

Influence of voxelization on finite difference time domain simulations of head-related transfer functions

Sebastian Prepelita^{a)}

Department of Computer Science, Aalto University, Otaniementie 17, FI-02150, Espoo, Finland

Michele Geronazzo and Federico Avanzini

Department of Information Engineering, University of Padova, Via G. Gradenigo 6/A, 35131, Padova, Italy

Lauri Savioja

Department of Computer Science, Aalto University, Otaniementie 17, FI-02150, Espoo, Finland

(Received 15 October 2015; revised 21 March 2016; accepted 13 April 2016; published online 9 May 2016)

The scattering around the human pinna that is captured by the Head-Related Transfer Functions (HRTFs) is a complex problem that creates uncertainties in both acoustical measurements and simulations. Within the simulation framework of Finite Difference Time Domain (FDTD) with axis-aligned staircase boundaries resulting from a voxelization process, the voxelization-based uncertainty propagating in the HRTF-captured sound field is quantified for one solid and two surface voxelization algorithms. Simulated results utilizing a laser-scanned mesh of Knowles Electronics Manikin for Acoustic Research (KEMAR) show that in the context of complex geometries with local topology comparable to grid spacing such as the human pinna, the voxelization-related uncertainties in simulations emerge at lower frequencies than the generally used accuracy bandwidths. Numerical simulations show that the voxelization process induces both random error and algorithm-dependent bias in the simulated HRTF spectral features. Frequencies f_r below which the random error is bounded by various dB thresholds are estimated and predicted. Particular shortcomings of the used voxelization algorithms are identified and the influence of the surface impedance on the induced errors is studied. Simulations are also validated against measurements.

© 2016 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

[<http://dx.doi.org/10.1121/1.4947546>]

[FCS]

Pages: 2489–2504

I. INTRODUCTION

One important acoustic wave scattering problem is the influence of the human body on a sound field in the vicinity of the ear canal since it contains all of the necessary cues for sound localization. Such influence can be modeled as linear time-invariant filters and are generally referred to as head-related transfer functions (HRTFs) or head-related impulse responses. These direction-dependent filters are usually measured in anechoic conditions such that the transmission of an incident plane wave generated by a point source from the far-field to a point close to the ear canal is captured for each ear.¹ This process is not without limitations and may result in inaccurate^{2–4} or costly and generally time-consuming HRTF measurements and protocols.^{5–8}

Computer simulated HRTFs represent an attractive alternative since they do not suffer from the drawbacks inherent in acoustic measurements. A variety of simulation methods have been employed to simulate HRTFs: the boundary element method,^{9–11} the finite element method,¹² the adaptive rectangular decomposition method¹³ and the finite-difference time-domain method.^{14–17} However, simulations also have specific limitations, constraints and parameters that affect the resulting

HRTFs such that the consistency with measured HRTFs is affected^{15,18} by inaccuracies in both frameworks.

In the case of Finite Difference Time Domain (FDTD) simulations, there are many factors influencing the computed sound field independent of the actual numerical simulation. One of them is the voxelization process, in which a given continuous geometry undergoes one or several stages of approximations until it results in a discrete, voxel-based geometry. For a staircase approximation of the boundary, different approaches exist: the continuous geometry is directly acquired into voxels and further processed (such as an MRI scan¹⁵) or first scanned at fixed, discrete locations (such as a mesh resulting from a laser scan¹¹) and then the resulting polyhedron mesh voxelized to be used in an FDTD simulation. The voxelization process of a three-dimensional (3D) geometry is neither a straightforward nor trivial task: depending on the constraints and the required properties of the resulting geometry, it is possible that it will not preserve the mesh topology or that it will not be unique.¹⁹ Different algorithms are designed for the voxelization process that have their own parameters, which in turn may generate additional errors in the resulting sound field, independent of the FDTD update scheme and boundary implementation.

It is unclear how important these voxelization-induced errors are within an FDTD simulation of the head-scattering problem. Some limitations of the staircase-like boundaries

^{a)}Electronic mail: sebastian.prepelita@aalto.fi

were shown²⁰ only for simple geometries and cases. The current work studies the influence of the voxelization process in the context of FDTD HRTF simulations for cases where the geometry is acquired through a laser scanning process. The propagated errors are quantified by measuring the induced variance in the sound field inside the concha volumes caused by minor translations of the acquired geometry. Nonetheless, the reported variance issue is expected to become insignificant as the voxel is greatly reduced in size. It is noted that similar issues may arise when the voxel-based geometry results from an MRI scan; although the geometry-grid misalignment of the geometry is less of an issue due to the reduced grid sizes generally used,^{15,17} either scanner parameters or voxel processing algorithms may influence the FDTD-simulated acoustic field significantly.

Various factors are analyzed: the voxel size is first examined alongside three different voxelization algorithms. In addition, the translation distance causing the sound field variance is also studied as a factor. The frequency is regarded as a parameter in the analysis. As the final factor, the influence of impedance is investigated. Both ears are considered in the investigation. Results are first analyzed based on all directions and frequencies and then a frequency-dependent variance analysis is performed for the same data. In addition, the current simulations are also validated against measurements.⁵

II. METHODS

This section describes the mesh acquisition process, the FDTD simulation parameters, and the used voxelization algorithms. A number of simulation parameters are set such that they do not introduce additional variance in the simulated HRTFs and the possible factors affecting the captured variance are identified. The coordinate system used in this work is the vertical-polar coordinate system⁵ whereby directions are given as (azimuth, elevation) = (θ, ϕ) pairs in degrees.

A. Mesh acquisition and voxelization

A laser-scanned version of the KEMAR head and torso simulator with large pinna²¹ was used since it minimizes variation in mesh acquisition and in HRTF measurements compared to a human subject. The head mesh was obtained using a high-resolution scanner (3D Scanner Ultra HD, NextEngine, Inc., Santa Monica, CA) at maximum resolution (dimensional accuracy²² of ± 0.127 mm). The acquired mesh scan had artifacts at the inner front walls of the cavum concha that had various parts occluded especially for the left ear. Since the voxelization is expected to represent a more coarse approximation of the geometry, the mesh acquisition process can be relaxed: the head mesh is initially decimated such that the resulting two-sided²³ Hausdorff distance was 0.31 mm (evaluated with an open-source tool²⁴). Subsequently, the frontal part of the cavum concha is manually corrected for both ears such that the measured distance from the surface of the microphone (at a blocked-ear location) to the tragus is the same on the mesh as it is on the physical KEMAR (~ 9 mm). Further, the anterior part of the concha is expanded and flattened in such a way

that a 1/2 in. microphone positioned to match lateral pictures of KEMAR could fit inside the cavity. In addition, the final shape is checked such that manually fitted, independent ear-scanned meshes have a similar shape in the problematic areas. However, some inaccuracies for the mesh in the concha are likely to persist.

Since a torso affects the spectrum of the resulting HRTFs especially at low frequencies,²⁵ a less-accurate torso mesh is also acquired using Kinect and ReconstructMe software.²⁶ The resulting torso mesh is scaled, aligned, and welded to the head mesh such that the resulting mesh matches the available anthropomorphic data of KEMAR^{5,21,27} (chest breadth, shoulder breadth, tragon to shoulder), the neck height corresponds to two neck rings,²⁷ and the horizontal plane and middle of the head match the reported data about KEMAR.^{28,29} A view of the resulting mesh that was used for the voxelization algorithms can be seen in Fig. 1.

As voxelization algorithms, the two surface voxelization algorithms and one solid voxelization described by Schwarz and Seidel³⁰ are implemented: *conservative* [“CONS”—an example that can be seen in Fig. 2(a)] voxelization that creates a unique supercover¹⁹ of the geometry, *6-separating* [“6SEP” example in Fig. 2(b)] that creates a 6-separating version of the voxelized mesh that is also suitable for a second order spatial stencil and a *solid* [“SOLID” example in Fig. 2(c)] voxelization that also voxelizes the interior of the volume. Since the solid voxelization is based on the Jordan curve theorem (starting from the center of the voxel), the polyhedron mesh needs to be watertight and free of other artifacts: orphan triangles, duplicate triangles, and overlapping triangles. Since this was not the case for the laser-scanned mesh, manual corrections were done to make the mesh itself suitable for the voxelization algorithm.

B. FDTD simulation

The update scheme used is the standard rectilinear explicit update with second order temporal and spatial



FIG. 1. 3D view of the used KEMAR mesh and the coordinate system.

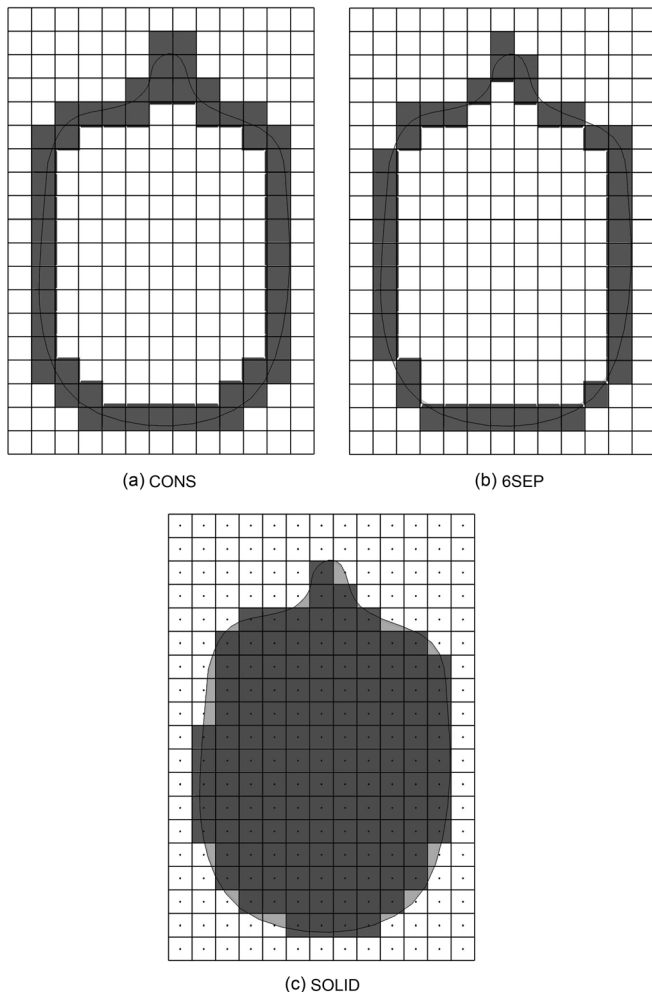


FIG. 2. Example of the voxelization algorithms used. A slice of the scanned mesh of KEMAR without ears is used with a voxel size of 13.49 mm. View from top. The slice for the solid voxelization in (c) is made watertight.

stencils run at maximum stable Courant number $\lambda = 1/\sqrt{3}$. This entails a regular cubic grid with a fixed voxel size of ΔX . To speed up simulation, the implementation is done on the Graphics Processing Unit (GPU) using CUDA kernels.³¹ The locally reacting staircase boundaries can be configured to have a frequency independent impedance using the one-sided difference operator for the spatial gradient at the boundary.³² Although the temperature during measurements is not reported,^{5,27} the temperature does not seem to drastically affect the resulting HRTFs (Ref. 18) and it is fixed to 20 °C, resulting³³ in a sound speed of 343.39 m/s. While all voxelization computations are done in double precision, all FDTD updates are done in single precision, which all have been confirmed to remain stable without any numerical accuracy problems.

The principle of reciprocity is used such that one simulation generates signals for all receivers. The source signal is a pressure injection (also referred to as “soft source”³⁴) discrete delta. The source and receiver positions in the continuous 3D space are transformed to FDTD integer lattice coordinates by setting the corresponding voxel index in which each position resides.

C. HRTF computation

Since a 3 ms long signal is enough to capture all the important attributes of HRTFs,^{35,36} the simulation domain is chosen so as no domain reflection affects a 3 ms long signal recorded 1 m from the center of the domain (excluding the propagation delay). The receiver locations are chosen 1 m away from the center of the head and the same positions as in the measurements⁵ are used with the addition of the lateral directions: $(\pm 90^\circ, 0^\circ)$. Consequently, the simulation domain is chosen as a cube of 3 m side with the center of the dummy head placed in the middle. The frontal direction matches the x axis and the interaural axis is parallel to the y axis. The mesh domain and FDTD lattice are axis-aligned and positioned such that the origin of the mesh domain corresponds to the corner of the minimum world space coordinates of the voxel with index $\{0, 0, 0\}$. The boundary for the voxelized geometry of KEMAR is set as fully rigid $\xi_{\text{KEMAR}} = 0$ and the exact admittance value is later considered as a factor in the analysis.

It is also reasonable to assume that as the voxel size decreases, the voxelization issues become less important. Therefore, two different voxel sizes are presented (and chosen such that the frequency resolution of the simulated HRTFs matched the measured ones -220.5 Hz): $\Delta X = 2.97$ mm and $\Delta X = 2.35$ mm, the latter limited by the available GPU memory. The corresponding sampling frequencies are $f_s = 200.214$ kHz and $f_s = 253.134$ kHz, respectively. Although the two chosen voxel sizes were found to be sufficient to show the influence of the voxelization process on the FDTD-simulated HRTFs, other voxel sizes are similarly analyzed to confirm the findings and to predict the behavior of the variance at even lower ΔX voxel sizes.

The computational nodes utilized each had either two Tesla M2090 GPU cards or two Tesla K40 GPU cards each having 6 GB of GPU memory or 12 GB of GPU memory, respectively. Thus, the maximum GPU memory available for a single FDTD computation was about 24 GB.

The HRTFs are obtained through the usual free-field equalization in the frequency domain¹ with a rectangular window applied to the time signals. The window size is chosen such that it matches the one in the measurements, i.e., ~ 4.53 ms. The time signals are low-pass filtered before the frequency division with a 150-tap linear-phase filter (designed with the window method) having a cutoff at 15 kHz.

III. ANALYSIS OF SIMULATIONS

The main objective is to quantify how the staircase boundaries resulting from the voxelization process in an FDTD simulation affect the sound field inside the concha. Here, this is achieved by computation of sound field variance.

A. Quantifying voxelization influence

The voxelization process entails different steps that can generate errors in the simulated sound field. First, the mesh must be super-imposed on an FDTD grid and second converted to voxels with a voxelization algorithm. For small translations of the mesh relative to the grid, the voxel-based

geometry resulting from a voxelization algorithm changes due to discretization effects. This possible algorithm-dependent error affecting the sound field in the concha can be quantified by assessing the changes in the sound field caused by small translations of the mesh. One way to quantify the sound field changes is to look at the variances in the pressure field caused by the voxelization process under similar conditions. The analysis is limited to magnitude changes in the spectrum caused by small translations of the polyhedron mesh and quantified through the use of HRTFs. No rotations are included in the study since additional variance would be introduced (either changes in HRTFs with direction if the source-receiver distances are kept constant or changes in source-receiver distances).

Albeit the voxelized geometry is composed of discrete cubic volumes, the relative position of the mesh to the FDTD grid can take any value. The variance of the magnitude of the HRTFs is quantified in the frequency domain based on the relative position of the mesh to the grid $\mathbf{X} = [x, y, z]^T$ for each frequency bin and each direction. For this, Monte-Carlo estimators are used, which are a simple and reasonable approach to a variance analysis problem.³⁷ The magnitude at each frequency bin can be written as a continuous scalar random variable y based on the continuous (up to machine epsilon) random input \mathbf{X} ,

$$|\text{HRTF}(\omega, \theta, \phi)| \triangleq y = f(\mathbf{X}) \text{ [dB]}, \quad (1)$$

where $f(\cdot)$ is deterministic. Since the position of a point relative to the grid is invariant to axial translations by ΔX , the input variable was initially chosen as uniformly distributed on the $[-\Delta X/2, \Delta X/2]^3$ cube.

Ideally, the variance should be quantified at the same location as measured HRTF, for example, at the blocked entrance of the ear canal. However, due to different voxelization algorithms, this is not possible since the source voxel may result in a solid (i.e., a boundary). In addition, for consistent comparisons, the source location voxel needs to be fixed, independent of the translation of the mesh and independent of voxelization type. This forces the location of the sound source at a certain minimal distance from any surface triangle so that even a diagonal displacement for a conservative voxelization (worst-case scenario) would not yield a solid voxel where the source has to be placed.

Therefore, for each voxel size, a sphere of radius $r = \Delta X \sqrt{3}(1 + 0.5)$ mm was manually fitted inside the cavum concha such that no mesh triangles intersected its volume, especially among the Cartesian diagonals (for example, the condition can be relaxed in axial directions to $r = \Delta X 1.5$ mm). The center of the sphere was chosen as the source location and the innermost locations satisfying such constraints are close to the exterior edge of the concha volumes, 11.87 mm (left ear) and 10.2 mm (right ear) away in the continuous space from the center of the 1/4 in. KEMAR blocked-ear microphone for $\Delta X = 2.97$ mm and 7.99 mm (left ear: see an example in Fig. 3) and 8.31 mm (right ear) away for $\Delta X = 2.35$ mm.

At such distances from the ear canal, although the scattered field inside the concha should be captured, all the localization cues are likely not.³⁸ However, it can be assumed

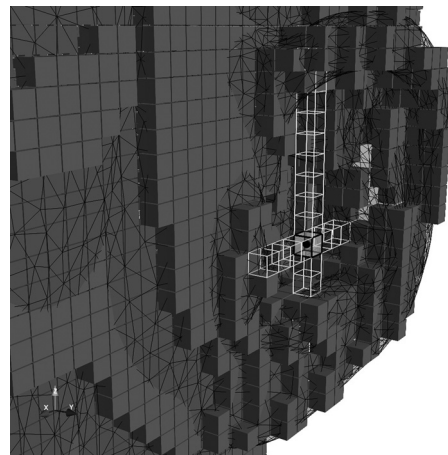


FIG. 3. Close-up of the voxel-based geometry: Left ear, $\Delta X = 2.97$ mm, SOLID. The exact location of the sound source is shown as a black dot and the corresponding voxel in white. The decimated polyhedron mesh is shown in a black wireframe and gray transparent triangles. The polyhedron mesh is in initial position (i.e., not translated relative to the grid). Two axial wireframe voxel columns are also displayed in white for ease of understanding. Note the hole in the voxel-based geometry around the antihelix.

that the variance in the sound field caused by each voxelization algorithm is similar inside the concha volume. This assumption could be confirmed by the similar variance estimates for the left and right ear since the source location relative to the ear is not equal. Nonetheless, the scanned pinnae are not identical and the left/right variance differences should not be considered an accurate indicator of a factor such as the exact source location.

In addition, the calculated free-field equalized filters in the variance analysis are referred to as HRTFs despite the fact that they are not calculated at a standard location (e.g., blocked ear). Note that although the HRTFs are calculated away from the concha walls, the ear canals of the mesh are still blocked.

Although implemented on the GPU, the simulations are rather time-consuming (~ 7.5 min for $\Delta X = 2.35$ mm, excluding voxelization for each simulation) and a reduced number of Monte-Carlo samples is desirable. The equiprobable stratified sampling method is chosen that provides estimates with lower variance³⁹ compared to uniform random sampling, independent of $f(\cdot)$: each dimension is equally divided into a number of u subintervals of length Δu creating a number of u^3 sub-volumes inside the input cube \mathbf{X} . Then, for each sub-volume of integer index $\{u_x, u_y, u_z\}$ a sample is uniformly drawn for each dimension from the corresponding subinterval (e.g., $[u_x \Delta u, (u_x + 1) \Delta u)$). A total of $u = 4$ strata per dimension with one sample per stratum are used such that the possible negative bias of the variance estimator

$$\bar{\sigma}^2 = \frac{1}{N_S} \sum_{i=1}^{N_S} (y_i - \bar{y})^2, \quad (2)$$

is bounded to $-0.016 \cdot \sigma_y^2 \leq (\mathbb{E}[\bar{\sigma}^2] - \sigma_y^2) \leq 0$ (according to McKay *et al.*³⁹) for the total of 64 strata, where $\mathbb{E}[\cdot]$ represents the expected value operator, N_S the number of samples, σ_y^2 the true variance of y , and \bar{y} the unbiased³⁹ estimator of the mean

$$\bar{y} = \frac{1}{N_S} \sum_{i=1}^{N_S} y_i. \quad (3)$$

The 64 Monte-Carlo runs (one for each stratum in the input \mathbf{X}) are referred to as a whole and named a “batch” run. Similarly, all the 64 samples in \mathbf{X} for each batch run are addressed as a “batch sample.”

The magnitude of the HRTFs varies with direction and frequency.⁴⁰ If the data are pooled over many directions or frequencies, the variance of HRTFs with direction or frequency will end up in the estimators. To exclude this variance in the conducted analysis, for each direction, frequency bin and Monte-Carlo batch run, the average magnitude is subtracted,

$$y(\omega, \theta, \phi, r) \leftarrow y(\omega, \theta, \phi, r) - \frac{1}{64} \sum_{r=1}^{64} y(\omega, \theta, \phi, r), \quad (4)$$

where r represents the Monte-Carlo run index inside a Monte-Carlo batch run.

Finally, the estimator in Eq. (2) can be calculated based on pooled data from all directions and/or frequencies. For example, a frequency-dependent variance estimator $\bar{\sigma}^2(\omega)$ is calculated over $N_S = 64 \cdot N_{\text{dir}}$ data points obtained from Eq. (4), where N_{dir} represents the total number of directions.

Although unclear if the estimated variances $\bar{\sigma}^2$ are normally distributed or if the central limit theorem can be applied to a small number of samples of the used estimator in Eq. (2), the uncertainty is quantified using a student distribution with $N_R - 1$ degrees of freedom,³⁷ where N_R represents the number of independent stratified sample batches used for the N_R -independent Monte-Carlo batch runs. The starting random seed is the same for each voxelization algorithm and each batch run. A total of $N_R = 5$ independent Monte-Carlo batch runs are used together with significance level $\alpha = 0.05$ such that the 95% confidence intervals (CIs) are reported for the final estimator

$$\bar{\sigma}^2 = \frac{1}{N_R} \sum_{b=1}^{N_R} \bar{\sigma}_b^2. \quad (5)$$

The translation distance or equivalently, the volume of the input cube \mathbf{X} , is expected to become a major factor in the variance estimates, especially with an increasing distance of translation. In addition, non-axial translations larger than a voxel side ΔX can be possible inside the used input cube \mathbf{X} $[-\Delta X/2, +\Delta X/2]^3$. Consequently, the side of \mathbf{X} is also changed to lower values than ΔX .

Although the induced variance through mesh translation is linked with both voxelization aspects (algorithm and relative mesh-grid position), for axial translations of ΔX , the changes in the sound field solely due to translation can be crudely estimated for a fixed voxel-based geometry. Therefore, for each voxelization algorithm, six axial translations from $\mathbf{X}_0 = [0, 0, 0]^T$ can be used to gain some insight into the variance caused by $\mathbf{X} \in [-\Delta X, \Delta X]^3$.

For ease of interpretation, the standard deviations are reported. The standard deviation for each Monte-Carlo batch run is obtained from the variance estimator as $\bar{\sigma} = \sqrt{\bar{\sigma}^2}$ and

then the CIs are calculated in the same manner as for the variance.

Following previous FDTD HRTF studies,¹⁶ the analysis is limited to frequencies for which the wavelength is spatially sampled at least 12 times. Therefore, the numerical dispersion is assumed to be negligible for frequencies below $f_{\text{max}} \approx 9.63$ kHz for $\Delta X = 2.97$ mm and $f_{\text{max}} \approx 12.18$ kHz for $\Delta X = 2.35$ mm. It is still noted that dispersion may affect the simulations even at the aforementioned f_{max} frequencies if the interference of planar wave packets is considered within a 1.62 ms time window when the scattering is analyzed (excluding the 1 m propagation delay from the 4.53 ms window length). The percentage of the group delay τ_g caused by numerical dispersion in the worst direction relative to the period of the frequency can be estimated as $E_{\tau_g} = \tau_g \cdot f_{\text{max}} \cdot 100$. The worst interferences appear at multipliers of $E_{\tau_g} = 50\%$. The group delay can be calculated numerically using a forward difference from the unwrapped phase of the grid transfer function.⁴¹ Thus, for a voxel of $\Delta X = 2.97$ mm, a wave packet containing a frequency of $f_{\text{max}} = 9.63$ kHz traveling in the axial direction will be delayed through numerical dispersion by $E_{\tau_g} = 38\%$ compared to the diagonal direction (unaffected by dispersion). For $\Delta X = 2.35$ mm, axial and diagonal wave packets with $f_{\text{max}} = 12.18$ kHz will almost be in antiphase ($E_{\tau_g} = 48\%$). While such scenarios are unlikely to substantially affect the not-fully-understood mechanism¹⁷ that generates spectral cues, uncertainty still persists. Therefore, the analyzed bandwidth is considered a parameter in the analysis. The dispersion caused during propagation should affect the magnitude of the simulated HRTFs to a very low degree⁴² below f_{max} .

B. Comparison to measurements

The simulations are validated against the CIPIC measurements⁵ of KEMAR with large pinna. The CIPIC database was chosen for the following reasons: it is a public database often used in HRTF-related studies, it contains a reasonable amount of directions (i.e., 1250) fairly distributed in space, a bare torso is used in the measurements that better matches the simulated hard surface impedance, and the measuring distance is 1 m, which requires a smaller simulation domain compared to larger distances.

For the purpose of comparison with measurements, the simulated HRTFs (see Sec. II C) are calculated at the blocked entrance of the ear canal. The voxel corresponding to the center of the KEMAR 1/4 in. microphone membrane at the blocked meatus is chosen as the sound source location. In case the voxel is solid due to the voxelization process, the first air voxel in the vicinity of a boundary is searched, with priority given to the outward direction (i.e., the normal to the median plane toward the ear) and then the backward direction (the sub-zero coordinates of the x axis). The position of the mesh relative to the grid is set to $\mathbf{X}_0 = [0, 0, 0]^T$ with the mesh aligned as described in Sec. II B. The FDTD HRTFs are computed for the same directions as in measurements.⁵ It is unclear below which frequency the low-frequency inaccuracies in measurements²⁷ dominate so as to restrict the bandwidth of the comparison. The lowest possible frequency for the comparison is set to 400 Hz.

As in the variance analysis, the comparison with measurements is limited to the magnitude of the HRTFs. A consistent positive gain mismatch between the simulated and the measured HRTFs is found throughout the entire analyzed frequency range. The reasons for this are unclear, but an unexpected negative magnitude is present in the measurements at low frequencies. To exclude such a difference from the error measurements, the average value of the magnitude of the diffuse-field (d.f.) HRTF is calculated for each ear and each HRTF set separately between 400 Hz and 3 kHz, and the corresponding HRTF magnitudes were shifted down in decibels by such an average. The magnitude of the d.f. HRTF is calculated as an average of all directions from the log-magnitude spectra of the HRTFs. The frequency interval over which the average magnitude is calculated is based on the inaccuracy of the measurements at low frequencies²⁷ and the frequency where the pinna starts influencing²⁵ the HRTFs at high frequencies.

The magnitude of the resulting HRTFs are finally compared for each direction and each frequency bin. For error reporting, the mean absolute error of all directions is calculated for each frequency in decibels,

$$\varepsilon(\omega) = \frac{1}{N_{\text{dir}}} \sum_{d=1}^{N_{\text{dir}}} |\text{HRTF}_S(d, \omega) - \text{HRTF}_M(d, \omega)|, \quad (6)$$

where N_{dir} represents the total number of directions, HRTF_S the magnitude-shifted FDTD-simulated HRTF, and HRTF_M the magnitude-shifted CIPIC HRTF. For each frequency bin, the standard deviation $\sigma_\varepsilon(\omega)$ is also estimated using the same estimator as in Eq. (2) calculated for all directions and each frequency bin.

For consistency with previous studies, the global unsigned error [also referred to as *spectral distortion*¹⁵ (SD)] is calculated for all of the directions and frequencies of interest ($400 \text{ Hz} \leq f \leq f_{\text{max}}$) based on the magnitude-shifted HRTFs.

IV. RESULTS

A. Artifacts caused by SOLID voxelization

The SOLID voxelization may miss thin layers of the geometry that pass between the test points. This can cause voxelized external ears with holes that may result in unreliable HRTF simulations. Therefore, the probability of a hole p_{hole} rounded to two decimal places is calculated after viewing and counting the resulting voxel-based geometries with holes. The same batch samples as in the variance analysis are used to determine the probability of holes. Since most solid voxelizations contained holes, the same voxelization sampling was done without running FDTD simulations for smaller voxels until the probability reached 0: only one hole out of the 320 total voxelizations was identified for the right ear and $\Delta X = 1.0 \text{ mm}$. Results are shown in Fig. 4 and show that SOLID voxelization is generally ill-suited for HRTF simulations unless a very small voxel is used. Note that p_{hole} depends on input cube \mathbf{X} : for $\Delta X = 2.35 \text{ mm}$, p_{hole} values drop to 0.82 (left) and 0.80 (right) when the sampling cube is restricted to 1 mm^3 . For the two analyzed voxel sizes, when

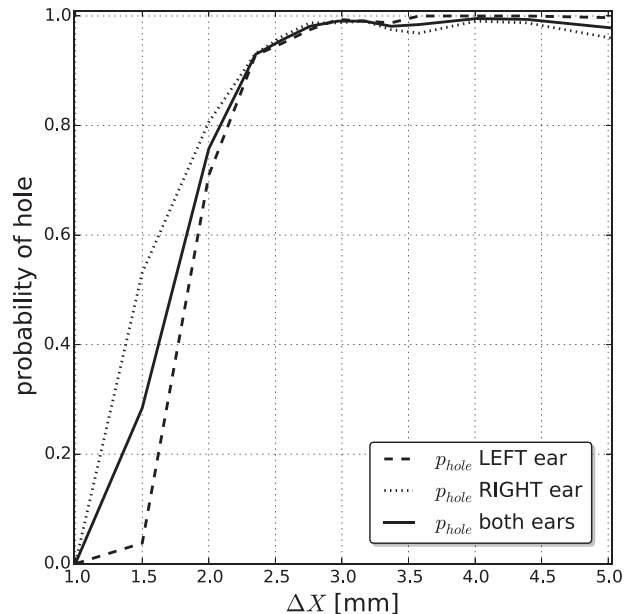


FIG. 4. Probability of a hole p_{hole} for various voxel sizes. p_{hole} is based on 320 samples for each ear and 640 samples for both ears. Translation cube $\mathbf{X} \in [-\Delta X/2, \Delta X/2]^3$.

the mesh is in initial position \mathbf{X}_0 , all SOLID voxelizations have holes except $\Delta X = 2.35 \text{ mm}$, right ear.

Holes are highly undesirable in the context of HRTF simulation and such high p_{hole} values may not justify the study of the solid voxelization algorithm for the two chosen voxel sizes. However, holes were found to greatly vary in topological complexity and location. This is expected to affect the high-frequency HRTF spectra in various degrees depending on mesh-grid misalignment. In addition, due to a different geometrical approach to the voxelization process, the analysis of the solid voxelization algorithm was found valuable to support the subsequent results. Consequently, the solid voxelization is included in the analysis and holes are considered an outcome of the algorithm.

B. Global variance

The estimated standard deviation for both ears across in directions for $\Delta X = 2.97 \text{ mm}$ and $\Delta X = 2.35 \text{ mm}$ are shown in Figs. 5(a) and 5(b), respectively. These variances/standard deviations that are calculated for all directions and frequencies are referred to as “global.”

Observing Figs. 5(a) and 5(b), numerical results show differences up to 1.3 dB in the reported global standard deviations across ears and voxelization algorithms. In addition, reducing the voxel size to 2.35 mm does not seem to drastically reduce the global standard deviation.

However, the maximal frequency of interest (i.e., f_{max}) for the smaller voxel is higher and extra variance between 9.63 and 12.18 kHz may affect the calculated estimators. For instance, if the global variance analysis for the smaller voxel $\Delta X = 2.35 \text{ mm}$ is restricted to the same frequency range as for $\Delta X = 2.97 \text{ mm}$, the standard deviation measure shows a more powerful decrease—up to a third of the original values [see Fig. 5(c)].

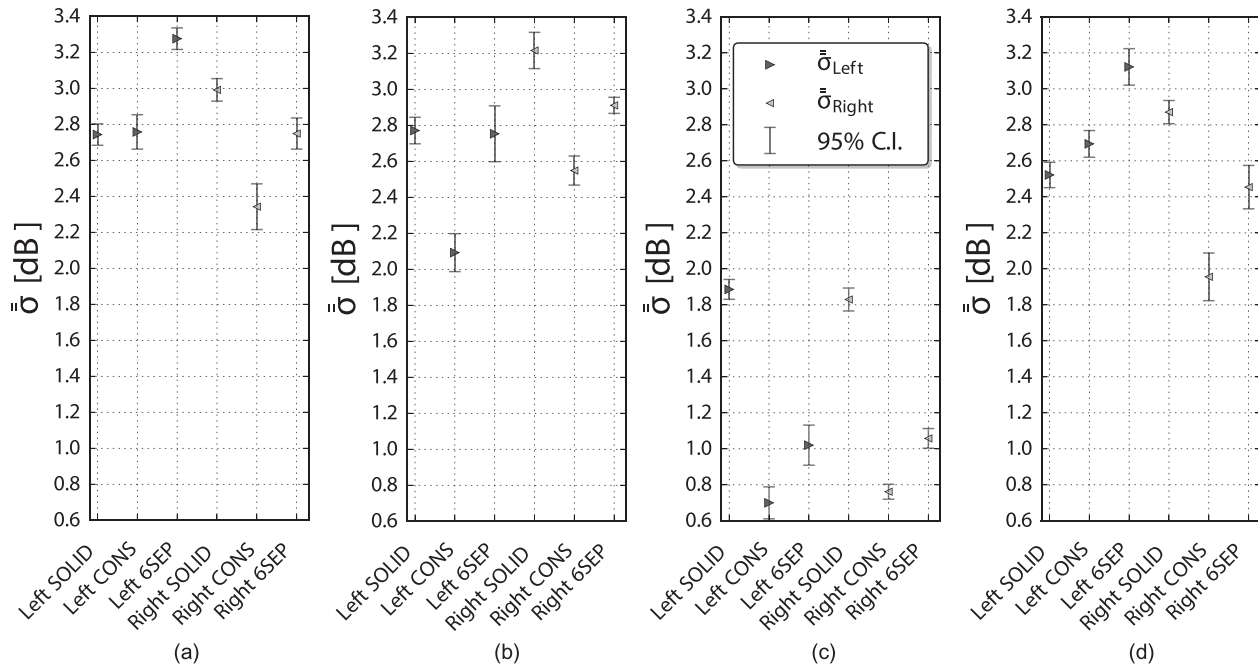


FIG. 5. Estimated standard deviation of all directions and specified frequencies for the used voxel sizes ΔX . The dimensions of the cube inside which \mathbf{X} is uniformly distributed are also mentioned. Plots (a) and (b) show the global standard deviation for the entire frequency of interest ($400 \text{ Hz} \leq f \leq f_{\max}$). Plot (c) depicts the global standard deviation for $\Delta X = 2.35 \text{ mm}$ when the analyzed frequency bandwidth is restricted to the one used for $\Delta X = 2.97 \text{ mm}$. Plot (d) depicts the global standard deviation for $\Delta X = 2.97 \text{ mm}$ when the translation cube \mathbf{X} is restricted to the one used for $\Delta X = 2.35 \text{ mm}$. (a) $\Delta X = 2.97 \text{ mm}, f \leq 9.63 \text{ kHz}, \mathbf{X} \in [2.97_{\text{mm}}]^3$. (b) $\Delta X = 2.35 \text{ mm}, f \leq 12.18 \text{ kHz}, \mathbf{X} \in [2.35_{\text{mm}}]^3$. (c) $\Delta X = 2.35 \text{ mm}, f \leq 9.63 \text{ kHz}, \mathbf{X} \in [2.35_{\text{mm}}]^3$. (d) $\Delta X = 2.97 \text{ mm}, f \leq 9.63 \text{ kHz}, \mathbf{X} \in [2.35_{\text{mm}}]^3$.

Since the translation distances in \mathbf{X} are also expected to influence the variance increase, it is likely that it will interact with the voxelization algorithm when analyzing the variance. For example, Fig. 5(d) shows for the larger voxel how the global standard deviation changes with voxelization algorithm for the same translation used in the $\Delta X = 2.35 \text{ mm}$ analysis. It can be seen that when analyzed up to f_{\max} , the translation distances do significantly affect the global standard deviations. However, the influence is much weaker compared to a factor such as the maximal frequency of interest f_{\max} [Fig. 5(c)]. Note that the exact source location is still inconsistent between the two voxel sizes.

Restricting the translation volume even further to a 1 mm cube (i.e., $\mathbf{X} \in [-0.5_{\text{mm}}, 0.5_{\text{mm}}]^3$), the global standard deviation for such small translations [Figs. 6(a) and 6(b)] drops compared to the maximal possible translation [Figs. 5(a) and 5(b)] even when analyzed up to f_{\max} . Figures 5(a) and 6(a) and Figs. 5(b) and 6(b) also suggest an interaction between the voxelization algorithm and the translation cube \mathbf{X} : for example, reducing the translation cube to 1 mm^3 for $\Delta X = 2.35 \text{ mm}$, the variance for the 6-separating voxelization and right ear becomes similar to other voxelization algorithms; by contrast, for the left ear, a significant decrease in variance can be seen for the same voxelization algorithm.

While assessing how the translation distance affects the global variance, it is still unclear how other sources of variance like the translation variance contribute to the values calculated by the estimators. Using only six axial translations of $\pm \Delta X$, a rough estimate for the translation variance is obtained (see Sec. III A) and the global standard deviations are shown in Figs. 6(c) and 6(d) for the two

voxel sizes used. No uncertainty measures can be estimated in such a case. Although the translation distance is higher in Figs. 6(c) and 6(d), it can be seen that the global variance drops below the CIs of Figs. 5(a) and 5(b) in half of the analyzed cases, while in three cases it rises marginally with values up to 0.2 dB. This suggests that the variance induced solely by translation generally affects the estimators to a smaller degree compared to voxelization-induced variance.

C. Frequency-dependent variance

Since the global variance drops when the analyzed frequency bandwidth is further restricted [Fig. 5(c)], it may be useful to analyze the variance based on frequency. Figure 7 shows the estimated standard deviation for each frequency and all directions for the maximal translation cube \mathbf{X} . A linear frequency scale is used since low frequencies are not highly affected.

It can be seen that the variation caused by slight translations of the mesh starts influencing the sound field at frequencies lower than f_{\max} . The variance is more or less independent of the voxelization algorithm utilized and voxel size at low frequencies ($\leq 5 \text{ kHz}$). Depending on voxel size, the variance starts increasing substantially around 6 kHz for $\Delta X = 2.97 \text{ mm}$ and around 8 kHz for $\Delta X = 2.35 \text{ mm}$. At higher frequencies, the variance becomes highly dependent on the voxelization algorithm and exact location in the concha (note the exact location inside the concha volume differs between ears and voxel sizes). However, the sudden increase in variance at high frequencies is consistent across the collected data. Figure 7 also explains why the global

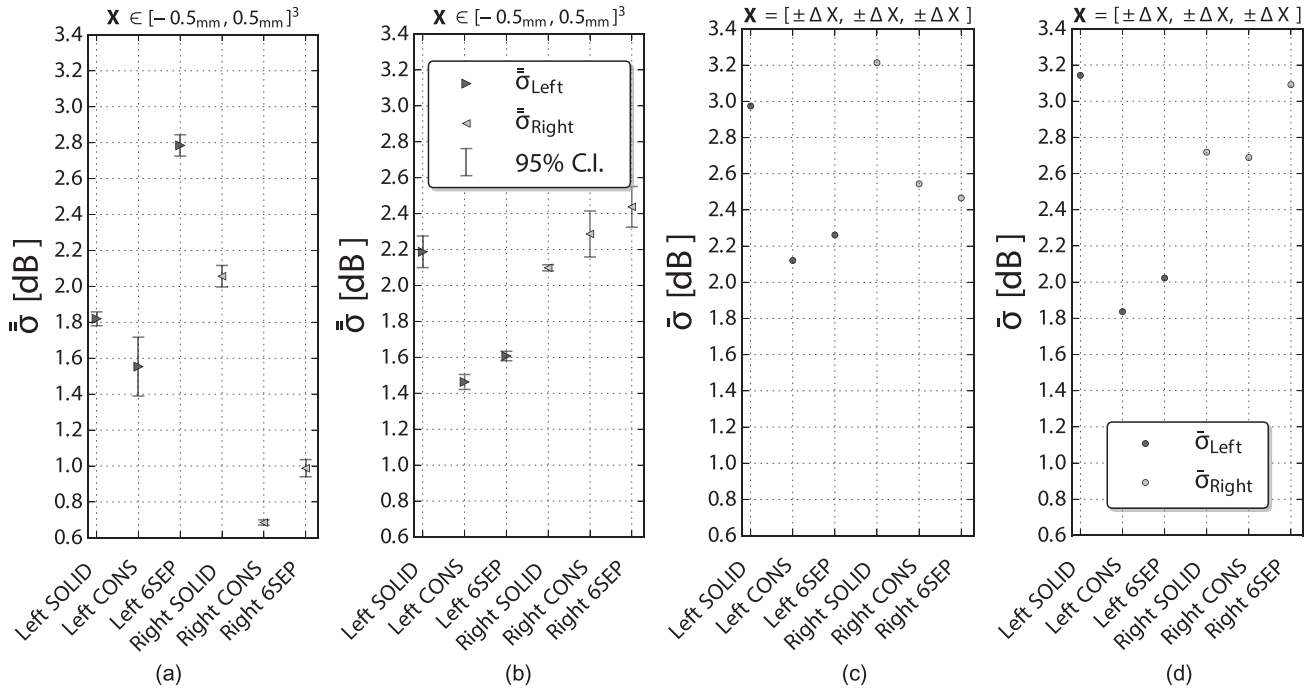


FIG. 6. Estimated standard deviation of all directions and frequencies of interest ($400\text{ Hz} \leq f \leq f_{\max}$) for the used voxel sizes ΔX . Plots (a) and (b) show the global standard deviation when $\mathbf{X} \in 1\text{ mm}^3$ cube. Plots (c) and (d) show a rough estimate of the standard deviation when the mesh is translated axially by $\pm \Delta X$. (a) $\Delta X = 2.97\text{ mm}$, $f \leq 9.63\text{ kHz}$. (b) $\Delta X = 2.35\text{ mm}$, $f \leq 12.18\text{ kHz}$. (c) $\Delta X = 2.97\text{ mm}$, $f \leq 9.63\text{ kHz}$. (d) $\Delta X = 2.35\text{ mm}$, $f \leq 12.18\text{ kHz}$.

variance values are about the same in Figs. 5(a) and 5(b): in Fig. 7 the standard deviation curves start to rise at a higher frequency at the finer spatial resolution, but high values of the standard deviation continue up to the start of the shaded area, which is also at a higher frequency at the finer spatial resolution.

Similar conclusions can be drawn when the translation cube \mathbf{X} for the large voxel $\Delta X = 2.97\text{ mm}$ is restricted as for the smaller voxel [same case study as in Fig. 5(d)]. The frequency-dependent standard deviation is shown in Fig. 8 and the sudden increase in variance slope around 6 kHz can still be observed.

So far, the variance analysis based on frequency appears to indicate that although the size of \mathbf{X} influences the global variance, the sudden increase in the slope of the variance is not affected to a high degree. On the other hand, when $\mathbf{X} \in [-0.5\text{ mm}, 0.5\text{ mm}]^3$, the frequency at which the variance increase is found can shift higher in frequency, especially for the conservative voxelization (see Fig. 9). In addition, the increasing rate of the variance appears to be affected by the size of the input cube \mathbf{X} , making the sudden increase in the slope of the variance less obvious.

Considering the standard deviation estimates for the six axial translations (i.e., for translation of fixed voxelized mesh done at \mathbf{X}_0), the frequency-dependent standard deviations are plotted in Fig. 10. The rate of change in variance lacks a clear sudden increase, showing a smoother transition compared to when the voxel-based geometry also changes (see Figs. 7 and 10). On the other hand, the frequency-dependent trend of the variance is similar, which shows that translation does induce variance *per se* that also increases with frequency and is captured in the estimators. A general increase in variance at lower frequencies compared to Figs. 7 and 8 and 9 is also seen.

Nevertheless, the axial-only translations yield smaller variances compared to Fig. 7, mostly at higher frequencies. The frequency-dependent variance reported in Fig. 7 is higher since it is a sum of translation variance, voxelization variance, and the covariance between the two. The lower variances at higher frequencies support the result that the translation-only variance does not dominate the variance estimators: despite larger translation distances, the variance is generally smaller than a combined variance from smaller translation and voxelization.

The local maxima in the frequency-dependent $\bar{\sigma}$ estimators shown in Figs. 7–9 after the sudden increase are expected to appear at HRTF features that are direction-dominant. Although the HRTF features are rather complex and vary strongly with direction, direction-dominant features should be captured in the spectrum of the d.f. HRTF (calculated as described in Sec. III B). While most of the modal frequencies were investigated close to the ear canal,⁴³ lower frequency concha modes are expected to be captured throughout the concha volume. The d.f. HRTFs for $\Delta X = 2.35\text{ mm}$ and $\Delta X = 2.97\text{ mm}$ and each of the 320 Monte-Carlo voxelization of the 5 batch runs are depicted in Fig. 11. First, for $\Delta X = 2.35\text{ mm}$, the peaks in the variance estimators in Fig. 7 generally correspond to features in the average d.f. HRTF. For $\Delta X = 2.97\text{ mm}$, the sudden increase in variance does generally match the first notch in the averaged d.f. HRTF after which a higher random error makes the match less clear. This is also consistent with the increased oscillatory shape of the estimators in Figs. 8 and 9 where the sampling space is reduced: more consistent HRTF spectra are expected for the Monte-Carlo samples in a batch run.

Observing the spectral features between the 4.2 and 9.5 kHz concha modes, Fig. 11 also shows an algorithm-dependent bias. For instance, the supercover generated by a

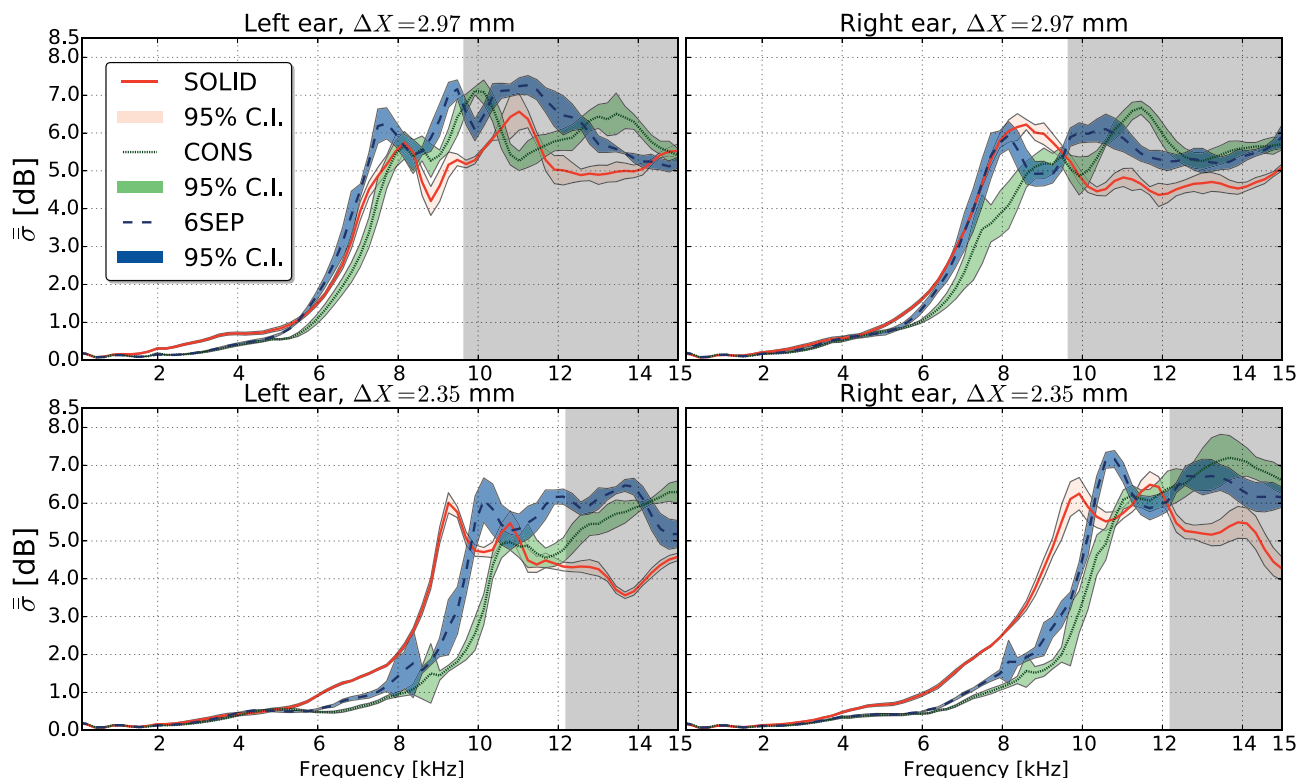


FIG. 7. (Color online) Estimated frequency-dependent standard deviation calculated for all directions. Spectrum shown for 100 Hz to 15 kHz. Gray background covers the frequency range $[f_{\max}, +\infty)$. Translation cube $\mathbf{X} \in [-\Delta X/2, \Delta X/2]^3$.

conservative voxelization generates smaller distances in the pinna volumes (or equivalently, creates less air voxels inside the cavities of the external ear) compared to other voxelization algorithms, especially the solid voxelization. This causes the resulting modes⁴⁴ or interference patterns⁴⁵ to shift up in frequency. By contrast, the solid voxelization shifts such features to lower frequencies due to probable voxel-based geometries with larger distances and the general convex geometry of the pinna. Considering the d.f. HRTF from the CIPIC measurements⁵ in Fig. 11, the two voxelization algorithms seem to induce opposite bias in the high-frequency spectral features of the HRTFs. The 6-separating voxelization induced volume bias is somewhere in between since the intersection volume is inscribed in each voxel (see work by

Laine⁴⁶) leaving parts of the polyhedron mesh close to the vertices of the Cartesian grid outside the voxelized mesh. The amount of bias depends on the voxel size: for example, the shifts in frequency for the first peak and first notch of the d.f. HRTF are higher for the bigger voxel in Fig. 11.

D. Surface impedance

The impedance value used in simulations is also expected to be a factor in the variance analysis. Since softer surfaces are expected to dampen the concha modes and interference patterns (effectively smoothing the resulting spectra in the process), lower variance estimates are expected. The frequency-dependent standard deviation for a specific acoustic

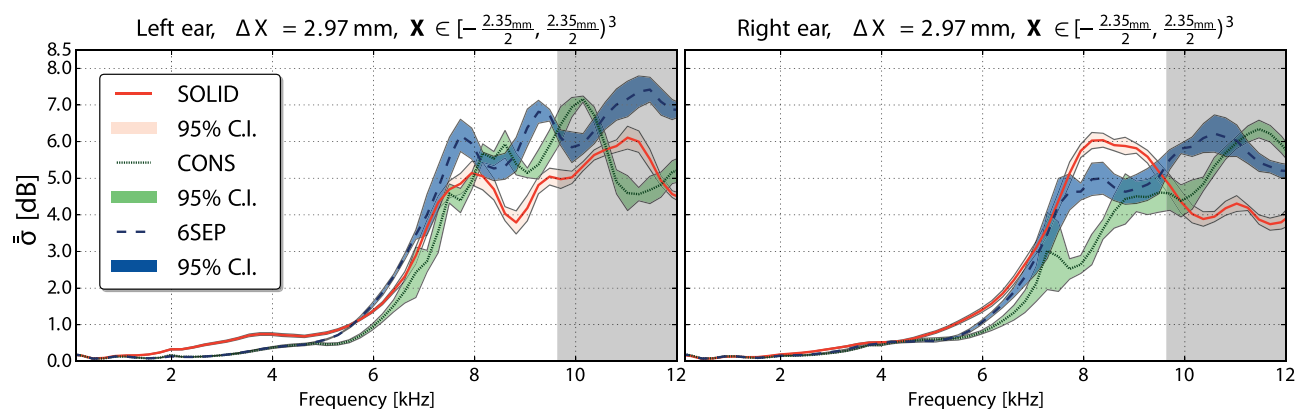


FIG. 8. (Color online) Estimated frequency-dependent standard deviation calculated for all directions with restricted translation of the mesh \mathbf{X} . Spectrum shown for 100 Hz to 12 kHz. Gray background covers the frequency range $[f_{\max}, +\infty)$.

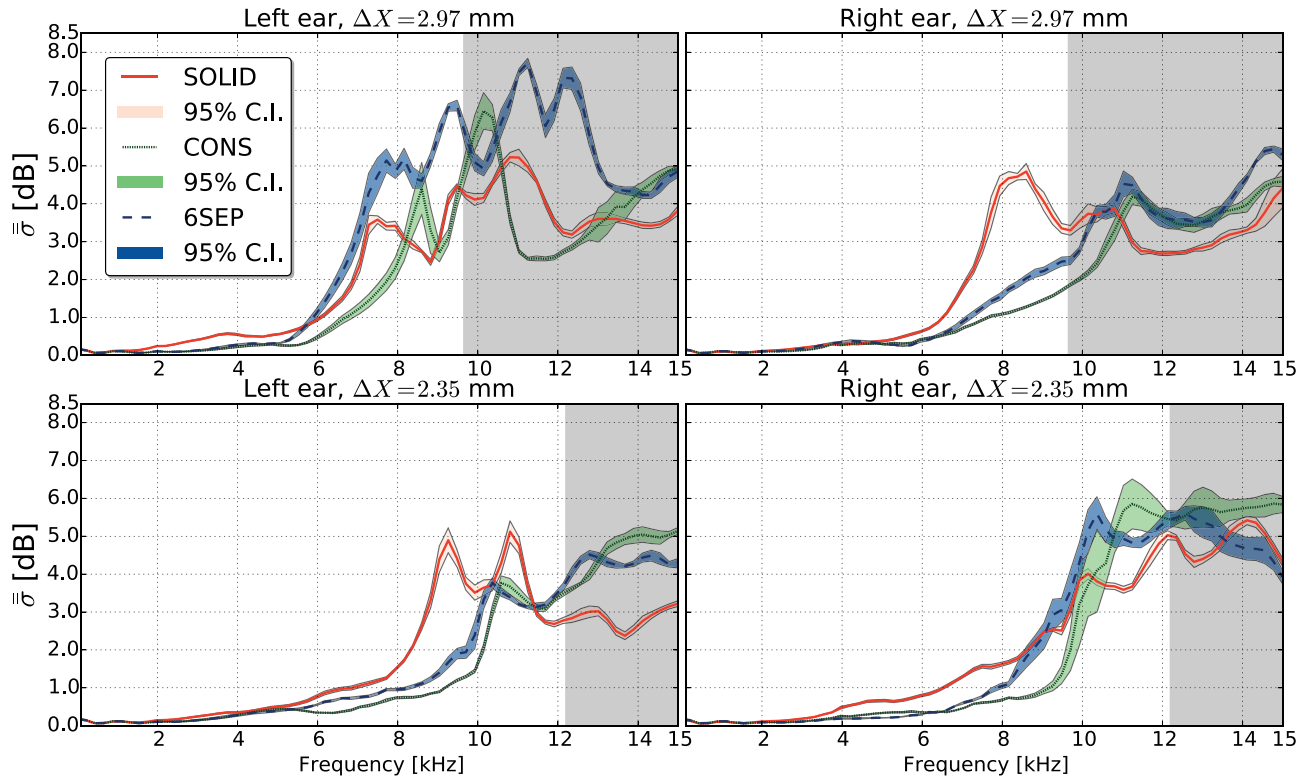


FIG. 9. (Color online) Estimated frequency-dependent standard deviation calculated for all directions. Spectrum shown for 100 Hz to 15 kHz. Gray background covers the frequency range $[f_{\max}, +\infty)$. Translation cube $\mathbf{X} \in [-0.5_{\text{mm}}, 0.5_{\text{mm}})^3$.

admittance value of $\xi_{\text{KEMAR}} = 0.015$ corresponding to the skin reflection coefficient⁴⁷ is plotted in Fig. 12. The reduction in the average frequency-dependent standard deviation from Fig. 7 is also shown. Results show a general reduction in the standard deviation with a smoother increase due to less pronounced HRTF spectral features. Although diminished, the peaks and dips in the standard deviation in Fig. 12 match the ones in Fig. 7 while the algorithm-dependent magnitude of the estimates remains in the same order. This shows that frequency-independent absorptive boundaries do only affect the random uncertainty induced by the voxelization algorithm and not the algorithm-induced bias shown in Fig. 11. Figure 12 also shows a higher decrease in standard deviation for the smaller voxel which indicates an increase in the total absorption.

E. Changes with voxel size

The first frequency f_r , at which the *random* error captured in the estimators, reaches a certain dB threshold value is defined. For each batch run, voxelization algorithm and voxel size, f_r is estimated through linear interpolation between the closest data points. By pooling the f_r data for both ears and all voxelization algorithms, the frequencies where the voxelization-induced random error reaches different threshold values can be estimated using an ordinary least squares prediction as shown in Fig. 13. Although not shown, the frequency-dependent standard deviation plots show a similar sudden increase in variance as can be seen in Fig. 7. In addition, the f_r values are strongly influenced by the bias induced by the voxelization algorithms (similar to Fig. 7), which makes the comparison of the captured variances for

each voxel difficult. The f_r frequencies increase with lower voxels and higher used thresholds and are lower than f_{\max} frequencies. Note that the curves in Fig. 13 corresponding to higher threshold values are less reliable: for example, dispersion is expected to increase.

F. Comparison with measured HRTFs

Figure 14 and Table I depict the error between the FDTD-simulated HRTFs (at blocked meatus location) and the CIPIC measurements for the two reported voxel sizes: 2.35 and 2.97 mm. The frequency-dependent error is shown in Fig. 14 where a threshold of 1 dB is considered for f_r . The error is small (≤ 2 dB) and roughly constant up to 4 kHz, after which it starts increasing gradually for most cases and remains at a higher value up to f_r . The variance of the reported error is highly similar among all analyzed voxelizations and voxel sizes: it is approximately constant up to 5.3 kHz, after which it starts increasing. The increase in error and the variance of the error above 4 kHz is likely due to the volume bias induced by the voxelization algorithms. The error may also have been inflated by an impedance mismatch. No reliable analysis can be conducted at higher frequencies (above the reported f_r frequencies) since the voxelization uncertainty dominates: slightly translating the mesh can easily cause modal shifts and affect the reported error.

The SD calculated up to f_{\max} are shown in Table I. Compared to previous studies,^{15,16} the reported SD are of similar magnitude. Since the most substantial contribution in error can be seen at high frequencies (Fig. 14), any difference is likely due to voxelization uncertainty. The solid

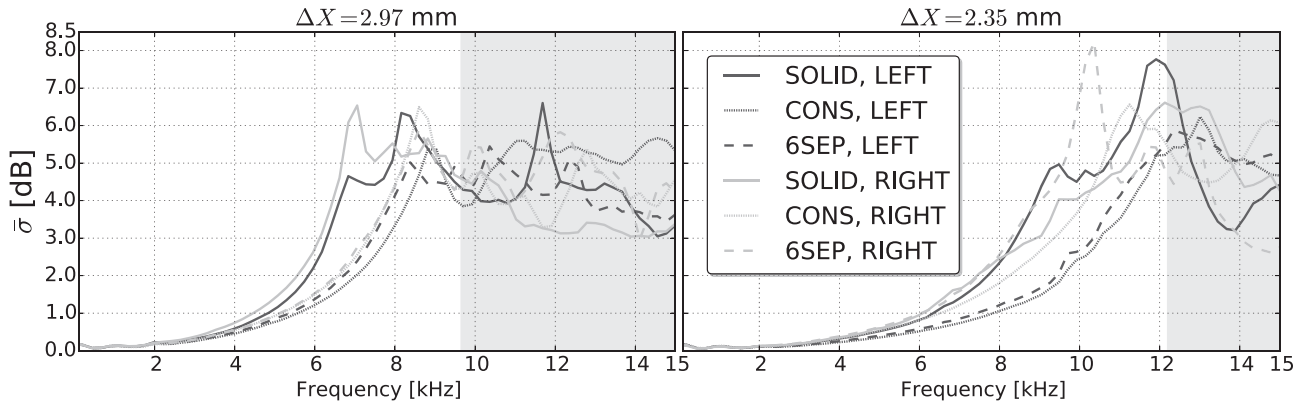


FIG. 10. Rough estimate of the change in the sound field for six axial translations of $\pm\Delta X$ and fixed voxel-based geometry (i.e., $\mathbf{X} = [\pm\Delta X, \pm\Delta X, \pm\Delta X]^T$). Gray background covers the frequency range $[f_{\max}, +\infty)$. Spectrum shown for 100 Hz to 15 kHz.

voxelization shows the smallest error for $\Delta X = 2.35$ mm, however such algorithm generates unrealistic holes in the pinna and most of the difference comes from frequencies above 6 kHz where it was shown that the voxelization process highly influences the result depending on mesh-grid alignment \mathbf{X} . Moreover, for $\Delta X = 2.35$ mm, since the frequency bandwidth for SD computation is larger, average results reported in Table I are even more influenced by the mesh-grid alignment \mathbf{X} . Figure 14 also provides some evidence that the topological changes caused by voxelization also play a role in the simulated HRTFs: the conservative voxelization has an additional error of approximately 1.5 dB compared to the 6-separating

voxelization for $\Delta X = 2.97$ mm, a value higher than the estimated voxelization standard deviation below f_r (see Fig. 7). The differences become insignificant for the smaller voxel (lower plots in Fig. 14). Note also the increase in error above 4 kHz especially for the 6-separating voxelization when the voxel size is decreased.

V. DISCUSSION

A. Errors introduced by voxelization

There are two types of error introduced by the voxelization process in the simulated HRTFs: the relative position of the mesh causes precision issues in the HRTF spectra (in the

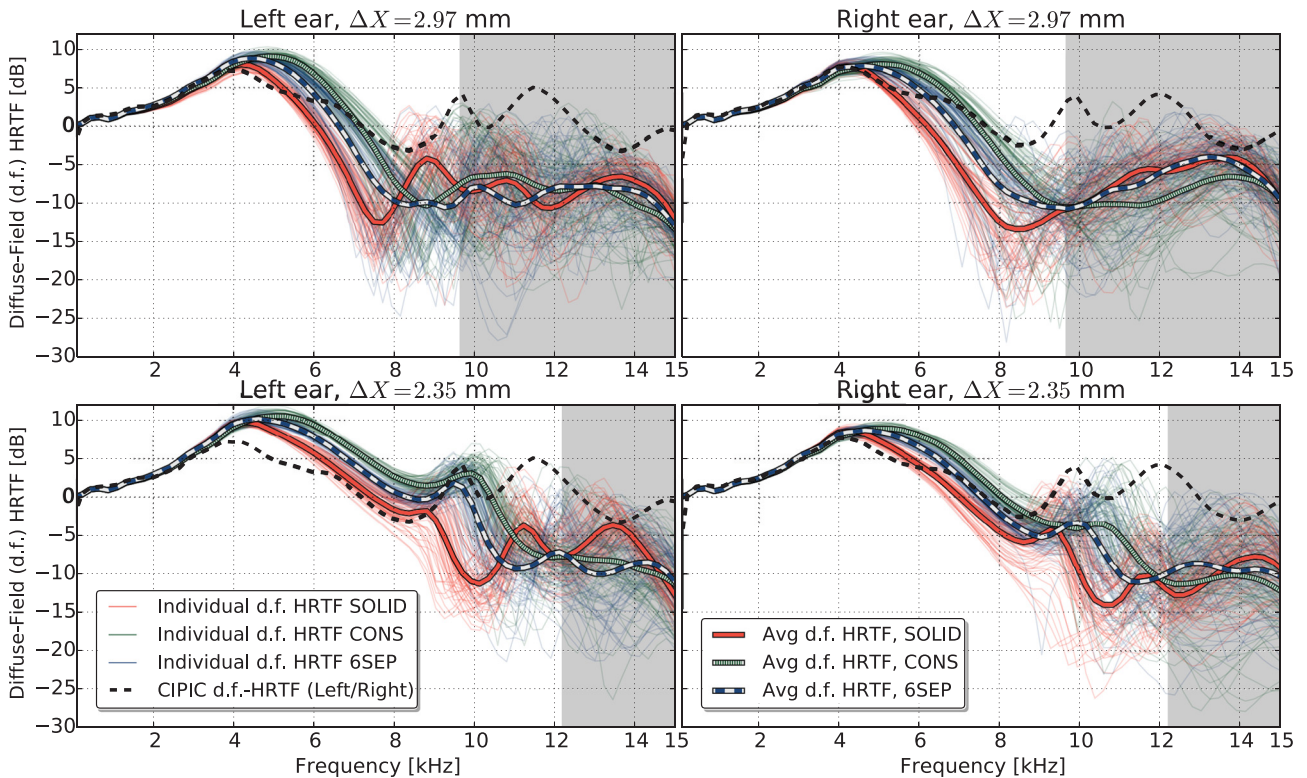


FIG. 11. (Color online) d.f. HRTFs calculated inside the concha volume for all of the 320 voxel-based geometries resulted from the 5 Monte-Carlo batch runs. Thick lines represent the averages of such d.f. HRTFs for each voxelization algorithm. The magnitude-shifted (see Sec. III B) CIPIC d.f. HRTF are also displayed. Translation cube $\mathbf{X} \in [-\Delta X/2, \Delta X/2]^3$. Gray background covers the frequency range $[f_{\max}, +\infty)$. Spectrum shown for 100 Hz to 15 kHz.

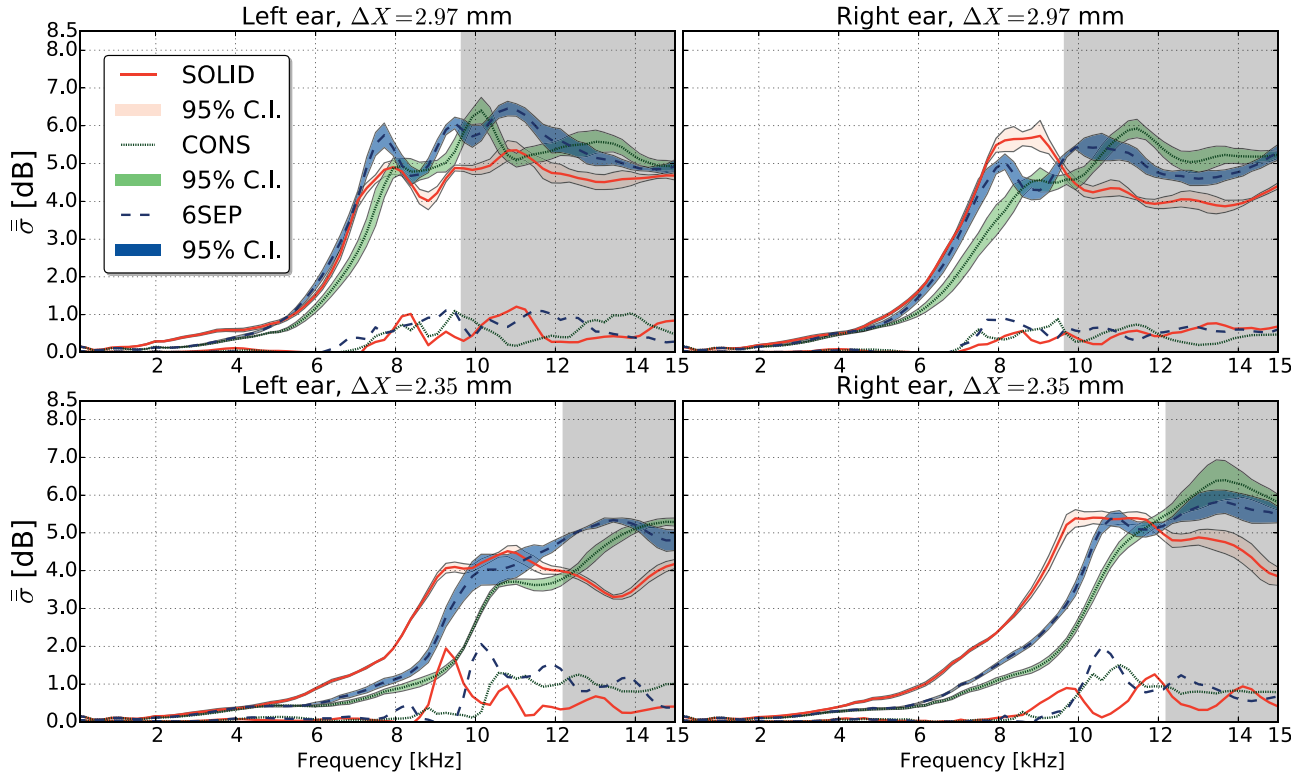


FIG. 12. (Color online) Frequency-dependent estimated standard deviation calculated for all directions for absorptive boundaries ($\xi_{\text{KEMAR}} = 0.015$). Lines without CI represent the difference between the averages for fully rigid boundaries (Fig. 7) and current averages. Spectrum shown for 100 Hz to 15 kHz. Gray background covers the frequency range $[f_{\text{max}}, +\infty)$. Translation cube $\mathbf{X} \in [-\Delta X/2, \Delta X/2]^3$.

present paragraph, accuracy and precision are referred to in a general sense in the context of resulting HRTF spectra, such as the taxonomy by Heffner and Heffner⁴⁸) that can be treated as a random error, while each voxelization algorithm causes biases in the simulated HRTF spectra due to corresponding topological biases of the resulting voxel-based geometry. The induced random error in the present study is generated by (i) the voxelization process, (ii) the translation, and (iii) the interaction of the two. The last two were shown not to dominate. The random error manifests itself in two ways in the HRTF spectra: first, the peaks and dips in the HRTFs get shifted in frequency (as previously reported⁴⁴) and second, the spectrum is shifted locally in amplitude (for instance, in Fig. 11). The frequency shifts cause the most severe errors and are the reason for the sudden increase in the captured estimators with frequency due to the onset of local frequency shifts (as seen, e.g., in Fig. 11). The algorithm-dependent bias is caused by a volumetric error of the voxel-based geometry and manifests itself as a general shift in frequency of HRTF spectral features: while the supercover of the conservative voxelization shifts the HRTF features up in frequency, the solid voxelization shifts such features down in frequency. Consequently, the bias depends on the voxel size, as shown in Fig. 11. It is noted that the voxelization bias is solely a volume bias for rigid boundaries. In the case of absorbing boundaries, the voxelization process also induces a surface mismatch⁴⁹ that will bias the simulated HRTF magnitudes compared to measurements with similar impedance values. Nevertheless, absorbing boundaries do reduce the amount of random error induced by the voxelization process, as shown in Fig. 12.

A volume bias in simulations compared to measurements is expected even when $\Delta X \rightarrow 0$ due to mismatches between the polyhedron mesh and the real geometry (such as in the present study due to laser scanning artifacts or due to

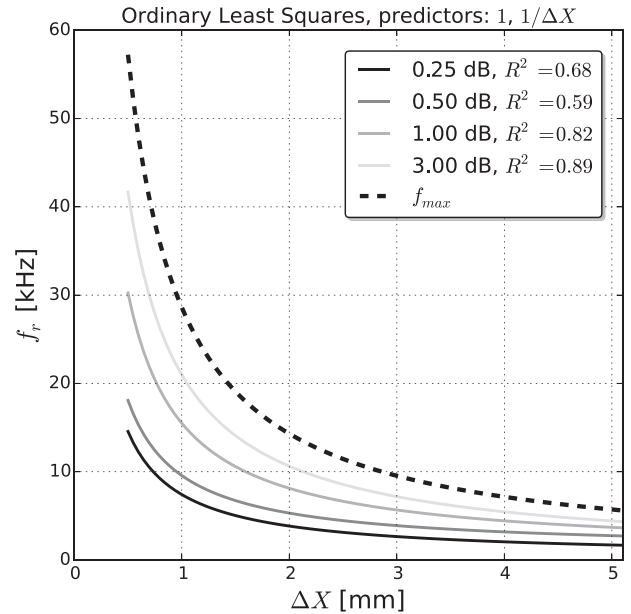


FIG. 13. Estimated and predicted frequencies f_r at which the random error reaches certain dB threshold values. $\Delta X \in \{2.35, 2.55, 2.76, 2.97, 3.17, 3.37, 3.58, 4.01, 4.39, 5.03\}$ mm. The curves are based on the pooled individual f_r values for each ear, voxelization algorithm, and batch run. Translation cube $\mathbf{X} \in [-\Delta X/2, \Delta X/2]^3$. Note that the source location is outside a tragus-helix-lobule plane from $\Delta X = 3.58$ mm. The f_{max} frequency that corresponds to 12 samples per wavelength is also displayed.

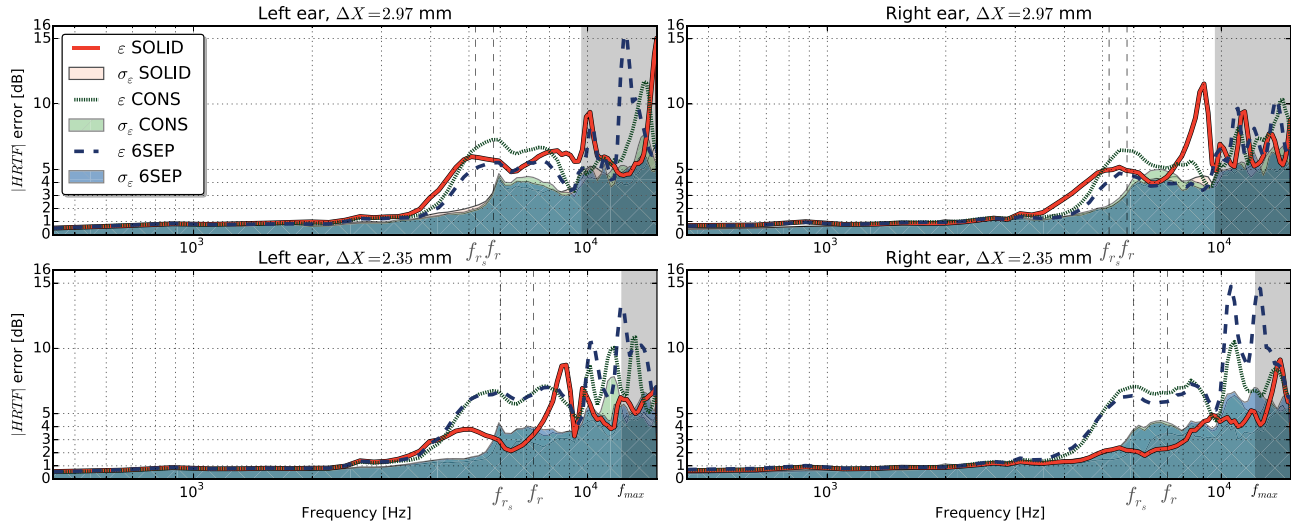


FIG. 14. (Color online) Frequency-dependent unsigned error between the simulated and measured HRTFs based on magnitude-shifted (see Sec. III B) log-spectra. The standard deviation of the error in all directions for each frequency bin in decibels is also displayed. HRTFs are simulated and measured at the blocked entrance of the ear canal. Spectrum shown for 441 Hz to 15 kHz. Gray background covers the frequency range $[f_{\max}, +\infty)$. f_r represents the frequency at which the voxelization process starts inducing a random error higher than 1 dB (due to close values and the ease of understanding the average f_r between 6SEP and CONS voxelizations is shown). $f_{r'}$ represents the same frequency as f_r , but for the solid voxelization.

exact fitting of the KEMAR silicone ears¹⁶). Such bias would interact with the voxelization bias as suggested by Fig. 14 (solid voxelization with $\Delta X = 2.35$ mm): the scanned mesh was likely not expanded enough, biasing the HRTF spectra upward in frequency while the bias of the solid voxelization shifts the HRTF features down in frequency. In addition, impedance mismatches may also add to the complexity of the comparison. This is why it is hard to quantify the magnitude of the algorithm-dependent volume bias in the present work (as compared to, for example, measurements). Nevertheless, as the voxel size approaches zero, the volume bias should converge to the one induced by the polyhedron mesh utilized.

B. Frequency dependence

Overall, results show that the global standard deviation analyzed up to f_{\max} differs slightly with ear and voxelization algorithm for the two analyzed voxel sizes. Differences are accentuated when the bandwidth of analysis is reduced. Possible causes for differences between ears can be asymmetries in the scanned geometry, inconsistent mesh corrections at the left/right concha volume, small number of samples used in estimating the CI, or different source location relative to the ear (especially for $\Delta X = 2.97$ mm). In addition, analyzing the introduced variance based on frequency shows similar behavior, with a sudden increase in variance at a

frequency that generally depends on voxel size, voxelization algorithm, and amount of mesh-grid mismatch.

When quantifying the amount of introduced random error in the simulated HRTFs in Fig. 13, it was found that the first frequency f_r , where the random error is limited to various threshold values, varies with voxel size. It is hard to quantify which voxelization algorithm generates less random error since the induced volumetric bias affects the estimated f_r frequencies and the comparison should be done for specific HRTF features. Nevertheless, the curves in Fig. 13 could be used as a general indicator of the random error introduced by the voxelization process.

Results generally show that the axis-aligned staircase boundaries induce three frequency regions in the simulated HRTF spectra: at very low frequencies, the simulated HRTFs are insignificantly influenced by the staircase approximation and the simulated HRTF spectra is in the asymptotic range; as the frequency increases the volume and surface bias created by the voxelization algorithm start dominating the simulated HRTF spectra with limited, algorithm-dependent random error. Then finally, the random voxelization-induced error dominates the HRTF spectra in the highest frequency region. These regions depend on voxel size, as can be seen from Figs. 11 and 13.

Solid voxelization tends to cause more variance at lower frequencies (i.e., $\leq f_r$) and less variance at higher frequencies. It is unclear why the solid voxelization generates less variance at high frequencies—it seems that the d.f. HRTFs are more consistent for this voxelization algorithm compared to the other two (see Fig. 11). Nevertheless, the high probability of holes in the resulting voxel-based geometries may be responsible: some direction-dependent HRTF features at high frequencies may be missing. At low frequencies, the volume bias pushes the random error toward lower frequencies and the estimated translation variance for the solid voxelization in Figs. 10 and 6(c) and 6(d) mostly appears to be higher. The

TABLE I. Average SD calculated in dB for all directions and frequencies: $f \leq 12.18$ kHz for $\Delta X = 2.35$ mm and $f \leq 9.63$ kHz for $\Delta X = 2.97$ mm.

Voxel size	$\Delta X = 2.35$ mm			$\Delta X = 2.97$ mm		
	SOLID	6SEP	CONS	SOLID	6SEP	CONS
Left ear [dB]	3.46	4.99	4.80	3.80	3.18	3.78
Right ear [dB]	2.57	5.36	4.83	3.91	2.64	3.39

results shown in Fig. 4 also suggest that unless a very small voxel is used, the solid voxelization may generate unrealistic HRTF spectra at high frequencies. In addition, any voxelization algorithm based on the Jordan curve theorem is the most sensitive algorithm to mesh artifacts and additional work is required to clean such artifacts of mesh acquisition.

C. Further analysis

Even in the continuous space, translating the geometry would create small changes in the sound field captured at a fixed location. This also generates additional variance of the sound field that is expected to inflate the calculated estimates. Unfortunately, this variance cannot be quantified and excluded from the current experimental method. It is also unclear how much this variance depends on the magnitude of the translation. Previous studies investigated the changes in the sound field at the entrance of the ear canal due to general head movements,² which is not necessarily related to the sound field variance in the concha caused by translations with a fixed receiver. Nevertheless, translation variance is expected especially at high frequencies: the translation problem is similar (yet not identical) to changing source location in HRTF simulations, which has already been reported to cause changes in the magnitude of the simulated field¹⁵ especially at high frequencies.¹⁸ In the context of FDTD simulations, the simulation parameters influencing such translation variance, together with possible interactions and covariances, are unknown. Still, results tend to show that the translation-induced variance and the covariance between translation-induced and voxelization-induced variance does not dominate the reported estimates.

The current framework was designed for a second-order spatial stencil operating on a Cartesian lattice. Other schemes could be studied as long as the voxelization algorithm matches the spatial update of the scheme at the boundary to avoid using pressure values on both sides of the auricle. For example, schemes employed on a non-Cartesian lattice (see for example, the work by Hamilton and Bilbao⁵⁰) would need a non-cubic voxelization. Nevertheless, any update could be used in the interior of the domain. Since higher-order spatial stencils are generally introduced to reduce dispersion⁵¹ and since voxelization errors are mainly caused by induced geometrical inaccuracies that dominate the HRTF spectra at frequencies with limited dispersion errors, similar results are expected for all schemes that use Cartesian grids as long as no additional error is introduced at the boundary.

Although a decrease in error relative to the measurement is generally expected as the voxel size is decreased, this is not always true as shown for the 6-separating voxelization above 5 kHz in Fig. 14. The bias created by the voxelization algorithm is partly responsible for such an increase. An impedance mismatch is also expected to influence the comparison since (within a finite volume interpretation²⁰ of the FDTD update) the amount of absorption given by the surfaces of the voxelized geometry can increase with smaller voxel size.⁴⁹ The error increase due to a higher impedance mismatch for the lower voxel is supported by the increase in absorption (equivalently, in surface bias) for the smaller

voxel in Fig. 12. Although most studies tend to indicate that impedance is not a major factor in the spectra of the HRTFs,^{10,18,21,52} some works support, to some extent, the importance of impedance in HRTF spectra both quantitatively^{53,54} and perceptually.⁵⁵

Since the volume bias is expected to decrease with voxel size, a grid convergence study (such as Roache's Grid Convergence Index⁵⁶) relevant to the HRTF problem should be developed and conducted to reveal any voxelization bias. Although convergence is expected for the volume bias, the error created by a surface bias cannot be eliminated⁴⁹ for absorptive boundaries. A superior modeling technique for the boundaries such as the fitted boundary cells²⁰ may provide a general solution to both topological biases. Further validation of the simulation with such boundaries over an unstructured grid is needed for a complex geometry such as the human pinna. Furthermore, it is hypothesized that the voxelization process is not such an important factor in simulations with simpler geometries and rigid boundaries. However, more complex geometries are expected to be affected to a greater extent by the voxelization process, as long as the voxel size is comparable to their topological characteristics.

VI. CONCLUSION

In the context of an FDTD HRTF simulation and staircase-like boundaries, results show that for complex geometries, errors are introduced in the FDTD-simulated sound fields solely due to the voxel-based, axis-aligned boundaries at frequencies lower than assumed accuracy limits.

Results show that a staircase approximation of the pinna boundaries can be highly problematic for the simulated high-frequency HRTF spectra. Both volume and surface errors are introduced by the voxelization process, which in turn bias the resulting spectra. In addition, the exact mesh-grid alignment adds random error to such bias in the HRTF spectra, dominating the simulated sound field at higher frequencies.

Considering the analyzed voxelization algorithms, the solid voxelization is the most problematic since it has a high probability of creating holes for voxel sizes greater than 1 mm. Conservative and 6-separating algorithms provide voxelizations that are more suitable for HRTF simulation but none of them is ideal as they all introduce some bias and shift the high HRTF spectra in frequency.

ACKNOWLEDGMENTS

This research has received funding from the Academy of Finland, Project No. 265824. It was also supported by the research project "PADVA—Personal Auditory Displays for Virtual Acoustics" (Project No. CPDA135702) of the University of Padova. The computational resources provided by the Aalto Science-IT project and the CSC-IT Center for Science, Espoo, Finland are also acknowledged.

¹H. Møller, "Fundamentals of binaural technology," *Appl. Acoust.* **36**(5), 171–218 (1992).

²T. Hirahara, H. Sagara, I. Toshima, and M. Otani, "Head movement during head-related transfer function measurements," *Acoust. Sci. Technol.* **31**(2), 165–171 (2010).

- ³D. Toledo and H. Møller, "Issues on dummy-head HRTFs measurements," in *126th Audio Engineering Society Convention*, Munich, Germany (2009), pp. 7727:1–7727:11.
- ⁴X.-L. Zhong and B.-S. Xie, "Consistency among the head-related transfer functions from different measurements," *Proc. Mtgs. Acoust.* **19**, 050014 (2013).
- ⁵V. Algazi, R. Duda, D. Thompson, and C. Avendano, "The CIPIC HRTF database," in *Proceedings of the IEEE WASPAA 2001*, New Paltz, NY (2001), pp. 99–102.
- ⁶J. G. Bolaños and V. Pulkki, "HRIR database with measured actual source direction data," in *133th Convention of the Audio Engineering Society*, San Francisco, CA (2012).
- ⁷F. Wightman and D. Kistler, "Measurement and validation of human HRTFs for use in hearing research," *Acta Acust. Acust.* **91**(3), 429–439 (2005).
- ⁸A. Andreopoulou, D. Begault, and B. Katz, "Inter-laboratory round robin HRTF measurement comparison," *IEEE J. Sel. Top. Signal Process.* **9**(5), 895–906 (2015).
- ⁹B. F. G. Katz, "Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation," *J. Acoust. Soc. Am.* **110**(3), 2440–2448 (2001).
- ¹⁰B. F. G. Katz, "Boundary element method calculation of individual head-related transfer function. II. Impedance effects and comparisons to real measurements," *J. Acoust. Soc. Am.* **110**(5), 2449–2455 (2001).
- ¹¹Y. Kahana and P. A. Nelson, "Boundary element simulations of the transfer function of human heads and baffled pinnae using accurate geometric models," *J. Sound Vib.* **300**(3–5), 552–579 (2007).
- ¹²T. Huttunen, E. T. Seppälä, O. Kirkeby, A. Kärkkäinen, and L. Kärkkäinen, "Simulation of the transfer function for a head-and-torso model over the entire audible frequency range," *J. Comput. Acoust.* **15**(4), 429–448 (2007).
- ¹³A. Meshram, R. Mehra, and D. Manocha, "Efficient HRTF computation using adaptive rectangular decomposition," in *Proceedings of the AES 55th International Conference*, Helsinki, Finland (2014), pp. 2-2:1–2-2:8.
- ¹⁴T. Xiao and Q. H. Liu, "Finite difference computation of head-related transfer function for human hearing," *J. Acoust. Soc. Am.* **113**(5), 2434–2441 (2003).
- ¹⁵P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, "Comparison of simulated and measured HRTFs: FDTD simulation using MRI head data," in *123rd Convention of the Audio Engineering Society*, New York (2007), pp. 7240:1–7240:12.
- ¹⁶P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, "Acoustic simulation of KEMAR's HRTFs: Verification with measurements and the effects of modifying head shape and pinna concavity," in *Proceedings of the International Working Principles and Applications of Spatial Hearing (IWPASH)*, Zao, Japan (2009).
- ¹⁷H. Takemoto, P. Mokhtari, H. Kato, R. Nishimura, and K. Iida, "Mechanism for generating peaks and notches of head-related transfer functions in the median plane," *J. Acoust. Soc. Am.* **132**(6), 3832–3841 (2012).
- ¹⁸W. Kreuzer, P. Majdak, and Z. Chen, "Fast multipole boundary element method to calculate head-related transfer functions for a wide frequency range," *J. Acoust. Soc. Am.* **126**(3), 1280–1290 (2009).
- ¹⁹D. Cohen-Or and A. Kaufman, "Fundamentals of surface voxelization," *Graph. Models Image Process.* **57**(6), 453–461 (1995).
- ²⁰S. Bilbao, "Modeling of complex geometries and boundary conditions in finite difference/finite volume time domain room acoustics simulation," *IEEE Trans. Audio, Speech, Lang. Process.* **21**(7), 1524–1533 (2013).
- ²¹M. D. Burkhard and R. M. Sachs, "Anthropometric manikin for acoustic research," *J. Acoust. Soc. Am.* **58**(1), 214–222 (1975).
- ²²NextEngine, "NextEngine 3d scanner technical specifications" (2015), URL <http://www.nextengine.com/assets/pdf/scanner-techspecs.pdf> (Last viewed September 26, 2015).
- ²³P. Cignoni, C. Rocchini, and R. Scopigno, "Metro: Measuring error on simplified surfaces," *Comput. Graph. Forum* **17**(2), 167–174 (1998).
- ²⁴P. Cignoni, <http://meshlab.sourceforge.net> (Last viewed April 1, 2015).
- ²⁵V. R. Algazi, C. Avendano, and R. O. Duda, "Elevation localization and head-related transfer function analysis at low frequencies," *J. Acoust. Soc. Am.* **109**(3), 1110–1122 (2001).
- ²⁶C. Heindl and C. Kopf, <http://reconstructme.net> (Last viewed September 10, 2014).
- ²⁷V. R. Algazi, "Documentation for the UCD HRIR files," Technical Report, CIPIC Interface Laboratory—University of California at Davis (1998).
- ²⁸G. F. Kuhn and E. D. Burnett, "Acoustic pressure field alongside a manikin's head with a view towards *in situ* hearing-aid tests," *J. Acoust. Soc. Am.* **62**(2), 416–423 (1977).
- ²⁹M. D. Burkhard, "Chapter 3. Anthropometric manikin for acoustic research, supplementary design information," in *Manikin Measurements: Conference Proceedings* (Industrial Research Products, Inc., Itasca, IL, 1978), pp. 13–16.
- ³⁰M. Schwarz and H.-P. Seidel, "Fast parallel surface and solid voxelization on GPUs," *ACM Trans. Graph.* **29**(6), 179:1–179:10 (2010).
- ³¹J. Saarelma and L. Savioja, "An open source finite-difference time-domain solver for room acoustics using graphics processing units," in *Proceedings Forum Acusticum*, Krakow, Poland (2014).
- ³²C. Webb and S. Bilbao, "Computing room acoustics with CUDA—3D FDTD schemes with boundary losses and viscosity," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Prague, Czech Republic (2011), pp. 317–320.
- ³³L. E. Kinsler, F. R. Austin, C. B. Alan, and S. V. James, "Chapter 5: The acoustic wave equation and simple solutions," in *Fundamentals of Acoustics*, 4th ed. (Wiley, New York, 2000), pp. 120–121.
- ³⁴J. Sheaffer, M. v. Walstijn, and B. Fazenda, "Physical and numerical constraints in source modeling for finite difference simulation of room acoustics," *J. Acoust. Soc. Am.* **135**(1), 251–261 (2014).
- ³⁵J. Huopaniemi, N. Zacharov, and M. Karjalainen, "Objective and subjective evaluation of head-related transfer function filter design," *J. Audio Eng. Soc.* **47**(4), 218–239 (1999).
- ³⁶J. Sandvad and D. Hammershøi, "Binaural auralization, comparison of FIR and IIR filter representation of HIRs," in *96th Convention of the Audio Engineering Society*, Amsterdam, The Netherlands (1994), pp. 3862:1–3862:16.
- ³⁷J. C. Helton and F. J. Davis, "Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems," *Reliab. Eng. Syst. Safe.* **81**(1), 23–69 (2003).
- ³⁸D. Hammershøi and H. Møller, "Sound transmission to and within the human ear canal," *J. Acoust. Soc. Am.* **100**(1), 408–427 (1996).
- ³⁹M. D. McKay, R. J. Beckman, and W. J. Conover, "Comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics* **21**(2), 239–245 (1979).
- ⁴⁰J. Blauert, "Chapter 2.2: The sound field at the two ears," in *Spatial Hearing: The Psychophysics of Human Sound Localization, Revised Edition* (MIT Press, Cambridge, MA, 1997), pp. 63–74.
- ⁴¹J. Schneider and C. Wagner, "FDTD dispersion revisited: Faster-than-light propagation," *IEEE Microwave Guided Wave Lett.* **9**(2), 54–56 (1999).
- ⁴²A. Southern, D. T. Murphy, T. Lokki, and L. Savioja, "The perceptual effects of dispersion error on room acoustic model auralization," in *Proceedings Forum Acusticum*, Aalborg, Denmark (2011), pp. 1553–1558.
- ⁴³Y. Kahana and P. A. Nelson, "Numerical modelling of the spatial acoustic response of the human pinna," *J. Sound Vib.* **292**(1–2), 148–178 (2006).
- ⁴⁴P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, "Acoustic sensitivity to micro-perturbations of KEMAR's pinna surface geometry," in *Proceedings of the 20th International Congress on Acoustics (ICA'10)*, Sydney, Australia (2010).
- ⁴⁵P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, "Pinna sensitivity patterns reveal reflecting and diffracting surfaces that generate the first spectral notch in the front median plane," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Prague, Czech Republic (2011), pp. 2408–2411.
- ⁴⁶S. Laine, "A topological approach to voxelization," *Comput. Graph. Forum* **32**(4), 77–86 (2013).
- ⁴⁷B. F. G. Katz, "Acoustic absorption measurement of human hair and skin within the audible frequency range," *J. Acoust. Soc. Am.* **108**(5), 2238–2242 (2000).
- ⁴⁸H. E. Heffner and R. S. Heffner, "The sound-localization ability of cats," *J. Neurophysiol.* **94**(5), 3653–3655 (2005).
- ⁴⁹S. Bilbao, B. Hamilton, J. Botts, and L. Savioja, "Finite volume time domain room acoustics simulation under general impedance boundary conditions," *IEEE Trans. Audio, Speech, Lang. Process.* **24**(1), 161–173 (2016).
- ⁵⁰B. Hamilton and S. Bilbao, "On finite difference schemes for the 3-D wave equation using non-Cartesian grids," in *Proceedings of the Sound and Music Computing (SMC) Conference*, Stockholm, Sweden (2013), pp. 592–599.
- ⁵¹S. Bilbao, "Optimized FDTD schemes for 3-D acoustic wave propagation," *IEEE Trans. Audio, Speech, Lang. Process.* **20**(5), 1658–1663 (2012).

- ⁵²F. Brinkmann, A. Lindau, M. Müller-Trapet, M. Vorländer, and S. Weinzierl, "Cross-validation of measured and modeled head-related transfer functions," in *DAGA 2015*, Nürnberg, Germany (2015).
- ⁵³G. F. Kuhn, "Model for the interaural time differences in the azimuthal plane," *J. Acoust. Soc. Am.* **62**(1), 157–167 (1977).
- ⁵⁴K. A. J. Riederer, "Repeatability analysis of head-related transfer function measurements," in *105th Convention of the Audio Engineering Society*, San Francisco, CA (1998), pp. 4846:1–4846:62.
- ⁵⁵A. Kärkkäinen, L. Kärkkäinen, O. Kirkeby, M. Pölönen, E. Seppälä, J. Turku, and M. Vilermo, "Perceptual evaluation of numerically simulated head-related transfer functions," in *124th Convention of the Audio Engineering Society*, Amsterdam, The Netherlands (2008), pp. 7489:1–7489:7.
- ⁵⁶P. J. Roache, "Chapter 5: Grid convergence studies and the grid convergence index," in *Verification and Validation in Computational Science and Engineering* (Hermosa Pub, Albuquerque, NM, 1998), pp. 107–136.