for each in-ear microphone. The measurement protocol was identical to measuring the ear signals, however, the chair, turntable, and the supporting structure were removed for this purpose. The reference was measured once before the actual HRTFs were measured and the playback level was identical for the reference and HRTF measurements. Note that the referencing also removes the (on-axis) frequency responses of the loudspeakers, microphones, AD/DA converters, and amplifiers. An example for a reference measurement of the loudspeaker at the equator is given in Fig. 3.

To discard reflections from opposing speakers that show up approximately 375 samples after the direct sound, the HRIRs and the reference were truncated to 256 samples before the spectral division. Reflections from neighboring speakers broaden the main impulse and cannot be discarded in post-processing. This detrimental effect, however, appears small compared to the benefits of using individual instead of non-individual HRTFs [31]. Subsequent to the spectral division, a circular shift was applied to assure causality as described for the simulated HRIRs. The HRTFs could not be reliably measured below approximately 200 Hz due to the limited bandwidth of the loudspeakers and environmental noise. To account for this, the low frequency content of the measured data was extrapolated with the numerically simulated data. While the magnitude spectrum could also have been extrapolated to 0 dB, using the simulated data also corrects the phase response where the extrapolation is more complicated. Simulated and measured HRTFs were combined using 4th order Linkwitz-Riley cross-over filters with a –6 dB cut-off frequency of 300 Hz. Before combining the data, the simulated HRIRs were aligned to their measured counterparts using fractional delays [32], i.e., by shifting the simulated HRIRs to the left or right to match the time of arrival of their measured counterparts (70th order Kaiser windowed sinc filters, 60 dB side lobe attenuation, magnitude and group delay distortions <0.1 dB and <0.01 samples, $\forall f < 20$ kHz). Ideally, onset times would be identical across measured and simulated HRTFs, which would make an alignment obsolete. However, differences in onset times were observed that stem from small imperfections in the positioning of the subjects during the acoustic measurements, and from the torso which influences the time of arrival for sources at low elevations but was removed before HRTF simulation. The amount of delay was estimated by finding the maximum of the cross-correlation between the 10 times upsampled measured and simulated HRIRs separately for each source position and the left and right ear. In a final step, a squared sine fade-in/out of 10/100 samples was applied. Post-processing and measurements were conducted using Matlab and routines from AKtools [33].

A spatially continuous HRTF representation was again achieved using SH processing. The 440 point spatial sampling grid allows for a SH transform of order $N = 16$. At this order, processing artifacts can be observed when directly using the HRTFs for the SH transform [34, 35]. To overcome this, HRTFs were pre-aligned as described by Brinkmann and Weinzierl [35] before SH processing. The SH transform was done based on the pseudo inverse $(\cdot)^\dagger$ [27]

$$\mathbf{f}_{nm} = \mathbf{Y}^\dagger \mathbf{f}\,, \tag{5}$$

with the SH coefficients $\mathbf{f}_{nm} = [f_{0,0},\, f_{1,-1},\, f_{1,0},\, f_{1,1},\, \ldots\, f_{N,N}]^\mathrm{T}$, and a frequency bin of pre-aligned HRTF spectra for all Q source positions $\mathbf{f} = [f_1, \ldots, f_Q]^\mathrm{T}$. The $Q \times (N+1)^2$ matrix $\mathbf{Y}$ contains the values of the SH functions for all Q sampling points, orders n, and degrees m. The condition of $\mathbf{Y}$ specifies how much the noise in $\mathbf{f}$ is amplified by the SH transform and thus should be close to one. For the 440 point sampling grid the condition of $\mathbf{Y}$ is 1.23 at $N = 16$. Interpolation to arbitrary source positions can be realized by the inverse SH transform given by Eq. (4).

1.4 Headphone Measurements

Headphone transfer functions (HpTFs) were measured for Sennheiser HD800 S and HD650 headphones directly after measuring the HRTFs without moving the in-ear

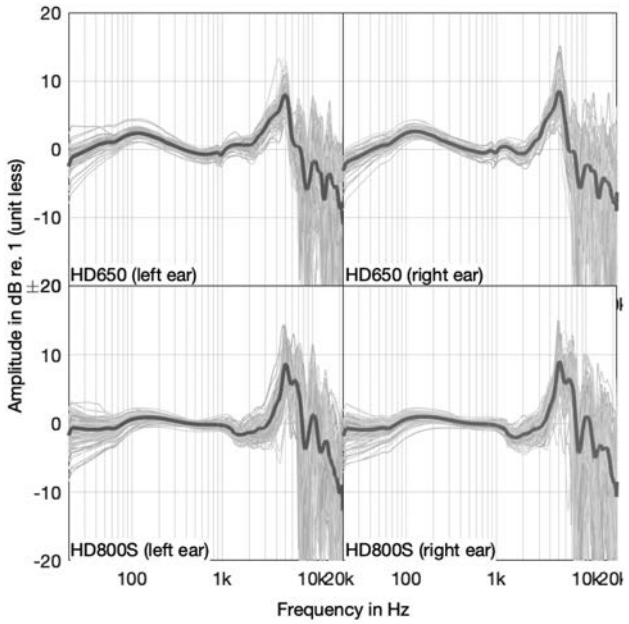

Fig. 4. Averaged HPTFs for all subjects and headphone models (light gray) and averaged HpTFs across subjects (dark).

microphones. To account for re-positioning variability, 10 HpTFs were measured per model after the subjects put the headphones off and on. The measurement level was set to a level that was comfortable for the subjects and where only moderate non-linear distortions below −60 dB were observed. A peak-to-tail signal to noise ratio of about 80 dB was achieved by using exponential sine sweeps [36] with a duration of 2^{16} samples.

In post-processing, HpTFs were truncated to 4096 samples, the microphone frequency responses were removed by means of an inverse filter, and the level was normalized to 0 dB between 300 Hz and 900 Hz. The gain for normalization was obtained by averaging the headphone magnitude spectra in the specified frequency range separately for each headphone and subject, i.e., a single gain was applied for each headphone and subject without touching level differences between the left and right ear. Since the HpTFs are intended to serve as a basis for inverse headphone filters, the normalization assures that the filters have approximately 0 dB gain in the mid-frequency range. The final HpTFs are shown in Fig. 4 for all subjects and headphones.

1.5 Anthropometric Measurements

Twenty-five anthropometric features of the torso, head, and pinnae were extracted from the 3D meshes following the definition from Algazi et al. [4] (cf., Fig. 5). To eliminate the bias of manual measurements, the features were extracted fully automatically if possible by finding characteristic points on the mesh outline [37]. In these cases outliers were manually identified and corrected based on a visual inspection of the distribution of each feature. The pinna rotation and flare angle were extracted semi-automatically using Python scripts for Blender that required the selection of characteristic points on the pinna. The rotation angle was calculated from the angle between the vertical axis (z-axis) and the axis through the highest point on the helix (largest z-coordinate) and the lowest point on the ear lobe (smallest z-coordinate) [38]. The flare angle is defined by the angle between the view-axis of the subject (x-axis) and the line that defines the tragus-to-helix distance [39]. In this case the outmost point on the tragus (largest absolute y-coordinate), and the point on the helix at the same height (z-coordinate) as the point on the tragus were selected to calculate the flare angle.

In some cases the original definitions [4] were modified either because they were blurry or to ease the automatic extraction: The head width (x_1) was taken as the average between points right above and below the ears, instead of between the outmost points of the head (largest absolute y-coordinate). The pinna height and width (d_5, d_6) were measured on the vertical and horizontal axis instead of oblique axes, and the pinna flare angle (Θ_2) was measured against the viewing direction (x-axis), instead of an axis

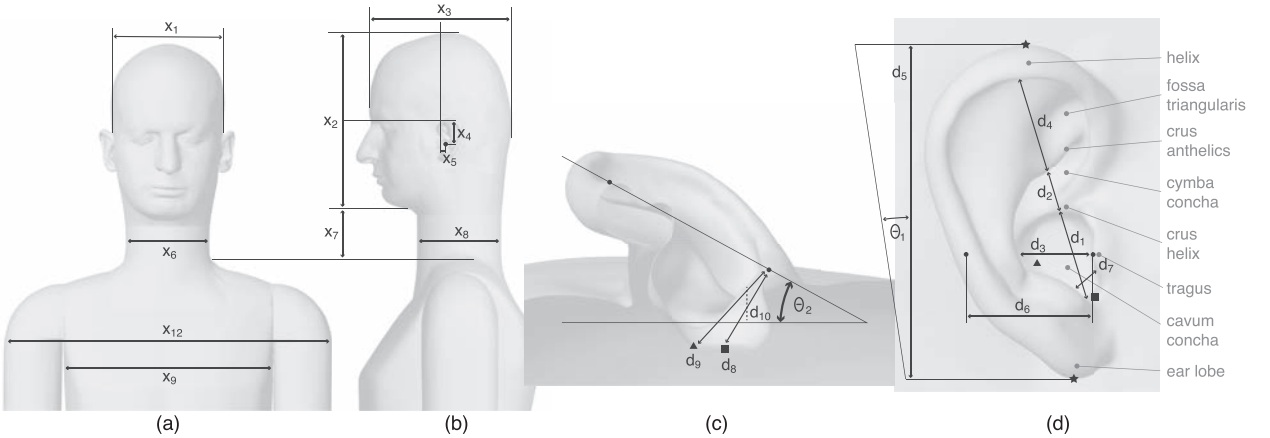

Fig. 5. Definition of anthropometric measures listed in Tab. 2 and characteristic parts of the pinna. The crus helix slant height d_{10} is marked by a dashed line to improve readability. The dot in (b) marks the center of the ear canal. The dots in (c) and (d) indicate the points that were used to estimate the pinna flare angle Θ_2 and the starting point for estimating the cavum concha depths d_8 and d_9. The stars in (d) show the points that were used to estimate the pinna height d_5 and pinna rotation angle Θ_1. The triangle and square in (c) and (d) mark the end points for estimating the cavum concha depth.

Table 2. Anthropometric features in centimeter and degree variation in percent for the 94 human subjects of the database ([a]Automatically extracted, [s]Semi-automatically extracted, [m]manually extracted). The variation is estimated from $100 \cdot 2 \cdot \mathrm{SD}/\mathrm{mean}$.

	feature	mean (SD)	var.
x_1	head width[a]	15.24 (0.72)	9.44
x_2	head height[a]	20.96 (1.04)	9.92
x_3	head depth[a]	20.24 (0.83)	8.20
x_4	pinna offset down[a]	0.27 (0.19)	140.74
x_5	pinna offset back[a]	0.39 (0.32)	164.10
x_6	neck width[a]	12.13 (0.91)	15.00
x_7	neck height[a]	7.02 (1.30)	37.03
x_8	neck depth[a]	11.96 (1.33)	22.24
x_9	torso top width[m]	33.64 (3.52)	20.92
x_{12}	shoulder width[m]	54.80 (4.46)	16.27
x_{14}	height[m]	181.03 (8.28)	9.14
x_{16}	head circumference[m]	56.80 (1.97)	6.93
x_{17}	shoulder circumference[m]	115.11 (8.67)	15.06
d_1	cavum conchae height[a]	1.81 (0.16)	17.67
d_2	cymba concha height[a]	1.01 (0.11)	21.82
d_3	cavum concha width[a]	1.75 (0.19)	21.71
d_4	fossa height[a]	2.10 (0.24)	22.85
d_5	pinna height[a]	6.14 (0.40)	13.02
d_6	pinna width[a]	2.97 (0.26)	17.50
d_7	intertragal incisure width[m]	0.63 (0.14)	46.39
d_8	cavum concha depth (down)[m]	1.15 (0.13)	23.08
d_9	cavum concha depth (back)[m]	1.19 (0.13)	22.07
d_{10}	crus helix slant height[m]	0.31 (0.06)	39.12
Θ_1	pinna rotation angle[s]	10.50 (4.60)	87.61
Θ_2	pinna flare angle[s]	25.35 (7.24)	57.12

defined by the head itself. Two additional features were included: The cavum concha depth was measured from the tragus downwards (d_8) *and* backwards (d_9) to the bottom of the cavum concha, whereas the original definition considered only a single cavum concha depth. The crus helix slant height (d_{10}) was measured from the point between the cavum concha and cymba concha height (d_1, d_2) to the transition of the crus helix to the cavum concha.

2 DATABASE DESCRIPTION

The HUTUBS database comprises 96 subjects—93 human subjects and the FABIAN head-and-torso-simulator [40]. For evaluation, FABIAN (subject IDs 1 and 96), and one human subject (IDs 22 and 88) were measured twice. Fifty-three students and employees of the Audio Communication Group, 32 employees of Sennheiser Electronics, and 8 employees of the HUAWEI Munich Research Centre with a mean age of 36 years (SD 9 years) participated in the acquisition of the database (10 female, 83 male).

The HRTFs are stored in SOFA files [41] at the positions specified by the spatial sampling grids. In addition, the spherical harmonics representation of the HRTFs is stored in Matlab files. To ease the access to this spatially continuous representation, the function `AKhrirDatabase.m` from AKtools [33] can be used to interpolate HRTFs at arbitrary azimuth and elevation angles. While the HpTFs are saved in SOFA files as well, the 3D head meshes are provided in the widespread PLY format, and the anthropometric features are stored in comma separated text files.

The data is available from https://doi.org/10.14279/depositonce-8487 under the free culture CC-BY license that grants unlimited access to everyone. More detailed information on the organization of the data and the interpolation can be found in the accompanying documentation.

3 EVALUATION

Measured and simulated HRIRs and HRTFs of subject 22 are shown in Fig. 6 in the horizontal and median plane. A general similarity between measured and simulated data can be observed and the measured data appears to be slightly noisier. Differences caused by the missing torso in the simulated data are best visible in the median plane HRTFs. The shoulder reflection induced comb filter appears as a u-shaped structure in the measured data, which cannot be seen in the simulated correspondants. In addition, the torso induced damping at low elevations can only be seen in the measured data. However, this effect might be unnaturally large due to the additional shadowing of the turntable and it's supporting structure. A more comprehensive analysis of the HRTFs is detailed in the next paragraphs.

3.1 Across Database Evaluation

It is known that the HRTF measurement system and it's operators can introduce considerable bias in the resulting HRTFs [42]. To asses this bias, the two HRTF sets of the FABIAN head and torso simulator measured with the system introduced above were compared to FABIAN's HRTFs measured with the Oldenburg two arc source positioning system [43, 44].

Absolute energetic spectral differences between HRTF sets were calculated in the horizontal plane using gammatone filters C from the auditory toolbox [45]

$$\Delta G(H_1, H_2, f_c) = \left| 10 \log_{10} \left(\frac{C(f, f_c)|H_1(f)|^2 \mathrm{d}f}{C(f, f_c)|H_2(f)|^2 \mathrm{d}f} \right) \right|, \tag{6}$$

with the complex HRTF spectra H, the filter center frequency f_c in Hz, and 200 Hz $\leq f, f_c \leq$ 20 kHz. For comparison, spectral differences were also calculated between six HRTF sets of the Neumann KU-100 dummy head taken from the inter-laboratory round robin on HRTF measurements[2] [42]. For this purpose, all KU-100 datasets that contain HRTFs for the horizontal plane with an azimuth resolution of 5° were selected (1, 4, 5, 7, 8, and 9). Results in Fig. 7 (top) show that differences between FABIAN datasets are below 1 dB up to 8 kHz, which is smaller than observed differences between most of the KU-100 datasets.

However, they increase to 3 dB at 20 kHz which is larger than most differences seen for the KU-100. An analysis of the signed spectral differences (not shown here) revealed that HRTF measured with the current system contain

[2] Available from https://www.sofaconventions.org/mediawiki/index.php/Files (checked Nov. 2018).

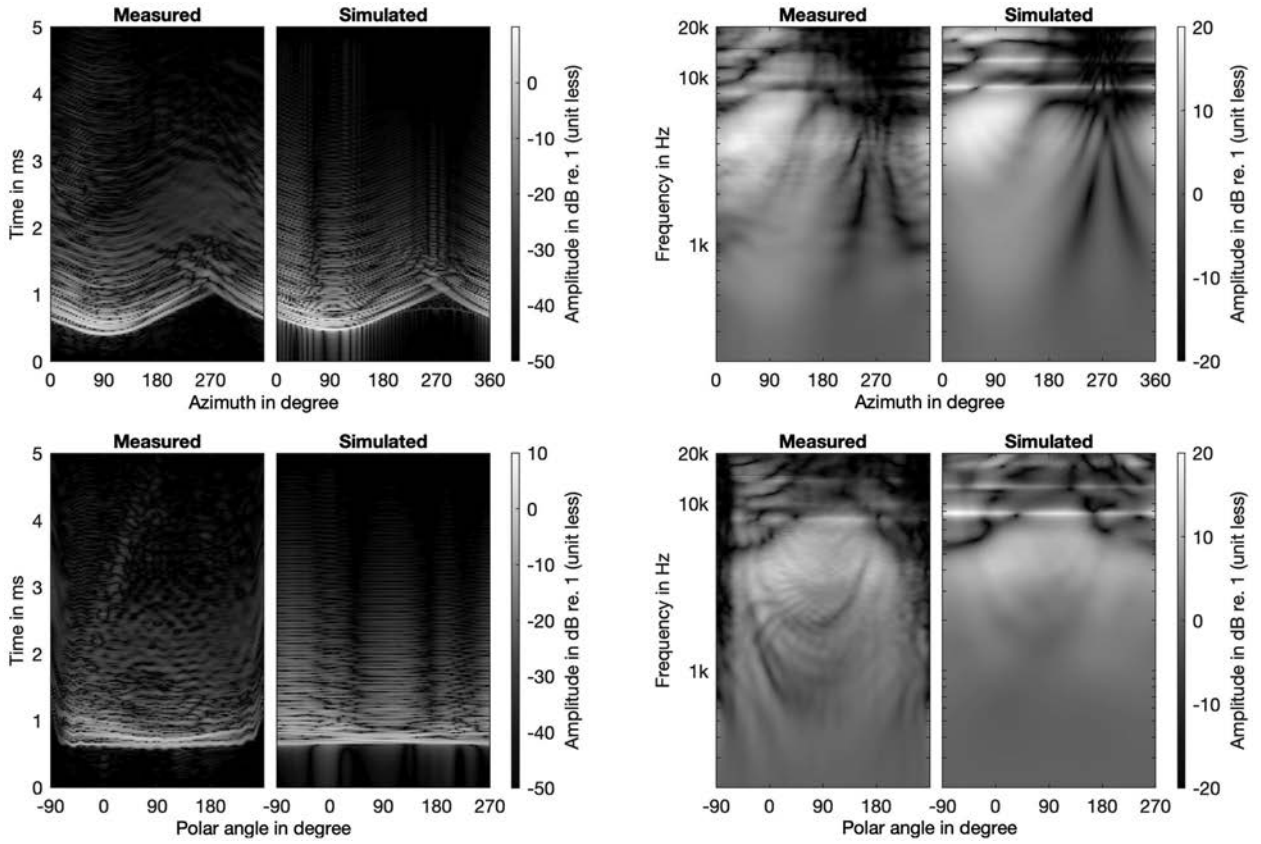

Fig. 6. Measured and simulated HRIRs (left) and HRTFs (right) of subject 22 in the horizontal (top) and median plane (bottom). Data were calculated according to Eq. (4). Azimuth angles of 0°, 90°, 180°, 270° denote sources in front, to the left, behind and to the right. Polar angles of −90°, 0°, 90°, 180° denote sources below, in front, above, and behind.

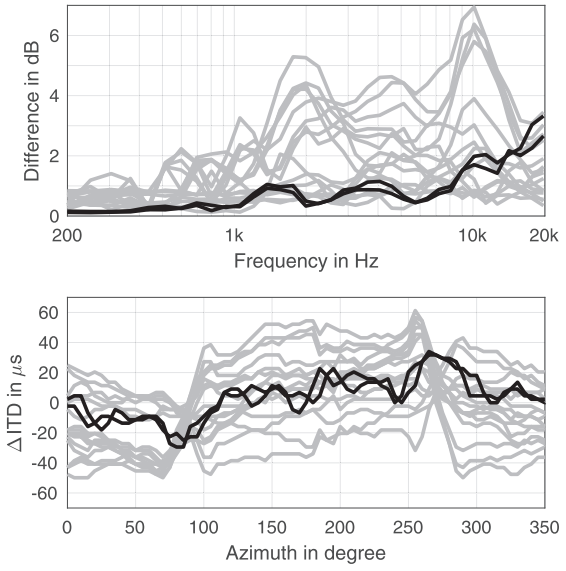

Fig. 7. Across database differences in auditory filter bands (top, averaged across source posistitions) and interaural time differences (bottom) in the horizontal plane. Results for the FABIAN head and torso simulator are given in black, results for the KU-100 in gray.

more energy at frequencies above 8 kHz. These differences might stem from reflections from the back and neck rest of the chair where the subjects were seated (cf., Fig. 2), which was not used in the Oldenburg measurement system (cf., [43], Fig. 1B).

Temporal differences between the HRTF sets were analyzed based on differences of the broadband interaural time difference (ITD). The ITD was calculated from the difference in onset times between the left and right ear. The onset times were estimated using a threshold based onset detection on the 10 times upsampled and low-passed HRIRs (8th order Butterworth at 3 kHz) by finding the first sub-sample with an absolute level of less than –20 dB below the absolute maximum (separately for each HRIR and the left/right ear). This ITD extraction method showed good agreement with the perceived lateralization of a source [46]. ITD differences between FABIAN datasets are below the just noticeable difference (JND) of 20 μs [47] for most azimuth angles, and smaller than most differences observed for the KU-100 (cf., Fig. 7, bottom). Differences above 20 μs only occur for sources to the left and right (azimuth around 90° and 270°) where the JND increases to about 100 μs [48]. These differences most likely stem from positioning inaccuracies of a few degrees and/or centimeters.

3.2 Within Database Evaluation

Besides the bias introduced by the measurement system, HRTFs that were measured with the same system at different points in time can differ due to positioning inaccuracy of the subjects and microphones, and long term variability of the system (e.g., loudspeaker aging) [49, 50]. To investigate this, HRTFs of the FABIAN head and torso simulator (id 1 and 96) and one human

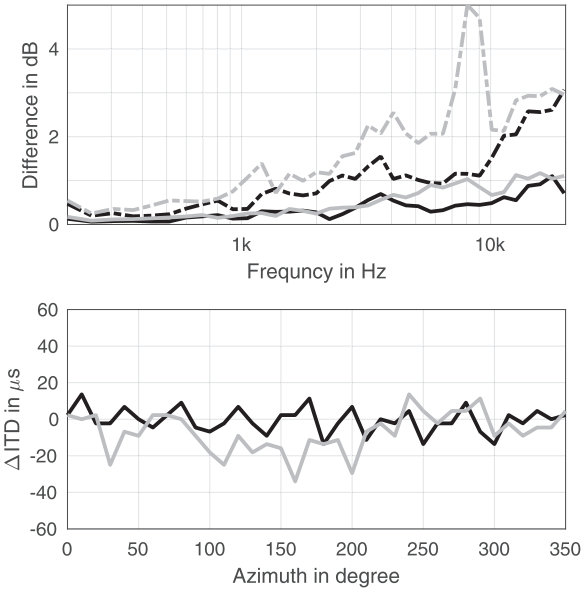

Fig. 8. Within database validation of the FABIAN head and torso simulator (black) and a human subject (gray). Top: Absolute energetic differences in auditory filter bands in the horizontal plane. Solid lines give the median across source positions, dashed lines the 95 percentile. Bottom: Differences in ITD.

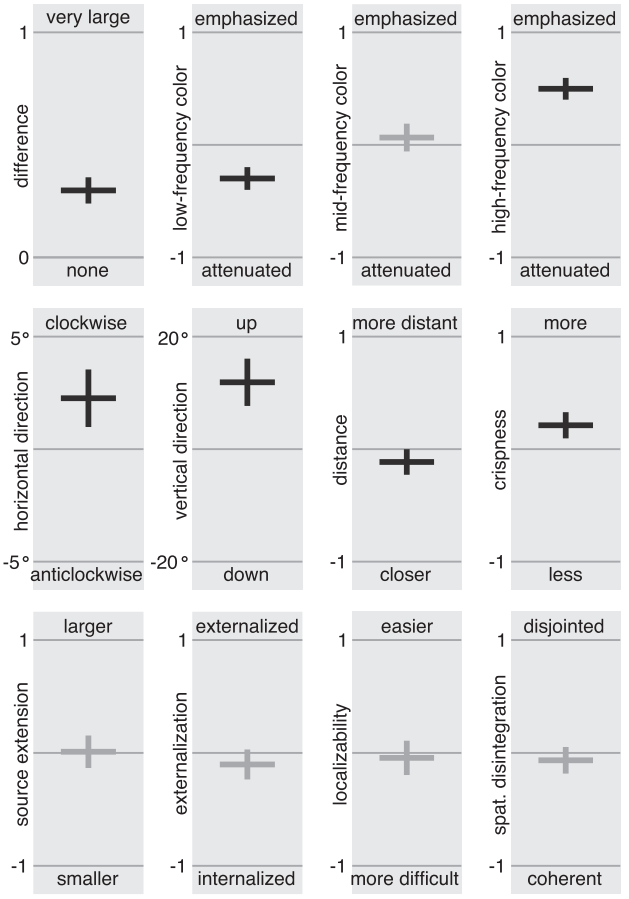

Fig. 9. Results from the perceptual comparison of measured vs. simulated individual HRTFs by means of the group mean (horizontal lines) and 95% confidence intervals (vertical lines). Cases where the confidence interval overlaps zero are shown in light gray to indicate the non-significance of the differences.

subject (id 22 and 88) were measured at the beginning and at the end of the measurement series (11 and 9 days apart), and spectral and temporal differences between the datasets were analyzed in analogy to the across database evaluation.

Results in Fig. 8 show a good average spectral fit between the repeated measurements with deviations below approximately 1 dB over the entire audible frequency range. Larger deviations of up to 5 dB can be observed for frequencies above 7 kHz and the 95 percentile. These differences most likely originate from shifted notches in the HRTF caused by slight positional changes across the repeated measurement sessions. An earlier study showed that these differences are perceptually less relevant, at least for the perceived source position [43]. Differences in ITD are below the JND of 20 μs for all tested source positions for FABIAN and only slightly exceed the JND for the human subject. Largest errors occur at the back of the subject, i.e., 180° azimuth, where the presence of the neck rest might have caused slightly audible ITD disturbances.

In theory, HRTFs exhibit an amplitude of 0 dB at low frequencies where the wave length is large compared to the human body. The measured HRTFs for frontal sound incidence showed a level of 1.25 dB at the left and 1.11 dB at the right ear prior to using the simulated HRTFs for extrapolating towards 0 Hz (averaged across subjects and between 200 and 250 Hz). The slight but systematic deviation from 0 dB can be explained by the fact that the heights of the sitting subjects are already in the range of the wave length where the level was assessed. Differences between the left and right ear were 0.14 dB on average, however, differences above the JND of 1 dB [51, Table 2.4] occurred for three subjects (IDs 39, 45, and 93).

3.3 Cross-Evaluation of Measured and Simulated HRTFs

Finally, a listening test was conducted to assess perceptual differences between measured and simulated HRTFs in direct comparison. For this purpose, 12 perceptual qualities were selected from the Spatial Audio Quality Inventory (SAQI) [52] with an eye on completeness and relevance (cf., Fig. 9). This detailed qualitative analysis was preferred over a simple forced choice test for detecting if any audible differences exist, because a priori informal listening by the authors showed that this was clearly the case.

The Matlab based open source tool WhisPER [53] was used to display the rating interface on a laptop computer and to randomize the order of the 12 qualities. WhisPER sent open sound control (OSC) messages to a second computer that rendered the audio. Dynamic binaural synthesis, accounting for the head orientation of the listener, of a source in front was realized with a modified version of the Sound Scape Renderer (SSR [54]) that loaded SOFA files [41] containing sets of measured and simulated HRTFs. The head rotation of the subjects was monitored with a Polhemus Patriot electro-magnetic tracker with an update rate of 120 Hz. HRTFs were exchanged in real time according the the subject's head position by cross-fading between succes-

sive frames of the block-wise convolution. The total system latency of about 43 ms is below the thresholds found in previous studies [55–58]. The audio signal was played back via an M-Audio Audiophile 192 sound card and Sennheiser HD800 S headphones for which individual HpTFs were measured during the acquisition of the database. To restrict the influence of the headphones to a minimum, individual inverse headphone filters were designed based on regulated inversion [59]. To limit excessive gains at frequencies where notches occur in the HpTFs, a regularization function was constructed from 1 to 3 narrow parametric equalizers per channel, whose center frequencies were manually tuned to coincide with the notches (termed PEQ regularizaton [29]). The gain at low frequencies and close to the Nyquist frequency was limited by a target Butterworth band pass consisting of a second order low-cut at $f_c = 30$ Hz and an eights order high-cut at $f_c = 20$ kHz.

The listening test was preceded by an instruction, in which the participants were informed about the nature of the experiment and the meaning of the 12 qualities. A training was conducted for familiarization with the stimuli and test procedure. For effectivity, different audio contents were presented that were thought to be most critical and suitable for the tested qualities: Anechoic male speech was used for *difference*, *distance*, and *externalization*, a dry drum recording was used for *crispness*, and pink noise was used for the remaining qualities. The user interface showed two buttons labeled *A* and *B* that triggered auralizations of the measured and simulated HRTF sets. A continuous slider was used for ratings that had the labels "±1" and "0" close to the end points and in the middle. For orientation, five intermediate steps were marked by dashes. When rating differences in perceived source positions, subjects reported an estimate of the angular offset in degree. In these cases the results are intended to give a rough estimate only and have to be interpreted with care because the verbal elicitation of angles can be distorted. All participants were instructed to listen to *A* and *B* as often and in any order they wanted, and to move their head within the common range of motion [60] of ±42° in horizontal and ±16° in vertical direction. The duration of the test including instructions and training was approximately 15 to 20 minutes. The test provided a blind quality assessment, i.e., the subjects did not know whether *A* triggered auralizations based on measured or simulated HRTFs. A bias due to different interpretation of the scale end labels (e.g., non – very large) can, however, not be avoided with this method and might result in increased variance and confidence intervals.

In total, 42 subjects participated in the experiment that are also part of the HUTUBS database. With the exception of 2, all subjects had participated in perceptual tests before. Results for all qualities are shown in Fig. 9, indicating the group mean and 95% confidence intervals.

Differences of the simulated HRTFs were rated with respect to their measured counterparts, i.e., a zero-rating indicates no perceivable difference for the respective quality, and a rating of 1 very large differences. While the *difference* indicates that measured and simulated HRTFs were generally distinguishable, 5 of the 11 remaining qualities

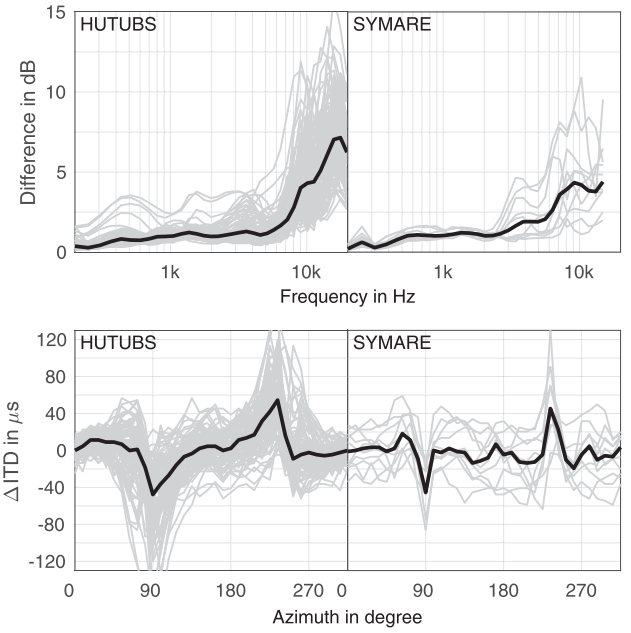

Fig. 10. Differences between measured and modeled HRTFs for the HUTUBS (left) and SYMARE database (right). Absolute energetic differences are shown in the top row, differences in the broadband ITD at the bottom. Median differences are given by black lines, individual differences for 96 subjects (HUTUBS) and 10 subjects (SYMARE) are given by gray lines.

show statistically non-significant differences between the two conditions. Largest differences were observed for the tone color, where simulated HRTFs were perceived as attenuated at low and emphasized at high-frequencies, and in the source position that was reported to be shifted clockwise for 2° and upwards for 12°. The means for the remaining qualities with significant differences were rather small (from −0.1 to 0.2).

To further analyze perceptual deviations in the tone color and localization, spectral and temporal differences between measured and simulated HRTFs were calculated in analogy to previous Sections, with the exception that spectral differences were calculated for all source positions above the horizontal plane (cf., Fig. 10, left).

Spectral differences are below approximately 1 dB up to 5 kHz and rise to 7 dB at 17.6 kHz on average. An analysis of the signed error (not shown here) revealed that the simulated HRTFs contained more high frequency energy as their measured counterparts. This not only explains the ratings for *low/high-frequency tone color*, but might also account for the fact that the simulated HRTFs were perceived as more crisp and closer to the head. Considering that an increase in energy at about 8 kHz can evoke the perception of a source above the listener [61], the coloration could also be partially responsible for the mismatch in the *vertical direction*, where the simulated HRTFs were perceived to be shifted up. Please note that differences at and below 200 Hz would be larger, if the simulated HRTFs would not have been used for extrapolating the measured data. Differences in the ITD are below the JND of 20 μs for most participants and source positions. Larger values were

observed for lateral source positions, where differences occasionally exceed the JND of 100 µs. This however does not account for the systematic shift in the *horizontal direction* of about 2°.

Lastly, differences between measured and simulated HRTFs were compared to the public part of the SYMARE database [9] (cf., Fig. 10, right). Average spectral differences are well comparable up to 5 kHz, above 5 kHz deviations are approximately 2.5 dB larger for the current database. An analysis of the signed errors (not shown here) revealed that simulated HRTFs of the SYMARE database contain more energy at higher frequencies as their measured correspondents. This general behavior can at least partially be explained by neglecting the absorbing effect of clothes and hair during numerical simulation [62]. Average differences in the horizontal plane ITD are also comparable across the two databases.

4 DISCUSSION AND SUMMARY

The publicly available HUTUBS HRTF database consists of measured and simulated HRTFs, measured HpTFs, anthropometric measurements, and high quality 3D surface meshes of 96 subjects. HRTFs were acquired on sampling grids that allow a perceptually transparent and spatially continuous representation in the spherical harmonics domain. The validation of the HRTFs showed a good agreement to HRTFs from a different measurement system, and between repeated measurements from the current system. Moreover, a perceptual comparison between measured and simulated HRTFs showed relatively small deviations in general, which were attributed to differences in the tone color in most cases.

Before numerically simulating the HRTFs, the shoulders and torso were removed from the 3D meshes. While the shoulder reflects incoming sound if it is approximately aligned with the source and ear, which results in a comb filter with a first minimum between 700 Hz and 1 kHz, the torso damps high frequencies for sources at low elevations [63]. Omitting the torso will increase localization errors outside the median plane [63, 64] and result in partly incorrect HRTFs in static rendering, e.g., a source directly below the listener will have too much high frequency energy. If including a fixed torso, however, errors will occur in dynamic rendering [65], e.g., an unnatural high frequency damping would be observed for a fontal source if the listener is looking up (because the HRTFs in dynamic rendering are picked according to the source-to-head orientation in a head-centric coordinate system). For this reason, the torsos were removed and a dynamic torso model that adds the comb filter and high frequency damping in dependency of the source position and listener view might be subject to further research. This could assure correct HRTF rendering under static and dynamic conditions.

5 ACKNOWLEDGMENTS

The authors would like to thank all participants of the HUTUBS database, Jan Joschka Wohlgemuth and Fabian Seipel for help during the setup of the measurement system and acquisition of the database, Alexis Baskind for sharing his ideas on tracking the turntable position with a potentiometer, and Martin Pollow for logistic help with the Artec scanning system.

6 REFERENCES

[1] H. Møller, M. F. Sørensen, C. B. Jensen, and D. Hammershøi, "Binaural Technique: Do We Need Individual Recordings?" *J. Audio Eng. Soc.*, vol. 44, pp. 451–469 (1996 Jun.).

[2] Y. Kahana, *Numerical Modelling of the Head-Related Transfer Function*, Ph.D. Thesis, University of Southampton, UK (2000 Dec.).

[3] H. Møller, "Fundamentals of Binaural Technology," *Appl. Acoust.*, vol. 36, pp. 171–218 (1992), doi:https://doi.org/10.1016/0003-682x(92)90046-u.

[4] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF Database," presented at the *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 99–102 (2001 Oct.), doi:https://doi.org/10.1109/ASPAA.2001.969552.

[5] O. Warusfel, "LISTEN HRTF Database," http://recherche.ircam.fr/equipes/salles/listen/index.html (2003).

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

[7] Austrian Academy of Sciences, "The ARI HRTF Data Base," https://www.kfs.oeaw.ac.at/index.php?view=article&id=608&lang=en (2014).

[8] K. Watanabe, Y. Iwaya, Y. Suzuki, S. Takane, and S. Sato, "Dataset of Head-Related Transfer Functions Measured with a Circular Loudspeaker Array," *Acoustical Sci. & Tech.*, vol. 35, no. 3, pp. 159–165 (2014).

[9] C. Jin, P. Guillon, N. Eapain, R. Zolfaghari, A. van Schaik, A. I. Tew, C. Hetherington, and J. Thorpe, "Creating the Sidney York Morphological and Acoustic Recordings of Ears Database," *IEEE Trans. on Multimedia*, vol. 16, no. 1, pp. 37–46 (2014 Jan.), doi:https://doi.org/10.1109/tmm.2013.2282134.

[10] R. Bomhardt, M. de la Fuente Klein, and J. Fels, "A High-Resolution Head-Related Transfer Function and Three-Dimensional Ear Model Database," *Proceedings of Meetings on Acoustics*, vol. 29, no. 1, p. 050002 (2016), doi:https://doi.org/10.1121/2.0000467.

[11] D. S. Brungart and W. M. Rabinowitz, "Auditory Localization of Nearby Sources. Head-Related Transfer Functions," *J. Acoust. Soc. Amer.*, vol. 106, no. 3, pp. 1465–1479 (1999 Sep.).

[12] P. Majdak, P. Balazs, and B. Laback, "Multiple Exponential Sweep Method for Fast Measurement of Head-Related Transfer Functions," *J. Audio Eng. Soc.*, vol. 55, pp. 623–637 (2007 Jul./Aug.).

[13] S. Weinzierl, A. Giese, and A. Lindau, "Generalized Multiple Sweep Measurement," presented at the *126th Convention of the Audio Engineering Society* (2009 May), convention paper 7767.

[14] P. Dietrich, B. Masiero, and M. Vorländer, "On the Optimization of the Multiple Exponential Sweep Method," *J. Audio Eng. Soc.*, vol. 61, pp. 113–124 (2013 Mar.).

[15] G. Enzner, C. Antweiler, and S. Spors, "Acquisition and Representation of Head-Related Transfer Functions," in J. Blauert (Ed.), *The Technology of Binaural Listening*, Modern acoustics and signal processing, 1st ed., pp. 57–92 (Springer, Heidelberg et al., 2013), doi:https://doi.org/10.1007/978-3-642-37762-4.

[16] M. Rothbucher, K. Veprek, P. Paukner, T. Habigt, and K. Diepold, "Comparison of Head-Related Impulse Response Measurement Approaches," *J. Acoust. Soc. Amer. (Express Letter)*, vol. 134, no. 2, pp. EL223–EL229 (2013 Aug.).

[17] E. A. Torres-Gallegos, F. Orduña-Bustamente, and F. Arámbula-Casío, "Personalization of Head-Related Transfer Functions (HRTF) Based on Automatic Photo-Anthropometry and Inference from a Database," *Applied Acoustics*, vol. 97, pp. 84–95 (2015).

[18] M. Dinakaran, F. Brinkmann, S. Harder, R. Pelzer, P. Grosche, R. R. Paulsen, and S. Weinzierl, "Perceptually Motivated Analysis of Numerically Simulated Head-Related Transfer Functions Generated by Various 3D Surface Scanning Systems," presented at the *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, pp. 551–555 (2018 Apr.), doi:https://doi.org/10.1109/ICASSP.2018.8461789.

[19] Z. Zhang, "Microsoft Kinect Sensor and its Effect," *IEEE MultiMedia*, vol. 19, no. 2, pp. 4–10 (2012 Feb.), doi:https://doi.org/10.1109/MMUL.2012.24.

[20] D. M. Ramírez, J. Jiménez, E. G. Ramírez, H. J. Paniagua, and V. C. Ruidíaz, "Discrepancies in Cephalometric Measurements in Relation to Natural Head Position," *Revista Mexicana de Ortodoncia*, vol. 1, no. 1, pp. 27–32 (2013 Oct.-Dec.).

[21] P. Cignoni, M. Callieri, M. Corsini, M. Dellepiane, F. Ganovelli, and G. Ranzuglia, "MeshLab: An Open-Source Mesh Processing Tool," presented at the *Eurographics Italian Chapter Conference* (2008), doi:https://doi.org/10.2312/LocalChapterEvents/ItalChap/ItalianChapConf2008/129-136.

[22] H. Ziegelwanger, W. Kreuzer, and P. Majdak, "Mesh2HRTF: An Open-Source Software Package for the Numerical Calculation of Head-Related Transfer Functions," presented at the *22nd International Congress on Sound and Vibration* (2015 Jul.).

[23] B. Bernschütz, C. Pörschmann, S. Spors, and S. Weinzierl, "SOFiA Sound Field Analysis Toolbox," presented at the *Proceedings of the ICSA International Conference on Spatial Audio* (2011).

[24] H. Ziegelwanger, W. Kreuzer, and P. Majak, "Effect of Element Size and Microphone Model on the Numerically Calculated Head-Related Transfer Functions," presented at the *AIA-DAGA 2013, International Conference on Acoustics*, pp. 600–603 (2013).

[25] J. Möbius and L. Kobbelt, "OpenFlipper: An Open Source Geometry Processing and Rendering Framework," presented at the *7th Int. Conf. Curves and Surfaces*, pp. 488–500 (2010 Jun.).

[26] H. Ziegelwanger, W. Kreuzer, and P. Majdak, "A-Priori Mesh Grading for the Numerical Calculation of the Head-Related Transfer Functions," *Appl. Acoust.*, vol. 114, pp. 99–110 (2016 Dec.), doi:https://doi.org/10.1016/j.apacoust.2016.07.005.

[27] B. Rafaely, *Fundamentals of Spherical Array Processing*, 1st ed. (Springer, Berlin, Heidelberg, Germany, 2015), doi:https://doi.org/10.1007/978-3-662-45664-4.

[28] A. Fuß, F. Brinkmann, T. Jürgensohn, and S. Weinzierl, "Ein vollsphärisches Multikanalmesssystem zur schnellen Erfassung räumlich hochaufgelöster, individueller kopfbezogener Übertragungsfunktionen," presented at the *Fortschritte der Akustik – DAGA 2015*, pp. 1114–1117 (2015).

[29] A. Lindau and F. Brinkmann, "Perceptual Evaluation of Headphone Compensation in Binaural Synthesis Based on Non-Individual Recordings," *J. Audio Eng. Soc.*, vol. 60, pp. 54–62 (2012 Jan./Feb.).

[30] M. Fallahi, F. Brinkmann, and S. Weinzierl, "Simulation and Analysis of Measurement Techniques for the Fast Acquisition of Head-Related Transfer Functions," presented at the *Fortschritte der Akustik – DAGA 2015*, pp. 1107–1110 (2015).

[31] J.-G. Richter and J. Fels, "Evaluation of Localization Accuracy of Static Sources Using HRTFs from a Fast Measurement System," *Acta Acust. united Ac.*, vol. 102, no. 4, pp. 763–771 (2016 Jul./Aug.).

[32] T. I. Laakso, T. Saramäki, and G. D. Cain, "Asymmetric Doph-Chebychef, Saramäki, and Transitional Windows for Fractional Delay FIR Filter Design," *Proc. 38th Midwest Symposium on Circuits and Systems (MWSCAS-95)*, pp. 580–583 (1995 Aug.).

[33] F. Brinkmann and S. Weinzierl, "AKtools – An Open Software Toolbox for Signal Acquisition, Processing, and Inspection in Acoustics," presented at the *142nd Convention of the Audio Engineering Society* (2017 May), e-Brief 309.

[34] C. Pike and T. Tew, "Subjective Assessment of HRTF Interpolation with Spherical Harmonics," presented at the *4th Int. Conf. Spatial Audio (ICSA)* (2017 Sep.), URL slideshare.net/ChrisPike21/subjective-assessment-of-hrtf-interpolation-with-spherical-harmonics-chris-pike-and-tony-tew.

[35] F. Brinkmann and S. Weinzierl, "Comparison of Head-Related Transfer Functions Pre-Processing Techniques for Spherical Harmonics Decomposition," presented at the *2018 AES International Conference on Audio for Virtual and Augmented Reality (AVAR)* (2018 Aug.), conference paper P9-3.

[36] A. Farina, "Simultaneous Measurement of Impulse Response and Distortion with Swept-sine technique," presented at the *108th Convention of the Audio Engineering Society* (2000 Feb.), convention paper 5093.

[37] M. Dinakaran, P. Grosche, F. Brinkmann, and S. Weinzierl, "Extraction of Anthropometric Measures from 3D-Meshes for the Individualization of Head-Related Transfer Functions," presented at the *140th Convention of the Audio Engineering Society* (2016 Jun.), convention paper 9579.

[38] J. Hall, J. E. Allanson, K. W. Gripp, and A. M. Slavotinek, *Handbook of Physical Measurements*, 2nd ed. (Oxford University Press, Oxford, UK, 2007).

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

[40] A. Lindau, T. Hohn, and S. Weinzierl, "Binaural Resynthesis for Comparative Studies of Acoustical Environments," presented at the *122th Convention of the Audio Engineering Society* (2007 May), convention paper 7032.

[41] AES Standards Comittee, *AES69-2015: AES standard for file exchange - Spatial acoustic data file format* (Audio Engineering Society, Inc.) (2015).

[42] A. Andreopoulou, D. R. Begault, and B. F. G. Katz, "Inter-Laboratory Round Robin HRTF Measurement Comparison," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 5, pp. 895–906 (2015 Feb.), doi:https://doi.org/10.1109/JSTSP.2015.2400417.

[43] F. Brinkmann, A. Lindau, S. Weinzierl, S. v. d. Par, M. Müller-Trapet, R. Opdam, and M. Vorländer, "A High Resolution and Full-Spherical Head-Related Transfer Function Database for Different Head-Above-Torso Orientations," *J. Audio Eng. Soc.*, vol. 65, pp. 841–848 (2017 Oct.), doi:https://doi.org/10.17743/jaes.2017.0033.

[44] F. Brinkmann, A. Lindau, S. Weinzierl, G. Geissler, S. van de Par, M. Müller-Trapet, R. Opdam, and M. Vorländer, "The FABIAN Head-Related Transfer Function Data Base," doi:https://doi.org/10.14279/depositonce-5718.2 (2017 Feb.).

[45] M. Slaney, "Auditory Toolbox. Version 2," Technical report #1998-010, Interval Research Corporation (1998).

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

[47] A. W. Mills, "On the Minimum Audible Angle," *J. Acoust. Soc. Amer.*, vol. 30, no. 4, pp. 237–246 (1958 Apr.), doi:https://doi.org/10.1121/1.1909553.

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

[49] K. A. J. Riederer, "Part IIIa: Effect of Microphone Position Changes on Blocked Cavum Conchae Head-Related Transfer Functions," presented at the *18th Intern. Congress on Acoustics*, pp. 787–790 (2004 Apr.).

[50] K. A. J. Riederer, "Part Va: Effect of Head Movements on Measured Head-Related Transfer Functions," presented at the *18th Intern. Congress on Acoustics*, pp. 795–798 (2004 Apr.).

[51] J. Blauert, *Spatial Hearing. The Psychophysics of Human Sound Localization*, revised ed. (MIT Press, Camebridge, MA, 1997).

[52] A. Lindau, V. Erbes, S. Lepa, H.-J. Maempel, F. Brinkmann, and S. Weinzierl, "A Spatial Audio Quality Inventory (SAQI)," *Acta Acust. united Ac.*, vol. 100, no. 5, pp. 984–994 (2014 Sep./Oct.), doi:https://doi.org/10.3813/AAA.918778.

[53] S. Ciba, A. Wlodarski, and H.-J. Maempel, "WhisPER – A New Tool for Performing Listening Tests," presented at the *126th Convention of the Audio Engineering Society* (2009 May), convention paper 7749.

[54] M. Geier, J. Ahrens, and S. Spors, "The Sound Scape Renderer: A Unified Spatial Audio Reproduction Framework for Arbitrary Rendering Methods," presented at the *124th Convention of the Audio Engineering Society* (2008 May), convention paper 7330.

[55] P. Mackensen, *Auditive Localization. Head Movements, an Additional Cue in Localization*, Dissertation, Technische Universität Berlin (2004 Apr.).

[56] D. S. Brungart, B. D. Simpson, and A. J. Kordik, "The Detectability of Headtracker Latency in Virtual Audio Displays," presented at the *Eleventh Meeting of the International Conference on Auditory Display (ICAD)*, pp. 37–42 (2005 Jul.).

[57] S. Yairi, Y. Iwaya, and Y. Suzuki, "Investigation of System Latency Detection Threshold of Virtual Auditory Display," presented at the *12th Int. Conf. on Auditory Display (ICAD)*, pp. 217–222 (2006 Jun.).

[58] A. Lindau, "The Perception of System Latency in Dynamic Binaural Synthesis," presented at the *NAG/DAGA 2009, International Conference on Acoustics*, pp. 1063–1066 (2009).

[59] S. G. Norcross, M. Bouchard, and G. A. Soulodre, "Inverse Filtering Design Using a Minimal Phase Target Function from Regularization," presented at the *121st Convention of the Audio Engineering Society* (2006 Oct.), convention paper 6929.

[60] W. R. Thurlow, J. W. Mangels, and P. S. Runge, "Head Movements During Sound Localization," *J. Acoust. Soc. Amer.*, vol. 42, no. 2, pp. 489–493 (1967), doi:https://doi.org/10.1121/1.1910605.

[61] J. Blauert, "Sound Localization in the Median Plane," *Acoustica*, vol. 22, pp. 205–213 (1969/70).

[62] B. F. G. Katz, "Boundary Element Method Calculation of Individual Head-Related Transfer Function. II. Impedance Effects and Comparisons to Real Measurements," *J. Acoust. Soc. Amer.*, vol. 110, no. 5, pp. 2449–2455 (2001 Nov.), doi:https://doi.org/10.1121/1.1412441.

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

[64] H. Møller, D. Hammershøi, C. B. Jensen, and M. F. Sorensen, "Evaluation of Artificial Heads in Listening Tests," *J. Audio Eng. Soc.*, vol. 47, pp. 83–100 (1999 Mar.).

[65] 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.), doi:https://doi.org/10.1109/jstsp.2015.2414905.

THE AUTHORS

Fabian Brinkmann

Manoj Dinakaran

Robert Pelzer

Peter Grosche

Daniel Voss

Stefan Weinzierl

Fabian Brinkmann received his M.A. degree (magister artium) in communication sciences and technical acoustics from TU Berlin, Germany. Since 2011, he has been a Research Associate at the Audio Communication Group from TU Berlin and is associated to the DFG research consortium SEACEN in which he completed his Ph.D. in the field of signal processing and evaluation approaches for spatial audio.

•

Manoj Dinakaran has received his M.Sc. degree in signal processing from Blekinge Institute of Technology, Sweden, in 2013. Since the end of 2013, he has been pursuing his Ph.D. degree at the Munich Research center, Huawei Technologies, Germany, and in Technical University of Berlin, Germany, in the field of individualizing head related transfer functions.

•

Robert Pelzer received an M.Sc. degree in audio communication and technology at TU Berlin in 2018. He wrote his thesis on perceptual analysis of head-related transfer functions for individualism and is also interested in music information retrieval and sound synthesis.

•

Peter Grosche received his B.S. and M.Sc. degrees in electrical engineering and information technology from Technical University of Munich in 2006 and 2008, respectively. From 2008 to 2012, he completed his Ph.D. in multimedia information retrieval and music processing group at Saarland University. He is currently working in Huawei European Research Center in Munich as Principle research engineer.

•

Daniel Voss received his Dipl.-Ing. degree in electrical engineering from Leibniz Universität Hannover in 2009. He wrote his diploma thesis on BEM simulations of microphone array enclosures. Since then he has been with Sennheiser's research department, focusing in the field of active noise cancelation and audio signal processing.

•

Stefan Weinzierl is head of the Audio Communication Group at the Technische Universität Berlin. His research activities are focused on audio technology, virtual acoustics, room acoustics, and musical acoustics. He is coordinating a master program in Audio Communication and Technology at TU Berlin. With a diploma in physics and sound engineering, he received his Ph.D. in musical acoustics from TU Berlin. He is currently coordinating research consortia in the field of virtual acoustics (SEACEN, DFG) and music information retrieval (ABC_DJ, H2020).