From Simulation to Reality

Enhancing HRTF Capture

Felix Klooss[1]**, Anja-Xiaoxing Cui**[2]**, Christoph Reuter**[2]**, Isabella Czedik-Eysenberg**[2,3]
[1]MediaLab of the Faculty of Philological and Cultural Studies, University of Vienna
[2]Department of Musicology, University of Vienna
[3]Institute of Musicology and Music Pedagogy, Osnabrück University

universität wien
MediaLab PHILKULT
UNIVERSITÄT OSNABRÜCK
ICMPC International Conference on Music Perception and Cognition

Background

The Head-Related Transfer Function (HRTF) is a function that describes the transfer from a sound source to the ear canal as a function of the direction of incidence. These changes can be categorized as interaural time differences (ITDs), interaural level differences (ILDs), and sonic cues (SCs) [1].
Individual ear and head shapes have the greatest impact on transmission [2]. Thus, for each person, each direction of incidence has an individual expression of these characteristics, based on which the brain constructs directional perception.
If an audio signal is filtered with a custom HRTF of one direction of incidence and presented to the individual (e.g. over equalized headphones), the individual will perceive the signal as coming from that direction [3].
HRTF can be measured or calculated [4]. Traditional measurements are costly because the measurement equipment is large and expensive and the person being measured must be physically present. Unintentional head movements are usually controlled and measurements are repeated if the deviation is too large, but an accurate measurement to the degree is not guaranteed [5]. Physical calculations are a valuable alternative because they can be performed in the absence of the person being measured and without involuntary movement. Scanning the shape of the head and ears is error-prone and non-trivial [cf. 6].

Goals

HRTFs based on the same digital model are determined in three different ways and then compared. For the physical calculations, the commercial software Comsol (**C**) [7] and the open source software Mesh2HRTF (**M2H**) [8] are used. The measurements, on the other hand, are performed on a 3D printed artificial head using a measurement method (M) patented by the University of Vienna [9].

Method

An already digitized head is used and 3D printed to eliminate variations in the scanning process. The variations that result from the printing process are very small. The printed head is placed on the measuring device (**M**), which aligns it in front of a sound source and performs the entire measuring process automatically.
Both physical calculations are based on the boundary element method, with **M2H** using Burton-Miller collocation coupled with the multistep fast multipole method. For easier comparison, all three results are put into sofa format. This format is generated natively by both **M** and **M2H**. The data export for **C** is automated using Matlab and converted to Sofa in Python using Numpy, Scipy and Sofar. For further processing, analysis and display, Matplotlib and Pyfar are used.
Measurements with **M** are performed in an acoustically untreated setting to demonstrate its capabilities in a conventional work environment. To minimize room influences, all impulse responses are filtered with a 4 ms wide Blackman-Harris window centered on the sample with the highest amplitude.
Next, the Directional Transfer Functions (DTFs, [10]) of all HRTFs (**C**, **M2H** & **M**) are extracted. To determine the differences in magnitude, they are summed according to the ERB(Equivalent Rectangular Bandwidth)-scale per measurement method. All comparisons are made in the horizontal plane with an angular resolution of one degree.

Results

The results are very similar for all three methods. The DTFs obtained from **M** and **C** show slightly smaller magnitude differences than DTF_M compared to DTF_{M2H} and DTF_{M2H} compared to DTF_C. In particular, the dominant side for localization facing the sound source, shows a high correspondence of the magnitudes of the frequency bins averaged over the ERBs.
50% of the magnitude differences of the entire horizontal plane between 1.2 kHz and 20.8 kHz are within -0.9 dB and -0.2 dB (**M** vs. **C**) and -1.7 dB and -0.3 dB (**M** vs. **M2H**). Outliers (gray circles) become more prominent as the frequency increases.
When the side facing away from the sound source is excluded from the analysis, 50% of the magnitude differences are between 0 dB and 0.4 dB (**M** vs. **C**) and -0.6 dB and 0.0 dB (**M** vs. **M2H**), respectively.

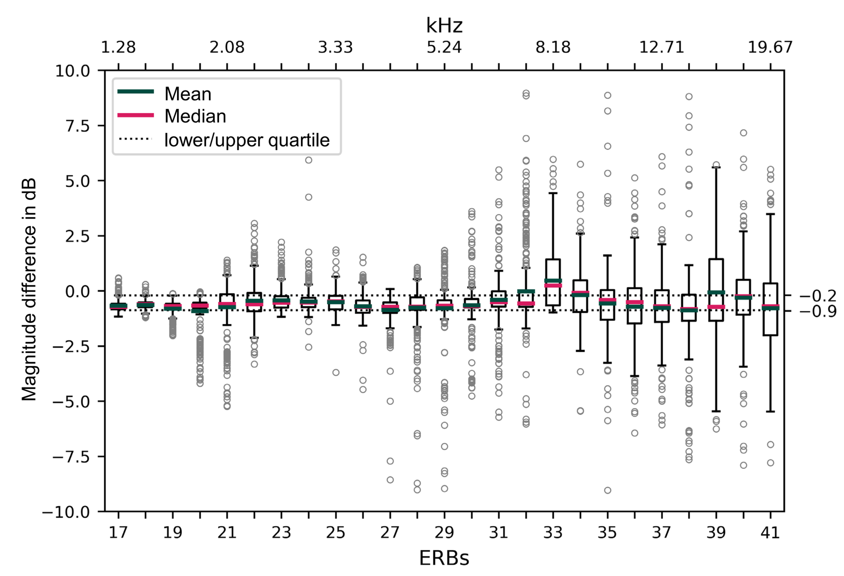

Fig. 1a: Horizontal plane, 0–359°, DTF_M vs. DTF_C

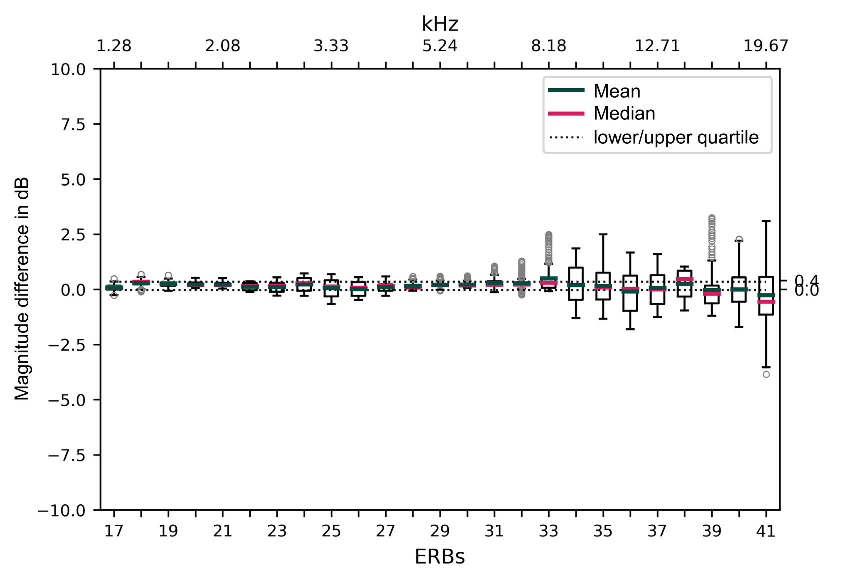

Fig. 1b: Horizontal plane, 0–180°, DTF_M vs. DTF_C

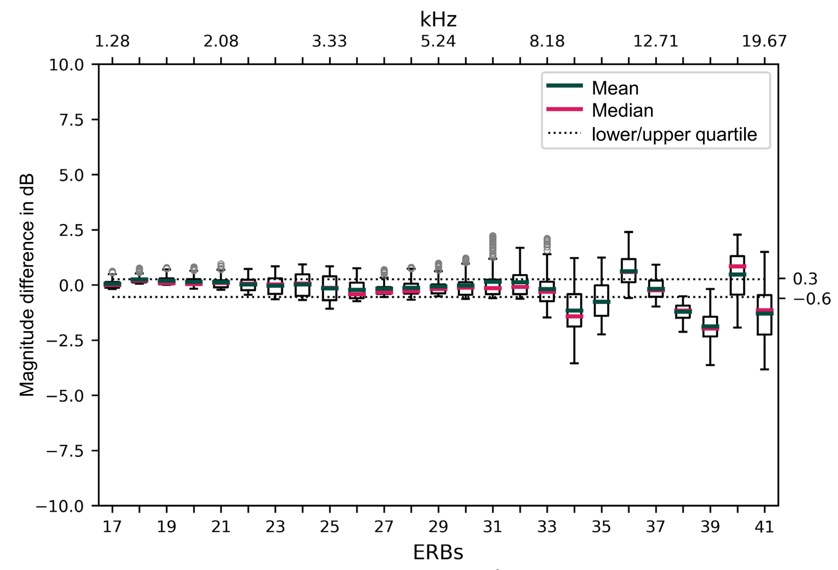

Fig. 1c: Horizontal plane, 0–180°, DTF_M vs. DTF_{M2H}

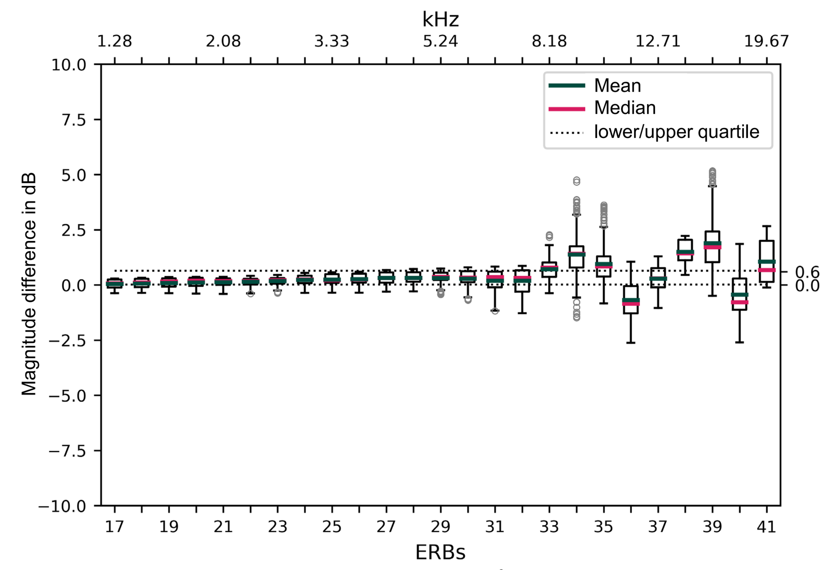

Fig. 1d: Horizontal plane, 0–180°, DTF_{M2H} vs. DTF_C

Figure 2 shows the magnitude responses for different directions of incidence. The DTFs obtained by **C**, **M2H**, and **M** show high agreement on the side facing the sound source (Figs. 2a, 2b). On the opposite side, the differences are visually obvious. Depending on the direction of incidence, the magnitudes vary greatly, as can be seen in Figures 2c and 2d with only one degree of angular deviation.

Despite the tendency for higher agreement between DTF_M and DTF_C compared to DTF_M and DTF_{M2H}, instabilities can be detected in DTF_C that cannot be observed in DTF_{M2H}. These can be seen as a notch in Figure 2a and also in the heat map of Figure 3a in a similar frequency range.

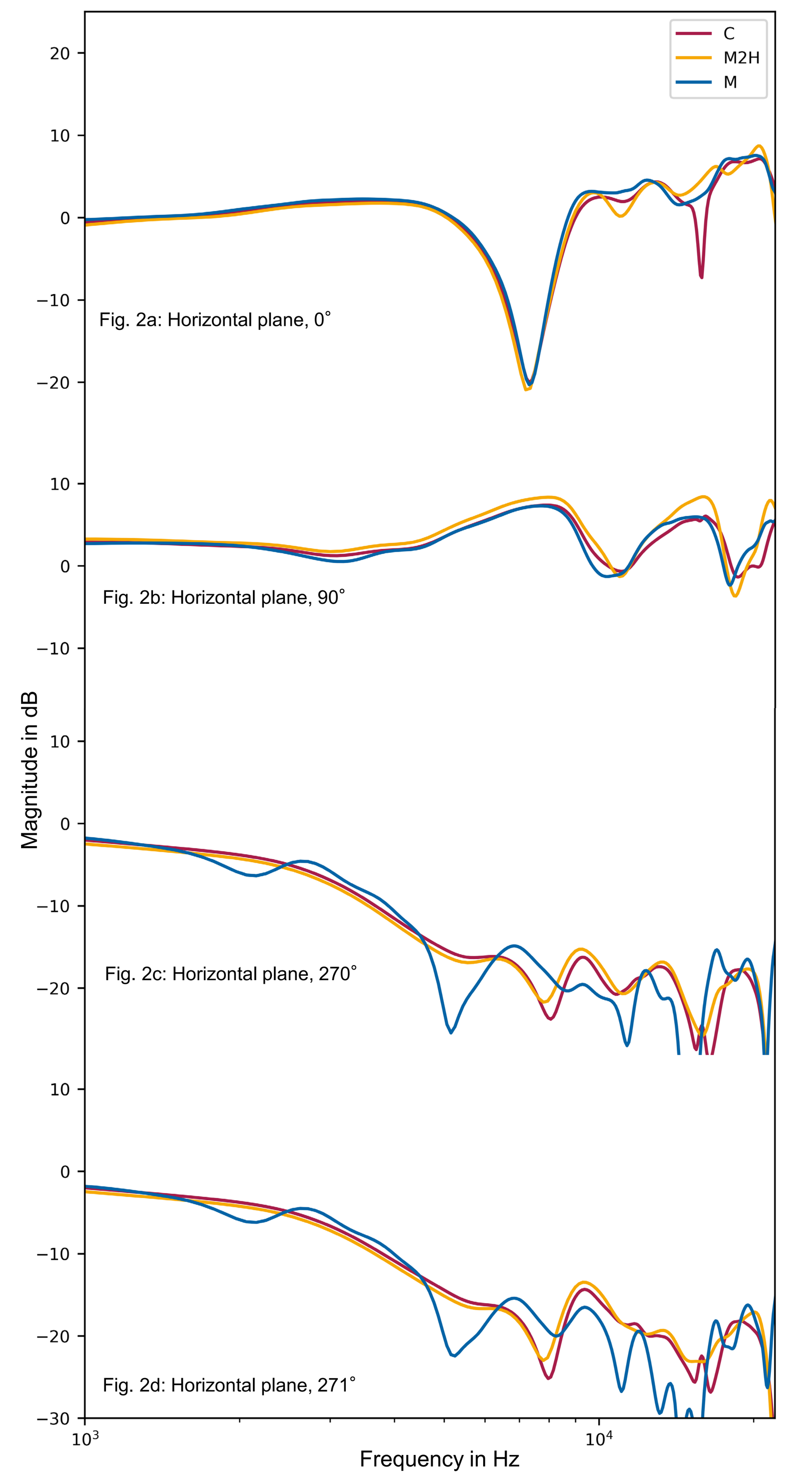

Fig. 2: Horizontal plane, DTF_C, DTF_{M2H}, DTF_M

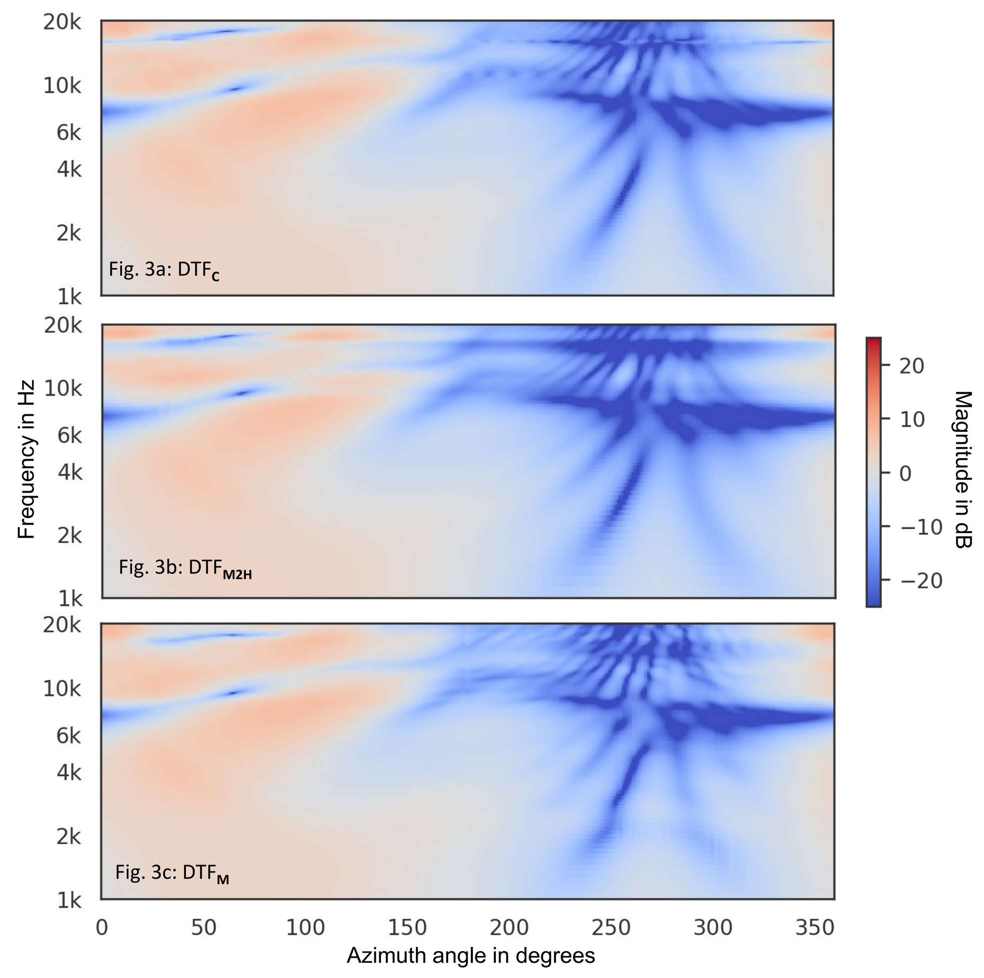

Fig. 3: Horizontal plane, 0 – 359° DTF Magnitude Heatmap

For a better visual overview of the magnitude responses as a function of the angle of incidence in the horizontal plane, Figure 3 spreads the magnitudes according to frequency and angle of incidence. The high level of agreement can also be observed here.

Conclusio

Both **C** and **M2H** and **M** produce DTFs with high agreement well below the threshold of perception. Since the smallest angular changes have large effects in the acoustic shadow, the most significant differences in the DTFs are also apparent there, but their perceptual influence is assumed to be small.

Without further investigation, it is not possible to determine which of these methods is closer to a real person's HRTFs. However, it should be noted that the high flexibility and complexity of **C** has more disadvantages than advantages in this context. Only in conjunction with Matlab can the necessary exports be automated. **M2H**, on the other hand, is much easier to use, delivers results faster, and is freely accessible..
M offers advantages especially when complex additional systems (e.g. hearing aids, headphones) are integrated into the measurement. Arbitrary measurement positions (e.g. Lebedev quadrature) can be approached in any resolution. Especially with few measurement points, **M** is much faster than **C** and **M2H** because they are based on the reciprocity principle.

Referenzen

[1] Hebrank, J. & Wright, D. (1974). Spectral cues used in the localization of sound sources on the median plane. The Journal of the Acoustical Society of Amer ica, 56(6), 1829–1834.
[2] Musicant, A. D. & Butler, R. A. (1984). The influence of pinnae-based spectral cues on sound localization. The Journal of the Acoustical Society of America, 75(4), 1195–1200.
[3] Wightman, F. L. & Kistler, D. J. (1989). Headphone simulation of free-field listening. I: Stimulus synthesis. The Journal of the Acoustical Society of America, 85(2), 858–867.
[4] Katz, B. F. (2001). Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation. The Journal of the Acous tical Society of America, 110(5 Pt 1), 2440–2448.
[5] ARI HRTF-Database detailed description (2023, 3. März). Austrain Academy of Science. https://www.oeaw.ac.at/en/isf/das-institut/software/hrtf-database/detailed-description
[6] sdx-lv (2023, 3. März). 3D scanning. Sourceforge. https://sourceforge.net/p/mesh2hrtf/wiki/Basic_tutorial_3d_scanning/
[7] COMSOL Multiphysics®, www.comsol.com. COMSOL AB, Stockholm, Sweden.
[8] Ziegelwanger, H. & Kreuzer, W. & Majdak, P. (2015). MESH2HRTF: An Open-Source software Package for the Numerical calculation of head-related transfer functions.
[9] Klooss, F. (2022). Robot for measuring HRTFs (Application No. 22208557.3). European Patent Office.
[10] Middlebrooks, J. C. (1999). “Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency,” J. Acoust. Soc. Am. 106, 1493–1510.