Published in final edited form as:

J Acoust Soc Am. 2015 July ; 138(1): 208. doi:10.1121/1.4922518.

Numerical calculation of listener-specific head-related transfer functions and sound localization: Microphone model and mesh discretization

Harald Ziegelwanger^{a)}, Piotr Majdak, and Wolfgang Kreuzer

Acoustics Research Institute, Austrian Academy of Sciences, Wohllebengasse 12-14, A-1040 Vienna, Austria

Abstract

Head-related transfer functions (HRTFs) can be numerically calculated by applying the boundary element method on the geometry of a listener's head and pinnae. The calculation results are defined by geometrical, numerical, and acoustical parameters like the microphone used in acoustic measurements. The scope of this study was to estimate requirements on the *size* and *position* of the microphone model and on the *discretization* of the boundary geometry as triangular polygon mesh for accurate sound localization. The evaluation involved the analysis of localization errors predicted by a sagittal-plane localization model, the comparison of equivalent head radii estimated by a time-of-arrival model, and the analysis of actual localization errors obtained in a sound-localization experiment. While the average edge length (AEL) of the mesh had a negligible effect on localization performance in the lateral dimension, the localization performance in sagittal planes, however, degraded for larger AELs with the geometrical error as dominant factor. A microphone position at an arbitrary position at the entrance of the ear canal, a microphone size of 1 mm radius, and a mesh with 1 mm AEL yielded a localization performance similar to or better than observed with acoustically measured HRTFs.

I. INTRODUCTION

Head-related transfer functions (HRTFs) describe the direction-dependent filtering of the incoming sound at the entrance of the ear canal due to sound reflection, shadowing, and diffraction caused by the head, torso, and the pinnae (Møller *et al.*, 1995; Wightman and Kistler, 1989). HRTFs contain listener-specific directional acoustic features, which are analyzed by the auditory system for sound-source localization. Particularly, two different types of features are relevant for the localization process. First, interaural differences, like interaural time differences (ITDs) and interaural level differences (ILDs), are relevant for localization in the lateral dimension (Macpherson and Middlebrooks, 2002). These differences are mostly defined by a listener's head dimensions due to sound scattering and shadowing (Algazi *et al.*, 2001b; Kuhn, 1977; Ziegelwanger and Majdak, 2014). Second, spectral features are relevant for vertical localization and front-back discrimination of sound-source positions. These spectral features are mostly¹ characterized by the fine

^{a)}harald.ziegelwanger@oeaw.ac.at.

geometrical details of a listener's pinnae (Bronkhorst, 1995; Hebrank and Wright, 2005; Middlebrooks, 1999a).

HRTFs are usually acquired in an acoustical measurement procedure (Møller *et al.*, 1995). Acoustically measured HRTFs yield localization performance similar to that obtained in free-field listening (Bronkhorst, 1995; Middlebrooks, 1999a). However, the acoustic measurement is a difficult and resource-demanding procedure. The measurement procedure might be uncomfortable for the listener because of the insertion of small in-ear-microphones and the requirement to sit still during the whole measurement.

HRTFs can also be calculated based on physical and geometrical models. While formerly, various physical models were developed to calculate temporal and spectral features analytically for simple geometrical objects and shapes (Kuhn, 1977; Lopez-Poveda and Meddis, 1996), in recent years, various techniques were used to numerically calculate HRTFs based on the head geometry and the complex shape of the pinna. HRTFs were numerically calculated using the *boundary element method (BEM)* (Katz, 2001a), the *ultra weak variational formulation* (Huttunen *et al.*, 2007), and the *finite-difference time-domain method* (Takemoto *et al.*, 2012). In general, these techniques aim at solving an acoustic scattering problem by numerically solving the Helmholtz equation. In the numerical calculation of HRTFs the scatterer is defined by a listener's geometry. Listeners' geometries were captured by laser scanning (Katz, 2001a; Kreuzer *et al.*, 2009), by magnetic resonance imaging (MRI, Jin *et al.*, 2014), by computerized tomography (CT, Mey *et al.*, 2008), and by photogrammetric reconstruction (Rébillat *et al.*, 2014; Reichinger *et al.*, 2013). When the listener's geometry is known, the numerical HRTF calculation is done by imitating the acoustic measurement procedure; a *virtual* sound source is placed at different positions around the discretized representation of the head and pinnae geometry, i.e., the three-dimensional (3D) polygon mesh, and the HRTFs are obtained from the ratio of the calculated sound pressure at the ear canal and the calculated sound pressure at the origin of the head without the presence of the head.

Much research effort has been done on the numerical calculation of HRTFs by the use of the BEM (Gumerov *et al.*, 2010; Kahana and Nelson, 2006; Katz, 2001a,b; Kreuzer *et al.*, 2009). The computational effort of the BEM increases with increasing frequency and, thus, the frequency range is limited by computer memory and computational power. While the first calculations (Katz, 2001a,b) were limited to frequencies below 6 kHz, the application of the multilevel fast-multipole method (ML-FMM) made the BEM feasible for the entire audible frequency range (Gumerov *et al.*, 2010; Kreuzer *et al.*, 2009).

HRTFs were numerically calculated for animals (Mey *et al.*, 2008; Rébillat *et al.*, 2014), for artificial heads (Greff and Katz, 2007; Gumerov *et al.*, 2010; Kahana and Nelson, 2006) and for human heads (Jin *et al.*, 2014; Katz, 2001a; Kreuzer *et al.*, 2009). For animal heads like those of a bat and for artificial heads like those of the KEMAR mannequin, the calculated HRTFs showed a good correspondence to acoustically measured HRTFs (Gumerov *et al.*,

¹Some spectral features below 5 kHz are caused by reflections at the torso (Algazi *et al.*, 2001b), however, these low frequency components do not contribute much to the localization of broadband signals (Bronkhorst, 1995).

2010; Mey *et al.*, 2008). Results for human heads are twofold. HRTFs calculated for meshes obtained from laser scans (Katz 2001a; Kreuzer *et al.*, 2009) showed obvious visible differences in the amplitude spectra when compared to measured HRTFs. HRTFs calculated for meshes obtained by MRI showed spectral correlation coefficients above 0.72 when compared to acoustically measured HRTFs (Jin *et al.*, 2014).

In general, quantifying the similarity of different HRTF sets and, thus, validating numerically calculated HRTFs is challenging. HRTFs can differ between measurements even for the same listener (Majdak *et al.*, 2013) or artificial heads (Andreopoulou *et al.*, 2015), thus, differences between the amplitude spectra of acoustically measured and numerically calculated HRTFs are unavoidable. While potential reasons for such differences can be the position of the microphone (Greff and Katz, 2007), the perceptual relevance, however, is unclear as a comparison of HRTF amplitude spectra, even when performed on a logarithmic frequency scale, is not well correlated with human perception (Middlebrooks, 1999a). Acoustically measured HRTFs were validated in sound-localization experiments (Bronkhorst, 1995; Middlebrooks, 1999a) by comparing listener-specific localization-performance parameters for real and virtual sound sources. Recently, various models have been developed to estimate such parameters (Baumgartner *et al.*, 2013; Ziegelwanger and Majdak, 2014) and they can be used to quantify the similarity of HRTF sets on a perceptual level (Ziegelwanger *et al.*, 2013).

While acoustically measured HRTFs can be affected by measurement noise, room reflections, and post-processing artifacts, results of the numerical HRTF calculation depend on geometrical, numerical, and acoustical parameters (Treeby and Pan, 2009). First, numerically calculated HRTFs depend on the accuracy and resolution of the geometric representation of the listener's head and pinnae (Kreuzer *et al.*, 2009). Problems originate from inaccuracies of the geometry acquisition method and of the mesh processing that is required for the numerical calculations. We refer to these problems as *geometrical errors*. Second, if the sound field on the boundary is not adequately approximated then the HRTF calculation suffers from large *numerical errors* (Marburg, 2002). Note that both geometrical and numerical errors are generally unavoidable because of the discretization of the head and pinnae surface as a 3D polygon mesh. Third, the parameters like the *virtual microphone*, i.e., the model simulating position and size of the microphone used in acoustic HRTF measurements (Greff and Katz, 2007; Kreuzer *et al.*, 2009; Mey *et al.*, 2008), and boundary conditions, e.g., admittance of skin and hair (Treeby *et al.*, 2007a), may result in *acoustical errors*. While results from Katz (2001b) and Treeby *et al.* (2007b) indicate a negligible effect of the boundary admittance on amplitude spectra in the median plane, requirements on the remaining parameters for numerical HRTF calculation and their impact on sound-localization performance, especially on *vertical* sound-localization performance, have not been clarified yet.

The aim of this study was to systematically address these issues, to separate the effects of the different error types, and to identify the relevant parameters required for *perceptually* valid numerically calculated HRTFs in terms of a good sound-localization performance. The effect of the virtual microphone and the effect of the mesh resolution were systematically evaluated. The separation of the different error types was not trivial. On the one hand,

position and size of the virtual microphone model parameters are limited by the resolution of the mesh. On the other hand, the results of the numerical HRTF calculation, when investigating the effect of geometrical and numerical errors, depend on the chosen parameters for the virtual microphone model. Thus, first the requirements for the virtual microphone model were evaluated for the best available mesh resolution and accuracy. Then, effects of geometrical and numerical errors were evaluated for a virtual microphone fulfilling these requirements. Both effects were evaluated using model predictions. For the most relevant conditions, numerically calculated HRTFs were finally verified in a sound-localization experiment.

II. GENERAL METHODS

A. Subjects

Three subjects participated in this study (NH5, NH130, and NH131 from the ARI database²). The subjects had different pinna sizes covering a wide range of sizes usually found in human listeners. Figure 1 shows the geometry of left and right ears of all subjects (for the geometry acquisition see Sec. II D). Table I shows the corresponding anthropometric data³ of the subjects together with those of KEMAR (G.R.A.S., Denmark) with normal pinna (DB-60). Note the different pinna sizes: The ears of NH130 are large compared to those of KEMAR and NH5. The ears of NH131 were the smallest and flattest in our study. All subjects were male and had different haircuts, particularly NH5 had an almost bald head, NH130 had shoulder long hair, and NH131 had a short haircut. The subjects had absolute hearing thresholds within a 20-dB range of the average normal-hearing population in the frequency range between 0.125 and 12.5 kHz.

B. Acoustic HRTF measurement

Acoustic HRTF measurement was done in a semi-anechoic chamber (Majdak *et al.*, 2010). HRTFs for 1550 directions were measured by placing in-ear-microphones in the subject's ear canals, i.e., blocked ear-canal method (Møller *et al.*, 1995). The horizontal and vertical range was 360° and -30° to 80°, respectively. The horizontal resolution was 2.5° and 5° for directions inside and outside of $\pm 45^\circ$, respectively. The vertical resolution was 5°. The measured transfer functions were equalized with the equipment's transfer functions by spectral division. Then, directional transfer functions (DTFs) were calculated (Middlebrooks, 1999b), i.e., the direction independent part of the HRTF was logarithmically subtracted from the HRTFs. The DTFs were temporally windowed with an asymmetric Tukey window to a length of 256 samples (5.33 ms at a sampling rate of 48 kHz).

C. Numerical HRTF calculation

HRTFs were calculated with the BEM using the collocation method with constant elements coupled with the ML-FMM (Chen *et al.*, 2008; Kreuzer *et al.*, 2009).⁴ In general, the BEM numerically solves the boundary integral formulation of the Helmholtz equation, i.e., the

²HRTFs and anthropometric data available from <http://www.kfs.oeaw.ac.at/hrtf/> (Last viewed June 24, 2014).

³Description available from http://www.kfs.oeaw.ac.at/research/experimental_audiology/hrtf/readme.pdf (Last viewed June 24, 2014).

⁴The software package Mesh2HRTF was used, available from <http://mesh2hrtf.sourceforge.net/> (Last viewed June 12, 2015).

boundary integral equation (BIE), in the 3D space (Sauter and Schwab, 2010). The BEM requires a 3D polygon mesh, i.e., a discrete representation of the boundary surface geometry which was the head and pinna in our case. A polygon mesh is usually defined by a set of nodes and a set of triangular or quadrilateral planar elements with these nodes as vertices. In our calculations triangular elements were used.

The sound pressure at the boundary surface is calculated by solving the BIE for a particular sound-source configuration, i.e., an incoming plane wave, a point source, a set of point sources, or radiating mesh elements. In our study, the calculation would take a considerable amount of time because many sound-source configurations had to be considered. In order to keep the computational effort feasible, two different approaches were used to simulate the sound-source configurations.

In the *direct* approach, a point source representing a virtual loudspeaker is placed at an arbitrary position outside the head, the sound pressure on the whole boundary surface is calculated, and the sound pressure at the receiver area, i.e., a mesh element or a set of mesh elements, representing the virtual microphone is evaluated. The BIE has to be solved separately for each virtual loudspeaker position, rendering the direct approach as inefficient for the simulation of many loudspeaker positions. However, many virtual microphone models can be evaluated by averaging the sound pressure on contiguous mesh elements.

In the *reciprocal* approach, the roles of the loudspeaker and microphone are interchanged (Morse and Ingard, 1986) by defining the virtual microphone as a radiating element at the blocked ear canal and evaluating the sound pressure at various positions around the head representing the virtual loudspeakers (Jin *et al.*, 2014; Kreuzer *et al.*, 2009). In contrast to the direct approach, only a single parameter setting for the virtual microphone can be evaluated at once, rendering the reciprocal approach as inefficient for the simulation of many microphone-parameter settings. However, many virtual loudspeaker positions can be evaluated at once.

HRTFs were calculated for 200 frequencies, which were linearly spaced between 0.1 and 20 kHz. The admittance of the boundary surface was set to zero, i.e., the head and pinnae were modeled fully reflective. The resulting HRTFs were inverse Fourier-transformed in order to obtain the head-related impulse responses (HRIRs), which were further re-sampled to a sampling rate of 48 kHz. Finally, DTFs were calculated following the same procedure as for the acoustically measured HRTFs.

In total, HRTFs for 78 conditions were numerically calculated (see details later). All calculations were done on a Linux cluster consisting of eight machines with Intel i7–3820 processors running at 3.6 GHz and 64 gigabyte of RAM each. Given the duration of a single calculation in the order of several hours and the huge number of calculations required, we limited the number of mesh elements to approximately 100 000 in order to finish the calculations in reasonable time. The calculation time for one particular condition was in the order of hours. In total, the duration of the pure calculation process for all conditions in this study was in the order of weeks.

D. Geometry

In the process of geometry acquisition for HRTF calculations the fine details of the pinna geometry, particularly its folds, undercuts, and deep cavities, were of special interest. However, these geometrical details posed technical difficulties during the geometry acquisition process. For instance, some parts of some ears are not always visible from the outside, and those details were difficult to access by surface scanning methods like laser scanners. In contrast, the shape of human head is usually smooth and easy to scan. Thus, two different methods were used for the head and the pinna.

The head geometry was captured by a laser scanner (ZScanner 700CX,⁵ ZCorp). A built-in optical tracking system of this device ensured a globally consistent and accurate scan, even when the subject moved the head during the scanning process. However, the laser scanner was not able to scan hairy surfaces. Thus, the subjects wore a custom tailored cap with holes for the ears (Ziegelwanger *et al.*, 2013). In a post-processing procedure, a 3D triangular polygon mesh was extracted from a volumetric representation with approximately 0.5 mm voxel size. In the final mesh the neck and the ear canal were cut and closed. The sharp cutting edge at the neck was smoothed to avoid edge effects in the BEM.

The pinna geometry was captured using a high-energy industrial computer-tomography scanner (v|tome|x c,⁶ GE Measurement & Control). Such a scanner is not limited in the amount of radiation by medical concerns and is capable to deliver a detailed representation of the pinna geometry. To prevent the subjects from the harmful x-ray radiation, silicone impressions of subjects' pinnae and surrounding area were made (viscosity: 2200 MPa-s, shore hardness: A 22, for more details on the silicon impressions see Ziegelwanger *et al.*, 2013). The impressions were then scanned using 1000 x-ray images spaced in 0.36° intervals (tube settings: 120 kV, 260 μ A, average of three images with 333-ms exposure time each). The resulting volumetric mesh consisted of approximately $2000 \times 2000 \times 650$ voxels and the voxel size was approximately 50 μ m. With the rotational resolution of the x-ray scanner, the accuracy at the outer edge of the largest pinna was 0.21 mm. For all other regions on the pinna, the accuracy of the x-ray scanner was smaller than 0.21 mm. The iso-surfaces separating air and silicone were extracted using automatic thresholding, i.e., a proprietary version of the marching cubes algorithm (Lorensen and Cline, 1987). The effective accuracy of this algorithm was in the range of two to four voxels, i.e., 0.1–0.2 mm.

The pinnae meshes from the CT scans were manually aligned with the pinnae meshes from the laser scans via tie points (Manual Registration, Geomagic⁷). Then the alignment was automatically refined (Best Fit Alignment, Geomagic). The pinnae meshes from the laser scan were removed and the seam-region, i.e., the surrounding area of the pinnae on the head surface, was re-sculpted, smoothed (Voxel-Sculpting, 3D-Coati) in a resolution of below 0.5 mm voxel-size, and water-tightened. Note that the pinnae meshes from the CT scans were not manipulated in this procedure. The final reference meshes consisted of approximately

⁵Specifications available from http://www.zcorp.com/documents/491_8749-ZScanner700_CX_SpecSheet.pdf (Last viewed June 24, 2014).

⁶Specifications available from <http://www.ge-mcs.com/en/radiography-xray/ct-computed-tomography/v-tome-x-c.html> (Last viewed June 24, 2014).

⁷Geomagic, available from <http://www.geomagic.com/> (Last viewed June 24, 2014).

2.5×10^6 elements and had an average edge length (AEL) and a maximum edge length of 0.47 and 0.77 mm, respectively (for more details on the geometry acquisition, see Reichinger *et al.*, 2013; Ziegelwanger *et al.*, 2013).

The meshes were aligned to match the coordinate system of the loudspeaker grid used in the acoustic HRTF measurement setup. The *auricular points*, i.e., the craniometric points at the center of the ear canal, were placed on the y axis with the left auricular point on the positive half-axis. The midpoint between the auricular points was aligned with the origin of the coordinate system. The match of the head rotation about the interaural axis, i.e., the *interaural rotation*, was more problematic because distinct markers indicating the exact rotation of the interaural horizontal plane, i.e., the horizontal plane with elevation angle of zero, were not available.⁸ This problem has been recently addressed by rotating the coordinate system after the HRTF calculations such that the correlation between the acoustically measured and numerically calculated HRTFs was maximized (Jin *et al.*, 2014). We assumed that a mismatch between the interaural horizontal plane of the acoustic measurements and the numeric calculations would result in a simple polar bias in the localization responses, which can be *post hoc* corrected (see the discussion later). Thus, we approximately aligned the interaural rotation to match photographs from the acoustic measurement, where the nasal point, i.e., the tip of the nose, seemed to point towards the interaural horizontal plane. In our meshes, the nasal point was placed on the positive half-axis of the x axis and the plane connecting the auricular points and the nasal point defined the interaural horizontal plane.

The reference meshes resulted in a detailed representation of the geometry (AEL < 0.5 mm), however, they had too many elements for feasible numerical calculations (more than 2.5×10^6 elements). For accurate BEM results, the length of edges in the mesh should be a fraction of the smallest simulated wavelength. In an empirical study on simple geometric objects, Marburg (2002) showed that the relative numerical error of the BEM is below 15% when the edge length is a sixth of the smallest simulated wavelength. We refer to the corresponding frequency as the critical frequency, f_{crit} . For a critical frequency of 20 kHz, the criterion from Marburg (2002) leads to an edge length of 2.7 mm. However, a simple increase of the AEL from 0.5 to 2.7 mm resulted in clearly visible geometrical deformations of the pinna. Since our goal was to systematically investigate the AEL required for perceptually valid HRTFs, the finest mesh in our conditions had to include all pinna details and to accurately represent the geometry.

In order to determine the largest possible AEL still accurately representing the geometrical details of a pinna, we used the Hausdorff distance (Cignoni *et al.*, 1998) implemented in *Meshlab*.⁹ The Hausdorff distance is a metric describing the deviation of parts of two meshes. We calculated the Hausdorff distance between the reference geometry and re-meshed geometries with AELs ranging from 0.5 to 20 mm. Figures 2(c), 2(d), and 2(e) show the left ear of NH5 with the Hausdorff distance for meshes with AELs of 1, 2.7, and 4 mm,

⁸Since a subject's head was not fixed for the acoustic measurement, the head alignment to the loudspeaker grid was not accurately controllable. Subjects were instructed to sit comfortable and to look at the loudspeaker located at 0° azimuth and 0° elevation. A subject's head was tracked during the measurement and a rotation of $\pm 2^\circ$ was allowed.

⁹Meshlab, available from <http://meshlab.sourceforge.net> (Last viewed June 24, 2014).

respectively (darker colors indicate larger Hausdorff distances). Figure 2(a) shows the maximum Hausdorff distance evaluated for all meshes of all subjects as a function of the AEL. The solid line represents a sigmoid fit to these data. For the AELs of 0.5 and 1 mm, the maximum Hausdorff distances were in the range of 0.5 mm. Beyond the AEL of 2 mm the maximum Hausdorff distance increased rapidly, with a distance of 2.25 mm for the AEL of 2.7 mm. Note that distances larger than the AEL indicate an inaccurate geometry representation due to the discretization as compared to the reference mesh. In order to find the largest AEL still representing the geometry accurately, we evaluated the normalized Hausdorff distance, i.e., the maximum Hausdorff distance divided by the corresponding AEL. Figure 2(b) shows the normalized Hausdorff distance as a function of the AEL, showing a clear minimum around the AEL of 1 mm. Thus, when compared to the reference geometry, it seems like meshes with an AEL of 1 mm would still accurately represent the pinna geometry and can be used as the finest meshes for further calculations.

Based on these findings, a mesh was created for each subject. The meshes had different resolutions inside and outside the pinna region, similar to the method used in Katz (2001a). The AEL was 1 mm in the pinna region and 2.5 mm at the rest of the head. This *1-mm mesh* satisfied both, the global numerical recommendation of six elements per wavelength on the head and the local geometrical accuracy requirement of 1 mm AEL on the pinnae. This mesh was obtained by applying three processing steps. First, the whole reference mesh was re-meshed to an AEL of 1 mm. Then, everything except the pinna region was re-meshed to an AEL of 2.5 mm with a transition region of four polygon rows and an AEL of 1.5 mm. Finally, the transition region was additionally relaxed to reduce degenerated non-equilateral triangles at the region seams. The 1-mm meshes consisted of approximately 110 000 elements and had the critical frequency of 24 kHz.

As a rough check for the methods described so far, we calculated HRTFs following the methods from Kreuzer *et al.* (2009). Spatial correlation coefficients, i.e., the correlation of acoustically measured and numerically calculated HRTFs over all directions as a function of the frequency, were evaluated as described in Jin *et al.* (2014). Figure 3 shows the spatial correlation coefficient for the left and the right pinnae of our subjects and gray areas indicating two ranges of the coefficients observed by Jin *et al.* (2014). The light gray area shows the range of the coefficients of all 61 subjects from Jin *et al.* (2014). The dark gray area shows the range given by the 20 listeners showing the best matches from Jin *et al.* (2014). The spatial correlation coefficients of our three subjects seem to be well-covered by the range from Jin *et al.* (2014) indicating promising results on the level of spectral comparison.

III. EFFECT OF THE VIRTUAL MICROPHONE

Different microphone configurations have been used to acoustically measure HRTFs, e.g., probe microphones were used for creating the CIPIC database (Algazi *et al.*, 2001c), the blocked-ear-canal method in combination with probe microphones was used in Møller *et al.* (1995), and the blocked-ear-canal method with electret microphones was used in Majdak *et al.* (2010). In the numerical calculation of HRTFs, these microphones are virtually represented by a model, i.e., a virtual microphone (Jin *et al.*, 2014; Kreuzer *et al.*, 2009).

Interestingly, the particular position of the virtual microphone has an effect on the amplitude spectra of calculated HRTFs because of interferences and, thus, on features like peaks and notches (Kreuzer *et al.*, 2009). These spectral features are mostly relevant for sagittal-plane sound-source localization, thus, the parameters of the virtual microphone may have a perceptual impact. On the one hand, one might argue that the position and the size of the actual microphone must be also represented in the numeric calculations. On the other hand, the HRTFs from various HRTF databases, all using different microphones, seem to be able to provide similar localization performance. Thus, it might be that the impact of the microphone is perceptually not significant. Hence, in this section, we aim at estimating the requirements on the virtual microphone in numeric calculations with respect to sound localization.

Acoustically measured HRTFs have been validated in terms of listener-specific localization performance, e.g., HRTFs measured with probe microphones were validated in Bronkhorst (1995) and HRTFs measured with electret microphones were validated in Majdak *et al.* (2010). Thus, in this study the acoustically measured HRTFs were assumed to correspond to the internal template of HRTFs a listener is calibrated to Langendijk and Bronkhorst (2002). Based on this assumption, the potential perceptual effect of the virtual microphone, namely, its position and size, on the numerically calculated HRTFs was evaluated. Since the acquisition of localization performance in psychoacoustic experiments would be too time intensive for such a large number of conditions, microphone parameters were evaluated by analyzing sound-localization performance predictions estimated by a sagittal-plane sound-localization model (Baumgartner *et al.*, 2013).

We assumed that the useful positions of the virtual microphone can be reduced to the surface surrounding the ear canal, i.e., the concha. HRTFs are usually measured for hundreds of source positions and the concha consists of approximately 2000 elements (in the 1-mm mesh). Since each mesh element represents a potential virtual microphone position, theoretically, hundreds of thousand of numerically calculated HRTFs would be needed. In order to keep the computational effort small, we used the direct approach, in which for a single sound-source position, the pressure at the whole boundary was calculated. Then, for that sound-source position the effect of the position and size of the virtual microphone model was evaluated by averaging the sound pressures on different contiguous elements. Assuming an impact on sagittal-plane sound-source localization only, the calculations were done for sound-source positions in the median plane.

In Secs. III A–III C, we show detailed analysis for subject NH5 (the same subject as in Kreuzer *et al.*, 2009). Based on the analysis for that listener, we derive quantitative parameters describing the perceptual effect of the virtual microphone, and then show these parameters for all three listeners to support our key findings.

A. Methods

1. Material—The complex-valued head-surface sound pressure for the 1-mm meshes of the three subjects (compare Sec. II C) was calculated. Point sources representing loudspeakers were placed in the median-plane at a distance of 1.2 m with a polar incident angle ranging from -30° to 210° and a resolution of 15° . In total, 17 sound-source positions were used.

After the inverse Fourier transform of the complex surface pressure spectrum, we obtained 17 HRIRs for each triangle in the meshes. All HRIRs were re-sampled to a sampling rate of 48 kHz. HRIRs for different microphone sizes were simulated by averaging HRIRs for contiguous sets of triangles. For the numerically calculated and acoustically measured HRIRs, median-plane DTFs were calculated following the procedure as described in Sec. II B with the exception of calculating the direction-independent part, i.e., the common transfer function (Middlebrooks, 1999a), from HRTFs in the median plane only.

2. Localization model—The perceptual similarity of the acoustically measured and numerically calculated DTFs was evaluated by means of the sagittal-plane sound-localization model (Baumgartner *et al.*, 2013). This model compares a given DTF to an internal template, i.e., an internal representation of the DTFs to which a listener's auditory system is thought to be calibrated to. In our case, the internal template was assumed to be represented by the listener-specific acoustically measured DTFs. The model is quite complex, consists of many stages, and was in detail described in Baumgartner *et al.* (2013) and Majdak *et al.* (2014).

For the calibration, the model requires an uncertainty parameter describing the sensitivity of the listener. The uncertainty parameter is assumed to distinguish between good and poor localizers and can be obtained from sound-localization experiments. We assumed our listeners to be average localizers and we calibrated the model with the uncertainty of 1.9.

The model describes the listener-specific localization performance by means of the *polar root mean square (RMS) error (PE)* and the *quadrant error rate (QE)*. We used definitions exactly as in Middlebrooks (1999a). The QE is a global parameter representing the amount of errors larger than 90° and can be interpreted as confusions between the hemifields (top, down, front, back). The PE is a local parameter representing the accuracy and precision of the localization when the hemifields are not confused (for more details, see Baumgartner *et al.*, 2013).

3. Conditions—The virtual microphone was an approximately circular area described by its position and size.

The size of the virtual microphone was defined by the circle radius, which was increased from 1 mm, i.e., a single mesh element, to 6 mm by incrementally adding rings of neighboring elements. Figure 4 shows four exemplary microphone positions A, B, C, and D and four microphone sizes for position C, for the left-ear mesh of NH5 zoomed to the region of the concha. A, B, C, and D represent the positions center of the ear canal, edge of the ear canal, outside the ear canal, and outside the concha, respectively. The 6-mm virtual microphone consisted of a center element encircled by six rings of elements.

The position of the virtual microphone was defined as the midpoint of the element in the case of the 1-mm microphone, and the midpoint of the central element in the cases of larger microphone sizes. The position was varied within a radius of 45 mm around the approximate center of the ear canal. In total, approximately 25 000 conditions for each ear were systematically evaluated.

B. Results and discussion

1. Spectral features—Figure 5 shows DTF amplitude spectra calculated for the microphone positions A, B, and C and a microphone size of approximately 1 mm which corresponds to a single element at our mesh. The lateral and polar angle of the sound source were 0° and 45° , respectively. The DTF amplitude spectrum from the acoustic measurement is shown for comparison. This figure directly addresses Fig. 11 from Kreuzer *et al.* (2009) where a frequency shift of spectral features was observed as the virtual microphone was moved from one element to a neighboring element. In Kreuzer *et al.* (2009), the largest frequency shift occurred for the same sound-source position. In our Fig. 5, a small frequency shift of the notch at 13.5 kHz, the absence of the peak 6 and 7 kHz, and the absence of the sharp notch at 9.5 kHz can be observed. Note that the acoustic measurement was band-limited by approximately 18 kHz.

In order to estimate the impact of the microphone size, DTFs for three microphone sizes at the positions A to D were evaluated. Figure 6 shows DTF amplitude spectra calculated for these microphone positions A to D with the size of the virtual microphone as parameter. The lateral and polar angle of the sound source was 0° and 45° , respectively. The radius of the virtual microphone was approximately 1, 3, and 6 mm. When compared to the acoustically measured DTFs, the DTFs calculated for the larger microphone sizes appear spectrally smeared, even for the position A, which most probably represents the actual microphone position from the acoustic measurement. Further, with larger size, additional spectral features arise. For instance, at the position D when the microphone size increased to 6 mm [Fig. 6(D)], an additional notch occurred at 10.5 kHz and a corresponding notch cannot be found in the acoustically measured DTFs. Such additional features probably result from including the areas outside of the ear canal in the average of the sound pressure over the microphone area. In summary, the larger sizes of the virtual microphone seem to yield DTFs with smeared features (as at position A, B, and C) and/or features, which are unreasonable in comparison to acoustically measured DTFs (as at position D). The microphone size of approximately 1 mm yielded the smallest spectral difference and, thus, seems to be the most appropriate choice.

As for the size, it seems that multiple elements for the virtual microphone are of no advantage. As for the position, a small shift in position of the virtual microphone placed near the ear canal results in frequency shifts of the spectral features. But do these shifts affect the sound-localization performance?

2. Sound-localization performance—In order to address this question, localization performance was predicted as a function of the microphone position and size. Figure 7(c) shows the left-ear concha of NH5 and two regions (denoted by vertical and horizontal lines) along which the microphone positions were investigated. Figures 7(a) and 7(b) show the relative PEs and QEs for the microphone positions from the vertical region. Figures 7(d) and 7(e) show the relative PEs and QEs for the microphone positions from the horizontal region. In each of the four panels (a, b, d, e), the different lines represent virtual microphone sizes with radii of 1, 3, and 6 mm. The relative PEs and QEs are PEs and QEs relative to

references, i.e., the smallest PEs and the QEs in the concha, respectively. These references are shown in Table II, labeled as SIM.

The predicted localization performance was generally within the range which was usually found in sound-localization experiments (Middlebrooks, 1999a). It can be seen in Fig. 7 that the position had a notable effect on the predicted localization performance. Receiver elements within the blocked ear canal resulted in small localization errors. Receiver elements at the border or outside the ear canal resulted in larger localization errors. The effect of the virtual microphone size depended on the position of the receiver elements. The specific microphone positions underline that effect. Positions A and C resulted in a similar predicted localization performance suggesting that a specific position is not required as long as being in the area of the ear canal. Positions B and D resulted in a deteriorated localization performance, confirming the intuitive assumption that a receiver element outside the ear canal is not recommended. Interestingly, the effect of the microphone size was small as long as the receiver element was approximately in the range of the blocked ear canal (position A), in contrast to the larger effect of the microphone size for microphone positions at the border of the ear canal (positions B and D). This indicates that consideration of the actual microphone size is not required in HRTF calculations.

In order to finally evaluate the optimal microphone positions within the ear for all three subjects, median-plane HRTFs were calculated for each single mesh element as the virtual microphone on the pinna, and localization performance was predicted for all six ears, see Fig. 8. The gray-value in the area of each element encodes the relative PEs and the QEs obtained for the corresponding ear. The references of the relative PEs and QEs are shown in Table II (SIM), along with the predicted individual localization performance for the listener-specific acoustically measured HRTFs (AC) and for mannequin HRTFs (KEMAR). The predictions for listener-specific HRTFs (SIM and AC) were in the range of localization performance usually observed for acoustically measured HRTFs (Middlebrooks, 1999a). Further, the predicted errors were significantly smaller than those predicted for non-individual KEMAR HRTFs.

For all ears, the localization errors were small when the virtual microphone was placed in the ear canal and increased rapidly when the virtual microphone was moved beyond the ear canal. Especially, the position of the smallest virtual microphone was not crucial. For that 1 mm size, the predicted PEs and QEs, averaged over the six ears, were in the range of the smallest $PE \pm 0.5^\circ$ and $QE \pm 0.5\%$ (see Table II, SIM) when the virtual microphone was placed within a range of 2.5 ± 1.8 mm at the blocked ear canal.

C. Conclusions

We aimed at estimating the requirements for the details of modeling the actual microphone used in the acoustic measurements. We could observe differences in the amplitude spectra for two neighboring receiver elements, similarly to differences found in Kreuzer *et al.* (2009). These difference were, however, not reflected in the predicted sound-localization performance.

All three sizes of the virtual microphone yielded predicted localization performance in the range of that found for acoustically measured HRTFs. This indicates that a detailed model of the microphone from the acoustic measurements is not required. The predicted performance, however, degraded significantly with increasing the distance of the virtual microphone to the ear canal and that dependence became larger with larger microphone sizes.

This indicates that a small receiver element, here with the size of 1 mm, placed approximately in the center of the blocked ear canal can sufficiently model the microphones for numeric HRTF calculations. It even seems that, permitting for a tolerance of $PE \pm 0.5^\circ$ and $QE \pm 0.5\%$, the exact position of the virtual microphone can deviate from the center by as much as 2.5 mm on average.

IV. EFFECT OF THE DISCRETIZATION: MODEL PREDICTIONS

In this section, the requirements on the boundary discretization were investigated. The discretization of the boundary surface by planar triangles may introduce both geometrical and numerical errors. Geometrical errors are caused by under-sampling the geometry and discarding geometric details. Numerical errors are caused by under-sampling the sound field. Both types of error are addressed in this section. The joint effect of geometrical errors and numerical errors was systematically investigated with *coarsened meshes*, i.e., re-meshed reference meshes with larger AELs. The effect of geometrical errors was separated from the effect of numerical errors by decreasing the AEL in the coarsened meshes, here referred to as *refined meshes*. These refined meshes had the same geometrical accuracy as the coarsened meshes but smaller AELs, hence they did not under-sample the sound field. The impact of the two types of errors on sound localization was evaluated in terms of comparing model predictions for HRTFs of the different meshes. Two models were used: The similarity of temporal features was evaluated by comparing equivalent head radii (Ziegelwanger and Majdak, 2014) and the similarity of spectral features was evaluated by comparing predicted sagittal-in Sec. III.

A. Conditions

For each subject, two sets of meshes were created. The first set was composed of the 1-mm mesh and of four coarsened meshes with AELs of 2, 3, 4, and 5 mm. The coarsened meshes were generated by re-meshing the reference meshes with a re-meshing algorithm (Botsch and Kobbelt, 2004) implemented in *OpenFlipper*.¹⁰ The statistics of the meshes are shown in Table III. The second set consisted of four refined meshes created by subdividing each triangle in the coarsened meshes to four triangles. For the 2- and 3-mm meshes, this subdivision was limited to the pinna region in order to limit the size of the meshes to approximately 100 000 elements.

B. Methods

For all three subjects, HRTFs were numerically calculated using the reciprocal approach. Based on the results from the previous section, the virtual microphone was a single,

¹⁰OpenFlipper, available from <http://www.openflipper.org> (Last viewed July 8, 2014).

uniformly vibrating, triangular element, placed at the approximate center of the ear canal. The velocity of the vibrating element was 0.1 mm/s. HRTFs were numerically calculated for the same loudspeaker grid as used in the acoustical measurement (Majdak *et al.*, 2010). HRIRs and DTFs were calculated as described in Sec. II.

The similarity of numerically calculated and acoustically measured HRTFs was evaluated by applying two models. The first model evaluated the similarity of spectral features by predicting listener-specific PEs and QEs. The model and its parameters were the same as in Sec. III A 2. In addition to listener-specific predictions for the three tested listeners, two benchmarks were calculated aiming to frame the predicted performance into adequate context. The benchmark “own” was the statistics of localization performance predicted for 172 listeners of the ARI,¹¹ CIPIC,¹² and LISTEN¹³ databases under the assumption of listening with own ears, i.e., localizing sound sources with their corresponding listener-specific HRTFs. The benchmark “other” was the statistics of localization performance predicted for NH5, NH130, and NH131 under the assumption of localizing with other’s ears, i.e., the PEs and QEs predicted for localizing sound sources with HRTFs of all 171 other listeners from the three databases.

The second model evaluated temporal features, in particular the broadband time-of-arrival (TOA) in HRIRs, which interaural difference is the relevant feature for localization along the lateral dimension (Macpherson and Middlebrooks, 2002). When considering the head as an approximation of a sphere, the TOA depends on the equivalent head radius (Algazi *et al.*, 2001a). The equivalent head radius was estimated by a geometrical TOA model (Ziegelwanger and Majdak, 2014). In a nutshell, the TOA model estimates broadband delays from an HRTF set of a listener by the use of the minimum-phase cross-correlation method¹⁴ and fits a spherical head model of the listener to them. The radius of the head model represents the equivalent head radius of the listener and defines the spatial ITD function in an HRTF set. The equivalent head radii were estimated for the numerically calculated HRTFs and for the acoustically measured HRTFs. As a rough estimation, we assumed any difference between equivalent head radii larger than 3.55 mm as evidence for differences in the temporal features because 3.55 mm corresponds to a half of the spatial sampling interval of 7.1 mm assuming a speed of sound by 343 m/s and a sampling rate of 48 kHz. Differences smaller than 3.55 mm were interpreted as insignificant.

C. Results and discussion

Figure 9 shows median-plane amplitude spectra for the acoustically measured [Fig. 9(a)] and numerically calculated DTFs for coarsened meshes [Figs. 9(b) to 9(f)] of NH5’s left ear. Note that the acoustic measurement was band-limited between 0.3 and 18 kHz. Further, the numerical calculations did not include a torso model, thus, spectral ripples at low

¹¹Retrieved from <http://sofacoustics.org/data/database/ari/> (Last viewed January 20, 2015), originating from the “(short)” HRTF set from <http://www.kfs.oaw.ac.at/hrtf> (Last viewed January 20, 2015).

¹²Retrieved from <http://sofacoustics.org/data/database/cipic/> (Last viewed January 20, 2015), originating from <http://interface.cipic.ucdavis.edu/-sound/hrtf.html> (Last viewed January 20, 2015).

¹³Retrieved from <http://sofacoustics.org/data/database/listen/> (Last viewed January 20, 2015), originating from the “compensated” HRTF set from <http://recherche.ircam.fr/equipes/salles/listen/download.html> (Last viewed January 20, 2015).

¹⁴Note that the minimum-phase cross-correlation method showed the best correspondence to actual radii in numerically calculated HRTFs (Ziegelwanger and Majdak, 2014), the results for other methods may vary (Katz and Noisternig, 2014).

frequencies arising from torso reflections are missing. For the acoustically measured DTFs and the numerically calculated DTFs of the 1-mm mesh, the spectral notches appear to be similar. For larger AELs, these notches change their spectrospatial position or even disappear, e.g., the notch at 13 kHz and 45° elevation completely disappeared in the 5-mm mesh. For this mesh, spectral features seem to be clearly different compared to those from the acoustic measurement, with differences appearing even at frequencies below 10 kHz.

Figure 10 shows the predicted localization performance in sagittal planes. The gray area shows the standard deviation range of the localization performance observed in Middlebrooks (1999a). For the acoustically measured HRTFs, for the 1- and 2-mm mesh, and for the benchmark “own,” predicted PEs and QEs were within that area. For numerically calculated HRTFs, the predicted performance degraded with increasing AEL. For the 5 mm mesh, the predictions were in the range of localization performance observed for the “other” benchmark. Thus, the range of AELs from 1 mm to 5 mm represents good candidates for further evaluation in actual sound-localization experiments.

For the coarsened meshes, both the PE and QE increased with AEL. A linear regression was fit to the predictions in order to roughly examine the relation between the AEL and predicted localization performance. For the PE, the slope of the linear regression was 1.51°, 1.79°, and 1.92°/mm for NH5, NH130, and NH131, respectively, and 1.74°/mm on average. For the QE, the slope was 0.99%, 2.13%, and 3.19%/mm for NH5, NH130, and NH131, respectively, and 2.10%/mm on average.

Table IV shows the equivalent head radii obtained from the TOA model for all conditions and subjects. For numerically calculated HRTFs, the equivalent head radii were similar to those for the acoustically measured HRTFs indicating that the AEL is not critical for accurate representation of broadband temporal features in numerically calculated HRTFs.

For refined meshes, the predictions were similar to the predictions for the coarsened meshes. For both PE and QE, the Pearson’s correlation coefficients between the predictions for the coarsened meshes and those for the refined meshes were 0.99, showing further evidence for a very similar performance for both coarsened and refined meshes. Also the equivalent head radii observed for the coarse and refined meshes were similar. This indicates that the performance degradation with increasing AEL was solely based on geometrical errors and the contribution of the numerical errors was negligible. Thus, the AEL effect observed in conditions with coarsened meshes appears to rely on poor discretization of the detailed geometry of the pinna with the consequence that the refined meshes are not required in further HRTF calculations.

V. EFFECT OF THE DISCRETIZATION: LOCALIZATION EXPERIMENTS

In this section, particular conditions from Sec. IV were tested in a sound-localization experiment in order to validate the predictions (and thus the numerical calculations) with respect to sound localization. Seven experimental conditions were tested: listener-specific acoustically measured HRTFs, acoustically measured KEMAR HRTFs, and numerically calculated HRTFs for the 1-mm mesh and the meshes coarsened up to the AEL of 5 mm.

A. Methods

All listeners from the previous sections participated in the sound-localization experiments. The procedure of the sound-localization experiment was based on that from Majdak *et al.* (2010) and identical to that from Majdak *et al.* (2013) with the exception of a small change in the visual virtual environment (see later).

Briefly, the stimuli were Gaussian white noises of 500-ms duration filtered with DTFs corresponding to the tested condition and position. The tested lateral positions ranged from -90° (right) to 90° (left) and the tested elevations ranged from -30° to 210° . The stimuli were presented via headphones (HD 580, Sennheiser) at the level of 50 dB above the individual absolute hearing threshold and level roving of ± 5 dB was applied. The listeners were immersed in a spherical virtual visual environment. They held a pointer in their hand. The pointer's projection was visualized and recorded as the perceived target position. In contrast to Majdak *et al.* (2013), the head-mounted display was an Oculus Rift (Developer Kit, OculusVR). For each eye, the display provides screens with a field of view of $110^\circ \times 90^\circ$ (horizontal $\times$ vertical dimensions) and a resolution of 800×640 pixels. In the virtual environment, the colors, brightness, and the grid were adapted to the field-of-view such that the immersed listeners were feeling comfortable and relaxed without any signs of dizziness, e.g., the background color was set to light gray, the grid lines were removed and replaced by balls at grid corners. The grid resolution was 45° and 11.25° in the azimuth and elevation, respectively. While balls indicated the horizontal and median plane for orientation were blue, the rest of the balls were black. A red ball indicated the position 0° azimuth and 0° elevation.

Prior to the main experiment, listeners participated in a visual and acoustic training procedure with the goal of using the equipment accurately and of performing stably. In the visual training, the listeners pointed to visual targets within a limited amount of time. They continued the training until they were able to point to 95% of the targets within 4 s with an average angular distance smaller than 2° (for more details see Majdak *et al.*, 2013, p. 2058). In the acoustic training, listeners were localizing sounds filtered with their own DTFs and visual feedback on the actual target position was applied. This training consisted of 300 trials (for more details, see Majdak *et al.*, 2010, p. 462).

In the main experiment, at the beginning of each trial, the listeners were asked to align themselves with the reference position and, then, the stimulus was presented. The listeners were asked to point to the perceived stimulus direction. Each condition was tested in three experimental blocks each. An experimental block consisted of 100 targets and lasted for approximately 20 min. The order of the 21 experimental blocks was randomized in such a way that within seven blocks all conditions were in a randomized order. The study was a double-blind experiment, i.e., none of the listeners were the authors, listeners were not enlightened as to the nature of the experiment, listeners and the experimenter did not know the tested conditions. As in Sec. IV, the PE and the QE were calculated for the analysis of localization performance. Additionally, the lateral RMS error (LE) and the signed polar bias were calculated.

B. Results and discussion

1. General—Figure 11 shows the localization performance parameters PE and QE obtained from the localization experiment (labeled as Original). Generally, an increase of the AEL yielded degraded performance (1–5 in Fig. 11), more or less approaching that obtained for the KEMAR HRTFs (K in Fig. 11).

For the numerically calculated HRTFs with AELs of 1 and 2 mm, both PE and QE were in the range usually found in localization experiment with listener-specific HRTFs (e.g., Middlebrooks, 1999a). These two conditions yielded also QEs very similar to those obtained in the condition with acoustically measured HRTFs (A in Fig. 11), for all three listeners. For the PEs, the situation was different: the PEs were similar for NH5 and NH131, but not for NH130.

NH130 showed actually a surprisingly large PE in the condition with acoustically measured HRTFs. Since the PEs in the conditions with HRTFs numerically calculated for AELs of 1 and 2 mm were within the expected range, we looked deeper into this issue. We identified four issues which might have yielded the poor PE in the condition with acoustically measured HRTFs. First, NH130 might generally have been a poor localizer. In that case, however, NH130 would not be able to perform such good in the conditions with numerically calculated HRTFs. Second, the acoustically measured HRTFs might have been problematic. We used exactly the same HRTFs in both training and main experiment, and the PE and QE from the training was 29.1° and 4.7%, respectively, showing that NH130 did much better in the training than in the later experiment. The performance from the training was in the range of usually observed localization performance, indicating that, in general, NH130 was able to localize sounds quite well even with the acoustically measured HRTFs. Third, in the main experiment, seven conditions were tested, potentially offering room for order effects like practice or fatigue. However, in the experiment, each of the seven conditions was split in three blocks, the blocks were presented in a randomized order, and the obtained PE was consistently poor in all three corresponding blocks. Thus, the order effect was unlikely the reason for poor performance with acoustically measured HRTFs. Fourth, we noticed that NH130 showed a large variability in PE across *all* tested conditions. In default of a better hypothesis, we can only speculate: While NH130, being a general good localizer, seems to have produced an outlier in that one particular condition. Anyway, from now on, we rely more on NH130's results from the acoustic training, which seem to be more reasonable (see the crosses in Fig. 11).

The performance of NH5 did not change much as a function of AEL. NH5 seems thus to be not that susceptible to AEL increase as NH130 and NH131 seem to be. However, the localization error obtained for NH5 in the acoustic and 1-mm conditions was larger than compared to the other subjects (with the PE of NH130 from the training as discussed before). This renders NH5 a generally less sensitive localizer. There is evidence that listener-specific sound-localization performance is linked to the ability to discriminate spectral shapes (Andéol *et al.*, 2013). Such a factor seems to be even more salient than the directional features provided by listener-specific HRTFs (Majdak *et al.*, 2014). Thus, for an insensitive subject as NH5 might be, the degradation of spectral features resulting from the

AEL increase might have little impact on the sound-localization performance. On the other hand, NH131 seems to be a sensitive subject because a small increase of AEL yielded large degradation of the localization performance. The sensitivity of NH130 seems to be between that of NH131 and NH5. As a common ground, HRTFs calculated with AELs of 1 or 2 mm seem to yield good localization performance in all our subjects.

In order to quantify the effect of AEL on the localization performance, a linear regression was fit to PEs and QEs obtained for AELs of 1–5 mm. For all subjects, the linear regression showed a positive slope, i.e., a degrading localization performance for increasing AEL in the mesh. The PE increased by 0.8°, 2.9°, and 2.9°/mm for NH5, NH130, and NH131, respectively, and 2.2°/mm on average. The QE increased by 1.3%, 4.2%, and 8.3%/mm for NH5, NH130, and NH131, respectively, and 4.6%/mm on average.

2. Polar bias and the interaural horizontal plane—The polar bias in localization responses was calculated for the conditions involving numerically calculated HRTFs. For NH5, NH130, and NH131, the bias was -8.1° , -20.1° , -11.6° , respectively, and -13.3° on average. This finding implies that our definition of the interaural horizontal plane as the intersection of the nasal and auricular points (see Sec. II D) did not correspond to the interaural horizontal plane used by the listeners in the experiment. This finding is actually not surprising because distinct markers of the interaural horizontal plane for a good match between the acoustical measurement and mesh generation were not available. For example, Jin *et al.* (2014) solved this problem *post hoc* by applying a rotation minimizing the spectral cross-correlation between the acoustically measured and numerically calculated HRTFs. The average rotation they had applied was 12.6° .

Given these findings, we were looking for a good definition of the interaural horizontal plane in the process of mesh generation. In the field of anatomy, besides the nasal points, the orbital points, i.e., the inferior margins of the orbits, can be used for alignment of the human head. Further, the plane intersecting orbital and auricular points, called the Frankfurt plane, is used to describe the position of head usually carried in the living subjects (Cheng *et al.*, 2012). The Frankfurt plane, together with the median plane can be used to define the three anatomical orientations: lateral, anterior–posterior (front–back), and superior–inferior (up–down). Thus, we checked the position of the Frankfurt plane in our meshes.

For NH5, NH130, and NH131, the Frankfurt plane deviated from our previous definition of the interaural horizontal plane by 10.5° , 15.9° , and 13.4° , respectively, with an average of 13.3° . This deviation might explain the polar bias in the localization responses.

In order to apply the Frankfurt plane as horizontal plane to our localization responses, for each sound target, we subtracted the deviation to the Frankfurt plane from the target angle. This polar shift corresponds to a rotation of the mesh in the numerical calculation similar to that from Jin *et al.* (2014). For NH5, NH130, and NH131, the resulting corrected polar bias was 2.4° , -4.2° , and 1.8° , respectively, and 0.0° on average.

We then recalculated the localization performance given the corrected target angles (see Fig. 11, data labeled as Corrected). The smaller corrected PEs and QEs indicate that the original

PEs and QEs were indeed subject of the polar bias resulting from the non-optimal definition of the interaural horizontal plane used in the first place.

In summary, both the vanishing polar bias and the smaller corrected PEs and QEs suggest that the Frankfurt plane is appropriate for the definition of the interaural horizontal plane in the process of mesh alignment for numerical HRTF calculation.

3. Lateral errors—Table V shows the lateral RMS error (LE) for all three subjects and statistics of the LE observed in virtual sound-localization experiments from Middlebrooks (1999a). For the 1-mm condition, the LE was in the range of LEs from Middlebrooks (1999a). It seems that the AEL had a negligible effect on LE, confirming predictions from the TOA model. The LE for NH131 was larger for numerically calculated HRTFs than for acoustically measured HRTFs but still was in the range of parameters found by (Middlebrooks, 1999a). Since impedance boundary conditions can affect the ITD (Katz, 2001b), this difference might be explained by the discrepancy of actual boundary conditions of a subject's head and boundary conditions used in the numerical calculation.

4. Predictions and actual performance—Looking back to the predictions from Sec. IV, similar degradation of localization performance with increasing AEL was observed in the predicted and the actual localization performance. The Pearson's correlation coefficients between predicted and actual localization-performance parameters were 0.73 and 0.83 for PE and QE, respectively.

Even though there seems to be a bias between the predictions and the actual performance, the trend in the performance degradation found in actual experiments seems to be reflected by the model predictions; an aspect important for the correct interpretation of predictions from the previous sections. If required in the future, with the actual responses from the localization experiment, the model can be calibrated to the actual listener-specific uncertainties, further improving the congruency between our predictions and the actual performance (Majdak *et al.*, 2014).

VI. CONCLUSIONS

We investigated requirements on numerical calculation of listener-specific HRTFs for accurate sound-localization performance. We focused on the size and position of the virtual microphone and on the discretization of the listener's head and pinna geometry.

A simple model of the microphone used in acoustic measurements seems to be sufficient. A single mesh element, approximately centered at the blocked ear canal yielded HRTFs comparable to those from acoustic measurements and more sophisticated virtual microphones were not required. With a tolerance of $PE \pm 0.5^\circ$ and $QE \pm 0.5\%$, any position of the virtual microphone within the distance of ± 2.5 mm around the ear-canal center seems to be acceptable. Larger virtual microphones were more susceptible to their exact position.

The resolution of the discretized listener's geometry had an obvious impact on the sound-localization performance in sagittal planes (top/down, front/back). The degradation of the geometrical detail level (geometrical error) was the dominant factor. The under-sampling of

the sound-field (numerical error) did not affect our results much. The localization performance degraded with increasing AEL, with an average rate of $2.2^\circ/\text{mm}$ and $4.6\%/\text{mm}$ (PE per AEL and QE per AEL, respectively). For AELs of 1 and 2 mm, the localization performance was in the range of that usually found when localizing virtual sound sources.

The resolution of the discretized geometry had a negligible effect on the sound-localization performance in lateral dimension (left/right). This result is plausible, because the change of the AEL was small compared to the dimension of the human head, thus, having a negligible effect on shadowing and diffraction caused by the head.

The analysis of the polar bias indicates that the perceived interaural horizontal plane (0° elevation) in the mesh generation process is better represented by the Frankfurt plane than by the tip of the nose.

Some of our findings are based on model predictions only. For conditions which were investigated by both model predictions and actual sound-localization experiments, we found that the predicted localization performance corresponded well (average correlation coefficient of 0.78) to the performance obtained in the actual localization experiments. This seems to be a good basis for drawing conclusions based on the predictions only.

In summary, we hope that our findings help in further establishing methods for the numerical HRTF calculation. While our study is clearly limited to sound localization and to three subjects only, further investigation of the requirements for other perceptual effects like sound externalization and timbre seem to be the next logical step.

ACKNOWLEDGMENTS

Special thanks go to Andreas Reichinger (VRVis GmbH, Austria) and his team for the work on geometry acquisition and generation of the reference mesh. We are also grateful to our subjects for their patience during the measurements and experiments. This study was supported by the Austrian Science Fund (FWF, P 24124-N13).

References

- Algazi VR, Avendano C, Duda RO. Estimation of a spherical-head model from anthropometry. *J. Audio Eng. Soc.* 2001a; 49:472–479.
- Algazi VR, Avendano C, Duda RO. Elevation localization and head-related transfer function analysis at low frequencies. *J. Acoust. Soc. Am.* 2001b; 109:1110–1122. [PubMed: 11303925]
- Algazi, VR.; Duda, RO.; Thompson, DM.; Avendano, C. The CIPIC HRTF database. *Proceedings of 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics*; New York. 2001c. p. 99-102.
- Andéol G, Macpherson EA, Sabin AT. Sound localization in noise and sensitivity to spectral shape. *Hear. Res.* 2013; 304:20–27. [PubMed: 23769958]
- Andreopoulou A, Begault D, Katz B. Inter-laboratory round robin HRTF measurement comparison. *IEEE J. Sel. Top. Signal Process.* 2015; PP(99):1–12.
- Baumgartner, R.; Majdak, P.; Laback, B. Assessment of sagittal-plane sound localization performance in spatial-audio applications. In: Blauert, J., editor. *The Technology of Binaural Listening*. Springer; Berlin: 2013. Chap. 4
- Botsch, M.; Kobbelt, L. A remeshing approach to multiresolution modeling. *Proceedings of the 2004 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing (SGP'04)*; New York: ACM; 2004. p. 185-192.

- Bronkhorst AW. Localization of real and virtual sound sources. *J. Acoust. Soc. Am.* 1995; 98:2542–2553.
- Chen Z-S, Waubke H, Kreuzer W. A formulation of the fast multipole boundary element method (FMBEM) for acoustic radiation and scattering from three-dimensional structures. *J. Comput. Acoust.* 2008; 16:303–320.
- Cheng, Y.; Leow, W-K.; Lim, TC. Automatic identification of Frankfurt plane and mid-sagittal plane of skull. *Proceedings of 2012 IEEE Workshop on the Applications of Computer Vision*, Breckenridge, CO; 2012. p. 233-238.
- Cignoni P, Rocchini C, Scopigno R. Metro: Measuring error on simplified surfaces. *Comput. Graph. Forum.* 1998; 17:167–174.
- Greff, R.; Katz, BFG. Round robin comparison of HRTF simulation systems: Preliminary results. *Proceedings of the 123th Convention of the Audio Engineering Society*; New York. 2007. Convention Paper 7188
- 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; 127:370–386. [PubMed: 20058984]
- Hebrank J, Wright D. Spectral cues used in the localization of sound sources on the median plane. *J. Acoust. Soc. Am.* 2005; 56:1829–1834. [PubMed: 4443482]
- Huttunen T, Seppälä ET, Kirkeby O, Kärkkäinen A, Kärkkäinen L. Simulation of the transfer function for a head-and-torso model over the entire audible frequency range. *J. Comput. Acoust.* 2007; 15:429–448.
- Jin CT, Guillon P, Epain N, Zolfaghari R, van Schaik A, Tew AI, Hetherington C, Thorpe J. Creating the Sydney York morphological and acoustic recordings of ears database. *IEEE Trans. Multimedia.* 2014; 16:37–46.
- Kahana Y, Nelson PA. Numerical modelling of the spatial acoustic response of the human pinna. *J. Sound Vib.* 2006; 292:148–178.
- Katz BFG. Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation. *J. Acoust. Soc. Am.* 2001a; 110:2440–2448. [PubMed: 11757933]
- Katz BFG. Boundary element method calculation of individual head-related transfer function II. Impedance effects and comparisons to real measurements. *J. Acoust. Soc. Am.* 2001b; 110:2449–2455. [PubMed: 11757934]
- Katz BFG, Noisternig M. A comparative study of interaural time delay estimation methods. *J. Acoust. Soc. Am.* 2014; 135:3530–3540. [PubMed: 24907816]
- 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; 126:1280–1290. [PubMed: 19739742]
- Kuhn GF. Model for the interaural time differences in the azimuthal plane. *J. Acoust. Soc. Am.* 1977; 62:157–167.
- Langendijk EHA, Bronkhorst AW. Contribution of spectral cues to human sound localization. *J. Acoust. Soc. Am.* 2002; 112:1583–1596. [PubMed: 12398464]
- Lopez-Poveda EA, Meddis R. A physical model of sound diffraction and reflections in the human concha. *J. Acoust. Soc. Am.* 1996; 100:3248–3259. [PubMed: 8914307]
- Lorensen, WE.; Cline, HE. Marching cubes: A high resolution 3D surface construction algorithm. In: Stone, Maureen C., editor. *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'87)*; New York: ACM; 1987. p. 163-169.
- Macpherson EA, Middlebrooks JC. Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited. *J. Acoust. Soc. Am.* 2002; 111:2219–2236. [PubMed: 12051442]
- Majdak P, Baumgartner R, Laback B. Acoustic and non-acoustic factors in modeling listener-specific performance of sagittal-plane sound localization. *Front. Psychol.* 2014; 5(319):1–10. [PubMed: 24474945]
- Majdak P, Goupell MJ, Laback B. 3-D localization of virtual sound sources: Effects of visual environment, pointing method, and training. *Atten. Percept. Psychophys.* 2010; 72:454–469. [PubMed: 20139459]

- Majdak P, Masiero B, Fels J. Sound localization in individualized and non-individualized crosstalk cancellation systems. *J. Acoust. Soc. Am.* 2013; 133:2055–2068. [PubMed: 23556576]
- Marburg S. Six boundary elements per wavelength: Is that enough? *J. Comput. Acoust.* 2002; 10:25–51.
- Mey FD, Reijniers J, Peremans H, Otani M, Firzlaff U. Simulated head related transfer function of the phyllostomid bat *Phyllostomus discolor*. *J. Acoust. Soc. Am.* 2008; 124:2123–2132. [PubMed: 19062853]
- Middlebrooks JC. Virtual localization improved by scaling non-individualized external-ear transfer functions in frequency. *J. Acoust. Soc. Am.* 1999a; 106:1493–1510. [PubMed: 10489706]
- Middlebrooks JC. Individual differences in external-ear transfer functions reduced by scaling in frequency. *J. Acoust. Soc. Am.* 1999b; 106:1480–1492. [PubMed: 10489705]
- Møller H, Sørensen MF, Hammershøj D, Jensen CB. Head-related transfer functions of human subjects. *J. Audio Eng. Soc.* 1995; 43:300–321.
- Morse, PM.; Ingard, KU. *Theoretical Acoustics*. Princeton University Press; Princeton, NJ: 1986. p. 1-949.
- 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; 135:2534–2544. [PubMed: 24815237]
- Reichinger, A.; Majdak, P.; Sablatnig, R.; Maierhofer, S. Evaluation of Methods for Optical 3-D Scanning of Human Pinnae. *Proceedings of the International Conference on 3D Vision (3DV 2013)*; 2013. p. 390-397.
- Sauter, SA.; Schwab, C. *Boundary Element Methods*. Springer; Heidelberg: 2010. p. 1-500.
- Takemoto H, Mokhtari P, Kato H, Nishimura R, Iida K. Mechanism for generating peaks and notches of head-related transfer functions in the median plane. *J. Acoust. Soc. Am.* 2012; 132:3832–3841. [PubMed: 23231113]
- Treeby BE, Pan J. A practical examination of the errors arising in the direct collocation boundary element method for acoustic scattering. *Eng. Anal. Boundary Elem.* 2009; 33:1302–1315.
- Treeby BE, Pan J, Paubally RM. An experimental study of the acoustic impedance characteristics of human hair. *J. Acoust. Soc. Am.* 2007a; 122:2107–2117. [PubMed: 17902848]
- Treeby BE, Pan J, Paubally RM. The effect of hair on auditory localization cues. *J. Acoust. Soc. Am.* 2007b; 122:3586–3597. [PubMed: 18247766]
- Wightman FL, Kistler DJ. Headphone simulation of free-field listening. I: Stimulus synthesis. *J. Acoust. Soc. Am.* 1989; 85:858–867. [PubMed: 2926000]
- Ziegelwanger H, Majdak P. Modeling the direction-continuous time-of-arrival in head-related transfer functions. *J. Acoust. Soc. Am.* 2014; 135:1278–1293. [PubMed: 24606268]
- Ziegelwanger H, Reichinger A, Majdak P. Calculation of listener-specific head-related transfer functions: Effect of mesh quality. *POMA*. 2013; 19:050017.

FIG. 1. Mesh renderings of the left (right panels) and right (left panels) ears of the subjects. Zoomed rendering with rendered edges for NH5 (top-left panel) and a photography of NH5's left ear (top-right panel).

FIG. 2.

Maximum Hausdorff distance (a) and normalized Hausdorff distance (b) as a function of the mesh resolution described by the AEL (dots: individual data for three subjects; line: sigmoid fit). Visualization of the Hausdorff distance for an AEL of 1 (c), 2.8 (d), and 4 mm (e) as a function of the location on the left reference ear of NH5. The color bar is limited to 0.5 mm.

FIG. 3.

Spatial correlation coefficient between the acoustically measured and numerically calculated HRTFs for left and right ears of NH5, NH130, and NH131. The gray areas show the range of correlation coefficients observed in Jin *et al.* (2014), for all their subjects (light gray) and for their 20 best subjects (dark gray).

FIG. 4.
Geometry of NH5's left concha represented as a triangular polygon mesh (1 mm AEL) with exemplary virtual-microphone positions (A to D) and sizes at position C.

FIG. 5.
 (Color online) DTF amplitude spectra for virtual-microphone positions A to C (compare Fig. 4) for a single triangle representing the virtual microphone.

FIG. 6.
(Color online) DTF amplitude spectra for virtual-microphone positions A to D (compare Fig. 4) and sizes of 1, 3, and 6-mm radius.

FIG. 7.

(Color online) Geometry of NH5's left concha (c) with horizontal and vertical lines representing horizontal and vertical regions, respectively, used for the evaluation of the relative PE [(a) and (e)] and the relative QE [(b) and (d)] for virtual-microphone sizes of 1, 3, and 6-mm radius. The references for the relative parameters are shown in Table II (condition SIM).

FIG. 8. Relative PEs and relative QEs as function of the virtual-microphone position for a microphone size of 1 mm. The references for the relative parameters are shown in Table II (SIM).

FIG. 9.
(Color online) Median-plane DTF amplitude spectra of NH5; acoustically measured (a) and numerically calculated for meshes with an AEL of 1–5 mm [(b)–(f)].

FIG. 10.

Localization performance predicted by the sagittal-plane localization model. Mostleft column: Average and ± 1 standard deviation of predictions for all subjects from the databases (ARI, CIPIC, and LISTEN) when listening with own-ear DTFs. Other columns: Predictions for NH5, NH130, and NH131 in conditions with acoustically measured own-ear DTFs (labeled as A), numerically calculated own-ear DTFs for the coarsened and refined meshes (1–5), acoustically measured manikin DTFs of the KEMAR (K), and acoustically measured other-ear DTFs from the databases (D). The gray area shows the average ± 1 standard deviation from Middlebrooks (1999a).

FIG. 11.

Localization performance observed for NH5, NH130, and NH131 in the localization experiment (Original). Conditions: acoustically measured own-ear DTFs (labeled as A), numerically calculated own-ear DTFs for the coarsened meshes (1–5), and acoustically measured manikin DTFs of the KEMAR (K). The crosses represent results from the acoustic training for NH130. Original (Fit): linear fit to the results. Corrected: Original results recalculated with corrected polar bias. Grey area as in Fig. 10.

TABLE I

Anthropometric measurements (x_1 : head width, x_2 : head height, x_3 : head depth, d_5 : pinna height, d_6 : pinna width) of the three subjects and the KEMAR mannequin.

Subject ear	NH5		NH130		NH131		KEMAR
	l	r	l	r	l	r	
x_1 (mm)	156		151		160		140
x_2 (mm)	233		230		240		221
x_3 (mm)	207		195		212		200
d_5 (mm)	61	60	66.9	65.4	56.2	57.1	55.4
d_6 (mm)	27.9	29.1	29.5	31.3	29.5	27.7	27.2

TABLE II

Predicted localization performance for acoustically measured HRTFs (AC), numerically calculated HRTFs (SIM), and HRTFs of KEMAR. L, R: left and right ear, respectively.

Subject	HRTFs	PE (°)		QE (%)	
		L	R	L	R
NH5	AC	28.7	29.2	7.9	14.8
	SIM	30.0	29.9	12.8	10.7
	KEMAR	35.1	34.7	24.9	21.2
NH130	AC	24.7	22.8	6.3	9.0
	SIM	25.9	26.6	10.7	9.3
	KEMAR	33.5	35.6	19.6	21.1
NH131	AC	30.7	27.4	13.5	10.0
	SIM	30.4	30.7	11.0	11.5
	KEMAR	38.0	33.3	25.8	18.9

TABLE III

Mesh statistics (standard deviation of edge lengths; minimum and maximum edge length; number of elements averaged over subjects) and critical frequencies.

AEL	<u>Edge length (mm)</u>			Number of elements	f_{crit} (kHz)
	Std.	Min.	Max.		
2	0.2	1.1	3.1	106 526	28.6
3	0.3	1.6	4.0	53 671	19.1
4	0.4	2.3	5.9	28 188	14.3
5	0.5	2.8	7.1	18 143	11.4

TABLE IV

Equivalent head radii for acoustically measured HRTFs (AC) and numerically calculated HRTFs of coarsened and refined meshes (AEL 1–5 mm).

	Equivalent head radius (mm)					
	Coarsened meshes			Refined meshes		
	NH5	NH130	NH131	NH5	NH130	NH131
AC	91.8	92.4	94.5	—	—	—
1	90.8	89.2	94.6	—	—	—
2	91.4	89.4	94.6	91.4	89.1	92.8
3	91.4	89.4	94.6	91.1	89.3	94.3
4	91.8	90.2	94.5	89.4	89.9	94.3
5	91.1	89.8	95.1	90.7	89.4	94.5

TABLE V

Lateral RMS error (LE). Middlebrooks: Average results from Middlebrooks (1999a).

	LE (°)		
	NH5	NH130	NH131
AC	15.8	17.6	9.3
1 mm	13.2	15.3	13.1
2 mm	16.4	15.5	12.1
3 mm	16.1	19.0	14.8
4 mm	15.5	17.3	15.6
5 mm	16.2	16.8	14.9
Middlebrooks	14.5 ± 2.2		
KEMAR	17.5	17.7	11.6