A High Resolution and Full-Spherical Head-Related Transfer Function Database for Different Head-Above-Torso Orientations

FABIAN BRINKMANN,[1] *AES Associate Member*, **ALEXANDER LINDAU**[1], **STEFAN WEINZIERL**[1],

(fabian.brinkmann@tu-berlin.de)   (alexander.lindau@aesthetics.mpg.de) (stefan.weinzierl@tu-berlin.de)

STEVEN VAN DE PAR,[2] *AES Member*, **MARKUS MÜLLER-TRAPET**[3],

(steven.van.de.par@uni-oldenburg.de)   (markus.mueller-trapet@nrc-cnrc.gc.ca)

ROB OPDAM[3]**, AND**   **MICHAEL VORLÄNDER**[3]

(rob.opdam@akustik.rwth-aachen.de)   (mvo@akustik.rwth-aachen.de)

[1]*Audio Communication Group, Technical University Berlin, Germany*
[2]*Acoustics Group, Cluster of Excellence "Hearing4all", Carl von Ossietzky University, Oldenburg, Germany*
[3]*Institute of Technical Acoustics, RWTH Aachen University, Aachen, Germany*

Head-related transfer functions (HRTFs) were acoustically measured and numerically simulated for the FABIAN head and torso simulator on a full-spherical and high resolution sampling grid. Moreover, HRTFs were acquired for 11 horizontal head-above-torso orientations, covering the typical range of motion of $\pm 50°$, making it possible to account for head movements of the listeners in dynamic binaural auralizations in a physically correct manner. In lack of an external reference for HRTFs, measured and simulated data sets were cross-validated by applying auditory models for localization performance and spectral coloration and by correlation analyses. The results indicate a high degree of similarity between the two data sets regarding all tested aspects, thus suggesting that they are free of systematic errors. The HRTF database is publicly available from https://doi.org/10.14279/depositonce-5718.2 and is accompanied by a wide range of headphone filters for use in binaural synthesis.

1 INTRODUCTION

Head-related transfer functions (HRTFs) capture the free field sound transmission from a sound source to the listeners ears. They incorporate all cues for sound localization such as interaural time and level differences (ITD, ILD) and spectral cues that originate from scattering, diffraction, and reflection on the human pinnae, head, and body [1]. Using binaural synthesis and room acoustic simulation [2], HRTFs can thus be used to simulate spatial hearing, and open up a wide range of virtual auditory display applications such as guiding systems [3], game and mobile sound [4], or room acoustic design [5], and acoustic recreation of historic spaces [6].

Algazi et al. [7] showed that the torso effects the HRTF by means of reflecting or shadowing sound waves travelling toward the listeners' ears. The reflection is strongest if source, shoulder, and ear are approximately aligned, and superimposes a comb filter to the HRTF with a magnitude of up to ± 5 dB. The first comb filter maxima occurs at approximately 700 Hz for sources at high elevations, and

gradually increases in frequency with decreasing elevation. Shadowing occurs for sources well below the horizontal plane and causes a high frequency damping of up to 25 dB that increases with decreasing sound source elevation. Perceptual investigations revealed that the cues induced by the torso and head are relevant for localizing the elevation of sources away from the sagittal median plane when pinna cues are absent [7]. Moreover, differences between head-above-torso orientations (HATOs) can be audible even for HRTFs that exhibit only weak torso effects [8]. Although the broadband interaural time and level differences (ITD, ILD) remain mainly unaffected by the HATO, it might be assumed that the HATO affects the ITD fine structure. This is known to be the case for the head and was assumed to provide additional elevation cues and to help resolve front-back confusion [9]. Besides the influence of the torso on localization and timbre, dynamic HRTF cues related to head movements and HATO also affect other aspects of spatial hearing. It was, for instance, observed that listeners naturally move their heads without moving the torso when judging perceptual sound field qualities such as source width, or
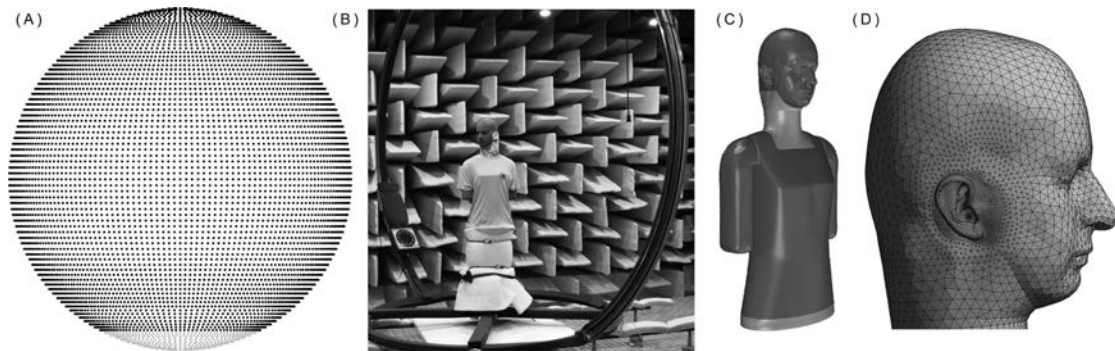
Fig. 1 (A) Spherical sampling grid. Grey points show source positions below −64°. (B) Two arc source positioning system with FABIAN set up in its geometrical center. (C) 3D model of FABIAN. Light gray areas were manually inserted in post-processing. (D) Detail of the fine 3D surface mesh used for numerical simulation. Shaded area marks the microphone position.

envelopment [10], and that head movements help to resolve front-back confusion and source elevation [11].

Nevertheless, currently available public HRTF sets—for an overview see [12]—were either measured for a fixed HATO or for dummy heads without torso. In the present study, we thus acquired HRTFs for multiple HATOs using acoustic measurements and numeric simulations as outlined in Sec. 2. In lack of an external reference for HRTFs, Sec. 3 details a cross-validation procedure that covers temporal and spectral aspects, as well as modeled localization performance. Please note that the current publication outdates the preliminary post-processing of the acoustic measurements [13] and extends the initial corss-validation [14] to all HATOs and localization performance. Last, Sec. 4 describes the publicly available HRTF database.

2 HRTF ACQUISITION

HRTFs of the FABIAN head and torso simulator [15] were acquired for 11,950 source positions with a dense spatial resolution (2° in elevation; 2° great circle distance in azimuth, cf., Fig. 1A) that makes it suitable for a high order spherical harmonic representation. Moreover, these high resolution data sets were obtained individually for 11 HATOs that covered the typical range of head rotation of ±50° to the left and right [16], and with a resolution of 10° allowing for their perceptually transparent interpolation [8].

2.1 Acoustic Measurements

Measurements were conducted in the anechoic chamber of the Carl von Ossietzky University Oldenburg ($V = 238$ m^3, cut-off frequency 50 Hz). To allow for subsequent HRTF identification, sine sweeps with a length of 2^{16} samples were recorded with DPA 4060 microphones at the entrances to FABIAN's blocked ear canals (*recorded HRTF*), and at the position of the interaural center in the absence of FABIAN (*reference*). HRTFs were then derived by spectral division of the recorded HRTFs and the reference, yielding a signal to noise ratio (SNR) of 80 dB for ipsilateral and 55 dB for contralateral sources after averaging across four measurements. The sweep was designed in the frequency domain between 100 Hz
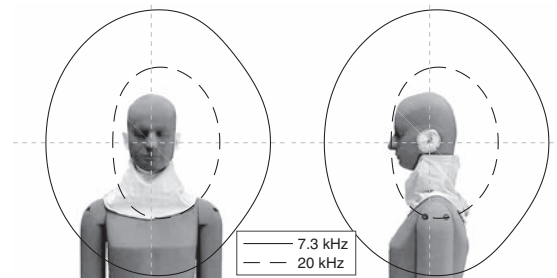

Fig. 2 Directivity of one speaker from the TASP projected onto FABIAN by means of the −3 dB main lobe at 7.3 kHz and 20 kHz. Dashed crosses mark the position of the interaural center.

and 20 kHz based on the group delay [17]. For achieving an almost frequency independent SNR, the energy of the sweep was set to be proportional to the background noise. The bandwidth was restricted according to the lower cut-off frequency of the loudspeakers used for measuring (Manger MSW bending-wave sound transducer in a custom made cubic closed box). *AKtools* [18] were used for sweep synthesis and deconvolution, as well as audio playback, and recording at a sampling rate of 44.1 kHz.

The two-arc-source-positioning system (TASP [19]), consisting of opposing semicircular arcs with a radius of 1.7 m, was used for positioning the sources with a precision of 0.1°. The two arcs could be rotated horizontally and were each equipped with a Manger MSW bending-wave sound transducer on vertically movable mounts (cf., Fig. 1B). Due to mechanical restrictions, HRTFs could not be obtained for elevations below −64°. Before the measurements, FABIAN's interaural center was carefully aligned to the geometrical center of the TASP using a self-leveling Bosch PCL10 cross-line laser with the frontal viewing direction being defined by a laser pointer attached to FABIAN's neck joint.

Before measuring HRTFs for one HATO, 5.000 warm-ups of the sweep signal were played back through each loudspeaker to reduce their time variability from ±4 dB to ±1 dB. Nevertheless, a comparison of HRTFs for different HATOs showed differences of up to 2.5 dB within narrow frequency bands centered around approximately 0.5 and 2 kHz, that were consistent across source positions. These
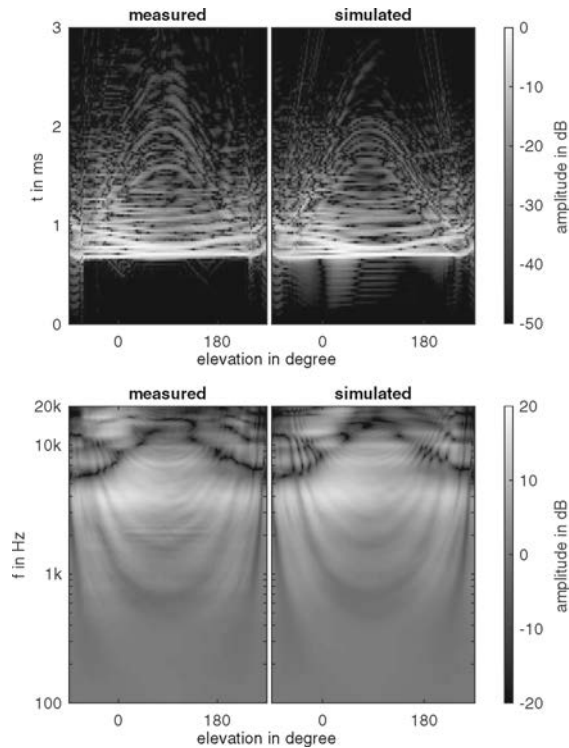
Fig. 3 Comparison of left ear median plane HRIRs (top), and HRTFs (bottom) for neutral HATO. Elevations of $0°$, $90°$, and $180°$ denote sources to the front, above, and in the back of FABIAN.

variations were attributed to long term changes in the loudspeakers' frequency responses caused by loudspeaker aging and are visible in Fig. 3 (bottom, left) by means of a horizontal line at 2 kHz for elevations between $14°$, and $166°$. The variations at 0.5 kHz are less distinct and not visible in Fig. 3. Although this frequency range is also affected by the comb-filter caused by the shoulder reflection (U-shaped arches in Fig. 3, bottom), the bandwidth of the observed artifacts is too narrow to be attributed to this effect.

By obtaining HRTFs through spectral division (deconvolution), the on-axis frequency response of the microphones, speakers, amplifiers, and converters cancels out, but the speakers directivity remains uncompensated. However, directivity measurements of the Manger transducers showed that FABIAN's shoulders and torso were within the –3 dB main lobe up to about 7 kHz. Consequently, the speakers directivities should be of negligible influence, because above 3–4 kHz, HRTFs are mainly influenced by the head and pinnae [7], which were within the main lobe up to 20 kHz (cf., Fig. 2). The directivities were initially measured on 5 x 5 equal angle sampling grid and were comparable across speakers. However, the oval shape at high frequencies, which might be caused by the custom box design, was less pronounced for the second speaker. The main lobe widths were estimated in the spatially continuous spherical harmonics domain after applying a spherical harmonics transform of order 25 [20].

Prior to deconvolution, a circular shift of 28 samples was applied to the recorded HRTFs to maintain causality and to ensure approximately 20 leading samples before the earliest peak in the head-related impulse responses (HRIRs). To discard reflections from opposing loudspeakers, HRIRs were truncated to 256 samples (5.8 ms). Finally, 10 (20) samples squared sine fade-ins (fade-outs) were applied.

To obtain an estimate of the mechanical reliability of the TASP, four ever identical HRTFs—neutral head orientation, sources to the front, back, left and righ—were measured before and after acquiring a set of HRTFs. Deviations in the HRTFs magnitude spectra within and between sets were generally below ±1.5 dB, but increased to ±10 dB at the contralateral ear, and in the vicinity of notches. This, however, is well comparable to previous dummy head HRTF measurements [21], while slightly larger variability was observed for human subjects [22].

2.2 Numeric Simulations

Numeric HRTF simulation was done by means of the boundary element method (BEM) that requires a 3D surface mesh representation of FABIAN. This was generated in the following way: an initial point cloud representation was measured with a *GOM ATOS I* structured light scanner. A resolution of approximately $1/_{100}$ mm for the head and pinnae, and $1/_{10}$ mm for the torso was achieved by automatic alignment of overlapping scans, relying on manually inserted reference points and conducted with the surface matching algorithm of *ATOS Professional*. A nonuniform rational basis spline (NURBS) representation was built from the point cloud using *Geomagix Studio 12*. Subsequently, *Rhino 4* was used to (I) design a cylindrical neck with a seamless transition between head and torso, (II) to extend the torso bottom to its original size, (III) close screw holes in the arms, and (IV) connect the separate arm scans to the torso (cf., Fig. 1C). At last, *Virtual.Lab Acoustics 13.1* was used for mesh generation and calculating complex HRTF spectra at frequencies between 100 Hz and 22 kHz (100 Hz resolution). For acceleration, different triangular meshes were generated: a coarse mesh with edge lengths of 2 mm (pinnae), and 10 mm (head and torso) was used for simulations up to 6 kHz, and a fine mesh with edge lengths of 2 mm (pinnae and head), and 5 mm (torso) was used for the fast multipole method (FMM) BEM above 2 kHz (cf., Fig. 1D). The chosen edge lengths fulfil the typical requirement of six elements per wavelength [23], and it was shown that a resolution of 1–2 mm sufficiently captures the details of the pinna geometry [24]. The overlapping region between 2 kHz and 6 kHz was used to verify that both simulations yielded identical results. Constant velocity boundary conditions were imposed to the mesh elements corresponding to the microphone at the entrances to the blocked ear canals. Otherwise, the mesh was assumed to be acoustically rigid. While this simplified assumption appears to be valid for human skin [25], it does not strictly hold for FABIAN's neck, which consists of a metal neck joint covered by a leather fabric, and was for simplicity modeled by a cylindrical shape that was smoothly extended towards the head and torso. Because the fabric of FABIAN's t-shirt with a thickness of less than 1 mm does not compare to existing experimental results for the effect of cloth in

HRTF [26, 27], it was also assumed to be rigid. Moreover, FABIAN's stand, which was wrapped in absorbing material during the acoustic measurements was not modeled due to computational restrictions.

As with the acoustic measurements, HRTFs were calculated by spectral division of the result at the sampling grid points by the analytical solution of a point source with the same volume velocity placed in the center of the coordinate system; the frequency bin at 0 Hz was set to 0 dB. HRIRs with a length of 441 samples and 44.1 kHz sampling rate were obtained by inverse Fourier transform after mirroring the single sided spectra. Finally, the simulated HRIRs were windowed in the same way as their measured counterparts.

3 CROSS-VALIDATION

A visual comparison of measured and simulated HRTFs showed a good agreement (cf., Fig. 3). In lack of an external reference for HRTFs, cross-validation between measured and simulated data was already suggested by Turku et al. [28], who perceptually tested differences in localization and preference. Moreover, Jin et al. [29] assessed differences in head radii and spatial correlation, however, without providing evidence for the perceptual relevance of the suggested measures. In the current study, we physically conducted the cross-validation by comparing the temporal and spectral structure as well as the modeled median plane localization performance.

3.1 Temporal Structure

In theory, the time of arrival (TOA), i.e., the onset in the HRIRs, should be identical across measured and simulated data sets. However, average (and maximum) differences of $\tau = 1.2$ ($\tau = 4$) samples (27 μs and 91 μs) were observed between the two conditions, which equals a displacement of 9 mm (31 mm), or $0.3°$ ($1°$) ($c = 339$ m/s according to the average temperature during the measurements of $11.4°$ C and the TASP radius of 1.7 m). Because the geometrical alignment of FABIAN was assumed to be close to perfect for the simulated HRIRs, differences in TOA can be caused by temperature fluctuations and positioning inaccuracy during the acoustic measurements. The latter was supported by an analysis of τ across source positions, revealing slight discontinuities of up to about three samples (not shown here) that were attributed to the start and end points of the TASP rotation and the transition between the two loudspeakers. Moreover, observed temperature fluctuations during the measurements of $3.1°$ C could induce an error of up to 1.2 samples (27 μs). The results of the TOA analysis suggest a high reliability of the setup and that there should be no audible differences between measured and modeled HRIRs caused by mechanical inaccuracy or temperature fluctuation.

Nevertheless, the simulated data were used for correcting the TOA of the measured HRIRs because time alignment was a prerequisite for the processing steps described in the next section. Alignment was done using fractional delays [30] (Kaiser windowed sinc filters of order 70, 60 dB side

lobe attenuation), with negligible magnitude and group delay distortions (<0.1 dB; <0.01 samples, $\forall f < 20$ kHz). As a result, the average cross-correlations between pairs of measured and simulated HRIRs increased from 0.51 to 0.93. Notably, the cross-correlation varied across HATO before the alignment (standard deviation $\sigma = 0.35$), showing the influence of the experimental setup but was almost constant afterward ($\sigma = 0.05$). The fractional delaying induced changes in the broad band interaural time difference (ITD) of up to 66 μs for lateral sources (cf., Fig. 4A). This, however, is below average discrimination thresholds for broad band noise reported by Mossop and Culling [31] and was thus assumed to be perceptually irrelevant.

Besides differences in TOAs, simulated HRIRs exhibit more pre-ringing (cf., Fig. 3, top right). Pre-ringing inevitably occurs in band limited digital signals if the TOA does not coincide with a sampling point of the impulse response. The fact that it is higher for the simulated data, implies that HRIRs are less damped, which might be a consequence of assuming rigid boundary conditions during numeric simulation.

3.2 Spectral Structure

Differences between measured and simulated HRTF magnitude spectra were analyzed in 40 auditory filter bands

$$\Delta G(f_c) = 10\log \frac{\int C(f, f_c)\, |\mathrm{HRTF}_{\mathrm{measured}}(f)|^2\, \mathrm{d}f}{\int C(f, f_c)\, |\mathrm{HRTF}_{\mathrm{simulated}}(f)|^2\, \mathrm{d}f}\,,$$

where $C(f, f_c)$ are filters from the auditory modeling toolbox [32] with center frequency f_c, and 50 Hz $\leq f, f_c \leq 20$ kHz. Results averaged across source positions and HATOs are shown as black lines in Fig. 4B. Deviations below 200 Hz, where the level of the measured HRTFs is systematically lower, can be attributed to the band limited frequency response of the loudspeakers. Above 200 Hz, the median deviation p_{50} is smaller than ± 1 dB, indicating that both HRTF data sets are free of systematic errors. Moreover, 90% of the differences given by the $p_{5, 95}$ percentile range are below ± 2 dB up to approximately 7 kHz, again emphasizing the very good agreement between measured and simulated data sets. The large difference between $p_{5, 95}$, and $p_{0, 100}$ in the range of 1–7 kHz indicates that $p_{0, 100}$ is dominated by occasional outliers. Higher differences above 7 kHz of ± 10 dB and more originate from contralateral source positions where the HRIRs carry less energy, and mismatched HRTF notches across data sets. They might thus be perceptually less relevant at least for source localization, that is assumed to be dominated by the ipsilateral ear and the overall spectral shape across auditory filters [33].

Observed differences at high frequencies are difficult to correct as they can either be caused by uncertainties in the measured HRTFs (e.g., positioning errors), or by simplifying assumptions of the BEM simulation (e.g., surface impedance). However, differences at low frequency can clearly be attributed to non-ideal loudspeaker characteristics. Given the good fit of measured and simulated HRIRs at mid frequencies, it seems reasonable to assume that the simulated data can be used to replace the measured data
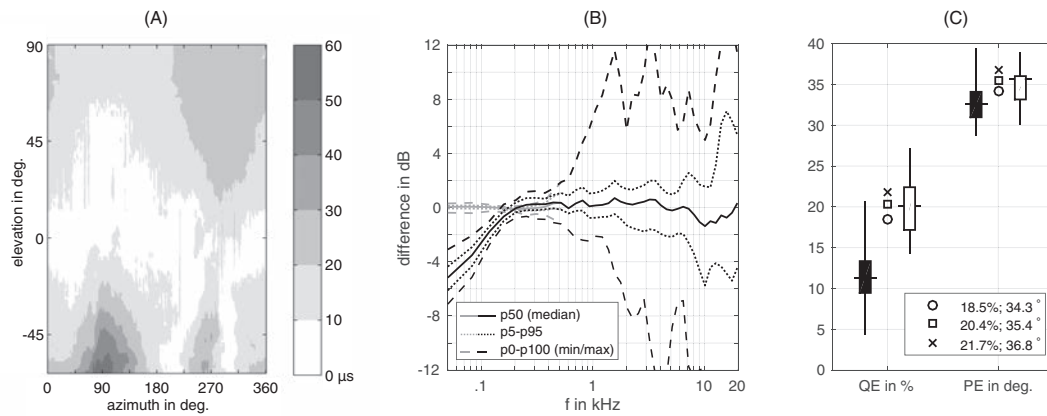
Fig. 4 Cross-validation of measured and simulated HRTFs: (A) Maximal changes in broad band ITD across HATOS due to fractional delaying of measured HRIRs. (B) Spectral differences between measured and simulated HRTFs (averaged across source positions and HATOS) given by selected percentile ranges $p_{i,j}$ in 40 auditory filter bands. Black lines show differences before replacing measured data, gray lines show differences afterwards. (C) Modeled localization performance by means of polar error (PE) and quadrant error (QE) for measured (circles), simulated (squares), and measured vs. simulated (crosses) FABIAN HRTFs, accompanied by results for the ARI database (black boxes), and a dummy head database (white boxes). Horizontal lines show the median, boxes show the interquartile range, and vertical lines span between the minimum and maximum values.

at low frequencies. Consequently, measured and simulated HRTF magnitude and unwrapped phase spectra were combined separately using a linear fade between 200 and 500 Hz (gray lines in Fig. 4B). Remaining differences below 200 Hz shown by $p_{0,100}$ are smaller than 0.5 dB and are caused by truncation and windowing.

The simulated data were also used to fill-up the missing HRTFs below $-64°$ elevation that could not be measured due to mechanical restrictions. To obtain a smooth transition between the measured and modeled data, a linear fade was applied by interpolating magnitude and unwrapped phase spectra for elevations between $-60°$ to $-64°$. This caused a slight discontinuity in the HRIRs (vertical line in Fig. 3, top, left) but was preferred to spherical harmonics based extrapolation [34] that resulted in low-passed HRTFs at the missing positions due to a low truncation order.

3.3 Median Sagittal Plane Localization

The localization performance in the median sagittal plane was assessed using the probabilistic model of Baumgartner et al. [33], which compares the spectral structure of a target HRTF set to a set of template HRTFs. Based on this comparison, it estimates quadrant errors (QE)—specifying the percentage of front-back/up-down confusion—and polar errors (PE)—specifying the angular rms error between actual and modeled source positions. Correspondingly, if target and template HRTFs are identical, the model indirectly assesses the uniqueness of an HRTF within the data set compared to the remainder, whereas, if passing different HRTF sets, it assesses the similarity between them. To be comparable to Baumgartner et al., we assumed a median listener sensitivity of $S_l = 0.76$ and considered only elevations above $-30°$.

The within-localization performance averaged across HATOS shows a very good agreement between data sets: QEs for simulated HRIRs (squares in Fig. 4C) are only 2% worse than those of their measured counterparts (circles in

Fig. 4C), and PEs differ by only $1°$. For external validation, localization performance was also modeled using HRTFs of all human subjects from the ARI database (solid lines in Fig. 4C), and all dummy heads (dashed lines in Fig. 4C) available from [12]. Average localization performance is better for human subjects than for dummy heads, a fact that deserves further investigation, however, the estimates for FABIAN are in excellent agreement with the dummy head results. Since the localization model relies on spectral structure, this result indicates a general validity of the FABIAN HRIRs. In addition, the between-localization performance, i.e., passing measured HRTFs as template and simulated HRTFs as target (crosses in Fig. 4C), is only marginally worse than the within performance. This supports the hypothesis that large spectral differences given by $p_{0,100}$ in Fig. 4B are perceptually less relevant regarding localization.

4 DATABASE

The FABIAN head and torso simulator database is publicly available from https://doi.org/10.14279/depositonce-5718.2. Measured and simulated head-related impulse responses (HRIRs) are given in the SOFA format [35]. They are accompanied by minimum phase common transfer functions (CTFs) and their inverse. Based on the assumption of a diffuse sound field, CTFs were computed by power averaging HRTF magnitude spectra

$$CTF(f) = \sqrt{\sum_i |HRTF_i(f)|^2 \, w_i} \,,$$

with frequency f and averaging weights w_i. The latter were computed according to the areas of spherical rectangles around each source position in Fig. 1A, normalized to $\sum_i w_i = 1$. The CTFs were calculated separately for measured and simulated HRTF sets, and averaged across HATOS.

For convenience, HRTF data sets were transformed into spherical harmonics (SH) coefficients separately for each HATO by applying the 35th order discrete spherical harmonics transform (DSHT) to the complex HRTF spectra [20, eq. (1.9), and (3.34)]. This converts the spatially discrete HRIR data sets into a continuous representation and reduces the the amount of data by a factor of about 10 (256 real HRIR values $\times$ 119050 source positions vs. 129 frequencies $\times$ $(35 + 1)^2$ complex SH coef.). HRIR interpolation between source positions (in the SH domain) and HATOs (in the frequency domain [8]) can be done using `AKhrirInterpolation` from the *AKtools* [18].

To foster the use of the database, headphone impulse responses (HpIRs) measured on FABIAN, and corresponsding inverse filters of about 35 models including widespread Sennheiser HD600, AKG K701, and Beyerdynamic DT770 headphones are part of the database. The filters were calculated by means of the regulated least mean square inversion [36]. Parametric equalizers (PEQs)—fitted to the notches in the headphone transfer functions—were used for regularization to avoid an exact inversion in these cases. They are intended for headphone equalization in binaural synthesis. In this context, the inverse CTFs could be used as a generic headphone compensation filter considering the diffuse field HRTF to be a common target curve in headphone development [37]. Additionally, all generated 3D meshes are included as well as example auralizations of fixed and moving sources.

5 SUMMARY

An HRTF data set of the FABIAN head-and-torso simulator was generated by measuring and simulating HRTFs for a high resolution, full spherical sampling grid, and 11 head-above-torso orientations. A detailed cross-validation showed a very good agreement in terms of temporal and spectral structure, as well as modeled localization performance. In turn, the simulated HRTFs were used to correct the time of arrivals and low frequency response in their measured counterparts. The data set is publicly available and comprises 11,950 HRIRs for each HATO, corresponding spherical harmonics coefficients, 3D surface meshes, numerous headphone filters for binaural synthesis, and auralizations of fixed and moving sources for comparing measured and simulated HRIRs [38]. A perceptually transparent interpolation between different HATOs can be done using *AKtools* [18]. Future work could evaluate the influence of the head-above-torso orientation on the ITD fine structure.

6 ACKNOWLEDGMENTS

7 REFERENCES

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

[2] M. Vorländer, *Auralization. Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*, 1st edition (Springer, Berlin, Heidelberg, Germany, 2008), DOI: https://doi.org/10.1007/978-3-540-48830-9.

[3] M. Bujacz, P. Skulimowski, and P. Strumiłło, "Naviton—A Prototype Mobility Aid for Auditory Presentation of Three-Dimensional Scenes to the Visually Impaired," *J. Audio Eng. Soc.*, vol. 60, pp. 696–708 (2012 Sep.).

[4] J. Sinker and J. Angus, "Efficient Compact Representation of Head Related Transfer Functions for Portable Game Audio," presented at the *AES 56th International Conference: Audio for Games* (2015 Feb.), conference paper 4-2.

[5] S. Pelzer, L. Aspöck, D. Schröder, and M. Vorländer, "Integrating Real-Time Room Acoustics Simulation into a CAD Modeling Software to Enhance the Architectural Design Process," *Building Acoustics*, vol. 4, no. 2, pp. 113–138 (2014 Apr.), DOI: https://doi.org/10.3390/buildings4020113.

[6] A. Pedrero, A. Díaz-Chyla, C. Díaz, S. Pelter, and M. Vorländer, "Virtual Restoration of the Sound of the Hispanic Rite," *Forum Acusticum*, Kraków, Poland (2014 Sep.).

[7] 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.), DOI: https://doi.org/10.1121/1.1349185.

[8] 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.

[9] V. Benichoux, M. Rébillat, and R. Brette, "On the Variation of Interaural Time Differences with Frequency," *J. Acous. Soc. Am.*, vol. 139, no. 4, pp. 1810–1821 (2016), DOI: https://doi.org/10.1121/1.4944638.

[10] C. Kim, R. Mason, and T. Brookes, "Head Movements Made by Listeners in Experimental and Real-Life Listening Activities," *J. Audio Eng. Soc.*, vol. 61, pp. 425–438 (2013 Jun.).

[11] K. I. McAnally and R. L. Martin, "Sound Localization with Head Movement: Implications for 3-D Audio Displays," *Frontiers in Neuroscience*, vol. 8, p. 210 (2014), DOI: https://doi.org/10.3389/fnins.2014.00210.

[12] P. Majdak et al., "Collection of Public Head-Related Impulse Response Data Bases," (last checked September 2017), URL https://www.sofaconventions.org/mediawiki/index.php/Files.

[13] F. Brinkmann, A. Lindau, S. Weinzierl, G. Geissler, and S. van de Par, "A High Resolution Head-Related Transfer Function Database including Different Orientations of Head above the Torso," *AIA-DAGA 2013, International Conference on Acoustics*, pp. 596–599, Merano, Italy (2013 Mar.).

[14] F. Brinkmann, A. Lindau, M. Müller-Trapet, M. Vorländer, and S. Weinzierl, "Cross-Validation of Measured and Modeld Head-Related Transfer Functions," *Fortschritte der Akustik – DAGA 2015*, pp. 1118–1121, Nürnberg, Germany (2015 Mar.).

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

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

[17] S. Müller and P. Massarani, "Transfer Function Measurement with Sweeps," *J. Audio Eng. Soc.*, vol. 49, pp. 443–471 (2001 Jun.).

[18] F. Brinkmann and S. Weinzierl, "AKtools— An Open Toolbox for Acoustic Signal Acquisition, Processing, and Inspection" (2016), URL www.ak.tu-berlin.de/AKtools.

[19] J. Otten, *Factors Influencing Acoustical Localization*, Dissertation, Carl von Ossietzky University Oldenburg, Germany (2001 Sep.).

[20] 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.

[21] B. P. Bovbjerg, F. Christensen, P. Minnaar, and X. Chen, "Measuring the Head-Related Transfer Functions of an Artificial Head with a High Directional Resolution," presented at the *109th Convention of the Audio Engineering Society* (2000 Sep.), convention paper 5264.

[22] H. Møller, M. F. Sørensen, D. Hammershøi, and C. B. Jensen, "Head-Related Transfer Functions of Human Subjects," *J. Audio Eng. Soc.*, vol. 43, pp. 300–321 (1995 May).

[23] R. Ciskowski and C. Brebbia, *Boundary Element Methods in Acoustics* (Elsevier Applied Science, London, UK, 1991).

[24] H. Ziegelwanger, P. Majak, 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.).

[25] 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. Am.*, vol. 110, no. 5, pp. 2449–2455 (2001 Nov.), DOI: https://doi.org/10.1121/1.1412441.

[26] G. F. Kuhn, "Model for the Interaural Time Differences in the Azimuthal Plane," *J. Acous. Soc. Am.*,

vol. 62, no. 1, pp. 157–167 (1977), DOI: https://doi.org/10.1121/1.381498.

[27] O. Kirkeby, E. T. Seppälä, A. Kärkkäinen, L. Kärkkäinen, and T. Huttunen, "Some Effects of the Torso on Head-Related Transfer Functions," presented at the *122nd Convention of the Audio Engineering Society* (2007 May), convention paper 7030.

[28] J. Turku, M. Vilermo, E. Seppälä, M. Pölönen, O. Kirkeby, A. Kärkäinnen, and L. Kärkkäinen, "Perceptual Evaluation of Numerically Simulated Head-Related Transfer Functions," presented at the *124th Convention of the Audio Engineering Society* (2008 May), convention paper 7489.

[29] 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.

[30] T. I. Laakso, V. Välimäki, M. Karjalainen, and U. K. Laine, "Splitting the Unit Delay," *IEEE Signal Processing Magazine*, vol. 13, no. 1, pp. 30–60 (1996 Jan.), DOI: https://doi.org/10.1109/79.482137.

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

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

[33] R. Baumgartner, P. Majak, 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.), DOI: https://doi.org/10.1121/1.4887447.

[34] J. Ahrens, M. R. P. Thomas, and I. J. Tashev, "HRTF Magnitude Modeling Using a Non-Regularized Least-Squares Fit of Spherical Harmonics Coefficients on Incomplete Data," *APSIPA Annual Summit and Conference*, Hollywood, CA, USA (2012 Dec.).

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

[36] 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.).

[37] H. Møller, D. Hammershøi, C. B. Jensen, and M. F. Sørensen, "Transfer Characteristics of Headphones Measured on Human Ears," *J. Audio Eng. Soc.*, vol. 43, pp. 203–217 (1995 Apr.).

[38] 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," https://doi.org/10.14279/depositonce-5718.2 (2016 Feb.).

THE AUTHORS

Fabian Brinkmann

Alexander Lindau

Stefan Weinzierl

Steven van de Par

Markus Müller-Trapet

Rob Opdam

Michael Vorländer

Fabian Brinkmann received his M.A. degree (magister artium) in communication sciences and technical acoustics from TU Berlin, Germany. Since 2011he 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 is currently pursuing the Ph.D. degree in the field of signal processing and evaluation approaches for spatial audio.

•

Alexander Lindau obtained an M.A. degree (magister artium) in communication sciences, electrotechnical engineering, and technical acoustics, and a doctoral degree (Dr. rer. nat.) from TU Berlin. From 2011 to 2015 he was a Research Associate at the Audio Communication Group of the TU Berlin and involved in the DFG research consortium SEACEN where he focused on new approaches towards the perceptual evaluation of spatial audio technologies. Since 2015, he heads the ArtLab at the Max Planck Institute for Empirical Aestetics.

•

Stefan Weinzierl received a diploma in physics and sound engineering and a two-year study in musicology at UC Berkeley, he received his Ph.D. from TU Berlin, where he now heads the Audio Communication Group. His research activities include audio technology, musical acoustics, room acoustics, and virtual acoustics. He is coordinating a master program in audio communication and technology at TU Berlin and teaching Tonmeister students at the University of the Arts (UdK) in audio technology and digital signal processing. He is currently coordinating a German research consortium on virtual acoustics at TU Berlin.

•

Steven van de Par studied physics at the Eindhoven University of Technology, Eindhoven, The Netherlands, and received the Ph.D. degree in 1998 from the Eindhoven University of Technology on a topic related to binaural hearing. As a Postdoctoral Researcher at the Eindhoven University of Technology, he studied auditory-visual interaction and was a Guest Researcher at the University of Connecticut Health Center. In early 2000, he joined Philips Research, Eindhoven, to do applied research in auditory and multisensory perception, low-bit-rate audio coding and music information retrieval. Since April 2010 he holds a professor position in acoustics at the University of Oldenburg, Germany with

a research focus on the fundamentals of auditory perception and its application to virtual acoustics, vehicle acoustics, and digital signal processing. He has published various papers on binaural auditory perception, auditory-visual synchrony perception, audio coding, and computational auditory scene analysis.

•

Markus Müller-Trapet graduated from RWTH Aachen University in Germany in 2009 with a Dipl.-Ing. in electrical engineering and information technology. He then became a Research Assistant at the Institute of Technical Acoustics at RWTH Aachen University to pursue his Ph.D. on the topic of surface reflection properties of architectural surfaces. After obtaining the Dr.-Ing. in 2015 he joined the Acoustics group of the ISVR as a Research Fellow, working on microphone arrays and array signal processing. As of April 2017, he is a Research Officer working on Building Acoustics at the National Research Council Canada in Ottawa.

•

Rob Opdam received his bachelors degree in applied physics from the Delft University of Technology, The Netherlands, in 2008 and completed his engineering degree in applied physics also there in 2010. Since 2011 he is a research assistant at the Institute of Technical Acoustics (ITA) at RWTH Aachen University, Germany. He works in the field of numerical acoustics, especially on the simulation and measurement of non-locally reacting materials.

•

Michael Vorländer graduated in physics in 1984, gained a doctor degree in 1989 at RWTH Aachen University, Germany, and a habilitation degree at Technical University Dresden, Germany, in 1995. He is now Professor at RWTH Aachen University, Germany, and the Director of the Institute of Technical Acoustics. His book *Auralization* (Berlin, Germany: Springer 2008) is a reference on the field of Acoustic Virtual Reality. His current research interest is auralization including simulation techniques and signal processing. Prof. Vorländer was president of the European Acoustics Association (EAA), in the term 2004–2006 and president of the International Commission for Acoustics (ICA), in the term 2010–1013. He is a Member of the German Acoustical Society (DEGA), of the German Physical Society (DPG), and a Fellow of the Acoustical Society of America (ASA).