www.akustik.rwth-aachen.de
DOI: 10.18154/RWTH-2020-06761

Technical Report

A High-Resolution Individual 3D Adult Head and Torso Model for HRTF Simulation and Validation: HRTF Measurement

Hark Braren, Janina Fels

Teaching and Research Area of Medical Acoustics,
Institute of Technical Acoustics, RWTH Aachen University, Germany
Corresponding author: janina.fels@akustik.rwth-aachen.de

Abstract: Head-related transfer functions (HRTFs) are highly individual, however they can be calculated based on 3D data of the head and torso. In an accompanying report, a high resolution 3D model of an adult male is provided. This report contains a detailed description of the same participant's continuous HRTF measurement, including the method for addressing motion artifacts. This can be used for validation of calculations based on the 3D model. The provided HRTF data sets contain equiangular elevation samplings at 2.5° from approximately 0° to 160° zenith angle. Different azimuth samplings are provided. The datasets are made available in *.ita* (used with the ITA-Toolbox) and *.sofa* (Spatially Oriented Format for Acoustics) file formats.

Keywords: Binaural Technologies, Anthropometric Model, HRTF Measurement

1 Introduction

Simulation of HRTFs using the boundary element method (BEM) is a well established practice for calculating the direction-dependent response of the pinna, head and torso geometry to incoming sound events. However, simulations need measurements for validation. For this purpose, a high resolution 3D model of an adult male was recorded and published [1] along with an HRTF measurement of the same individual available under doi.org/10.18154/RWTH-2020-06761. This report details the continuous HRTF measurement procedure[2] used to capture the participants individual HRTF. To reduce artifacts caused by involuntary motion of the participant during the measurement, a tracking and feedback system described in Section 2.3, was implemented. The required post processing steps are detailed in Section 2.5 and the results are discussed in Section 2.6.

2 HRTF Measurement

2.1 Setup

To be able to validate HRTF simulations of the provided 3D model an HRTF measurement of the same subject was needed. The measurement took place in the hemi-anechoic chamber at the Institute of Technical Acoustics, RWTH Aachen University, using the fast measurement system evaluated by Richter et al. [3]. The room measures 12.6 m × 7.6 m × 4.5 m with 80 cm mineral wool absorber placed on all four walls and the ceiling. It provides free-field conditions down to a lower cutoff frequency of 100 Hz, below which room modes start to show. The measurement system consisted of an arc containing 64 loudspeaker at 2.5° increments, that can rotate a full 360° around a participant standing in its center.

Figure 1: HRTF measurement arc and feedback monitor with height adjustment and head rest. The headrest was removed for the final measurement.

Figure 1 shows the arc including the height adjustment plates and a head rest, that helped to stabilize the subject during measurements. However, it could be removed once the feedback system detailed in Section 2.3 was implemented, to further reduce measurement artifacts from reflections on the foreign geometry. The height adjustment was needed to ensure that the subject's inter-aural axis intersected the center of the sphere created by the rotating measurement arc. Correct positioning was ensured using a laser alignment system.

Figure 2: Sennheiser KE3 microphone in a silicone dome.

To be able to measure the transfer-function from each loudspeaker to the ear-canal entrance, the participant hat to Sennheiser KE3 microphones in his ears. They were mounted inside silicone domes, as shown in Figure 2, blocking the ear canals.
A detailed list of the measurement hardware and settings is presented in Table 1.

Table 1: Measurement Hardware and Settings

Device	Description
Audio Interfaces	
Out	RME HDSP MADI
In	RME Multiface II
DAC & Amplifier	2x innosonix MA32-D
Loudspeakers	Tang Band W1-2025SA
Microphones	Sennheiser KE3 in silicone domes
Pre-amplifier & ADC	RME Octamic
Sampling Rate	48 kHz
Frequency Range	500 Hz–20 kHz
N_{Samples}	362

2.2 Measurement

The measurement utilized the continuous rotation approach presented by Richter and Fels [2], as it greatly reduces measurement times at the cost of increased post-processing complexity. Therein, the arc, mounted on a silent stepper motor straight above the center, rotated continuously around the participant at a speed of $1.38\,^{\circ}\,\mathrm{s}^{-1}$. Interleaved swept sines [4, 5] were used as the measurement signals presented through the 64 loudspeaker in the arc.
Figure 3 shows the raw measurement data for the full duration of the measurement. The rotation of the loudspeakers around the subject can be observed in the varying levels at the left and right ear. In addition to the in-ear microphone signals (blue and red), the signal of a switch located in a fixed location on the fulcrum of the rotating arc is recorded as a positional reference. This allows easy reconstruction of the arcs location at all times during the measurement.
During the measurement, the average relative humidity was 45.85% at an average temperature of 24.1 °C.

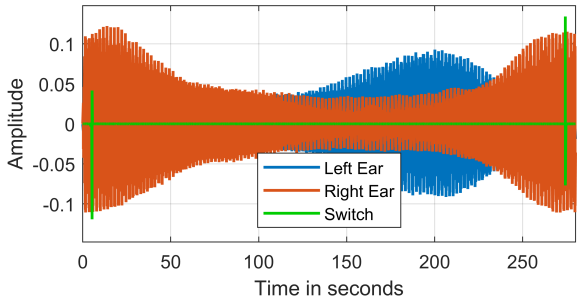

Figure 3: Raw measurement signals of the two in-ear microphones and the locating-switch signal.

2.3 Optical Feedback System to reduce movement

To further reduce the movement during the measurement, a feedback system, as presented by Richter [6], was used to show the participant his own movements and how to compensate for it during the measurement. The system, shown in Figure 4 used an optical, infrared (IR) tracking system by OptiTrack (NaturalPoint Optitrack, Corvallis, OR, USA) [7] to capture both translate and rotary motion. The movement was presented to the participant in real time using a crosshair-style indication system, that shows the zero position crosshair and the current position crosshair overlayed on top of one another along with arrows indicating how he should move to compensate for his movement. In addition smilies indicate the amount of compensatory motion needed for all 6 degrees of freedom. The system used a tracking body placed on the participants head using a head band as the reference. Prior to the measurement it needed to be calibrated, so that the movement is calculated relative to the center of his inter-aural axis, which was easy to understand and follow.

Optical Feedback Calibration

The OptiTrack-system assumes that the tracking body is rigidly attached to the head. To calibrate the motion relative to the inter-aural axis, two tracked pointers

Figure 4: The feedback system consisting of a head tracker on the subject's head and a monitor outside the critical radius, giving feedback on the head movement.

are positioned the participant's left and right tragus. With their positions relative to the tracking body on the head known, their center is used as the new center of motion for the tracked IR-body, making the compensation more intuitive. Preliminary observations during the development of the feedback system showed, that whilst it is possible for participants to keep up the concentration needed for the active motion compensation for about 10 to 15 minutes, longer measurements are not recommended. Thus, the fast-continuous measurement method as presented by Richter [2, 6] offers a good compromise of high angular resolution and short measurement times.

2.4 Reference Measurement

To calculate the HRTF, the impulse response measured at the ear canal entrances is referenced to an impulse response, measured at the center of the interaural axis with the participant absent from the measurement. This measurement is used to eliminate the transfer function of the microphone (H_{Mic}) and that of the loudspeaker (H_{LS}), which is delayed by the time sound needs to travel from the source to the receiver $t_{\mathrm{delay}} = r/c$ located at a distance r, with c denoting the speed of sound, from the HRTF. The reference measurement can thus expressed as the convolution, denoted by a $*$ symbol, yielding:

$$H_{Ref} = H_{\mathrm{LS}} * \delta(t - t_{\mathrm{delay}}) * H_{\mathrm{Mic}}.$$

Because of the small size of the used Sennheiser KE3 microphones, the free-field and pressure-field response don't differ in the frequency range of interest, where larger diameter microphones would show pressure buildup effects due to diffraction as it would be the case for typical 1/2" measurement microphones. The transfer function of each loudspeaker is measured individually (to compensate for variations of the transducers and enclosures) at the starting position of the arc and is assumed to be constant over the whole rotation. For the measurement, the KE3 microphones are placed side by side in a microphone holder with their membranes facing upwards as shown in Figure 5.

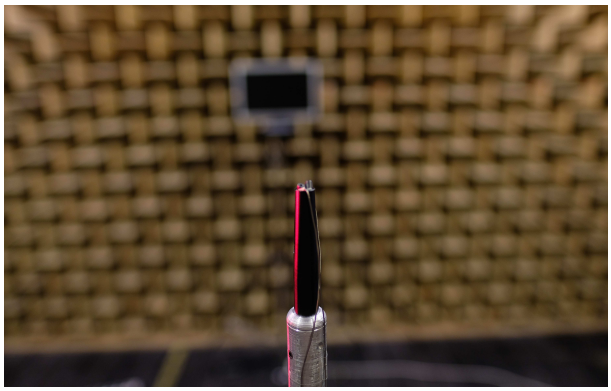

Figure 5: For the reference measurement, the microphones were placed in the center of sphere circumscribed by the HRTF arc. For alignment, a laser system was used.

2.5 Post-processing

While the continuous measurement offers the advantage of a reduced measurement time, and thus less strain on the subject, who has to concentrate on compensating his own movements throughout the measurement, it comes at the cost of an increased complexity in post processing, which is described in the following. In total 129 sweeps were played back through each loudspeaker in the arc in the time it took the system to rotate the full 360°. This equates to an angular resolution of 360°/129 repetitions = 2.78° in azimuth.

Cropping and Referencing

With the measurement taking place in a hemi-anechoic chamber, the reflection from the non-absorbing floor needed to be taken care of by time windowing as seen in Figure 6.

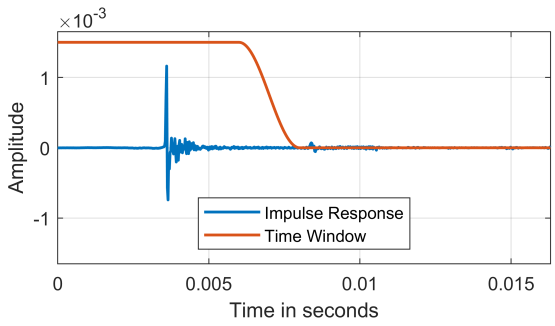

Figure 6: Reference measurement and scaled time window used remove off unwanted reflections from the floor or other equipment in the room.

A one-sided Hann-window with 2 ms fade-out between 6 ms and 8 ms was applied to the time domain response to remove any unwanted reflections, including those from the monitor of the feedback system. The impulse response was then cropped after the fade out leaving 362 samples. The reference was measured and compensated for each speaker individually using division in the frequency domain. This lead to anti-causal impulse responses for direction where the microphone in the ear was located closer to the loudspeaker than the reference position in the center of the head. To avoid acausalities, the compensated impulse responses were circularly shifted in time by 3.5 ms.

Reconstruction of Physical Loudspeaker Locations

Because the arc moved throughout the measurement, the location of each loudspeaker had to be calculated for each measured impulse response. As the loudspeaker was not stationary for the duration of a measurement sweep, each frequency bin had to be regarded separately. To compute the arc's location, the switch installed in the mounting of the arc, was used as a angular reference. Angles in between these two known locations 360° apart can be linearly interpolated, as all acceleration took place before running into the switch the first time, and deceleration starts only after going through the switch the second time, therefore a constant angular velocity was assumed. The switch signal was recorded using the same equipment as for the microphone signals. Thus a sample-accurate alignment with the arc location was possible. Each loudspeakers location at the start of each measurement sweep could be calculated based on the time the sweep started.
To calculate the location for each frequency bin, the time-offset was calculated based on the parameters of the exponential sweep, namely the sweep rate R_{sweep} and the starting frequency of the sweep f_0 as

$$t(f) = \log(f/f_0) \cdot \log_2(e)/R_{sweep}.$$

The angular offset was then calculated based on the angular velocity ω_{ang} and the time offset as

$$\phi_{offset}(f) = t(f) \cdot \omega_{ang}.$$

ITD Alignment

To better align the HRTF data with the orientation on the subject, the zero crossing of the inter-aural time difference (ITD), evaluated via cross-correlation between 200 Hz and 2 kHz [8], was used to calculate an overall azimuth offset. It is assumed, that the ITD is zero right in front and behind the subject, due to the symmetric locations of the ears relative to the median plane. This zero crossing is evaluated in the ITD data in the horizontal plane and all azimuth angles are compensated so that the zero crossing in front of the subject is defined as 0° azimuth. This results in a coordinate system, where the y-axis lies on the inter-aural axis pointing out through the left ear. The x-axis points in the direction the subject is looking and the z-axis points straight up. The origin lies in the center of the inter-aural axis.

Low-Frequency Extrapolation

Content below the arc's lower frequency limit of 500 Hz needed to be extrapolated based on two factors. First, the head and torso become small with regard to the wavelength and their effect becomes negligible at very low frequencies. This yields an HRTF magnitude of unity. Secondly, the HRTF in this frequency range are nearly constant and the phase is a linear function of frequency[9]. Thus a linear interpolation between the known HRTF magnitude and phase above 500 Hz towards unity at 0 Hz was applied.

Per-Frequency Interpolation

As a last step, the movement of the loudspeaker for the duration of each individual sweep signal was taken into account. From the known position at the start of the sweep, the sweep duration and the momentary frequency during the sweep, the azimuth angle offset was calculated at each frequency bin, resulting in a frequency-dependent spiraling sampling grid [6]. In most use-cases however, the broadband HRTF at a single known location is needed. The HRTF was thus spatially interpolated along the azimuth direction for each frequency bin to an equiangular grid. An equiangular spacing of the same ∼2.5° elevation angle sampling as before down to ∼160° zenith angle and 360°/129 repetitions = 2.78° in azimuth, resulting in 8256 direction, is provided. For the interpolation a linear interpolation scheme was used, as more advanced interpolation approaches, such as spherical harmonics based methods, were found to introduce larger errors, especially at higher frequencies.

2.6 Results

In total, three versions of the measured HRTF are provided in the data set under the following names:

1. HRTF_raw
   The original measurement where the coordinates are set at the start of each sweep.

2. HRTF_perFrequencyInterpolation
   A per frequency-domain interpolated data set using the interpolation scheme described above.

3. HRTF_5DegInterpolation
   A version interpolated to a more commonly used 5° azimuth resolution, which introduces more

errors, but may be easier to reference.

Figure 7 shows a time and frequency domain comparison of the uncorrected measurement and the per-frequency interpolated version. It becomes clear, that the interpolation affects predominantly the high frequencies, as they show a greater spatial variance and occur later in the sweep, thus the angular error is greater. Also, the selected interpolation does not significantly alter the HRTF magnitude, neither in time nor frequency domain and seems to be appropriate for the given data set.

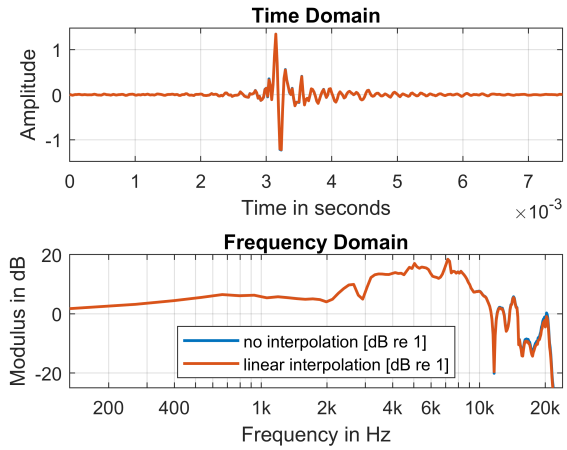

Figure 7: Comparison of the direct measurement and the per-frequency interpolated version of the left-ear HRTF in the horizontal plane at 70° azimuth angle.

All further plots are based on the second version of the dataset, as it is assumed to be best suited.

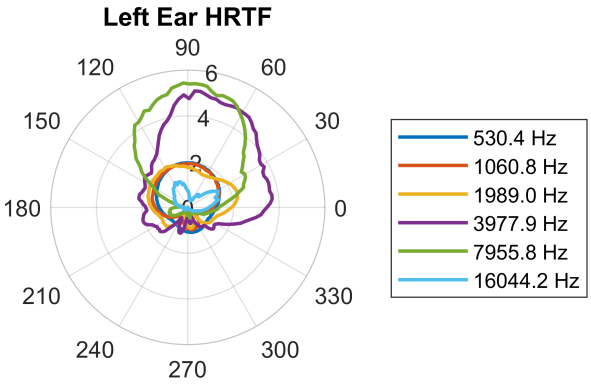

Figure 8: Polar plot of the left ear HRTF in the horizontal plane at approximated octave band frequencies. The actual bin frequencies are stated for better comparison with simulations.

Figure 8 shows a polar plot of the linear magnitude in the horizontal plane at approximated octave-band frequencies. The effectiveness of the feedback system is reinforced by the fact, that no distinct change in magnitude at the start and finish position at 310° azimuth is visible, which have been observed in prior measurements due to the subject's position drifting over the duration of the measurement.

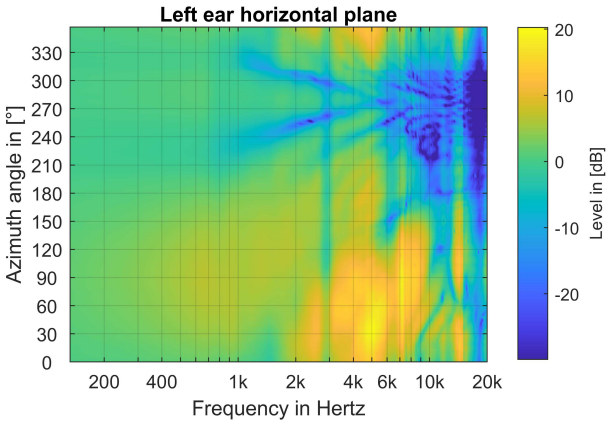

Figure 9: Frequency slice of the horizontal plane of the left ear.

This effect is however still present along with a presumed time-variant noise level of the power amplifier, which leads to higher noise floor at the end of the rotation compared to the beginning. This is particularly evident in Figure 9, where a distinct cut is visible at the starting angle, especially in areas of low magnitude, the notches in the HRTF, which get filled with noise. The measurement started at ∼310° azimuth, the location seen in Figure 4 and rotated in negative azimuth (clockwise) direction.

3 Summary

The process of a continuous HRTF measurement of a participant, of whom a high resolution 3D-scan is available [1] been detailed. A method to reduce motion during the measurement was described and its effect showed in the results. The post processing steps needed for the complex measurement method and the resulting HRTF were discussed and evaluated.

Supplementary Material

The HRTF dataset can be downloaded from doi.org/10.18154/RWTH-2020-06761 in both *.ita* [10] and *.sofa* [11] format.

As accompanying material, a high-resolution 3D data set of the same participant with measurement descriptions [1] is provided under doi.org/10.18154/RWTH-2020-06760.

All datasets related to this report are published with consent and a review by the medical ethics committee at the RWTH Aachen University ([EK 218/18]).

The data collection was are in line with the Declaration of Helsinki [12]. Participation was voluntary and sub-

ject to extensive information; personal data is treated confidentially (cf. DSGVO, [13]) and physical or mental harm to the participants is excluded.

References

[1] H. Braren and J. Fels, "A High-Resolution Individual 3D Adult Head and Torso Model for HRTF Simulation and Validation: 3D Data," Medical Acoustics Group, Institute of Technical Acoustics, RWTH Aachen, 2020. DOI: 10.18154/RWTH-2020-06760.

[2] J. Richter and J. Fels, "On the influence of continuous subject rotation during high-resolution head-related transfer function measurements," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 4, pp. 730–741, 2019.

[3] J.-G. Richter, G. Behler, and J. Fels, "Evaluation of a fast HRTF measurement system," in *AES 140th Convention*, Paris: Audio Engineering Society, 2016.

[4] P. Majdak, P. Balazs, and B. Laback, "Multiple exponential sweep method for fast measurement of head-related transfer functions," *Journal of the Audio Engineering Society*, vol. 55, pp. 623–637, Jul. 2007.

[5] P. Dietrich, B. Masiero, and M. Vorländer, "On the optimization of the multiple exponential sweep method," *Journal of the Audio Engineering Society*, vol. 61, pp. 113–124, Mar. 2013.

[6] J.-G. Richter, "Fast measurement of individual head-related transfer functions," Dissertation, RWTH Aachen University, Berlin, 2019. DOI: 10.18154/RWTH-2019-04006.

[7] Natural Point, Inc., *Optitrack*. [Online]. Available: https://optitrack.com/.

[8] B. F. G. Katz and M. Noisternig, "A comparative study of interaural time delay estimation methods.," *The Journal of the Acoustical Society of America*, vol. 135, no. 6, pp. 3530–3540, 2014. DOI: 10.1121/1.4875714.

[9] B. Xie, "On the low frequency characteristics of head-related transfer function," *Chinese Journal of Acoustics*, vol. 28, no. 2, pp. 116–128, 2009.

[10] M. Berzborn, R. Bomhardt, J. Klein, J.-G. Richter, and M. Vorländer, "The ITA-Toolbox: An Open Source MATLAB Toolbox for Acoustic Measurements and Signal Processing," 43th Annual German Congress on Acoustics, Kiel (Germany), 2017. [Online]. Available: http://ita-toolbox.org/.

[11] P. Majdak and M. Noisternig, "Aes69-2015: Aes standard for file exchange-spatial acoustic data file format," in *Audio Engineering Society*, 2015.

[12] "World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects.," eng, *Bulletin of the World Health Organization*, vol. 79, no. 4, pp. 373–374, 2001.

[13] Verordnung (EU) 2016/679, "Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung)," 2016.