Fabian Brinkmann, Alexander Lindau, Stefan Weinzierl, Gunnar Geissler, Steven van de Par

A high resolution head-related transfer function database including different orientations of head above the torso

Conference paper | Published version
This version is available at https://doi.org/10.14279/depositonce-8768

Brinkmann, Fabian; Lindau, Alexander; Weinzierl, Stefan; Geissler, Gunnar; van de Par, Steven (2013): A high resolution head-related transfer function database including different orientations of head above the torso. In: AIA-DAGA 2013 : proceedings of the International Conference on Acoustics ; 18 - 21 March 2013 in Merano. Berlin: Deutsche Gesellschaft für Akustik e.V. pp. 596–599.

A high resolution head-related transfer function database including different orientations of head above the torso

Fabian Brinkmann[1], Alexander Lindau[1], Stefan Weinzierl[1], Gunnar Geissler[2], Steven van de Par[2]
[1] *Audio Communication Group, TU Berlin, Einsteinufer 17c, D-10587 Berlin, Germany*
Email: {fabian.brinkmann, alexander.lindau, stefan.weinzierl}@tu-berlin.de
[2] *Acoustics Group, University Oldenburg, Carl von Ossietzkystr. 9-11, D-26129 Oldenburg, Germany*
Email: {gunnar.geissler, steven.van.de.par}@uni-oldenburg.de

Introduction

An extensive database of head-related transfer functions (HRTFs) was measured for the FABIAN head and torso simulator (HATS) which is suitable for the fast acquisition of complete datasets of binaural impulse responses in natural acoustic environments [1]. In total, 124,795 HRTFs were measured with high spatial resolution, covering source elevations from -64° to 90° and including 11 different azimuthal orientations of head above torso. Special care has been taken to ensure the accuracy of the measurements, and verification procedures are described in detail.

The HRTF dataset was generated in the context of a joint research project dedicated to the analysis and synthesis of both real and simulated acoustical environments [2]. Here it will facilitate the perceptual evaluation of room acoustical modeling algorithms by comparing measured and modeled binaural datasets of corresponding acoustical environments with consistent HRTFs. Moreover, it will be possible to study the effect of the torso on localization and spatial hearing within dynamic binaural environments, which so far has been assesed numerically [3] and perceptually [4] for the static case only.

Measurements

By definition, the HRTF is given by the ratio of the complex spectra observed at the blocked ear canal entrances ('binaural response') and the spectra observed at the center of head with the HATS absent ('reference response') [5]. Hence, we measured binaural responses for each of the 11 head orientations above the torso using an ever-identical spherical measurement grid. Additionally, and once for all directions of the spherical grid, reference responses were measured as well. HRTFs could then be derived by spectral division of binaural and respective reference responses.

The measurement grid included elevations from -64° to 90° in two degree steps. The azimuthal resolution was chosen not to exceed two degrees great circle distance between neighboring points of the same elevation and to include the horizontal, median and frontal plane (Figure 1). Hence, except for the limited range of elevations the grid is identical to the one used by Bovbjerg et al. [6]. The grid was measured repeatedly for 11 different horizontal orientations of the dummy head above its torso, ranging from -50° to +50° in 10° steps, thus covering a typical range of horizontal head movements while listening [7].

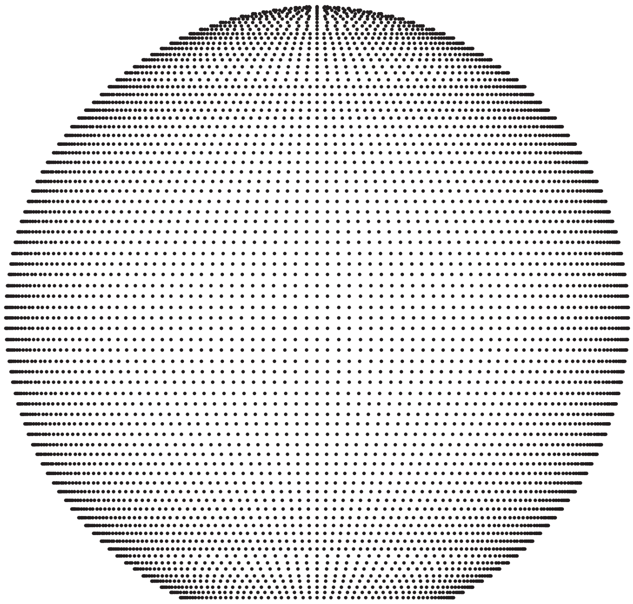

Figure 1: Side view of the HRTF grid.

In total, 124,795 (11 x 11,345) HRTFs were measured in the anechoic chamber of Carl von Ossietzky University Oldenburg ($V = 238\,\mathrm{m}^3$, cutoff frequency 50 Hz). As measurement stimulus, we used sine sweeps of FFT order 16, which were band-limited between 100 Hz and 21 kHz with a spectral coloration according to the background noise to increase signal-to-noise ratio (SNR) [8]. The Oldenburg two-arc-source-positioning system (TASP) [9] consisting of two opposing semicircular arcs with a radius of 1.7 m, was used for source positioning. The two arcs could be rotated above the vertical axis and were each equipped with a Manger MSW bending-wave sound transducer on vertically movable mounts (Figure 2). Matlab© routines were used for audio playback and recording at a sampling rate of 44.1 kHz and for controlling the movement of arcs and speakers with a precision of 0.1 degree.

The FABIAN HATS, equipped with DPA4060 microphones blocking the entrances to the ear channels, was set up with its interaural center aligned to the geometrical center of the TASP. The vertical axis of the TASP was

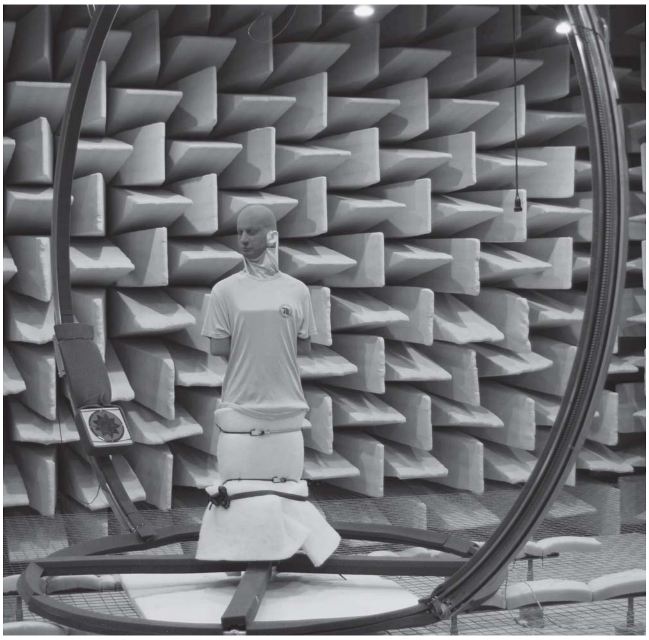

Figure 2: FABIAN HATS centered inside the TASP.

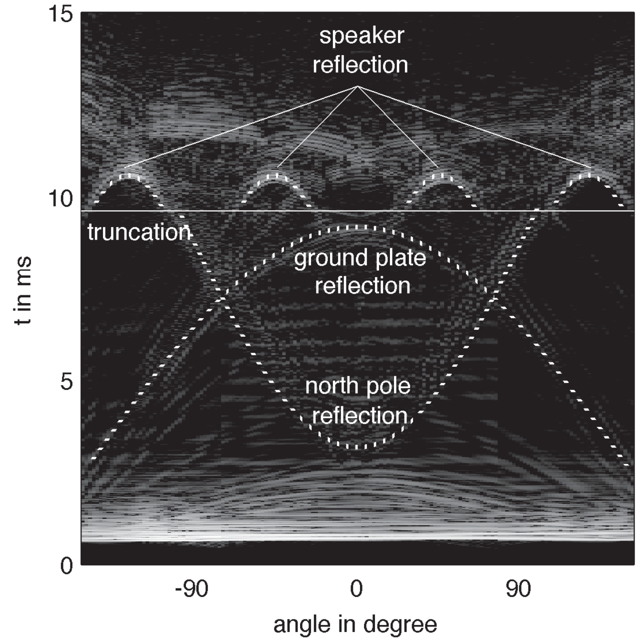

Figure 3: HRIRs for neutral head orientation in the median plane. Positive angles denote source position in front of FABIAN; zero degree denotes north pole; gray scale is level in dB.

defined by dropping a perpendicular from the 'North Pole' onto its bottom plate where FABIAN was to be installed on. FABIAN's interaural center had been determined geometrically and projected as a small drill hole onto the foot plate of the HATS' height adjustable stand. Then, the marks of the projected 'North Pole' and the drill hole marking the interaural center were aligned, and FABIAN was fine-tuned to be horizontally leveled. Next, a self-leveling Bosch PCL10 cross-line laser, aligned to the equator of the TASP, was used to adjust FABIAN's height until the laser line coincided with the ear canal entrances. Since FABIAN's head and ears are casts of a human subject they are of slightly asymmetrical shape. Thus, the frontal viewing direction could not be aligned by establishing an interaural time difference (ITD) of zero for sources in the median plane as proposed for instance by [6]. Instead, a laser was attached to the neck joint to point at a desired target point parallel to FABIAN's viewing direction. After the positioning procedure, ITDs between -30 and 20 μs were observed in the median plane (14 μs for frontal sound incidence) which have to be attributed both to the mentioned asymmetries of FABIAN and to mechanical shortcomings of the TASP.

Before starting the measurements, all parts of the measurement equipment inside the anechoic chamber were wrapped in absorbing material thus considerably reducing the level of reflections from the arcs and mounts of the TASP (Figure 3). Then, the loudspeakers were subjected to a durability test consisting of 20,000 measurements. We observed a temporal variability of the magnitude response of $\pm$4 dB, which was addressed in the measurement procedure by applying 5,000 pre-sends each time before measuring an HRTF set. To obtain an estimate for the temporal variability that remained after the pre-sends, two reference measurement series covering the 11,345 source positions were conducted. Spectral division of corresponding measurements showed that the loudspeakers' temporal variability could be reduced to $\pm$1 dB between 200 Hz and 20 kHz (Figure 4), when measurement series were conducted directly one after another.

To obtain an estimate of the mechanical reliability of the setup, before measuring each set of HRTFs four ever identical HRTFs were measured (neutral head orientation, sources to the front, back, left and right in the horizontal plane). Next, FABIAN's head was turned to the desired orientation above torso and 5,000 pre-sends were sent through both speakers to reduce their temporal variability. Finally, a complete grid of HRTFs was measured with four averages per measurement position. From later analysis of the reliability measurements it was found that deviations between individual HRTF measurement runs were generally in the range of $\pm$1.5 dB, however, for the contralateral ear and in the vicinity of notches deviations of about $\pm$10 dB could be observed. Hence, all reported deviations are within the same order as those reported for similar measurements [6]. With four averages, an SNR of 80 dB for ipsilateral and 55 dB for contralateral sources was achieved.

For the reference measurement both FABIAN microphones were mounted simultaneously on an acoustically minimally invasive stand at the geometrical center of the TASP while pointing to the left and right, respectively. The microphones were geometrically adjusted using the same procedure as described above.

In order to enable an ex post correction of the temperature influence on the velocity of sound, temperature values were recorded using a Hygrosens CON-TSIC-LABKIT-USB and a TSIC 306 TO92 sensor throughout all measurement runs.

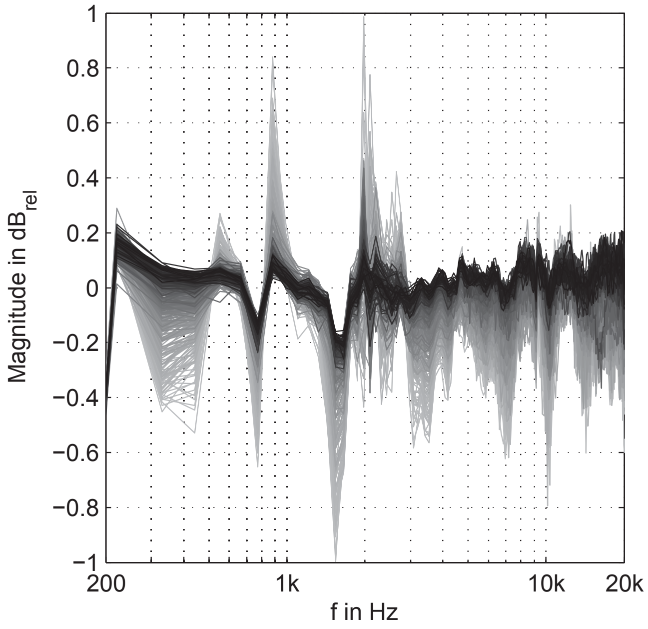

Figure 4: Remaining temporal variability after 5000 presends, obtained from spectral division of two successive reference measurement sessions. Brighter lines denote earlier measurements.

Post-processing

HRTFs were obtained by spectral division (deconvolution) of the recorded HRTF and the reference measurements, thus cancelling out the frequency response of microphones and speakers, and the transfer functions of involved amplifiers, as well as A/D and D/A converters. 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 deconvolved signal.

The dual microphones setup chosen for the reference measurement slightly affected the microphones' nearly ideally omnidirectional directivity. Balloon measurements revealed a slight figure of eight shape with attenuations starting at around 5 kHz and increasing to 5 dB at 20 kHz. From these measurements corresponding 256 tap minimum phase compensation filters were designed in order to restore a perfect omnidirectional directivity.

During the measurements which in total lasted more than two months, temperatures between 9.9 and 13 degree Celsius were observed resulting in differences of acoustic times of flight of up to 1.2 samples (27 μs). The influence of the temperature fluctuations on the arrival time in HRTFs was corrected by applying a fractional delay depending on the temperature observed during reference and binaural measurement.

$$\Delta_{corr} = f_s \left(\frac{r}{c_{ref}} - \frac{r}{c_{binaural}} \right) \tag{1}$$

with c_{ref} and $c_{binaural}$ being the respective sound velocities, r the radius of the TASP as estimated from times of arrival in the latency compensated reference measurements and f_s the sampling rate. Fractional delays were realized using sinc filters of order 70 weighted by a Kaiser window with a side lobe attenuation of 80 dB [10].

In the next step, head-related impulse responses (HRIRs) were truncated to 425 samples (approx. 10 ms) to discard prominent acoustic reflections mainly caused by the measurement speakers (Figure 3). Faint reflections which resulted from the TASP's north pole and ground plate though being wrapped in absorbing material, could not be eliminated as they occurred very early in the HRIRs.

Windowing and the limited low frequency response of the Manger sound transducer led to a loss of information below 200 Hz. Algazi et al. [11] and Gumerov et al. [12] suggested restoring low frequency data either by using a simple analytical head and torso model or by using data obtained from boundary element method (BEM) simulations. A method based on measured data was suggested by Xie [13], who extrapolated the low frequency response by constant magnitude and group delay values obtained from valid regions in the HRTFs below 400 Hz. All three studies showed that the magnitude responses of HRTFs are nearly constant up to a few hundred Hz, where the dimensions of head and torso reach the order of a wavelength resulting in effects such as shadowing, diffraction and reflection. Thus, the first two frequency bins of each magnitude response (0 and 103 Hz) were set to 0 dB, while the phase information was left unchanged.

Finally, a 15 sample squared-sine fade-in/out was applied to the HRIRs to avoid discontinuities at beginning and end. The final database includes original phase HRIRs as well as minimum phase HRIRs obtained from the cepstrum method. In conjunction with numerical ITD values, these allow for ex post ITD individualization [14].

Remaining sources of error

Although the frequency responses the loudspeakers were removed from the HRTFs, any deviation from omnidirectionality is a possible source of error. The Manger sound transducers exhibit a membrane diameter of 8" and balloon measurements revealed distinct main and side lobes emerging above 3 kHz. However, analysis showed that up to 7 kHz, shoulders and torso of FABIAN were within the speakers' main lobe, as defined by its -3 dB points. As above 3-4 kHz HRTFs are mainly influenced by head and pinnae [4, 15] which were within the main lobe up to 20 kHz, the speakers' directivity should be of negligible influence.

An estimation of the TASP radius based on onset detection in the reference measurements showed that the TASP deviates from a perfect spherical shape, with radii ranging between 1.62 and 1.72 m. The reference measurement, however, compensates for varying radii, and proximity effects in HRTFs were shown to become negligible for measurement distances above 1 - 1.5 m [16, 17].

Outlook

The new FABIAN HRTF dataset will be applied to compare virtual acoustic environments based on measured binaural room impulse responses (BRIRs) directly to model-based auralizations. With identical HRTFs under both conditions, an important source of variance will be eliminated. The more so after the missing data below -64° elevation level will have been filled using numerically modeled HRTFs, based on a detailed 3D model of the FABIAN HATS. In future investigations it will be interesting to evaluate the perceptual relevance of different head orientations above torso, and in how far these can be interpolated without perceptual degradation.

The complete HRTF database comprises a comprehensive documentation of the measurement procedure and all applied post-processing steps, as well as figures of the resulting HRTF spectra, time signals, and estimated interaural time and level differences. Thus, the creation of the final data set can be fully reproduced starting from the raw measurement data and using our original Matlab© scripts.

Acknowledgement

The work is part of the *Simulation and Evaluation of Acoustical Environments* (SEACEN) project funded by the German Research Foundation (DFG WE 4057/3-1). The authors like to thank Christoph Scheicht for assistance in setting up the measurement system and Pascal Dietrich and Benedikt Krechel for providing balloon measurements of speakers and microphone setup used during the HRTF acquisition.

References

[1] Lindau, A.; Weinzierl, S.: FABIAN - Schnelle Erfassung binauraler Raumimpulsantworten in mehreren Freiheitsgraden. Fortschritte der Akustik: Tagungsband d. 33. DAGA. Stuttgart (2007)

[2] Simulation and Evaluation of Acoustical Environments (SEACEN, 2011), URL: `http://www.seacen.tu-berlin.de`

[3] Guldenschuh, M.; Sontacchi, A.; Zotter, F.: HRTF modelling in due consideration variable torso reflections. Acoustics, Paris (2008)

[4] Algazi, V.R.; Avenado, C.; Duda, R. O.: Elevation localization and head-related transfer function analysis at low frequencies. J. Acoust. Soc. Am. **190**(3), 1110-1122 (2001)

[5] Møller, H.: Fundamentals of binaural technology. Appl. Acoust. **36**, 171-218 (1992)

[6] Bovbjerg, P.B.; Christensen, F.; Minnaar, P.; Chen, X.: Measuring the head-related transfer functions of an artificial head with a high directional resolution. 109^{th} Audio Eng. Soc. Convention (2000)

[7] Thurlow, W.R.; Mangels, J.W.; Runge, P.S.: Head movements during sound localization. J. Acoust. Soc. Am. **42**(2), 489-493 (1967)

[8] Müller, S.; Massarani, P.: Transfer-function measurement with sweeps. J. Audio Eng. Soc. **49**(6), 443-471(2001)

[9] Otten, J.: Factors influencing acoustical localization. Doct. Thesis, Physiks Group, University of Oldenburg, URL: `http://oops.uni-oldenburg.de/volltexte/2001/365` (2001)

[10] Laakso, T.I.; Välimäki, V.; Karjalainen, M.; Laine, U. K.: Splitting the unit delay. IEEE Signal Proc. Mag. **13**(1), 30-60 (1996)

[11] 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**(5), 2053-2064 (2002)

[12] 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**(1), 370-386 (2010)

[13] Xie, B.: On the low frequency characteristics of head-related transfer function. Chinese J. Acoust. **28**(2), 1-13 (2009).

[14] Lindau, A.; Estrella, J.; Weinzierl, S.: Individualization of dynamic binaural synthesis by real time manipulation of the ITD. 128^{th} Audio Eng. Soc. Convention (2010).

[15] Algazi, V.R.; Morrison, P.R.; Thompson, D.M.: Structural composition and decomposition of HRTFs. IEEE WASPAA, 103-106 (2001)

[16] Brungart, D.S.; Rabinowitz, W.M.: Auditory localization of nearby sources. Head-related transfer functions. J. Acoust. Soc. Am. **106**(3), 1465-1479 (1999)

[17] Wierstorf, H.; Geier, M.; Raake, A.; Spors, S.: A free database of head-related impulse resonse measurements in the horizontal plane with multiple distances. 128^{th} Audio Eng. Soc. Convention, Engineering Brief (2011)