

Published in final edited form as:

Comput Aided Des. 2016 June ; 75-76: 39–46. doi:10.1016/j.cad.2016.02.006.

A framework for geometry acquisition, 3-D printing, simulation, and measurement of head-related transfer functions with a focus on hearing-assistive devices

Stine Harder^{a,*}, Rasmus R. Paulsen^a, Martin Larsen^b, Søren Laugesen^c, Michael Mihocic^d, and Piotr Majdak^d

^aTechnical University of Denmark, DTU Compute, DK-2800 Lyngby, Denmark

^bOticon A/S, DK-2765 Smørum, Denmark

^cEriksholm Research Centre, DK-3070 Snekkersten, Denmark

^dAcoustics Research Institute, Austrian Academy of Sciences, Vienna, Austria

Abstract

Individual head-related transfer functions (HRTFs) are essential in applications like fitting hearing-assistive devices (HADs) for providing accurate sound localization performance. Individual HRTFs are usually obtained through intricate acoustic measurements. This paper investigates the use of a three-dimensional (3D) head model for acquisition of individual HRTFs. Two aspects were investigated; whether a 3D-printed model can replace measurements on a human listener and whether numerical simulations can replace acoustic measurements. For this purpose, HRTFs were acoustically measured for four human listeners and for a 3D printed head model of one of these listeners. Further, HRTFs were simulated by applying the finite element method to the 3D head model. The monaural spectral features and spectral distortions were very similar between re-measurements and between human and printed measurements, however larger deviations were observed between measurement and simulation. The binaural cues were in agreement among all HRTFs of the same listener, indicating that the 3D model is able to provide localization cues potentially accessible to HAD users. Hence, the pipeline of geometry acquisition, printing, and acoustic measurements or simulations, seems to be a promising step forward towards in-silico design of HADs.

Keywords

3D head model; CAD modeling; 3D printing; Acoustical measurements; Acoustical simulations; Head-related transfer functions

1 Introduction

Human listeners are able to localize sounds in space in terms of assigning direction and distance to the perceived auditory image [1]. This ability is an essential function of spatial

*Corresponding author. sthar@dtu.dk (S. Harder).

hearing, which involves further perceptual effects like the estimation of the apparent source width, spatial unmasking of speech (cocktail party effect), and externalization (out-of-head perception) of sounds. Generally, spatial hearing relies on physical directional acoustic features which are the consequence of acoustic filtering of the sound by the pinna, head, and torso. The filtering can be described by the head-related transfer functions (HRTFs), which represent the transfer functions from sound sources to a sound receiver, usually placed at the entrance of the ear canal. For far-field sources, HRTFs depend on the direction of sound incidence [2]. For near-field sources, HRTFs additionally depend on the distance between the source and the ear [3].

An HRTF encodes monaural spectral cues, which are used by human listeners to estimate the sound-source position along sagittal planes (top, down, front, back). For the localization of the source along horizontal planes (left, right), interaural cues are used, namely, interaural time and level differences (ITDs, ILDs). In particular, ITD cues in the frequency range below 1.5 kHz are important for the sound localization in the horizontal planes. The interaural cues are encoded in a binaural pair of HRTFs, thus, a binaural set of HRTFs can be used to describe all directional cues required for spatial hearing.

HRTFs can be used for many purposes, e.g., for models of spatial hearing [4–6], for fitting of hearing aids [7,8], and for presenting virtual binaural audio signals via headphones in so-called virtual auditory displays [9,10]. As HRTFs depend on the individual geometry of the listener's head and ear, HRTFs are listener-specific [11]. Listener-specific HRTFs are often acoustically measured by placing small microphones at the entrance of the listener's ear canals. During a measurement session, often consisting of measurements for many spatial positions, the listener must sit still for tens of minutes depending on the measuring facilities [12–14]. Thus, it is not surprising that in some applications, generic HRTFs, i.e., HRTFs of a manikin representing an average of the human population [15–17], are used. However, in applications like fitting hearing aids to children [18] or providing accurate sound localization performance via headphones [19], it might be important to consider listener-specific HRTFs. As an alternative to demanding acoustic measurements, HRTF can also be numerically calculated from a three-dimensional (3D) representation of a human geometry [20–23] and established methods have shown good congruence between measured and simulated HRTFs for frequencies below 7 kHz [23,24].

In this study, a method for obtaining listener-specific HRTFs with the focus on virtual product design is evaluated. An application could be optimization of the directional microphone using the Directivity Index as an evaluation criteria, as is performed in.¹ Virtual product design is an emerging discipline that has the potential of reducing production costs and creating more comfortable and better functioning wearables as for example clothes, helmets, and in our case, hearing-assistive devices (HAD), which are commonly used to treat hearing impairment. An example of virtual product design is the size-China project that aims at creating a population statistics on human heads for product design [25]. In [26], a parametric model of the entire human body was computed based on 250 full body scans.

¹Optimizing hearing-aid directionality from measurements and simulations, S. Harder, R.R. Paulsen, M. Larsen, M.S. Pedersen, S. Laugesen, M. Mihocic, and P. Majdak, manuscript in preparation.

This model was then used to synthesize plausible body shapes as input to product design. In another example, foot shape was investigated in several studies driven by the large footwear industry and recently, 50 surface scanned feet were used as input to a statistical analysis of shape with the goal to produce optimized shoe lasts [27]. Similarly, in the design of HADs, the position of the microphones is essential, not only for the design of the casings, but also for capturing spatial acoustic cues. Thus, different types of microphones at various places have been proposed, for a recent review, see [28]. The measurement of HRTFs with HADs is even more demanding than the measurement done in the ear canal because of the much larger degree of freedom: a simple positioning of the HAD might yield different HRTFs, and many microphone positions have to be considered in the individualization of the HAD. On the other hand, since most of the HADs focus on transmitting frequencies below 8 kHz [29], the evaluation of HRTFs can be limited to the upper frequency 8 kHz.

This paper presents yet another step towards full in-silico design of HADs. The goal is to evaluate whether simulated HRTFs can replace expensive acoustic measurements in the future. We simulate listener-specific HRTFs based on a reconstruction of the listeners' geometry from a set of 3D surface scans with missing data. An evaluation of methods for optical 3D scanning of human pinnae is available in [30]. HRTF measurements are expensive with respect to measurement equipment and measuring time for the listener, in particular when measuring HRTFs for different microphone positions in an HAD. In a simulation re-positioning of the microphones does not require additional participation of the listener. HRTF simulations can have a further impact on the future product design: by using statistical shape modeling (as in for example [26,27]), HRTFs of future HADs could be obtained for hundreds or even thousands of plausible virtual heads, in contrast to the current practice where one or a few generic manikin heads are used because of the time-consuming measurement procedure.

In the following, the three stages of investigations are described. First, a high-quality 3D geometrical model of a human listener was used to numerically simulate corresponding HRTFs. Second, computer-aided-design modeling and 3D printing were used to obtain a listener-specific 3D printed head model, from which acoustically measured HRTFs were obtained. Third, acoustical measurements were performed to obtain HRTFs of the corresponding actual human listener. This three-step pipeline allowed for a thorough evaluation of the differences among acoustically measured and simulated HRTFs of the model and the listener. The acoustically measured HRTFs for the printed head model made it possible to distinguish between HRTF deviations caused by differences in shape (human versus obtained 3D model) and differences in method (measurements versus simulations).

2 Material and methods

2.1 The surface model

A complete human head model was generated using a framework which combines a number of 3D surface scans captured from different angles surrounding the individual. Listener NH167 from the ARI database (<http://sofaconventions.org>) was scanned. The method was presented in [31] and it consists of both scanning, stitching, and meshing. Here, the method is briefly reviewed.

Sixteen 3D surface scans were acquired using a 3D facial scanner (Canfield scientific Vectra M3) and subsequently aligned with a rigid-body transformation based on a few manually annotated landmark points. The result was a roughly aligned point cloud where many parts of the head were well covered by samples, but several regions lacked well-defined sample points. The regions which lack sample points were in particular hairy areas that were not covered by a cap, such as eyebrows or sideburns.

Surface reconstruction was applied on the point cloud. It was based on the Markov Random Field reconstruction in [32], which is well-suited for human head scans because it (1) is robust to noise, (2) interpolates in areas with missing data in an anatomically plausible way, and (3) accurately aligns the partially overlapping scans. An implementation of the higher-order mathematical model are found at (<http://www2.imm.dtu.dk/image/MRFSurface/>) along with examples. The scanning process lasted for approximately 15 min, while the stitching required half a day's work. The stitching process has however not been optimized or completely automated yet, which could decrease the required time dramatically.

By visual inspection it can be seen that the anatomy is correctly described by the final surface model. All anatomical structures are present and the surface contains no holes, see Fig. 1(a). Further information on the surface reconstruction accuracy can be found in [32].

2.2 Adapting 3D surface models to simulation and printing

By nature, the 3D surface model acquired from range scanning is a two-dimensional (2D) surface embedded in 3D space. Technically it consists of vertices and triangles, which also applies to the results of the surface merging and stitching. Several numerical simulation packages exist that can use triangulated surfaces as input but most programs, such as Ansys, Comsol Multiphysics, or Abaqus, work better with surfaces represented as non-uniform rational B-splines (NURBS) [33]. Thus, the surface model was converted to NURBS (Geomagic Studio, www.geomagic.com), see Fig. 1(b).

The head model represented in NURBS was further processed using a computer-aided design software package (Creo Parametric 2.0, <http://www.ptc.com/cad/3d-cad/creo-parametric>). Because the surface model has no thickness per se, an artificial thickness was added to the surface. This is typically done by creating an inward offset surface of several millimeters. However, with the complicated geometry of the human head and ears this approach did not give satisfactory results. Instead, a balloon-like object was created inside the head and manually reshaped to turn the head into a hollow shell. The overall minimum thickness was 5 mm and certain parts like the nose remained solid.

In order to mount onto a torso simulator (Type 4128, Brüel & Kjær) after the printing, a flange was added to the model at the bottom of the head. Further, in order to be able to manufacture the head and ears separately using different materials, boxes around the ears were created. After the printing, the ears were attached to box-shaped elements that were inserted in the corresponding box-shaped recesses in the head.

2.3 3D print specifications for the head and ears

With the rapid development in layered manufacturing (in the following referred to as 3D printing) new possibilities are opening up. Previously, creating an accurate replica of a human head would have required massive amounts of work. As demonstrated in this article it is still not a trivial task, but with the development of software tools and the decreasing costs of 3D printing, generating flexible human replicas for product testing is becoming increasingly accessible.

In the current work, the focus is not on creating a full head with material properties that exactly match a living human. It would be very complicated and it is believed that the impact on the acoustical measurements is of less significance than other factors in the measurement setup, such as geometrical modeling errors. The head was therefore printed in hard plastic except the ears which were printed in a soft material. The reason for choosing a soft material for the ears was mainly the need of placing hearing aids behind the ears in a way that is as close as possible to the placement of hearing aids on human ears.

The ears were printed on a polyjet Objet500 Connex 3D printer (www.stratasys.com) using two materials and a layer thickness of 0.03 mm. The core was printed in a hard, acrylic material (Verowhite-FullCure830) and the outer part was printed in a soft material (TangoPlus-FullCure930). The hard part of the head was printed on an SLS printer EOS P395 (www.eos.info) in hard acrylic (PA2200) using 0.12 mm layers. Further material properties can be located at (www.damvig.dk).

2.4 HRTF simulation

The simulation was set up in ANSYS 15.0 (<http://www.ansys.com/>). The head geometry from Section 2.2 was mounted on a virtual model of the Brüel & Kjær torso simulator for the simulations. The head geometry was meshed with first-order acoustic elements (ANSYS type 30). The mesh was refined at the ears to an average edge length of 2 mm, while the rest of the model used a length of 4 mm. The re-meshing method was 'Hex dominant', which produced a mesh with less nodes compared to a tetrahedral mesh. The final mesh, which represented the model of the head and torso, contained 1.9 million nodes. The mesh was, with a maximum edge length of 4 mm, valid up to 8.5 kHz using 10 elements per wavelength as recommended in [34]. The mesh was however also refined to an edge length of 2 mm at the ears and results are shown for frequencies up to 10 kHz.

The final model was placed inside a box with dimensions 420 mm × 700 mm × 250 mm, which in the simulation was set to be filled with air. The model was subtracted from the box, leaving the air surrounding the model. A 40 mm perfectly matched layer (PML) was added to the outside of the inner air-box, see Fig. 2. The purpose of the PML was to absorb radiated sound from the inner air part. The default setting for attenuation in the PML region was used (0.001 or -60 dB), yielding 120 dB total attenuation of the sound field reflected from the PML.

In the acoustic measurement setup (see Section 2.5), loudspeakers surrounding the head generated the sound, and the microphones placed in the ear canal captured the sound pressure. In the simulation, the reciprocal approach was used [13,35]: the sound source was

placed at the microphone position, and the sound pressure level was calculated on a sphere surrounding the head at 1.2 m distance. The reciprocal approach yields equivalent results to the direct approach, but the simulation can be performed for all sound-source directions in a single simulation step [36].

The simulation was performed for a total of 54 linearly spaced frequencies from 187.5 to 10,125 Hz and for two vibrating surfaces positioned in the ear canals. Circular vibrating surfaces with a radius of 0.25 mm were used. The resulting far field was evaluated in vertical and horizontal angle steps of 2.5° at a distance of 1.2 m, however, only the 1550 source positions, that were acquired for the measurements (see Section 2.5), were used for this paper.

Furthermore, a free-field transfer function was simulated and used as a reference for the other simulations, resulting in a set of simulated HRTFs comparable to the measured HRTFs described in Section 2.5. In this way both measured and simulated HRTFs were ratios of sound pressures at a microphone on the head and at a microphone in the center of the head (head absent).

2.5 HRTF acoustic measurement

2.5.1 Human listeners—HRTFs of human listeners were acoustically measured in a semi-anechoic chamber. Twenty-two loudspeakers (custom-made boxes with VIFA 10 BGS as drivers; the variation in the frequency response was ± 4 dB in the range from 0.2 to 16 kHz) were mounted on an arc at fixed elevations from -30° to 80° . They were driven by amplifiers adapted from Edirol MA-5D active loudspeaker systems. The loudspeakers and the arc were covered with acoustic damping material to reduce the intensity of reflections. The total harmonic distortion of the loudspeaker-amplifier systems was on average 0.19% (at 63-dBSPL and 1 kHz). The human listener was seated on a chair in the center of the arc and was wearing in-ear-microphones (Sennheiser KE-4-211-2). The Sennheiser KE-4 was fixed by wrapping medical tape (brand: Leukoplast) around the curved surface of the microphone to reach a subject- and ear canal-dependent individual thickness that blocks the ear canal after inserting the microphone. The Sennheiser membrane is the only side of the microphone that is visible, after the insertion. 1–2 short strips of tape help to fix the cable on the cheek and to reduce the tension. HRTFs were measured at the entrance of the blocked ear canal. The microphones were connected via amplifiers (RDL FP-MP1) to the digital audio interface. A 1728.8-ms exponential frequency sweep beginning at 50 Hz and ending at 20 kHz was used to measure each HRTF. Then a system identification procedure was used to obtain raw HRTFs [12].

The HRTFs were measured for one azimuth and several elevations at once (see below) by playing the sweeps and recording the signals at the microphones. Then the listener was rotated by 2.5° to measure HRTFs for the next azimuth. In the horizontal interaural plane, the HRTFs were measured with 2.5° spacing within the azimuth range of $\pm 45^\circ$ and with 5° spacing outside this range. The positions of the HRTFs were distributed with an approximately constant spherical angle, which means that the number of measured HRTFs in a given horizontal plane decreased with increasing elevation. For example, at the elevation of 80° , only 18 HRTFs were measured. In total, 1550 HRTFs were measured for each

listener. To decrease the total time required to measure the HRTFs, the multiple exponential sweep method (MESM) was applied [12]. This method allows for a subsequent sweep to be played before the end of a previous sweep, but still reconstructs HRTFs without artifacts. The MESM uses two mechanisms, interleaving and overlapping and both depend on the acoustic measurement conditions (for more details see [12]). The facilities allowed the interleaving of three sweeps and overlapping of eight groups of the interleaved sweeps. During the HRTF measurement, the position and orientation of the listener's head were captured via an electromagnetic tracker (Flock of Birds, Ascension) in real time. The tracking sensor was mounted on the top of the listener's head. The tracking device was capable of measuring all 6 degrees of freedom (x , y , z , azimuth, elevation, and roll) at a rate of 51.5 measurements/sec. The tracking accuracy was 1.7 mm for positions and 0.5° for orientation. If the head was outside the valid range, the measurements for that particular azimuth were repeated once the listener was back in the range for 500 ms. The valid ranges were set to ± 2.5 cm for the position, $\pm 2.5^\circ$ for the azimuth, and $\pm 5^\circ$ for the elevation and roll. On average, measurements for three azimuths were repeated per listener and the complete measurement procedure lasted for approximately 20 min.

For each raw HRTF, head-related impulse responses (HRIRs), i.e., the inverse Fourier transform of HRTFs, was calculated, yielding raw HRIRs. The raw HRTFs were affected by loudspeaker, room, and microphone used for the measurements. These effects can be described by the equipment transfer functions (ETFs), which were captured by placing the in-ear microphones in the center of the arc and measuring the room impulse response for all loudspeakers. ETFs were measured each time before the HRTF measurement of an object, thus five times in total.

The impulse responses of the ETFs showed a strong reflection of at least 20 dB below the level of the direct sound, delayed by at least 6.9 ms. This reflection can be attributed to the floor reflection as an effect of using a semi-anechoic room for the measurements. Since HRIRs do not contribute much beyond first 5 ms [37], such reflections were removed from both the raw HRIRs and the impulse responses of ETFs by windowing in the time domain.

The remaining loudspeaker and microphone responses were compensated by dividing an HRTF for a given microphone, m , and loudspeaker position, (θ, ϕ) , with its corresponding ETF. The set of compensated HRTFs for all directions and for both microphones, will from now on be referred to as the *human* HRTFs.

HRTFs of four listeners (NH166, NH167, NH168, and NH170 from the ARI database, <http://sofaconventions.org>) were measured. In addition, the measurement was repeated for NH167. Thus, five sets of human HRTFs were available for the analysis.

2.5.2 Printed model—HRTFs of the printed head fixed on a torso simulator (Type 4128, Brüel & Kjær) were measured using the same methodology as for the human listeners with one exception: The torso was fixed with tape in order to prevent any movements during the measurements. Also, all other aspects of post processing were identical to those of the post processing done for the human listeners, resulting in *printed* HRTFs.

2.6 HRTF evaluation

The monaural spectral features were first analyzed by visual comparison of the amplitude spectra of the corresponding HRTFs. Similar comparisons were performed in studies like [18,21,24,38]. Further, comparisons across various pairs of conditions were performed by means of a spectral distortion [22]. To this end, for each condition and spatial position, HRTF spectral magnitudes were calculated in 22 frequency bands from 729 Hz to 9 kHz, each band spaced by one equivalent rectangular bandwidth (ERB). Our spectral distortion was then the standard deviation of the spectral differences between magnitudes of conditions under the test, similarly to [39]. Note that in contrast to [22], who used absolute spectral average, our spectral distortion was based on standard deviation in order to obtain a metric robust against the confounding effect of broadband differences between HRTFs. Also note that the model from [4], while showing good predictive power for normal-hearing listeners, was not applied in our study because it has neither been calibrated nor applied to hearing-impaired listeners yet. Finally, pairs of conditions were compared by means of the spatial correlation analysis [23]. For each condition, spectral magnitudes were calculated for a single frequency band and considered for all spatial positions. Then, for each band, a correlation coefficient was calculated between the magnitudes (as functions of position) for the two conditions under the test.

For the evaluation of the ITD cues, the ITD was evaluated as a function of the azimuth angle in the horizontal plane by using a similar approach to that used in [40]. A comparison of different ITD definitions are found in [41]. For each azimuth ϕ , the ITD was derived from the interaural phase difference of the corresponding HRTFs. In particular, given a pair of HRTFs, $H_{l,\phi}$ and $H_{r,\phi}$ for a given source position in the frequency domain, the interaural phase difference, $\Phi_{\phi}(f)$, was

$$\Phi_{\phi}(f) = \text{unwrap} \left(\frac{H_{l,\phi}}{H_{r,\phi}} \right), \quad (1)$$

where $\text{unwrap}(\cdot)$ denotes the “unwrap-phase” operator along the frequency axis. An implementation of the unwrap algorithm by the commercial software package (MATLAB and Statistics Toolbox Release 2015a, The MathWorks, Inc., Natick, Massachusetts, United States) was used. Then, the frequency-dependent ITD, $\tau_{\phi}(f)$, is

$$\tau_{\phi}(f) = \frac{-\Phi_{\phi}(f)}{2\pi f}. \quad (2)$$

The estimated ITD, τ_{ϕ} , was then the $\tau_{\phi}(f)$ averaged in the frequency range from 164 to 890 Hz for the human and printed HRTFs, and from 188 to 938 Hz for the simulated HRTFs, respectively. These frequency ranges were selected to be as wide as possible while including only reliable data for all test subjects.

The estimated ITDs were accompanied by measured head sizes of the four individuals. The head sizes were measured using a custom-made vernier calliper, as the distance between a position in the bottom of concha next to the ear canal.

3 Results

3.1 Monaural spectral features

Fig. 3 shows the human HRTFs for the frontal position in the horizontal plane for the right ear. Even though the general shapes are similar, variations up to 10 dB can be observed among the listeners. The repetition of the HRTF measurement yielded similar HRTFs for frequencies below 1.7 kHz but less similar HRTFs for higher frequencies. Further investigations of HRTF reliability for our data can be found in [42].

Fig. 4 shows the HRTFs of NH167, namely the human HRTFs, the repeated human HRTFs, printed HRTFs, and the simulated HRTFs for the right ear. Five positions in the horizontal plane are shown with azimuth angles of 90, 45, 0, -45, and -90°. As expected, the shapes of the HRTFs change with the azimuth angle for all types of HRTFs. For the ipsilateral directions (negative azimuth angles), the human, printed, and simulated HRTFs show a good similarity, especially in the frequency range up to 4 kHz. Notice that the printed HRTFs and simulated HRTFs are based on both identical head and torso geometries. A larger difference between the HRTFs can be observed for the contralateral directions (positive azimuth angles). Generally, the differences are as small as a few dB at lower frequencies and increase with frequency.

When comparing across the conditions, the two human HRTFs and the printed HRTFs seem to be in a good agreement. While the repeated human HRTFs appear to be very similar to the human HRTFs, the printed HRTFs seem to be less similar, and the simulated HRTFs only roughly appear to correspond with the human HRTFs. Interestingly, a large difference between the human and printed HRTFs can be observed at approximately 1.1 kHz for azimuth angles of 45 and 0°, where a notch is present in both human HRTFs, but is missing in the printed HRTFs. The 1.1-kHz notch does not seem to be an artifact of the measurement setup, because the same artifact would be expected in human and printed HRTFs. Thus, it seems to be an aspect of the human measurement and was apparently not captured by the geometry acquisition.

Spectral distortions were calculated for a comparison of the three conditions with the human HRTFs, for both the five source positions in Fig. 4 and as an average over all source positions, see Table 1. The comparison between the re-measured HRTFs showed spectral distortions of 5.2 and 4.6 dB for the right and left ear, respectively, when averaged across all tested positions. This seems to define the level of similarity which has been obtained when acoustically remeasuring HRTFs of the same listener. The comparison between human and printed HRTFs showed spectral distortions of 4.8 and 5.4 dB. This is similar to the spectral distortions obtained for the measurement repetitions, thus, measurements on the printed and human heads seem to have yielded equivalent HRTFs. On the other hand, the comparison between human and simulated HRTFs showed spectral distortions of 14.3 and 14.4 dB. These are much larger than the spectral distortions obtained for measurement repetition,

indicating that the simulation yielded much more different HRTFs than the measurement repetition. These spectral distortions are also much larger than the spectral distortions obtained for the printed HRTFs indicating that the simulation yielded more different HRTFs than the measurement on the printed head. Since both printed and simulated HRTFs were based on identical geometries, the larger spectral distortions for the simulated HRTFs indicate that the origin of this deviation is not based on the geometry acquisition and post processing.

Further analysis was based on spatial correlation coefficients. Fig. 6 shows the spatial correlation coefficients for both left and right ear, and the same conditions as those for the analysis based on spectral distortions. The repetition of the acoustic measurement yielded correlation coefficients larger than 0.9 for most of the frequencies, which is in agreement with most of the results from [23]. In contrast, the correlation coefficients representing the similarity between the human and simulated HRTFs were much lower, indicating not much similarity between the simulation and acoustic measurements. These results clearly show the tendency that the human repeated HRTFs were most similar, the printed HRTFs were less similar, and the simulated HRTFs were least similar as compared to the human HRTFs. This is in agreement with the observations made for the spectral distortions.

Finally, HRTF magnitudes were plotted as a function of the elevation angle in the median plane, i.e., for all azimuth angles of zero degrees, see Fig. 5. Human, printed, and simulated HRTFs are shown, the color encodes the relative magnitude (in dB). The spectra at elevation of zero degrees correspond to the spectra shown in the panel for azimuth angle of zero in Fig. 3. In all four panels of Fig. 5, spectral modulations can be observed, beginning at frequencies around 2 kHz and polar angles of -30° , and then decreasing in frequencies with increasing polar angle up to 0° , with an opposite pattern in the rear hemifield. There is, however, a striking difference in these patterns between the human and printed HRTFs and the simulated HRTFs: the modulations in the human and printed HRTFs appear to be more dense showing more spectral fluctuations within the same frequency range.

3.2 ITD cues

Fig. 7 shows ITDs in the horizontal plane calculated from human, printed, and simulated HRTFs. For NH167, the ITD showed similar patterns when compared across the two acoustic measurements, with a maximum ITD of $790 \mu\text{s}$ (head size of 12.9 cm). These ITD patterns also seem to be similar when compared to those obtained from the printed and simulated HRTFs. The ITDs obtained for the remaining three listeners exhibit clear differences with maximum ITDs of $760 \mu\text{s}$ (head size of 12.8 cm), $810 \mu\text{s}$ (head size of 13.4 cm) and $830 \mu\text{s}$ (head size of 14.0 cm), respectively. These differences can be attributed to different head sizes of the test subjects.

4 Discussion

The goal of the present study was to investigate the use of an individual head model in the acquisition of individual HRTFs. One aspect was to investigate whether a 3D printed version of a listener's head can replace the human listener in measurements. Another aspect was to evaluate the quality of simulated HRTFs when using an individual head model.

Generally, there was a variability in magnitude spectra of HRTFs across the four human listeners, particularly for higher frequencies. A variability (however smaller than among subjects) was also noticed for repeated measurements on the same listener. This variability can be attributed to small changes in the measurement setup, namely, (1) microphone position (recall that microphone was re-positioned between the two measurements), and (2) position and orientation of the listener (recall the criteria for accepting a measurement, see Section 2.5.1). The acoustic consequences of these geometrical changes naturally become greater at high frequencies where the wavelength of sound is shorter and thus in the range of the magnitude of the geometrical changes. The variability captured by the repeated measurement sets a reference for further evaluation of the differences among the human, printed, and simulated HRTFs of that listener. Interestingly, similar variability between two HRTF measurements in the same listeners was found in ([43], Fig. 3 conditions binOwn and binOwnB, the diffuse part removed from those HRTFs). Despite the differences in the HRTFs, sound-localization performance obtained with these two conditions in localization experiments was similar. This indicates that also the repeated measurements are similar in terms of providing similar directional cues.

The correlations between HRTF magnitudes for the human HRTFs and repeated, printed, and simulated HRTFs, show that the repeated HRTFs were most similar, the printed HRTFs were less similar, and that the simulated HRTFs were least similar across all source positions. These results are further supported by the analysis based on spectral distortion, indicating that the printed HRTFs but not the simulated HRTFs are within the range of measurement repetition.

The correlation coefficients were low between 1 and 4 kHz, which can possibly be attributed to a poor match of torso and sitting position in the measurements. The high correlation between 4 and 7 kHz can be attributed to a good match of the actual and printed head dimensions. The low correlation above 7 kHz can be attributed to a poor match of the ears between the human and the geometric model.

A variability was also found for the ITDs, the most salient cue for localizing sound sources in the horizontal plane. However, the ITD variability showed a different pattern than the monaural amplitude spectra: very similar ITDs were found for all HRTFs of the same listener NH167 (smaller differences than 15 μ s at 90° and below 5 μ s at 180°, which is below the ITD discrimination threshold found at those directions [44]), but different ITDs were found across the listeners. Since the shape of the ITD as a function of azimuth theoretically depends on the shape and size of the listener's head, it was not surprising that across-listener differences were found.

The question of whether a 3D printed version of a listener's head can replace repeated measurements of the human listener was addressed in the comparison between the human and printed HRTFs. The comparison of the amplitude spectra revealed some differences. Some of these differences, when compared to those obtained by the repeated measurement on the actual human, have been attributed to small differences in microphone position and head orientation. Given the limited frequency range of the HADs, these differences can be

considered as minor. The spectral distortions being in the range of measurement repetition seem to validate the 3D printed version of a listener for our use.

The quality of simulated HRTFs was addressed in the comparison between the measured and simulated HRTFs. Generally, the human and printed HRTFs appear to be more similar than the simulated HRTFs in terms of spectral magnitude and spectral distortions. The differences in ITDs found across both measured and simulated HRTFs were below the ITD discrimination threshold, indicating a good match between the acoustical measurements and simulations, when considering the ITD cue only. Spectral deviations between the simulated and measured HRTFs were observed especially at frequencies above 3 kHz. While the head and ear model used for printing was exactly translated to the simulation, some errors might have been introduced by the printing process. The layer thickness and thereby printing accuracy was, as stated in Section 2.3, 0.03 mm and 0.12 mm for the ears and head, respectively. Thus, the printing errors rather unlikely contributed to the noticeable HRTF differences between measurement and simulation. The microphone position, however, might have been more critical. Even though photo documentation of the microphone position was captured for the measurements and used to guide the positioning of the microphone in the simulation, the position of the microphone might have been translated with a deviation in the range of millimeters. Further differences between the HRTFs can also arise due to differences between measuring and simulating HRTFs. For example, the measurements were performed in a semi-anechoic chamber with electro-acoustic equipment involved, whereas the simulations had a PML that absorbed all radiated sound. Further, the printed head and torso simulator was attached to a chair during the measurements, whereas they were perfectly (virtually) placed in the simulations.

5 Conclusions

All three kinds of HRTFs were acquired for a single subject, NH167. Very similar spectral cues were observed between the repeated human measurements and between the human and printed-head measurements. More substantial differences were observed between the measurements and the simulations. The ITDs revealed a good agreement among the three different kinds of HRTFs for the same subject, whereas meaningful differences were observed among listeners. The very similar ITDs are encouraging and it is possible that the simulations are useful for virtual prototyping despite the spectral differences observed between simulations and measurements. Applications like optimization of directionality in HADs, where the spatial dependency of HRTFs is more important than the individual monaural cues, are the focus of our future work. The process of printing a head-and-ear geometry seems to be promising considering both the ITDs and the spectral cues.

References

- [1]. Strutt JW. Our perception of the direction of a source of sound. *Nature*. 1876; 14:32–3.
- [2]. Blauert, J. MIT Press; 1996. Spatial hearing-revised edition: the psychophysics of human sound localization.
- [3]. Brungart DS, Rabinowitz WM. Auditory localization of nearby sources. Head-related transfer functions. *J Acoust Soc Am*. 1999; 106:1465–79. [PubMed: 10489704]

- [4]. Baumgartner R, Majdak P, Laback B. Modeling sound-source localization in sagittal planes for human listeners. *J Acoust Soc Am*. 2014; 136:791–802. [PubMed: 25096113]
- [5]. Gaik W. Combined evaluation of interaural time and intensity differences: Psychoacoustic results and computer modeling. *J Acoust Soc Am*. 1993; 94:98–110. [PubMed: 8354765]
- [6]. Jelfs S, Culling JF, Lavandier M. Revision and validation of a binaural model for speech intelligibility in noise. *Hear Res*. 2011; 275:96–104. [PubMed: 21156201]
- [7]. Bentler RA, Pavlovic CV. Transfer functions and correction factors used in hearing aid evaluation and research. *Ear Hear*. 1989; 10:58–63. [PubMed: 2721830]
- [8]. Bentler R, Pavlovic C. Addendum to “Transfer functions and correction factors used in hearing aid evaluation and research”. *Ear Hear*. 1992; 13:284–6. [PubMed: 1397773]
- [9]. Hartmann WM, Wittenberg A. On the externalization of sound images. *J Acoust Soc Am*. 1996; 99:3678–88. [PubMed: 8655799]
- [10]. Wightman FL, Kistler DJ. Headphone simulation of free-field listening. I: Stimulus synthesis. *J Acoust Soc Am*. 1989; 85:858–67. [PubMed: 2926000]
- [11]. Wightman FL, Kistler DJ. Factors affecting the relative salience of sound localization cues. *Binaural Spat Hear Real Virtual Environ*. 1997; 1:1–23.
- [12]. Majdak P, Balazs P, Laback B. Multiple exponential sweep method for fast measurement of head-related transfer functions. *J Audio Eng Soc*. 2007; 55:623–37.
- [13]. Zotkin DN, Duraiswami R, Grassi E, Gumerov NA. Fast head-related transfer function measurement via reciprocity. *J Acoust Soc Am*. 2006; 120:2202–15. [PubMed: 17069316]
- [14]. Algazi, VR., Duda, RO., Thompson, DM., Avendano, C. The CIPIC HRTF database. *Appl Signal Process to Audio Acoust 2001 IEEE Work*; p. 99-102.
- [15]. Gardner WG, Martin KD. HRTF measurements of a KEMAR. *J Acoust Soc Am*. 1995; 97:3907–8.
- [16]. Evans MJ, Angus JAS, Tew AI. Analyzing head-related transfer function measurements using surface spherical harmonics. *J Acoust Soc Am*. 1998; 104:2400–11.
- [17]. Møller H, Sørensen MF, Hammershøi D, Jensen CB. Head-related transfer functions of human subjects. *J Audio Eng Soc*. 1995; 43:300–21.
- [18]. Fels J, Buthmann P, Vorlander M. Head-related transfer functions of children. *Acta Acust United with Acust*. 2004; 90:918–27.
- [19]. Middlebrooks JC. Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency. *J Acoust Soc Am*. 1999; 106:1493–510. [PubMed: 10489706]
- [20]. Kahana Y, Nelson PA. Boundary element simulations of the transfer function of human heads and baffled pinnae using accurate geometric models. *J Sound Vib*. 2007; 300:552–79.
- [21]. Katz BFG. Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation. *J Acoust Soc Am*. 2001; 110:2440–8. [PubMed: 11757933]
- [22]. Mokhtari P, Takemoto H, Nishimura R, Kato H. *Audio Eng Soc Convention*. 2007; 123
- [23]. Jin CT, Guillon P, Epain N, Zolfaghari R, van Schaik A, Tew A, et al. Creating the sydney york morphological and acoustic recordings of ears database. *IEEE Trans Multimedia*. 2014; 16:37–46.
- [24]. Kreuzer W, Majdak P, Chen Z. Fast multipole boundary element method to calculate head-related transfer functions for a wide frequency range. *J Acoust Soc Am*. 2009; 126:1280–90. [PubMed: 19739742]
- [25]. Luximon Y, Ball R, Justice L. The 3D Chinese head and face modeling. *Comput Des*. 2012; 44:40–7.
- [26]. Baek SY, Lee K. Parametric human body shape modeling framework for human-centered product design. *Comput Des*. 2012; 44:56–67.
- [27]. Rodrigo AS, Goonetilleke RS, Witana CP. Model based foot shape classification using 2D foot outlines. *Comput Des*. 2012; 44:48–55.
- [28]. McCreery RW, Venediktov RA, Coleman JJ, Leech HM. An evidence-based systematic review of directional microphones and digital noise reduction hearing aids in school-age children with hearing loss. *Am J Audiol*. 2012; 21:295–312. [PubMed: 22858614]
- [29]. Valente, M. *Hearing aids: standards, options, and limitations*. Thieme; 2002.

- [30]. Reichinger, A., Majdak, P., Sablatnig, R., Maierhofer, S. Evaluation of methods for optical 3-D scanning of human pinnas. 3D Vision-3DV 2013 International Conference on IEEE; 2013. p. 390-397.
- [31]. Harder S, Paulsen RR, Larsen M, Laugesen S. A three dimensional children head database for acoustical research and development. Proc Meet Acoust. 2013; 19:050013.
- [32]. Paulsen RR, Bærentzen JA, Larsen R. Markov random field surface reconstruction. IEEE Trans Vis Comput Graphics. 2010; 16:636-46.
- [33]. Foley, JD., Van Dam, A., Feiner, SK., Hughes, JF., Phillips, RL. Introduction to computer graphics. Vol. 55. Addison-Wesley Reading; 1994.
- [34]. Kohnke, P. ANSYS mechanical APDL theory reference. 2012.
- [35]. Morse, PM. Theoretical acoustics. Princeton University Press; 1986.
- [36]. Jacobsen, F., Juhl, PM. Fundamentals of general linear acoustics. John Wiley & Sons; 2013.
- [37]. Senova MA, McAnally KI, Martin RL. Localization of virtual sound as a function of head-related impulse response duration. J Audio Eng Soc. 2002; 50:57-66.
- [38]. Gumerov NA, O'Donovan AE, Duraiswami R, Zotkin DN. Computation of the head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation. J Acoust Soc Am. 2010; 127:370-86. [PubMed: 20058984]
- [39]. Baumgartner, R., Majdak, P., Laback, B. Assessment of sagittal-plane sound localization performance in spatial-audio applications. Technol binaural list. Springer; 2013. p. 93-119.
- [40]. Rébillat M, Benichoux V, Otani M, Keriven R, Brette R. Estimation of the low-frequency components of the head-related transfer functions of animals from photographs. J Acoust Soc Am. 2014; 135:2534-44. [PubMed: 24815237]
- [41]. Katz BFG, Noisternig M. A comparative study of interaural time delay estimation methods. J Acoust Soc Am. 2014; 135:3530-40. [PubMed: 24907816]
- [42]. Harder S, Paulsen R, Larsen M, Laugesen S, Mihocic M, Majdak P. Reliability in measuring head related transfer functions of hearing aids. Acta Acust United Acust. 2015; 101:1064-6.
- [43]. Majdak P, Masiero B, Fels J. Sound localization in individualized and non-individualized crosstalk cancellation systems. J Acoust Soc Am. 2013; 133:2055-68. [PubMed: 23556576]
- [44]. Domnitz R. The interaural time jnd as a simultaneous function of interaural time and interaural amplitude. J Acoust Soc Am. 1973; 53:1549-52. [PubMed: 4719250]

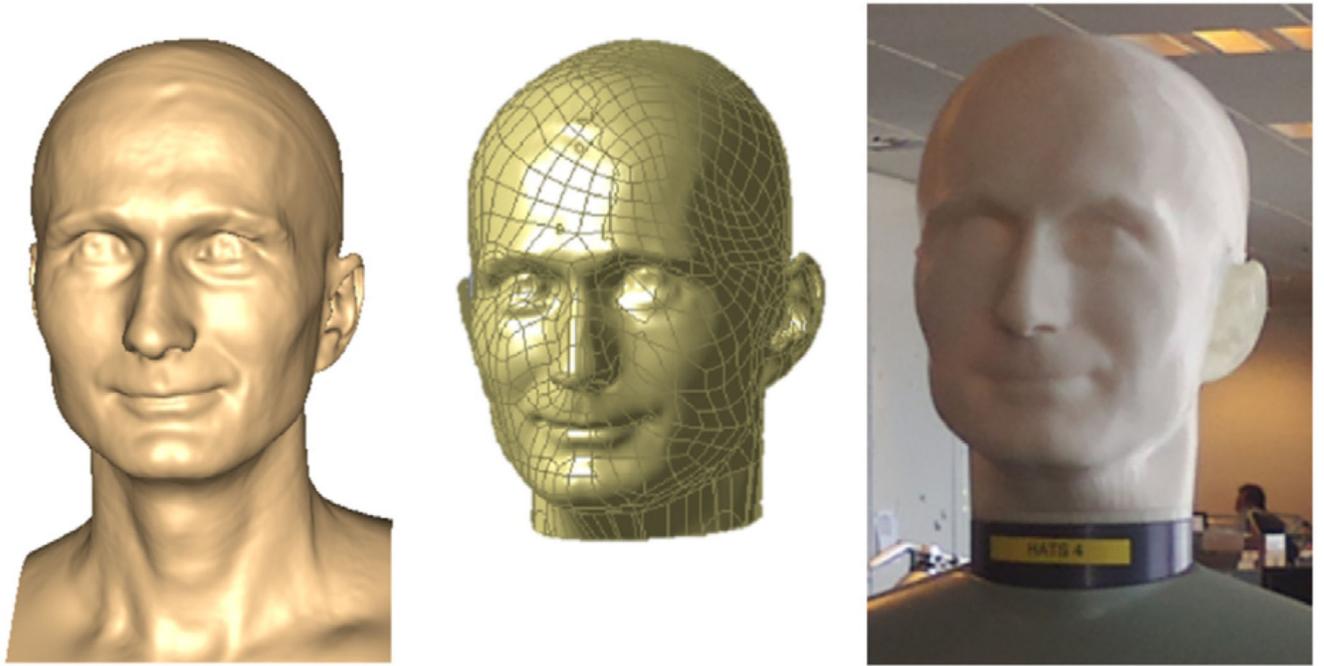


Fig. 1. Head-and-ear model. (a) Acquired geometry (b) Model used for printing (c) Printed model mounted on a torso simulator.

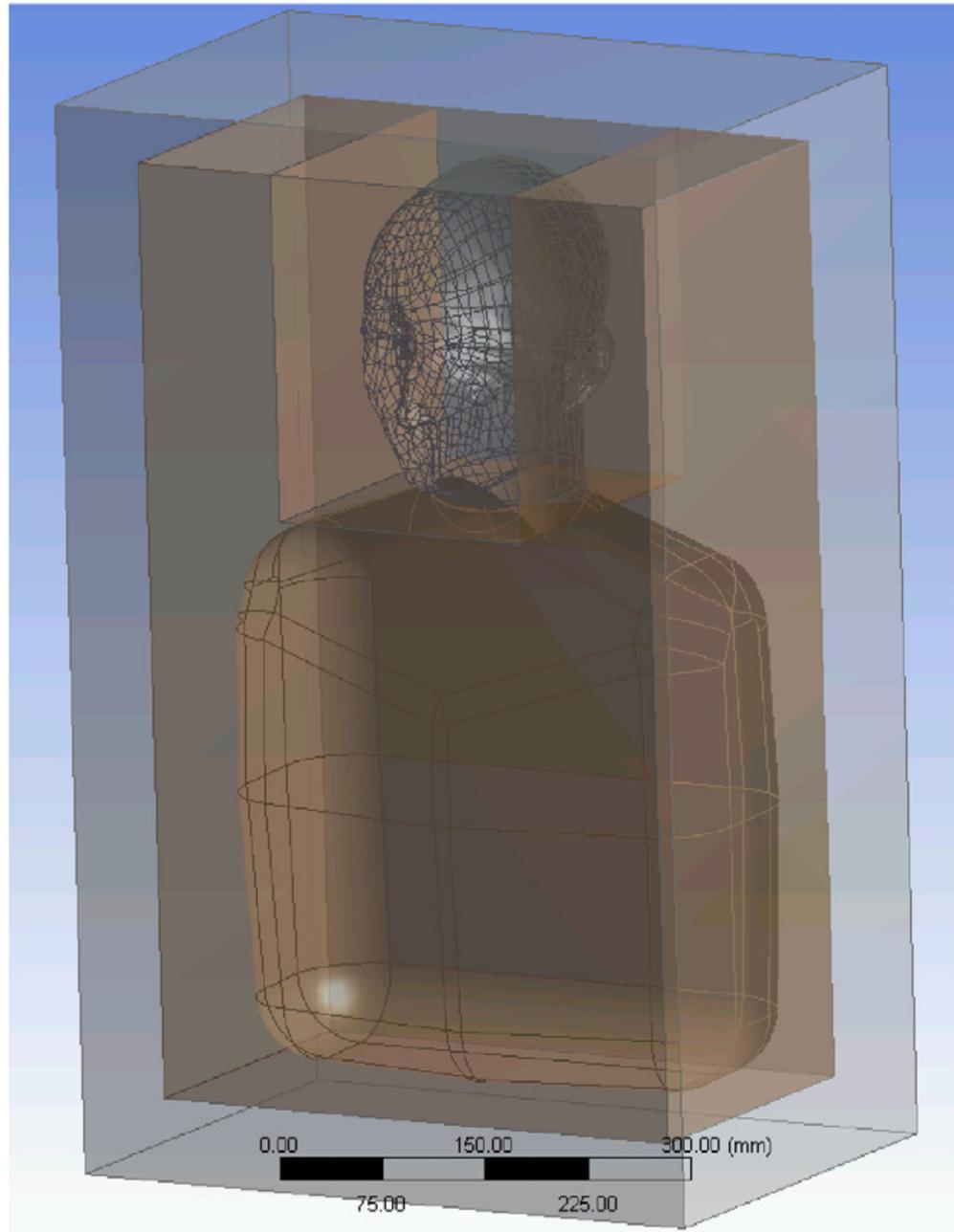


Fig. 2. Geometry layers used for HRTF simulation. From inside to out: empty space shaped as the head and torso, inner air box, and perfectly matched layer.

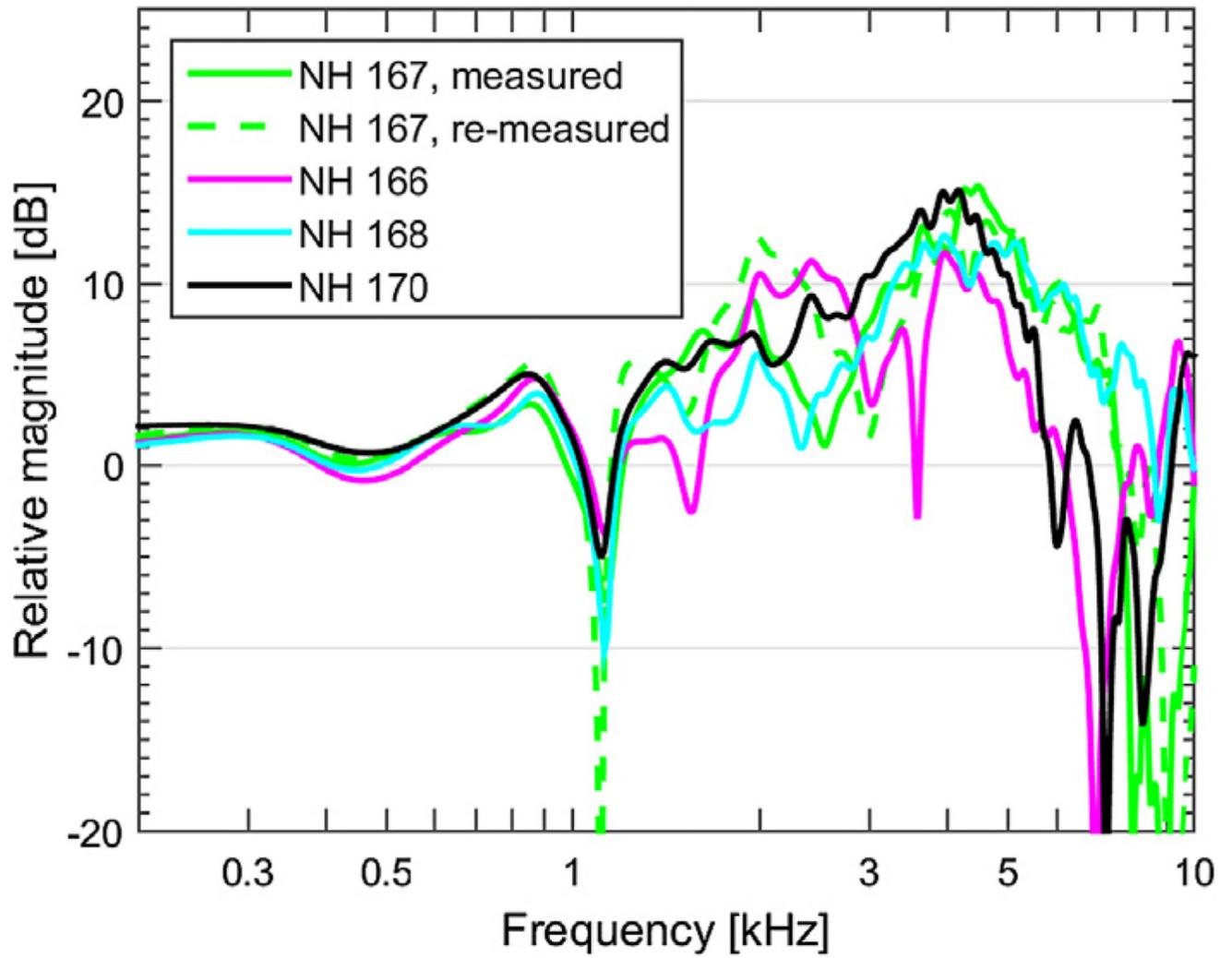


Fig. 3.
Right-ear HRTF magnitudes for the frontal position in the horizontal plane.

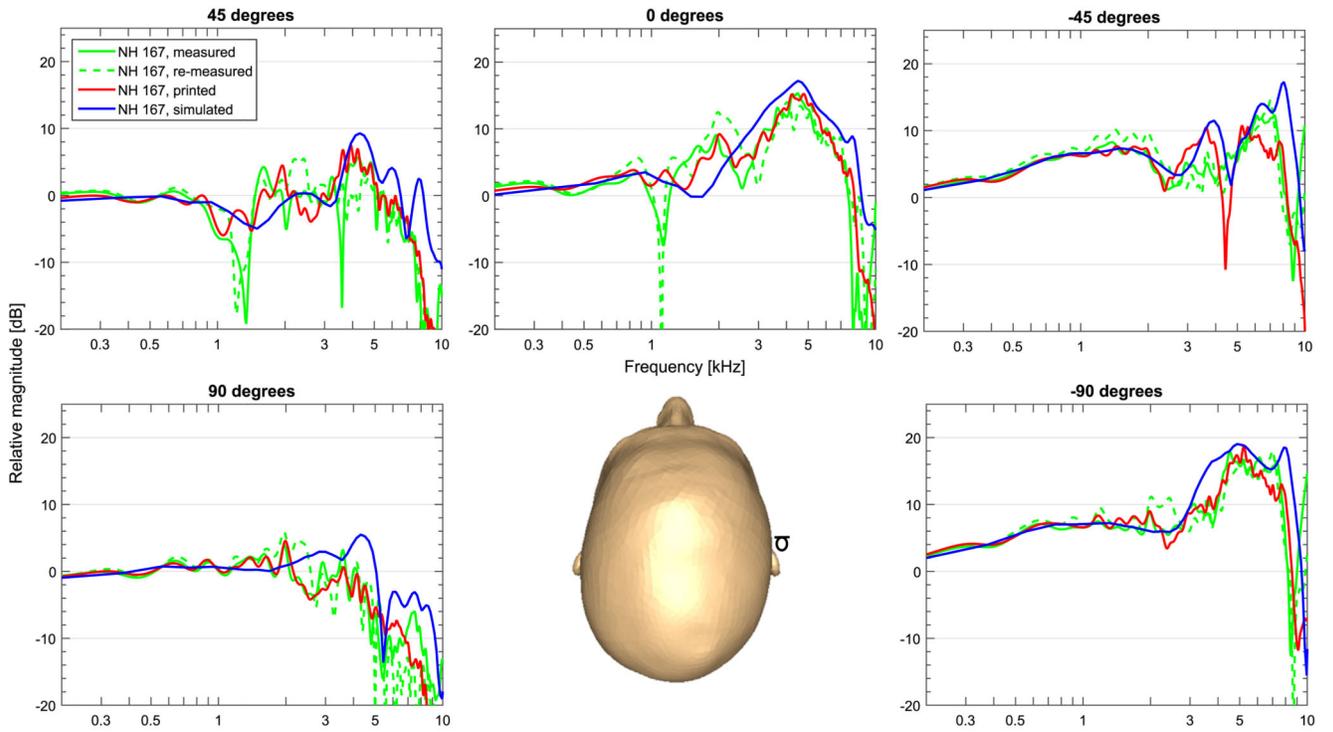


Fig. 4. Right-ear HRTF magnitude spectra in the horizontal plane (elevation angle of 0°) and the azimuth angle of 90, 45, 0, -45 and -90° .

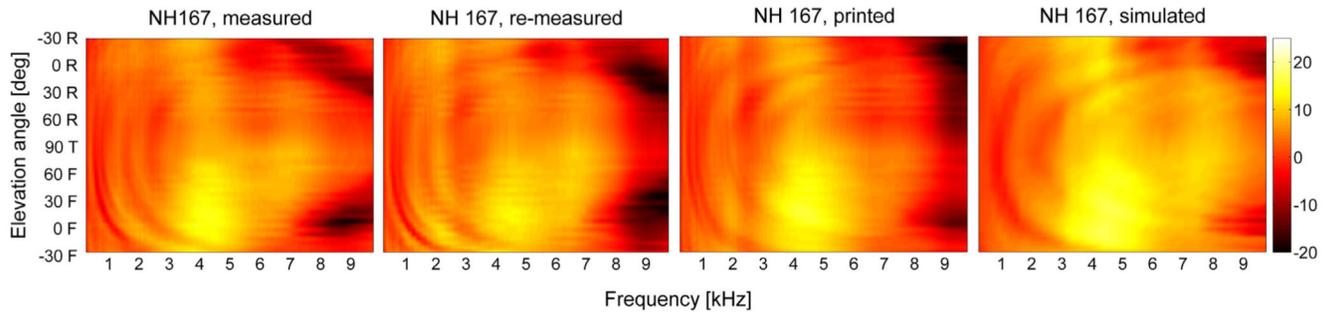


Fig. 5. HRTF magnitude spectra in the median plane (as function of elevation angle). Color encodes the relative magnitude in dB. F, T, R describe directions in front, top, and rear of the listener. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

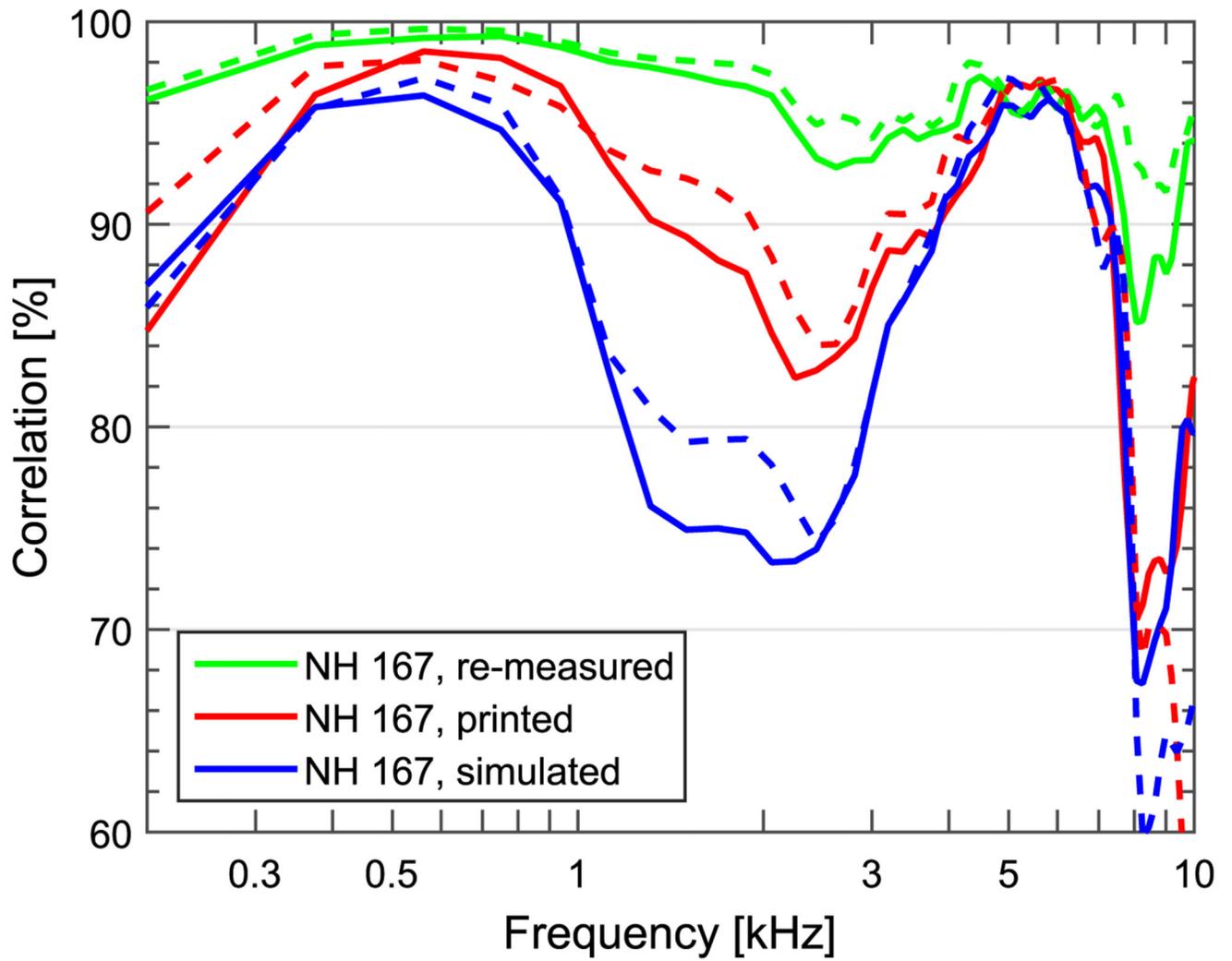


Fig. 6. Spatial correlation between HRTF magnitudes for NH167 and NH167 re-measured, NH167 printed and NH167 simulated, respectively. Solid line: right ear, dashed line: left ear.

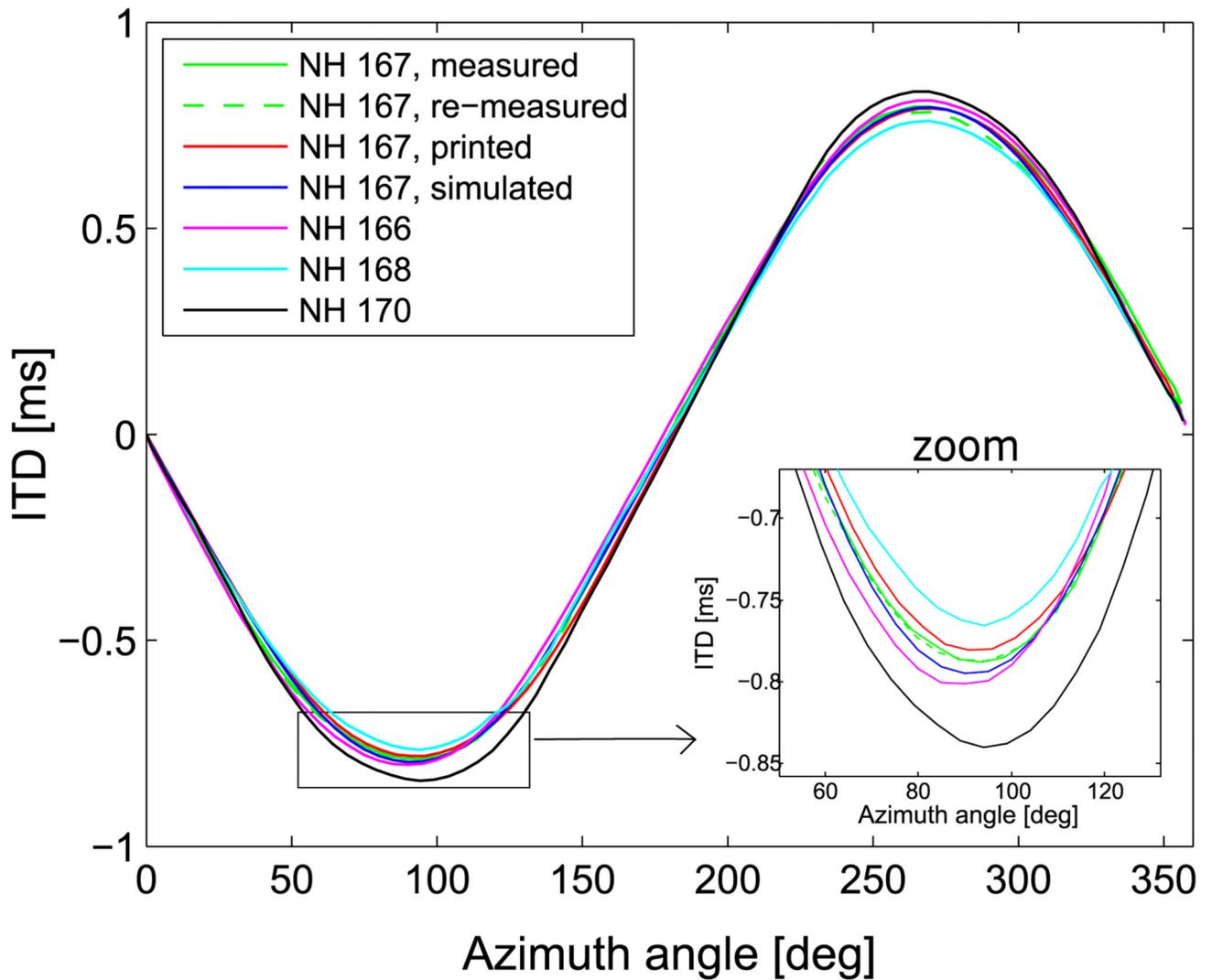


Fig. 7. ITDs in the horizontal plane (as function of azimuth angle). The right panel shows a zoom around the most lateral directions shown in the left panel.

Table 1

Spectral distortions (in dB) of re-measured human (R), printed (P), and simulated (S) HRTFs for five positions in the horizontal plane (90°, 45°, 0°, -45°, and -90°) and the average with ± 1 standard deviations across all tested positions (Mean). The basis for comparisons was the human HRTF, NH167, right (R) and left (L) ear, respectively.

	90°	45°	0°	-45°	-90°	Mean
R-R	2.9	2.6	2.6	4.7	2.8	5.2 \pm 3.7
P-R	2.1	2.6	2.7	1.6	2.2	4.8 \pm 2.7
S-R	11.9	13.3	14	16.2	11.8	14.3 \pm 3.4
R-L	1.5	3.7	3.3	2.6	2.3	4.6 \pm 2.9
P-L	2.0	4.4	1.8	4.2	1.8	5.4 \pm 3.2
S-L	13.9	16.6	14.9	12.6	13.4	14.4 \pm 3.1