

Cross-validation of measured and modeled head-related transfer functions

Fabian Brinkmann¹, Alexander Lindau¹, Markus Müller-Trapet²,
Michael Vorländer², Stefan Weinzierl¹

¹ *Audio Communication Group TU Berlin, Einsteinufer 17c, D-10587 Berlin, Germany.*

E-Mail: {fabian.brinkmann, alexander.lindau, stefan.weinzierl}@tu-berlin.de

² *Institute of Technical Acoustics, RWTH Aachen University, Kopernikusstr. 5, D-52074 Aachen, Germany.*

E-Mail: {mmt, mvo}@akustik.rwth-aachen.de

Abstract

In the current study, we present full-spherical high resolution head-related transfer function (HRTF) datasets of the head and torso simulator FABIAN including data for multiple head-above-torso orientations within a typical horizontal rotation range of ± 50 degrees. While the first dataset was measured acoustically using sequential swept sines, the second one was modeled numerically using the boundary-element method accelerated with the fast-multipole method (FMM-BEM). Comparison of magnitude spectra revealed a good agreement at mid and high frequencies, but increasing differences at frequencies below 200 Hz which could be attributed to the limited frequency range of the measurement loudspeaker. Additionally, an analysis of times of arrival (TOA) revealed differences which were presumably caused by mechanical inaccuracies of the physical measurement setup. Consequently, by using the modeled data as a reference, in the measured head-related impulse responses (HRIRs), we corrected TOAs by fractionally delaying, and extrapolated the low frequency range.

Introduction

Virtual acoustical environments may be created through reproduction of measured or modeled binaural signals using headphones or cross-talk compensated loudspeaker setups – a method which is called binaural synthesis [1]. It was shown that a plausible acoustical simulation – i.e., “a simulation in agreement with the listener’s expectation towards a corresponding real event” [2, p. 804] – may be achieved already with non-individual dynamic binaural synthesis. In this case, head rotations of the listener are accounted for by real time exchange of binaural transfer functions which were measured using a dummy head. When using individual transfer functions and less critical audio contents such as speech, simulations can be authentic – i.e., indistinguishable from a given reference [3].

The potentially high degree of realism and the technical simplicity of binaural synthesis suggest its usage as a reference when benchmarking other approaches of spatial audio reproduction such as loudspeaker array based methods [4] or room acoustic simulation [5]. This in turn implies that in order to allow for a reliable comparison between measured and modeled binaural simulations, the acquisition of binaural transfer functions (binaural room impulse responses – BRIRs, HRTFs) needs to be thor-

oughly validated as otherwise measurement errors could bias the results.

Hence, we suggest a thorough cross-validation using HRTFs both from acoustical measurements, and numerical simulations. This ‘two-sided’ validation approach is founded in inherent restrictions that are related to both acquisition approaches: In case of measurements, through inaccuracies in microphone, subject, and loudspeaker positioning, as well as non-ideal characteristics of electro-acoustic transducer (e.g. time variance; deviation from omni-directionality [6]) may introduce errors, whereas in case of the simulation trade-offs in the discretization of the subjects’ geometry, and assumptions about the acoustical surface impedance may give room for speculations [7].

Head-related transfer function acquisition

HRTFs for the head and torso simulator FABIAN [8] were acquired for 11950 source position and 11 azimuthal head-above-torso orientations (head rotations to the left and right) covering the typical range of motion of $\pm 50^\circ$ [9]. The resolution of source positions (2° in elevation; 2° great circle distance in azimuth, cf. Fig. 1) was chosen to allow for perceptually transparent HRTF interpolation, and aliasing-free high order spherical harmonic representation. Head-above-torso orientations were measured in distances of 10° to assure that artifacts due to their interpolation remain below the threshold of perception [10]. In the following, we briefly describe the methods used for measuring and modeling HRTFs.

Measurements

HRTFs were measured in the fully anechoic chamber of the Acoustics Groups at Carl von Ossietzky University Oldenburg with a lower cut-off frequency of 50 Hz above which free field conditions may be assumed. The used Two Arc Source Positioning system (TASP) consisted of two semicircular arcs with a radius of 1.7 m which could be rotated about the vertical axis. Each arc was equipped with a Manger MSW bending-wave sound transducer that could automatically be moved to different elevations. Due to mechanical restriction, HRTFs could not be measured for elevations below -64° , resulting in approximately 5 % of missing data (11345 HRTFs). FABIAN’s interaural center was aligned to the center of the measurement system using a cross-line laser and a laser pointer attached to FABIAN neck joint. Measure-

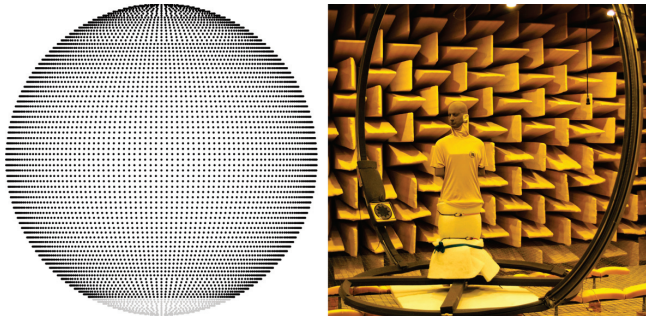


Figure 1: Left: Source positions used for HRTF acquisition. Gray dots denote missing data in measured HRTFs. Right: Setup for measuring HRTFs.

ments were conducted using sequential swept sines at a sampling rate of 44.1 kHz. HRTFs were obtained by spectral division of sweeps recorded at FABIAN's blocked ear canals and sweeps recorded in the center of the measurement system in the absence of FABIAN. The measurement setup is depicted in Fig. 1, more information can be found in [11].

Numerical modeling

The numerical HRTF simulation by means of the BEM required a mesh representation of FABIAN which was obtained as follows: In a first step, a point cloud representation of FABIAN was obtained using a *GOM ATOS I* structured light scanner. A point spacing of approximately $1/100$ mm for the head and pinnae, and $1/10$ mm for the torso was achieved by aligning overlapping scans from multiple viewing directions. FABIAN's neck, and torso bottom plate were excluded from the scan because of their reflecting metallic surfaces. The alignment of point clouds from different scans was done using reference points that were marked on FABIAN beforehand and surface matching as implemented in the *ATOS Professional* software (precision of approx. $1/100$ mm).

In the next step, a non-uniform rational basis spline (NURBS) representation was generated from the point cloud data using the *Geomagic Studio 12* software. Subsequently, the CAD software *Rhino* was used (a) to design a cylindrical neck with seamlessly transitions between head and torso, (b) to close holes in the NURBS representation, (c) to extend the torso bottom to its original size, and (d) to connect the arms to the torso as these were scanned separately before (cf. Fig. 2, left).

In a last step, *Virtual.Lab Acoustics 13.1* was used for meshing the NURBS data and calculation of complex HRTF spectra at frequencies between 100 Hz and 22.2 kHz with a resolution of 100 Hz. To speed up the calculation, two triangular meshes with different resolutions were generated: a coarse mesh with edge lengths of 2 mm, 10 mm, and 10 mm for pinnae, head, and torso, respectively was used for BEM calculations up to 6 kHz, and a fine mesh with edge lengths of 2 mm, 2 mm, and 5 mm was used for the FMM-BEM above 2 kHz. The overlapping region between 2 kHz and 6 kHz was used for verifying that both calculations yielded identical results. The edge lengths were chosen to fulfil the typical

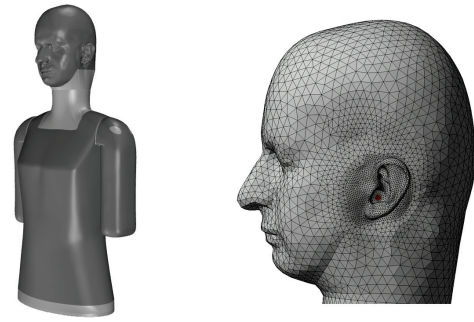


Figure 2: Left: NURBS representation of FABIAN. Light gray surfaces were manually inserted into the model. Right: Coarse mesh of FABIAN's head. Red area denotes microphone position.

requirement of six elements per wavelength in the frequency range under investigation [12]. Simulations were then carried out by imposing a constant velocity boundary condition on the part of the mesh corresponding to the microphone at the ear canal entrance of FABIAN. Otherwise, the mesh was assumed to be sound-hard, i.e., with an admittance of zero. The HRTFs were calculated by dividing the result at the fieldpoints by the analytical solution of a point source with the same volume velocity placed in the center of the coordinate system. Finally, HRIRs with a sampling rate of 44.1 kHz were obtained by inverse Fourier transform after mirroring the single sided spectrum considering the symmetry properties of the discrete Fourier transform [13]. The frequency bin at 0 Hz was set to 1 (0 dB) beforehand. HRIRs for three different models were calculated: (a) A sound-hard head and torso model, (b) A head and torso model with an impedance boundary condition on the torso bottom corresponding to a porous absorber with a thickness of 30 mm, and (c) a head, torso, and legs model with legs modeled by an elliptical cylinder with the surface impedance of the porous absorber as in case (b). (cf. Fig. 2, right).

Cross-validation of measured and modeled head-related transfer functions

Samples of measured and modeled HRIRs/HRTFs for neutral head-above-torso orientation in the median plane showed a good first visual agreement (cf. Fig. 3). A more detailed analysis of spectral and temporal differences and their implications will be given in the following – restricted to HRTFs for neutral head-above-torso orientation due to page limitations.

Temporal cross-validation

Ideally, TOA should be identical in measured and modeled HRIRs. However, differences in the range of 3 samples (equaling a displacement of approx. 2.3 cm) were observed between the two conditions. As the geometrical alignment of the sound sources and FABIAN's interaural center was believed to be almost perfect for the BEM simulation, we used the modeled data as a reference for correcting the TOA of the measured HRIRs. This was done separately for the left and right ear HRIRs us-

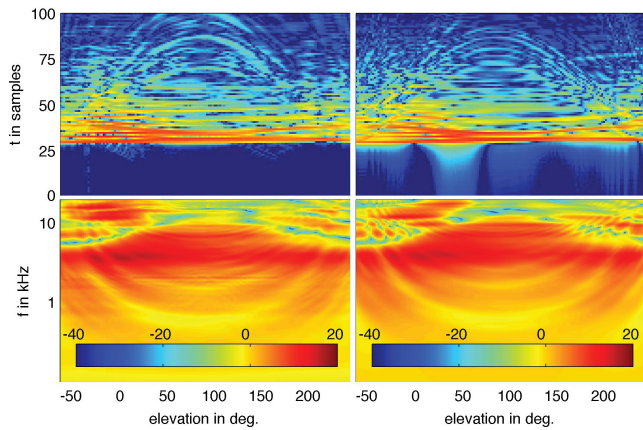


Figure 3: Measured (left) and modeled (right) HRIRs (top) and HRTF magnitude spectra (bottom) in the median plane. Elevation of 0° , and 180° denote sources in the front, and back. Color denotes magnitude in dB.

ing fractional delaying [14]. The amount of delay τ was estimated by maximizing the cross-correlation between pairs of ten times up-sampled measured and modeled HRIRs, hence $\tau = \arg \max_{\tau} \varrho_{xy}(\tau)$. For the fractional delay we used Kaiser windowed sinc filters of order 70, exhibiting negligible magnitude and group delay distortions (< 0.1 dB; < 0.01 samples, $\forall f < 20$ kHz). As a result the average (and minimum) cross-correlations in our data increased from 0.7 (-0.18) to 0.94 (0.56). Fig. 4 shows two HRIRs before and after TOA correction.

An analysis across source positions revealed slight TOA discontinuities in measured data – stemming from start and end points of the 360° rotation of the TASP, and from the two differently mounted measurement loudspeakers – which were also removed by the fractional delays. The TOA treatment caused changes in the broad band interaural time difference (ITD) of about ± 0.5 samples ($\approx 11 \mu\text{s}$) in the proximity of the horizontal plane. Differences of about -1.1 to 2.9 samples ($\approx 66 \mu\text{s}$) occurred for lateral sources where, however, the auditory system is less sensitive to changes in the ITD [15]. Induced ITD changes were thus believed to be below the threshold of perception, and the applied fractional delays were considered to be perceptually non-critical.

Spectral cross-validation

Spectral differences between measured and modeled HRTFs were evaluated in 40 auditory filter bands between 50 Hz and 20 kHz [16]. The results for the three different FABIAN models (cf. Sec. *Numerical modeling*) were almost identical, and thus the non-rigid surfaces models were discarded for simplicity. Results for the rigid surface model are shown in Fig. 5 for the left ear HRIRs averaged across source positions. Results can be discussed with regard to three frequency ranges showing distinct error pattern: Deviations as seen below 200 Hz are caused by the limited frequency response of the loudspeaker. Here, the levels of measured HRIRs are systematically below their modeled counterpart. Between 200 Hz and 6 kHz the median deviation is approx.

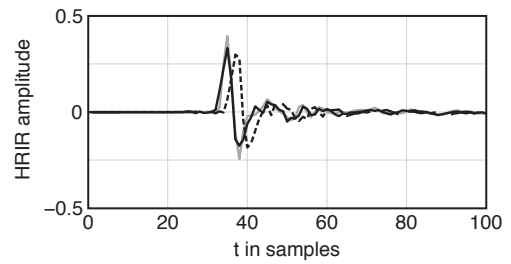


Figure 4: Modeled (gray) and measured left ear HRIRs before (dashed) and after (solid) fractional delaying. For illustration only the HRIR with the lowest cross-correlation before the alignment is shown.

0 dB, while 90% of the errors are smaller than ± 2 dB, and the maximum error remains below ± 8 dB. Above 5-6 kHz notches occur in the HRTF magnitude response, that are caused by pinnae (anti) resonances [17] and are highly sensitive to small changes of sound incidence, or microphone position [6]. Hence, the median difference between measured and modeled HRTFs increases to approx. ± 3 dB. For 90% of the source positions the error remains well below ± 10 dB, whereas the maximum error occasionally exceeds this range. Results for the right ear were of comparable magnitude.

High frequency differences in the HRTFs could be either induced by mechanical inaccuracies of the measurement setup, or by simplifying assumptions underlying the BEM simulation, making them difficult to correct. Observed low frequency differences, however, are clearly caused by non-ideal loudspeaker characteristics and could therefore be corrected by extrapolation: Xie [18] proposed a linear interpolation of magnitude and phase responses, Bernschütz [19] coupled a time and level aligned low pass filter to the measured HRIRs, Algazi et al. [20] fitted analytically modelled transfer functions derived from fitted spherical/elliptical head and torso models, and Gumerov et al. [21] suggested using data obtained from BEM simulations. The latter was applied in our case, too, assuming the modeled data to be valid for low frequencies. To avoid discontinuities, measured and modeled HRTF magnitude and unwrapped phase spectra were combined separately using a linear fade between 200 and 500 Hz.

Discussion and outlook

In the current study, we presented high-resolution HRTF datasets for various head-above-torso orientations which we measured acoustically, and simulated numerically using the BEM. We presented results from temporal and spectral analyses of both types of data. Thereby, we came to the conclusion that both, measured and modeled data suffered from specific shortcomings in so far that none of them alone could provide us with a general ground truth. However, from our analyses we could identify frequency ranges in which each of the data sets appeared to be more reliable than the other one, in turn using the more reliable data for a correction (cross-validation) of the other. Thus, small TOA deviations which we observed in measured data and which could be ascribed to asymmetries

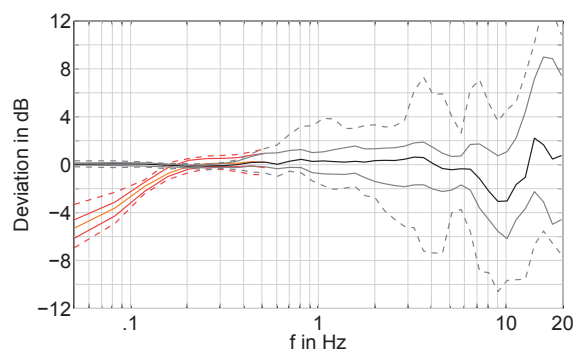


Figure 5: Deviations between measured and modeled HRTFs for 40 auditory filter band averaged across source positions before (colored) and after low frequency extrapolation (gray scale; Median errors given by black and orange solid lines; 10%-90% percentile range by solid gray/red lines; 0%-100% percentile range by dashed gray/red lines).

of the mechanical measurement setup were corrected by aligning them to the modeled data by means of fractional delay correction. Additionally, we investigated the differences between measured and modeled HRTF magnitude spectra in 40 auditory filter bands. Below 200 Hz, measured HRTFs were found to be invalid due to the non-ideal loudspeaker frequency response, hence, missing low-frequency information was recreated using the modeled data. Above 200 Hz, magnitude spectra were found to be reasonably comparable and differences were of an order as reported, e.g., by Gumerov et al. [21]. Above 6 kHz, however, spectral differences increased presumably caused by slightly mismatched pinna notches. As in this case, causes for discrepancies could not be satisfactorily identified, we refrained from correcting either data.

For the future we plan to extrapolate the missing measured HRTFs at low elevations using an approach as suggested by Ahrens et al. [22]. Furthermore, we aim at psycho-acoustically validating the localization performance obtainable with our HRTF sets using a model suggested by Baumgartner et al. [23].

Acknowledgement

The work is part of the Simulation and Evaluation of Acoustical Environments (SEACEN) project funded by the German Research Foundation (DFG WE 4057/3-2).

References

- [1] Møller, H.: Fundamentals of binaural technology. *Applied Acoustics* 36 (1992), 171-218.
- [2] Lindau, A., Weinzierl, S.: Assessing the plausibility of virtual acoustic environments. *Acta Acustica united with Acustica*, 98 (2012), 804-810.
- [3] Brinkmann, F., Lindau, A., Vrhovnik, M., Weinzierl, S.: Assessing the authenticity of individual dynamic binaural synthesis. *Proc. of the EAA Joint Symposium on Auralization and Ambionics* (2014), 62-68.
- [4] Spors, S., Wierstorf, H., Raake, A., Melchior, F., Frank, M., Zotter, F.: Spatial sound with loudspeakers and its perception: A review of the current state. *Proc. of the IEEE* 101 (2013), 1920-1938.
- [5] Vorländer, M.: *Auralization. Fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality*. Springer, Berlin, Heidelberg, Germany, (2008).
- [6] Andreopoulou, A., Begault, D.R., Katz, B.F.G.: Inter-laboratory round robin HRTF measurement comparison. *IEEE J. Sel. Topics Signal Process.* (2015), in print.
- [7] Kahana, Y.: *Numerical modelling of the head-related transfer function*. Ph.D. Thesis, University of Southampton, United Kingdom, (2000).
- [8] Lindau, A., Hohn, T., Weinzierl, S.: Binaural resynthesis for comparative studies of acoustical environments. *122th AES Convention* (2007), Vienna, Austria.
- [9] Thurlow, W.R., Mangels, J.W., Runge, P.S.: Head movements during sound localization. *J. Acoust. Soc. Am.* 42 (1967), 489-493.
- [10] Brinkmann, F., Roden, R., Lindau, A., Weinzierl, S.: Audibility and interpolation of head-above-torso orientation in binaural technology. *IEEE J. Sel. Topics Signal Process.* (2015), in print.
- [11] Brinkmann, F., Lindau, L., Weinzierl, S., Geissler, G., van de Par, S.: A high resolution head-related transfer function database including different orientations of head above the torso. *AIA-DAGA 2013, Int. Conf. Acoust.*, 596-599.
- [12] Ciskowski, R. and Brebbia, C.: Boundary Element Methods in Acoustics, *Elsevier Applied Science*, (1991).
- [13] Oppenheim, A.V., Schaffer, R.W., Buck, J.R.: *Discrete-time signal processing*. Prentice Hall, Upper Saddle, USA, 2nd ed. (1999).
- [14] Laakso, T.I., Välimäki, V., Karjalainen, M., Laine, U.K.: Splitting the unit delay. *IEEE Sig. Proc. Mag.*, 13 (1996), 30-60.
- [15] Blauert, J.: *Spatial hearing. The psychophysics of human sound localization*. MIT Press, Revised ed. (1997).
- [16] Slaney, M.: Auditory toolbox. Version 2. *Tech. Rep. #1998-010, Interval Research Corporation* (1998).
- [17] Takemoto, H., Mokhtari, P., Kato, H., Nishimura, R., Iida, K.: Mechanism for generating peaks and notches of head-related transfer functions in the median plane. *J. Acoust. Soc. Am.*, 132 (2012), 3832-3841.
- [18] Xie, B.: On the low frequency characteristics of head-related transfer function. *Chinese J. Acoust.*, 28 (2009), 1-13.
- [19] Bernschütz, B.: A spherical far field HRIR/HRTF compilation of the Neumann KU 100. *AIA-DAGA 2013, International Conference on Acoustics*, Merano, Italy, 592-595.
- [20] Algazi, V.R., Duda, R.O., Duraiswami, R., Gumerov, N.A., Tang, Z.: Approximating the head-related transfer function using simple geometric models of the head and torso. *J. Acoust. Soc. Am.*, 112 (2002), 2053-2064.
- [21] Gumerov, N.A., O'Donovan, A.E., Duraiswami, R., Zotkin D.N.: Computation of the head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation. *J. Acoust. Soc. Am.*, 127 (2010), 370-386.
- [22] Ahrens, J., Thomas, M.R.P., Tashev, I.J.: HRTF magnitude modeling using a non-regularized least-squares fit of spherical harmonics coefficients on incomplete data. *AP-SIPA Annual Summit and Conference*, (2012).
- [23] Baumgartner, R., Majdak, P., Laback, B.: Modeling sound-source localization in sagittal planes for human listeners. *J. Acoust. Soc. Am.*, 136 (214), 791-802.