

On the accuracy of finite-difference time-domain simulations of head-related transfer functions as a function of model complexity

Julie Meyer^{a,*}, Sebastian Prepeliță^b, Lorenzo Picinali^a

^a Dyson School of Design Engineering, Imperial College London, London, SW72DB, England, United Kingdom

^b Reality Labs Research, Meta, Redmond, 98052, WA, USA

ARTICLE INFO

Keywords:

Head-related transfer function
Finite-difference time-domain method
Accuracy
Solution verification

ABSTRACT

Wave-based numerical tools such as finite-difference time-domain (FDTD) solvers are useful for modeling several acoustic properties and interactions. While these numerical tools are widely used in acoustics, there seems to be less attention to assessing the quality of the produced outputs. However, in order to ensure that the obtained results are reliable, the quantification of the errors present in the simulation results is an essential step. There exists a mathematical process known as solution verification which aims at assessing the accuracy of the computed solutions. A relevant application for the FDTD method is the simulation of head-related transfer functions (HRTFs), since these are relatively complex to acoustically measure on humans. This paper aims at applying the solution verification process on HRTF modeling using the FDTD method to evaluate the accuracy of the simulated HRTF magnitudes with increased human head/torso model complexity. The FDTD-simulated HRTFs are also compared with respect to the similarity/dissimilarity of their spectrum and with respect to the relevance of these spectral variations on sound source localization. The results show that asymptotically extrapolating the FDTD-simulated HRTFs from a series of simulations provides more accurate HRTF predictions when compared to using single FDTD simulations ran on sub-millimeter grids, regardless of the model complexity. Results also demonstrate that the accuracy of the FDTD-simulated HRTF predictions decreases with increased model complexity. The localization performance predictions showed that the largest localization errors were obtained with models with the lowest complexities. Significant differences in predicted sound source localization performance were found between FDTD-simulated results.

1. Introduction

The head-related transfer function (HRTF) describes the response of a listener's head (including the ears) and torso (including the shoulders) to a sound source located in the free field. It is usually measured at the listener's ears' canal entrances and it essentially characterizes the auditory cues employed by the hearing system for sound source localization. As such, integrating an HRTF in the audio rendering system is essential for creating immersive augmented/virtual reality (AR/VR) applications. While there exist several publicly available HRTF datasets to select these filters from (e.g., [1,2]), one hindrance to creating a binaural rendering with a high degree of fidelity relates to the individual nature of these filters, since they depend on the listener's outer ears' and head/torso morphological features. For example, it has been shown that poorer localization performance [3,4] and decreased perceived externalization [5] are usually achieved when using a non-individual HRTF instead of

someone's own. This is why, when an acoustically-measured individual HRTF is not available, various strategies have been designed and evaluated to personalize the rendering, for example selecting or modifying non-individual HRTFs, or training the listener to use a non-individual filter [6]. Such methods though are not always effective, nor the personalization process is successful enough to improve the overall rendering quality (e.g., [7]). Evidently, measuring one's set of HRTFs is possible. However, such a task is time consuming and sensitive to the listener's movements.

In contrast to acoustically measuring HRTFs, which is subject to multiple sources of variability, wave-based numerical modeling techniques are highly controllable and repeatable since only the 3D model of the listener needs be embedded in the simulation domain. As such, using these techniques to simulate HRTFs is an attractive alternative to acoustic measurements. Commonly used numerical methods for this purpose include the boundary element method (BEM) [8–12], the fast

* Corresponding author.

E-mail address: jmeyer@ic.ac.uk (J. Meyer).

multipole BEM (FMBEM) [13,14], and the finite-difference time-domain (FDTD) method [15–17]. Although BEM is more frequently used than FDTD for HRTF modeling, the latter has several advantages, the first being the fact that it can be efficiently parallelized, greatly improving its computational performance. Thus, FDTD simulations are relatively fast considering that the entire volume of the simulation domain is discretized. In addition, and unlike BEM which necessitates one simulation per frequency bin, with FDTD a single simulation is required for obtaining a solution at all frequencies. Finally, FDTD can directly employ the input mesh (which usually results from 3D scans) without additional mesh processing [18], thus eliminating one source of geometrical/anthropometric error in the simulation pipeline. On the other hand, the mesh gets voxelized, a process which introduces additional geometrical errors, which can though mostly be eliminated by employing smaller grids or by doing a convergence study.

Despite the wide use of such numerical tools, less attention is usually given to assessing the accuracy of the simulated results. For example, in [14] the FMBEM outputs were generated using a single set of criteria to run the solver, which does not provide sufficient insight on the quality of the simulated data. The quality of the measured and simulated data was then assessed by comparing their extracted interaural time difference (ITD) values, as well as by comparing the correlation between their spatial frequency response surfaces as a function of frequency, either of which has a relatively small contribution in the output quality, especially from a perceptual standpoint. Similarly, the FDTD-simulated results were generated using two accuracy criteria per subject in [16], and the mean spectral distortion (SD) across all the HRTF directions considered was taken as an overall measure of the match between simulation and measurement. While most of the studies focusing on HRTF modeling using wave-based methods include comparisons between measured and simulated HRTFs, fewer comprise a systematic quantification of the numerical errors associated to these numerical modeling techniques. However, in order to be sufficiently confident that the simulated results can be used reliably with respect to a given operational validity [19], two mathematical processes known as code verification and solution verification [20, Parts II & III, pp. 145–366] should typically be first undertaken. The code verification process aims at determining if the implemented discrete mathematical model is free of inconsistencies (or bugs), and the solution verification process has for objective the assessment of the accuracy of the computed solutions.

Previous related work on the verification of simulated head-related transfer functions using the FDTD method have investigated the single sphere [21] and the snowman models [22]. The accuracy of pinna-related transfer functions (PRTFs) was also evaluated through a similar process in [23]. In [24], a solution verification process was applied on three 3D models with increased complexity (i.e. increasing the level of details and shape of the models): a sphere, a snowman, and a human head/torso model. This work complements and extends [24] by also evaluating the change in accuracy in the FDTD-computed HRTFs as a function of model complexity. Furthermore, it employs not only significantly more FDTD grids in the simulation series but, more importantly, smaller grids, as well as an additional intermediate model complexity (a human head/torso with the ears removed). In addition, a mannequin head/torso model and a second human head/torso model are included to investigate potential individual differences in the results. Finally, a Bayesian model for the estimation of the perceived direction of a broadband sound source is used as a means to evaluate the perceptual relevance of the differences between the simulated outputs. In short, this paper aims at exploring the use of FDTD for HRTF modeling by 1) rigorously quantifying the quality of the FDTD-simulated HRTFs through a solution verification process, as well as through standard HRTF numerical assessment measures and a Bayesian-based computational model estimating human sound source localization responses, and 2) evaluating the impact of changing the 3D model complexity on the above quantitative assessments.

2. Simulations

2.1. 3D models

Seven different 3D models with increased complexity were considered. By model complexity, it is meant the level of detail (including more or less parts of the head/torso features typically-considered in HRTF measurements) and shape (higher or lower accuracy in the representation of a human head/torso morphology) of the 3D models. With this in mind, the following 3D models were considered (see Fig. 1): a single sphere representing a simplified head model, a two-sphere model also known as the snowman model [25] representing a simplified head/torso model, two human head/torso models without the ears, a mannequin head/torso model (KEMAR with the “large” ears, GRAS Sound & Vibration, Denmark), and two human head/torso models with the ears.

The single sphere and the snowman models were directly generated from solid spheres using the 3D modeling software Rhinoceros [26]. The solid spheres were further converted into meshes (which is a collection of vertices and polygons defining the shape of a polyhedral object) by using the software option *Polygon Mesh Options* with the largest possible number of polygons. As in previous work ([21,22,24]), the single sphere radius was 8.25 cm, the same as the smallest sphere of the snowman model representing the head. The ratio of the two sphere radii of the snowman model was 1.3253, as in [27]. For the remaining 3D models, the meshes were directly taken from the publicly available SONICOM dataset [2]. These “watertight” meshes were generated from 3D scans taken with a EinScan Pro device (Shinning 3D, China). As mentioned in [2], small unscanned patches remained from the scanning process, thus limiting the accuracy of the resulting meshes to an extent that was not explicitly indicated.

The human #1 and #2 meshes corresponded to these of subjects P0001 and P0006 from the SONICOM dataset, respectively. The horizontal plane (x-y plane) of the meshes was re-aligned on the Frankfurt plane, which is defined in anatomy as the plane that passes through the lower margin of the eye socket to the upper margin of the ear canal [28]. Furthermore, the torsos of the mannequin and all the human models (i.e. (c-d) and (f-g) from Fig. 1) were cut such that their height corresponded to the one of the snowman model. Finally, in order to avoid capturing each ear canal entrance resonance, two additional meshes aligned with the median plane and blocking the ear canal entrances were manually added to the human #1 and #2 models. Except cutting the torso and removing the ears of the human #1 and #2 models (the cuts were filled by smooth surfaces) to obtain the corresponding human #1 no ears and human #2 no ears models, no additional changes were made to the original scans from [2].

The spherical coordinates (r, θ, ϕ) were employed, where r is the radial distance (or radius), θ is the azimuth defined on the interval $[0^\circ, 360^\circ)$, and ϕ is the elevation angle defined on the interval $[-90^\circ, 90^\circ]$. The center of the coordinate system was the head center of the 3D models. It is important to underline that an HRTF is defined as the ratio of the Fourier-transformed pressure at the entrance of the ear canal over the pressure measured (or, as here, simulated) in the free field at the head center, and can be expressed as

$$H(r, \theta, \phi, f) = \frac{P_{\text{ear}}(r, \theta, \phi, f)}{P_{\text{inc}}(r = 0, f)}, \quad (1)$$

where f denotes the frequency. P_{ear} is the Fourier-transformed pressure at the entrance of the ears' canals for an incident wave coming from the position (r, θ, ϕ) , and P_{inc} denotes the Fourier-transformed pressure of the incident wave simulated in the free field at the head center.

2.2. FDTD parameters

An FDTD solver implementing the standard rectilinear (SRL) scheme [29, p. 332] was employed to run the simulations in single floating point precision and at the stability limit of the scheme, i.e. when the Courant

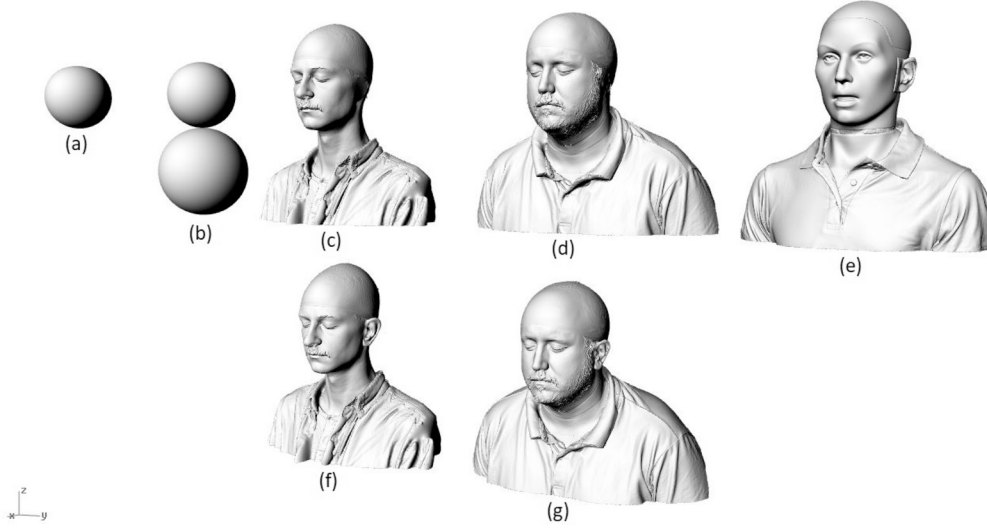


Fig. 1. Perspective view of the simulated 3D models: (a) sphere, (b) snowman, (c) human #1 no ears, (d) human #2 no ears, (e) mannequin, (f) human #1, and (g) human #2.

number λ is set to $1/\sqrt{3}$ for the SRL. The simulations were run on a number of nodes (DGX A100) varying from 5 to 17, containing 8 graphics processing units each. This solver previously underwent several code verification processes [23,21,22].

The source excitation signal was a Gaussian pulse defined as follows

$$G(t) = e^{-(t-\mu)^2/2\sigma^2}, \quad (2)$$

where $\sigma^2 = 6.5 \times 10^{-11}$, $\mu = 2.24 \times 10^{-4}$, and $t = (0:N-1) \times (1/f_s)$ denotes the discrete time vector with N representing the number of time steps and f_s the temporal sampling frequency of the simulation. This Gaussian pulse was designed such that the magnitude of its discrete Fourier transform was below machine epsilon ϵ in double precision (i.e. $\epsilon = 2.2 \times 10^{-16}$) at 135 kHz, which represents the Nyquist frequency of the smallest temporal sampling frequency used in the series of simulations. Machine epsilon is usually defined as an upper bound on the relative round-off error. The round-off error is caused by computations done using finite-precision representation of numbers.

The simulation time was 6.6 ms with a speed of sound set to 343.38 m/s, assuming a temperature in dry air of 20°C. A box with dimensions 4 m $\times$ 4 m $\times$ 4 m was constructed around the 3D models since no perfectly matched layer was implemented in the FDTD solver. More precisely, the box dimensions were defined such that no reflection from its surfaces was captured by the receivers during the entire simulation time. The 3D models were modeled with rigid boundaries and the center of their heads was placed at the center of the box.

All the 793 unique HRTF directions from the SONICOM dataset were simulated. These directions were measured at a distance of 1.5 m from the heads' center in the SONICOM dataset, which was also utilized in the simulation setup. Typically, when using the FDTD method a single source position is used at a time per simulation. However, to reduce the number of simulations to be ran, the reciprocity principle was invoked. As a result, the two source positions were placed at the blocked entrance of the two ear canals for the mannequin, human #1 and #2 models, and were run separately. As for the human #1 no ears and human #2 no ears models, the two sources were located such that they corresponded to the blocked entrance of the two ear canals from human #1 and #2, respectively. The sources were placed such that they converged, as the FDTD spatial grid spacing $X \rightarrow 0$, to the surface of the blocked meatus. For the sphere and snowman models, the two sources were directly placed on the head surface in the directions $(\theta = 90^\circ, \phi = 0^\circ)$ and $(\theta = 270^\circ, \phi = 0^\circ)$. Each of these two simulations contained 793 receivers corresponding to the totality of the HRTF directions from the SONICOM dataset. No spatial interpolation was applied to the receiver nor to the source posi-

tions, as it was shown in [22] that it does not improve the asymptotic predictions from the solution verification procedure. Also note that due to the different grid spacings employed in the simulation series, the resulting final positions of the receivers and sources within the simulation domain had a maximal deviation from r of $r_i = r \pm X_i \sqrt{3}/2$ since the sources and receivers were positioned at the center of the cubic air voxels of the simulation domain. Thus, the location of the sources and/or receivers converges to the continuous locations $r_i \rightarrow r$ in a first order form as $X_i \rightarrow 0$. The voxelization, which is the process of transforming a given continuous geometry into a voxel-based geometry, was made using a surface-conservative voxelization algorithm which creates a supercover of the continuous surface consisting of all the voxels meeting the surface of the meshes [30].

In order to obtain both P_{ear} and P_{inc} for the calculation of Eq. (1), and due to the fact that FDTD simulation outputs are time-domain responses, a Fourier transform was applied separately to both the outputs from the simulations generated in the presence of the 3D models and these generated in the free field (i.e. in the absence of the 3D models in the simulation domain). A frequency bandwidth in the range of [151.5 Hz, 20 kHz] was considered for the analysis of the resulting HRTFs, covering most of the human audible range, and most of the frequencies relevant for sound sources localization. The frequency range of analysis included the upper limit of the audible range (i.e. 20 kHz) since it represents the most conservative case to simulate considering that the discretization error increases as a function of frequency. The smallest temporal sampling frequency of the simulation series (corresponding to the coarsest FDTD spatial grid spacing) was set such that the phase velocity error [31] (in the axial direction of propagation which is the direction with the most discretization error for the SRL scheme [32]) at 20 kHz was 2%. Following this criterion, the ratio of the highest frequency of the bandwidth considered (i.e. 20 kHz) over the sampling frequency of the simulation was set to always be below 0.0757. Furthermore, the FDTD spatial grid refinement ratio between two consecutive simulations, defined as $\tau = X_{i+1}/X_i$, where $X_i < X_{i+1}$, was chosen to be 1.1 (minimum recommended in practice [33]) so as to include a large number of grids. These criteria resulted in a total of 19 simulations to run. A summary of the most relevant FDTD parameters for these 19 simulations are reported in Table 1.

While the 2% phase velocity error criterion has been explored in the context of room acoustics modeling (e.g., as in [34,35]), it may be irrelevant in the context of HRTF modeling. In fact, due to the voxelization error, which was evaluated in [36] in the context of HRTF modeling including ears, HRTFs calculated with FDTD employing grids

Table 1

Summary of the FDTD parameters used in the series of simulations. Sorted in descending order of the temporal sampling frequency f_s . X is the FDTD spatial grid spacing, rounded to two decimal places for its unit. The running time is given for the sphere model and one of the computing nodes as an indication of the approximate time taking for each simulation.

Grid #	f_s (kHz)	X (mm)	N	Running time (hh:mm:ss)
1	1490	0.40	9834	02:12:34
2	1355	0.44	8943	01:38:42
3	1230	0.48	8118	01:13:47
4	1120	0.53	7392	00:56:15
5	1020	0.58	6732	00:43:26
6	925	0.64	6105	00:32:42
7	840	0.71	5544	00:25:46
8	765	0.78	5049	00:19:16
9	695	0.86	4587	00:14:44
10	635	0.94	4191	00:12:12
11	575	1.03	3795	00:09:14
12	525	1.13	3465	00:07:08
13	475	1.25	3135	00:06:01
14	435	1.37	2871	00:04:23
15	395	1.51	2607	00:03:14
16	360	1.65	2376	00:02:51
17	325	1.83	2145	00:02:11
18	295	2.02	1947	00:01:45
19	270	2.20	1782	00:01:30

whose spatial spacing is above ≈ 1 mm were found to be sensitive to the voxelization error [36]. Thus, grid spacings X_i larger than about 1 mm can be considered outside of the asymptotic range for models containing human ears. Based on these findings, the effect of including grids with $X < 1$ mm on the accuracy of the FDTD-simulated HRTF magnitudes will be investigated in Sec. 3.2 to corroborate these results.

3. Solution verification

3.1. Asymptotic predictions

Physical models are mathematical models representing physical phenomena or events. In order to study these phenomena, computational models can be derived using computational methods such as FDTD. Essentially, these computational models are approximations of the physical ones. These approximations infer that there will be a set of numerical errors associated to the computational model. One example of such numerical errors associated to FDTD is the discretization error, which is due to the approximation of continuous differential operators by discrete operators. In order to gain trust on the results emanating from a computational model, it is essential to quantify the numerical errors underlying it. This can be done through the solution verification process. One approach to performing this is to choose an error model based on the order of accuracy which is expected to dominate the computational model. The expected order of accuracy can be determined by identifying the set of numerical errors present in the desired simulation setup and their behavior, i.e. through knowledge of error behavior. Here, the first-order asymptotic model, which is essentially a linear regression model, based on the magnitude was chosen and is expressed as follows

$$|H_i(r_i, \theta, \phi, f, X_i)| = |H_{\text{asympt}}(r, \theta, \phi, f)| + \beta(r, \theta, \phi, f)X_i, \quad (3)$$

where β represents the coefficient of the principal error term, $|H_i|$ denotes the simulated HRTF magnitude on the spatial grid spacing X_i and $|H_{\text{asympt}}|$ is the sought asymptotic HRTF magnitude prediction. It can be noted from Eq. (3) that in practice the sought asymptotic prediction corresponds to the intercept resulting from the least-squares fit of the first-order asymptotic model.

Furthermore, the weights introduced in [37] were applied to the first-order asymptotic model to result in a weighted least-squares regres-

sion. As it was shown in [22], this results in more accurate asymptotic predictions when compared to the non-weighted linear regression model based on Eq. (3). These weights are likely to result in a decrease of the variance of the residuals in the linear regression, since more importance is given to the finer grid solutions (assumed to be more accurate and more likely in the asymptotic range) compared to the coarser grid ones. These weights are defined as follows

$$w_i = \frac{\frac{1}{X_i}}{\sum_{i=1}^{n_g} \frac{1}{X_i}}, \quad (4)$$

where $n_g = 19$ is the total number of FDTD grids used in the first-order asymptotic model, and X_i is the spatial FDTD grid spacing corresponding to the i -th grid.

The accuracy on the asymptotic prediction was assessed by computing 95% confidence intervals (CIs). These latter were estimated using the bias-corrected and accelerated (BCa) bootstrap method [38] with 500 replicates. In essence, bootstrap methods create multiple “bootstrap samples” (500 here) by sampling with replacement from the original data (here, $(X_i, |H_i|)$ pairs). Then, the linear regression model is applied for each bootstrapped sample and the intercept is retained. Finally, the 95% CI is estimated based on the resulting 500 intercepts. The specificity of the BCa method is that CI estimates try to account for the bias and skewness in the distribution of the bootstrap samples. The failure of the first-order asymptotic model was assumed when the predicted $|H_{\text{asympt}}|$ from Eq. (3) was below machine epsilon ϵ in single precision (i.e. $\epsilon = 1.2 \times 10^{-7}$) in which case the log-magnitude of H_{asympt} was fixed to $20 \log_{10}(\epsilon) \approx -138.5$ dB. Failure of the model was also propagated to the 95% CIs calculation.

3.2. Results and discussion

The results, shown in Fig. 2 for the left ear, one HRTF direction and four 3D models, demonstrate that a better accuracy is achieved for the asymptotic predictions in comparison to most of the single individual FDTD grid simulation results, regardless if these latter are in the sub-millimeter range. For the sphere model, for which there exist an analytic solution, this is demonstrated in Fig. 2(a) by comparing the obtained asymptotic prediction and the 19 individual FDTD simulations with the HRTF magnitude of the analytic solution computed based on the method from [39]. The accuracy of the asymptotic predictions is also shown by the width of the 95% CIs, which is smaller than the magnitude of most of the individual FDTD grid simulations. The results for the right ear (not shown) were similar to these of the left ear.

Furthermore, the mean bias, defined as the width of the 95% CIs for each frequency bin averaged across all the 793 HRTF directions and shown in Fig. 3(a) for the left ear, was calculated for each 3D model to compare the results obtained as a function of model complexity. The mean bias for the right ear (results not shown) was similar to this of the left ear. It can be seen in Fig. 3(a) that the mean bias increases as a function of both frequency (which is expected due to the discretization error behavior) and model complexity. More precisely, for the sphere, the snowman and the two models without ears (labeled “human #1 no ears” and “human #2 no ears”) the average mean bias between the two ears was below 2.6, 3.4, 9.8 and 8.3 dB across the entire frequency range considered, respectively, while the 3D models containing ears (mannequin, human #1 and #2) had extremely large mean bias at higher frequencies. The failure of the first-order asymptotic model was also observed for a larger number of HRTF directions as a function of frequency. In fact, the mean bias for both ears for human #1 and #2 was within 10 dB for frequencies up to 7121 and 7272 Hz, respectively. Interestingly, the mean bias for both ears of the mannequin model was slightly smaller than this of the two other models containing ears and was within 10 dB for frequencies up to 10303 Hz.

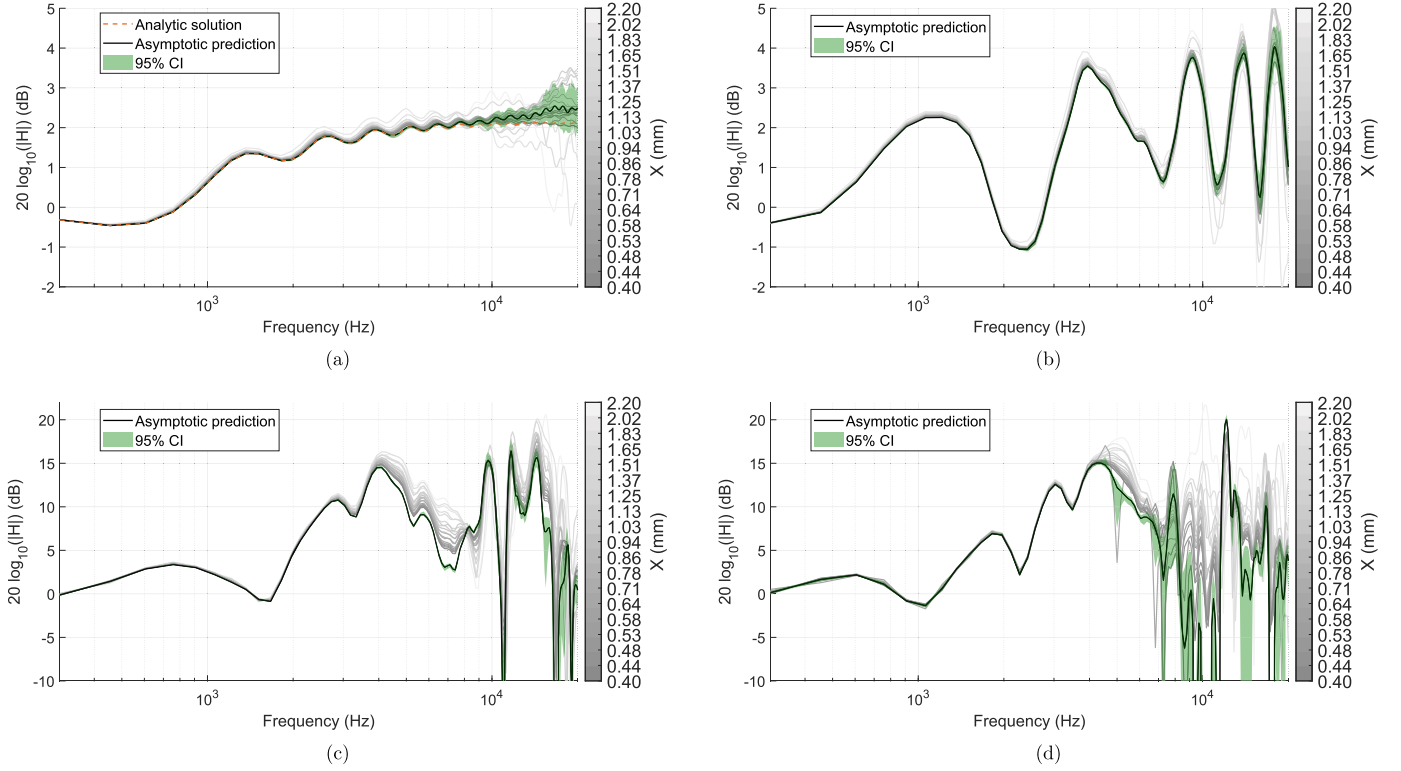


Fig. 2. Individual FDTD grid simulation results and asymptotic predictions for the left-ear HRTF with the receiver positioned in the direction ($\theta = 0^\circ$, $\phi = 0^\circ$) for the (a) sphere, (b) snowman, (c) mannequin, and (d) human #1. The asymptotic predictions resulted from Eq. (3) using all the 19 FDTD spatial grids listed in Table 1.

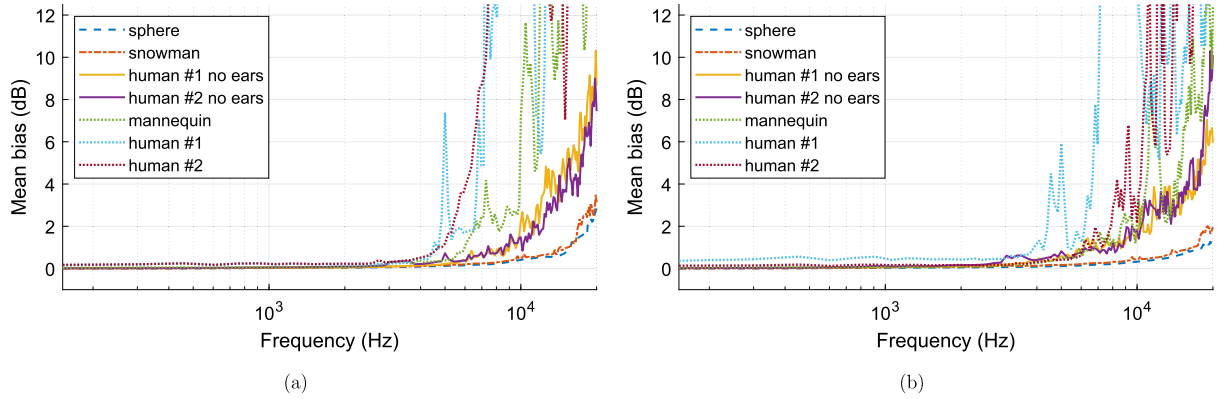


Fig. 3. Mean bias (bias averaged across all the 793 HRTF directions considered) for the left ear as a function of frequency for the seven 3D models when (a) all the 19 grids from Table 1 were included in Eq. (3), and (b) the 10 finest grids from Table 1 were included in Eq. (3).

In order to test whether reducing the number of the coarser grids in the series of simulations would improve the first-order asymptotic model, Eq. (3) was re-run using the 10 first (i.e. finest) FDTD grids listed in Table 1. While there is a risk that the linear regression model be less reliable if too little grids are included in the model, there is also a likelihood to improve the first-order asymptotic model since finer FDTD grids are considered. Choosing the 10 first grids listed in Table 1 seems to be a reasonable trade-off between the number of grids included in the linear regression model and the size of the grid spacings included (which are all below 1 mm). The choice of limiting the grid spacing to $X < 1$ mm in the first-order asymptotic model was also motivated by the results from [36] in which FDTD results from these grids can be considered outside of the asymptotic range in the HRTF modeling context.

The results of including the 10 first FDTD grids in Eq. (3) are shown in Fig. 4 for the sphere model, the left ear and the same HRTF direction

as displayed in Fig. 2. These results are to be visually compared with Fig. 2(a) specifically. It can be seen that not only the analytic solution is included in the 95% CIs across the entire frequency range but also that the resulting asymptotic prediction is closer to the analytic solution. The mean bias results are shown in Fig. 3(b) for the left ear and the seven 3D models. The results (not shown) were similar for the right ear. As can be seen in Fig. 3(b), including less, albeit finer, FDTD grids in the first-order asymptotic model improved the mean bias for all the 3D models except for human #2 no ears for which the worsening was marginal. In fact, for the sphere, snowman, human #1 no ears and human #2 no ears models, the average mean bias across the two ears was below 1.4, 2.0, 6.6 and 9.8 dB within the entire frequency range considered, respectively. The improvement in mean bias was the largest for mannequin, and resulted in an average mean bias across the two ears below 12.3 dB within the entire frequency range of analysis. For the two remaining models including ears (human #1 and #2), the first-asymptotic model

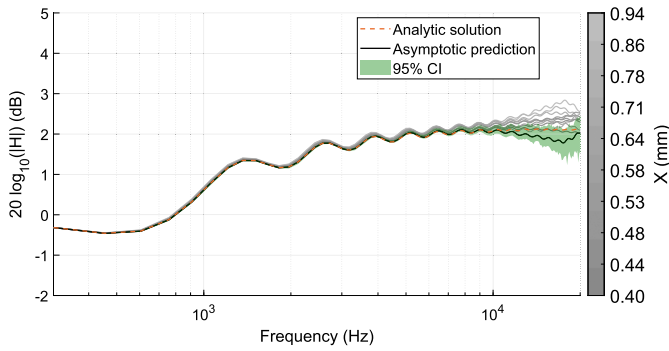


Fig. 4. Individual FDTD grid simulation results and asymptotic prediction for the left-ear HRTF with the source direction ($\theta = 0^\circ$, $\phi = 0^\circ$) for the sphere. The asymptotic prediction resulted from Eq. (3) using the 10 finest FDTD spatial grids listed in Table 1.

failed for a smaller number of HRTF directions per frequency bin compared to when all the 19 grids were included in the regression model, and resulted in generally smaller mean bias values with respect to the same comparison point, as can be seen when comparing Figs. 3(a-b).

As a result of this improvement in mean bias, the asymptotic predictions resulting from the first-order asymptotic model which only included the 10 finest FDTD grids are considered for all the 3D models in the remainder of the paper. Despite the mean bias improvement resulting from the inclusion of the 10 finest FDTD grids in the first-order asymptotic model, it is worth pointing out that in the HRTF modeling context, a mean bias of 2 to 5 dB is likely to become a threshold value above which the accuracy of the asymptotic predictions is too poor, indicating a low reliability in the simulated results. This was also observed in [22] for the snowman model, where the asymptotic prediction departed from the analytic solution as the 95% CIs increased. More precisely, Table I from [22] reports a maximum deviation of 4.64 dB between the log-magnitude of the asymptotic prediction and the log-magnitude of the analytic solution for an average 95% CIs width of 1.40 dB across the seven HRTF directions considered and the frequency range of analysis which covered [250 Hz; 20 kHz].

While it is unlikely that the failure of the first-order asymptotic model for a single frequency bin be of perceptual relevance, it is likely that its failure for several frequency bins in a row be a risk of perceptually impacting negatively the resulting asymptotic prediction. As such, the asymptotic predictions were corrected when the first-order asymptotic model failed for one or several frequency bins in a row, that is when the log-magnitude of H_{asympt} was below or equal to -138.5 dB. This correction consisted of substituting the log-magnitude of H_{asympt} by that of the smallest FDTD grid simulation result, when this latest fell within the 95% CIs. However, for each frequency bin where the HRTF magnitude from the smallest FDTD grid was outside of the 95% CIs, the log-magnitude of H_{asympt} was replaced by the upper bound of the 95% CIs. The rationale behind this decision is based on the assumption that the 95% CIs encompass the true solution. As such, in the case of large 95% CIs, indicating a poor accuracy of the first-order asymptotic model, using either the asymptotic prediction (resulting from the first-order asymptotic model) or the result from the smallest FDTD grid falling outside of the 95% CIs is not strongly founded. Note that following this correction, none of the asymptotic predictions for the sphere model were corrected.

4. Perceptually-guided numerical evaluation

4.1. Error metrics and computational model

In order to investigate the effect of varying the model complexity on objective error metrics, the difference in interaural level difference (ILD_{diff}) and the log-spectral distortion (LSD) between the corrected

asymptotic prediction and two of the individual FDTD grid simulations (more specifically, the grids #1 and #10 from Table 1) of each 3D model were calculated as follows

$$\text{ILD}_{\text{diff}}(r, \theta, \phi, f) = \frac{1}{123} \sum_{k=10}^K 20 \log_{10} \left| \frac{H_{\text{LasympC}}(r, \theta, \phi, f)}{H_{\text{RasympC}}(r, \theta, \phi, f)} \right| - \frac{1}{123} \sum_{k=10}^K 20 \log_{10} \left| \frac{H_{\text{Li}}(r_i, \theta, \phi, f)}{H_{\text{Ri}}(r_i, \theta, \phi, f)} \right|, \quad (5)$$

where $K = 132$ denotes the total number of frequency bins covering the frequency range of analysis. Note that for the calculation of Eq. (5), the frequency range was limited to the range [1515.2 Hz; 20 kHz] since the ILD cue is mostly useful for sound localization above approximately 1500 Hz [40]. $|H_{\text{LasympC}}|$ and $|H_{\text{RasympC}}|$ are the corrected asymptotic HRTF magnitude predictions resulting from Eq. (3) (using grids #1 to #10 from Table 1) for the left and right ear, respectively. $|H_{\text{Li}}|$ and $|H_{\text{Ri}}|$ denote the HRTF magnitudes simulated on the FDTD spatial grid spacing of $X_i = 0.94$ mm or 0.40 mm for the left and right ear, respectively.

$$\text{LSD}(r, \theta, \phi, f) = \sqrt{\frac{1}{N} \sum_{n=1}^N \left(20 \log_{10} \left| \frac{H_i(r_i, \theta, \phi, f)}{H_{\text{asymptC}}(r, \theta, \phi, f)} \right| \right)^2}, \quad (6)$$

where $N = 793$ denotes the total number of HRTF directions. The LSD was computed separately for each ear and each frequency bin.

Furthermore, the Bayesian computational model described in [41] was used to evaluate the perceptual relevance of the spectral variations in the FDTD-simulated HRTF magnitudes for the different model complexities, specifically looking at sound source localization. This model has previously been used to simulate behavioral localization test performances and quantifying the perceptual implications of employing different processing strategies and choices. More explicitly, the model was used in [42] for evaluating the quality of novel HRTF spatial up-sampling methods, and an older version was employed in [43] when assessing different methods for binaurally rendering Ambisonics signals.

The Bayesian computational model (implemented as *barumerli2023* and available as part of the auditory modeling toolbox (AMT) [44] in Matlab/Octave) takes a “template” and a “target” set of HRTFs as inputs and predicts the localization response of a subject (whose set of HRTFs is the template) for a given target HRTF set. The model first spatially interpolates the two provided input sets of HRTFs at uniformly distributed directions (to account for non-uniform HRTF acquisition grids), then computes the following features on the interpolated grid of directions for both inputs: the ITD and ILD cues, as well as the spectral amplitudes of the directional transfer functions (DTFs). From these calculated features, the model, which is based on a template-matching procedure, outputs the predicted localization performance of the subject whose set of HRTFs was provided as the template for the given target set of HRTFs.

In practice, the FDTD-simulated HRTFs (including the asymptotic predictions) were first converted to DTFs by using the *SOFAhrtf2dtf* function (provided in the SOFA Toolbox [45]). Note that because the *SOFAhrtf2dtf* function takes a SOFA object with the “Simple-FreeFieldHRIR” convention, Eq. (3) was re-run using the real and imaginary part of each FDTD-simulated HRTFs separately (as previously done in [22] where the approach is referred to as “indirect”), so as to extract the complex asymptotic HRTFs (as opposed to the HRTF magnitudes as indicated in Eq. (3)). The complex asymptotic HRTFs were then transformed into time-domain responses by applying an inverse Fourier transform (to obtain their corresponding HRIRs) and saved into a SOFA object with the “SimpleFreeFieldHRIR” convention (using both functions *SOFAgetConventions* and *SOFAsave*) prior to be fed to the *SOFAhrtf2dtf* function. Second, the *barumerli2023_featureextraction* function (also available in the AMT) was used to extract the interaural time and level differences, as well as the spectral amplitudes of the DTFs for both the template and the target inputs separately. Finally, these features extracted from both the template and target sets of HRTFs were

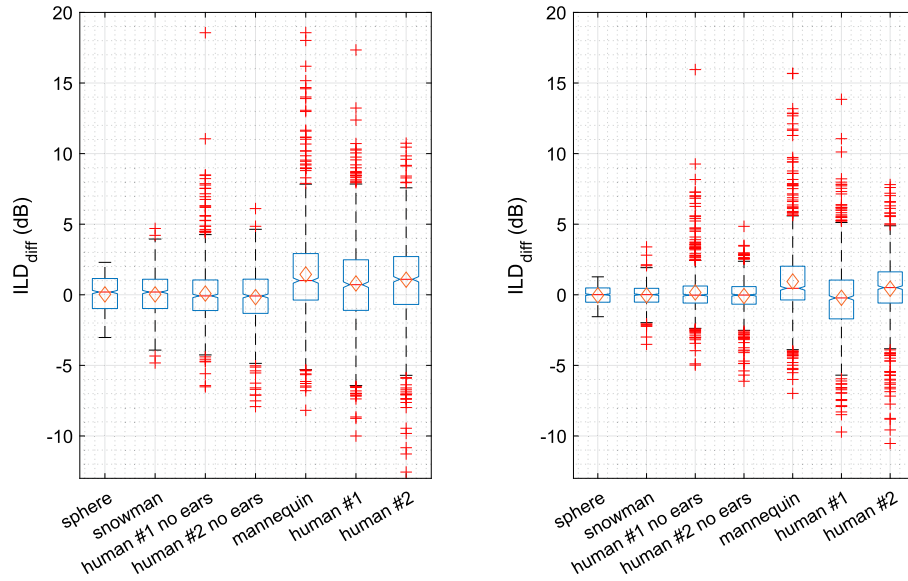


Fig. 5. Boxplots showing the ILD_{diff} between the corrected asymptotic HRTF magnitude prediction $|H_{asymptC}|$ and the FDTD-simulated HRTF magnitude $|H_i|$ using (left) the 10th smallest (i.e. 0.94 mm), and (right) the smallest (i.e. 0.40 mm) grid spacing of the simulation series. The mean of each boxplot values is indicated as a diamond symbol and reported in Table 2.

fed to the main *barumerli2023* function, constituting the Bayesian computational model which provided the final localization responses of the target set of HRTFs.

For each 3D model, the computational model was first provided with the simulated HRTFs from grids #1 (the finest grid of the simulation series) as the template HRTF set and two target HRTF sets which corresponded to the FDTD-simulated HRTFs from grid #10 and that from the asymptotic prediction. Unlike in [43] where the reference set of HRTFs was the original measured one, no ground truth or reference are available in the present case, except for the human #1 and #2 models. As such, and because the computational model is not commutative (it will output different localization performance results when swapping the provided target and template HRTFs), it was re-run separately with the asymptotic prediction as the template HRTF set as well as the FDTD-simulated HRTFs from grid #10 from Table 1, thus considering all possible combinations of template - two targets triplets. The model input parameters were the same as the medians indicated in Table 1 from [46]. In addition, due to the stochastic nature of the computational model and in order to obtain stable results, the localization response was estimated 300 times for each sound target direction.

The predicted localization responses were further analyzed using the *localizationerror* function from the AMT Toolbox to calculate the following error metrics: mean absolute lateral angle error, local polar RMS error (unweighted RMS error in the polar dimension considering target sounds around the median plane only ($\pm 30^\circ$ lateral angles) and excluding quadrant errors (i.e. polar error $> 90^\circ$), and quadrant error (i.e. frequency of polar errors larger than 90° as per defined in [47]).

4.2. Results and discussion

The ILD_{diff} (which is an average across frequencies ranging from [1515.2 Hz; 20 kHz]) calculated between the corrected asymptotic prediction and the individual simulation results are shown in Fig. 5 for all HRTF directions and each model as boxplots. The range of ILD_{diff} , whose minimum and maximum values across all the HRTF directions are reported in Table 2 for each 3D model, increased as a function of model complexity. More precisely, for the 3D models including ears, larger maximum values of ILD_{diff} were obtained compared to the other models. The mean ILD_{diff} across HRTF directions, reported both in Table 2 and in Fig. 5, also increased as a function of model complexity.

Table 2

Minimum and maximum ILD_{diff} values obtained for each 3D model across all the 793 HRTF directions considered. Each minimum and maximum value is indicated as X / Y where X and Y represent the ILD_{diff} values calculated from Eq. (5) using $|H_i = 0.94 \text{ mm}|$ and $|H_i = 0.40 \text{ mm}|$, respectively. The mean ILD_{diff} across directions is also reported as $\overline{ILD_{diff}}$.

3D model	min. ILD_{diff} (dB)	max. ILD_{diff} (dB)	$\overline{ILD_{diff}}$
sphere	-3.0 / -1.6	2.3 / 1.3	0.0 / 0.0
snowman	-4.8 / -3.5	4.7 / 3.4	0.0 / 0.0
human #1 no ears	-6.5 / -5.0	18.6 / 16.0	0.1 / 0.2
human #2 no ears	-7.9 / -6.1	6.1 / 4.8	-0.2 / -0.1
mannequin	-8.2 / -7.0	18.6 / 15.7	1.4 / 0.9
human #1	-10.0 / -9.7	17.3 / 13.8	0.8 / -0.2
human #2	-12.6 / -10.5	10.7 / 7.8	1.1 / 0.4

The threshold for ILD detection at 1000 Hz is ~ 1 dB [48], and even smaller (0.5 dB) at higher frequencies [40]. As such, the mean ILD_{diff} across HRTF directions calculated between the corrected asymptotic prediction and the individual simulation results are within this detection threshold for the sphere only and using the smallest FDTD grid of the simulation series. For the rest of the models, larger ILD_{diff} values are obtained for the 3D models without ears and even larger ones for the three remaining models containing ears (i.e. mannequin, human #1 and #2).

As for the LSD, it is commonly used as a means to objectively compare two transfer function spectra. These two spectra could be resulting, e.g., from a measured HRTF and from a simulated one. Typically considered small LSD values fall within the range of 3.0 to 5.0 dB while large ones are around 10.0 dB and above. For example, comparing an individual HRTF with a non-individual one, the average LSD across frequencies is about 7.5 to 10.0 dB (see Fig. 6 from [42]). In Fig. 2 of [43], even larger spectral differences of up to about 32.0 dB at around 15 kHz are reported for 1st order Ambisonics. From the results shown in Fig. 6, it can be seen that overall, the smallest LSD values were obtained for the sphere and the snowman models with maximum values of 3.0 dB and 6.1 dB across the entire frequency range and considering both ears, respectively, regardless of which simulated data ($|H_i = 0.94 \text{ mm}|$ or $|H_i = 0.40 \text{ mm}|$) was used. The average LSD values across frequencies and (left and right) ears was also small for the models with low complexities, regardless of which simulated data ($|H_i = 0.94 \text{ mm}|$ or $|H_i = 0.40 \text{ mm}|$) was

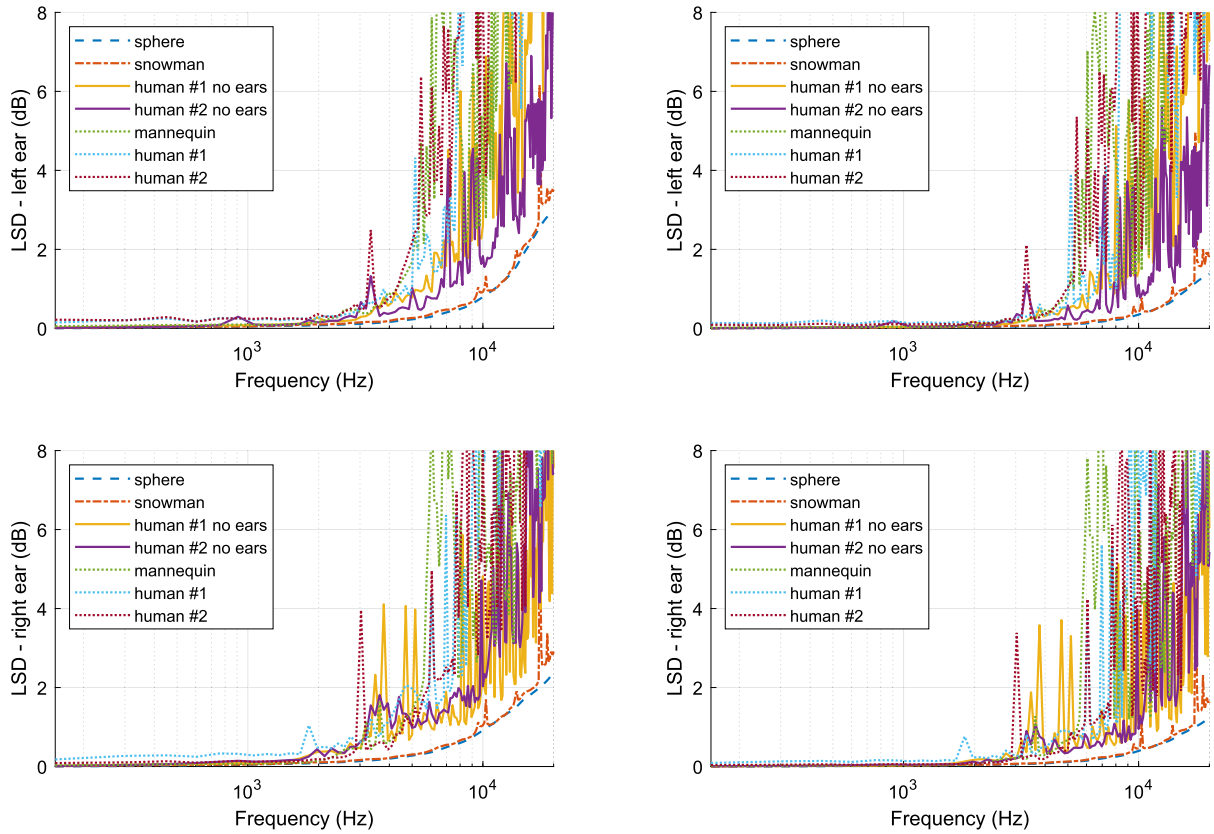


Fig. 6. LSD for the left- (top) and right-ear (bottom) HRTF magnitude for the seven 3D models comparing the FDTD-simulated HRTF magnitude $|H_i|$ which ran on the (left) the 10^{th} smallest grid spacing of the simulation series (i.e. 0.94 mm) and (right) the smallest grid spacing of the simulation series (i.e. $X_i = 0.40$ mm) with the corrected asymptotic HRTF magnitude prediction $|H_{\text{asympt}}|$.

Table 3

Mean LSD (dB) across frequencies and ears, when using $|H_i = 0.94 \text{ mm}|$ and $|H_i = 0.40 \text{ mm}|$.

3D model	$ H_i = 0.94 \text{ mm} $	$ H_i = 0.40 \text{ mm} $
sphere	1.1	0.5
snowman	1.2	0.6
human #1 no ears	3.8	3.0
human #2 no ears	3.1	2.1
mannequin	7.7	6.4
human #1	9.8	7.8
human #2	7.9	6.3

used. More precisely, the LSD averaged across frequencies and left and right ears is reported in Table 3 for each 3D model. As can be seen in Table 3, for the three models with higher complexity (i.e. these including ears), the average LSD using both $|H_i = 0.94 \text{ mm}|$ and $|H_i = 0.40 \text{ mm}|$ was higher than that of the models with lower complexity.

Overall, relatively acceptable mean LSD values were obtained for all the 3D models and both ears when using $|H_i = 0.40 \text{ mm}|$ in Eq. (6), though higher mean LSD values were observed for the 3D models with ears (mannequin, human #1 and #2) in comparison to the other five 3D models. However, the LSD averaged across frequencies and ears obtained using $|H_i = 0.94 \text{ mm}|$ was above acceptable values when considering the 3D human models with ears (human #1 and #2). This highlights the need to employ a very refined FDTD grid spacing of $X \approx 0.40 \text{ mm}$ for the 3D models with the lowest complexity such as the sphere, the snowman and the two models without ears. However, for 3D models with higher complexity, even smaller grids spacings (i.e. $X < 0.40 \text{ mm}$) would be necessary to satisfy the dB range for typical acceptable LSD values between the asymptotic predictions and the FDTD-simulated results.

The results from the computational model predicting a human localization performance based on the three different provided template HRTFs are shown in Fig. 7 for the two target HRTFs. As can be seen in Fig. 7 and despite the non-commutativity of the Bayesian computational model, the predicted localization performance is similar between the three used template HRTFs, though some slight differences are observed for the quadrant error. In addition, the overall predicted localization performance appears to be similar between the two target HRTFs, regardless of which template HRTF set is used. It can also be seen that the sphere and the snowman models exhibit the two largest errors for all the error metrics, except for the mean absolute lateral angle error which appears similar compared to the other 3D models. This is in line with the fact that these models both exhibit the necessary geometrical feature (a spherical head geometry) to provide accurate estimates of the ITD/ILD cues which are predominantly represented by the lateral angle error. In addition, the quadrant error is at chance level for the sphere and the snowman models which contrasts with the rest of the models which are below chance level. This highlights the importance of the degree of model complexity to obtain localization responses comparable with human performance.

Shapiro-Wilk tests were run on the data shown in Fig. 7 for each error metric and target HRTF. The statistical tests (results not shown) failed to reject the null hypothesis stating that the sample comes from a normal distribution at the 1% significance level in all cases (i.e. for all error metrics and both target HRTFs sets for each template HRTFs set), thus indicating that the data for all the error metrics was normally distributed. In order to determine whether the predicted localization performance was significantly different within the two target HRTFs as well as between the two template HRTFs, a two-way mixed ANOVA test was performed with the target HRTF sets as the within-subjects factors and the template HRTF sets as between-subjects factors. The test re-

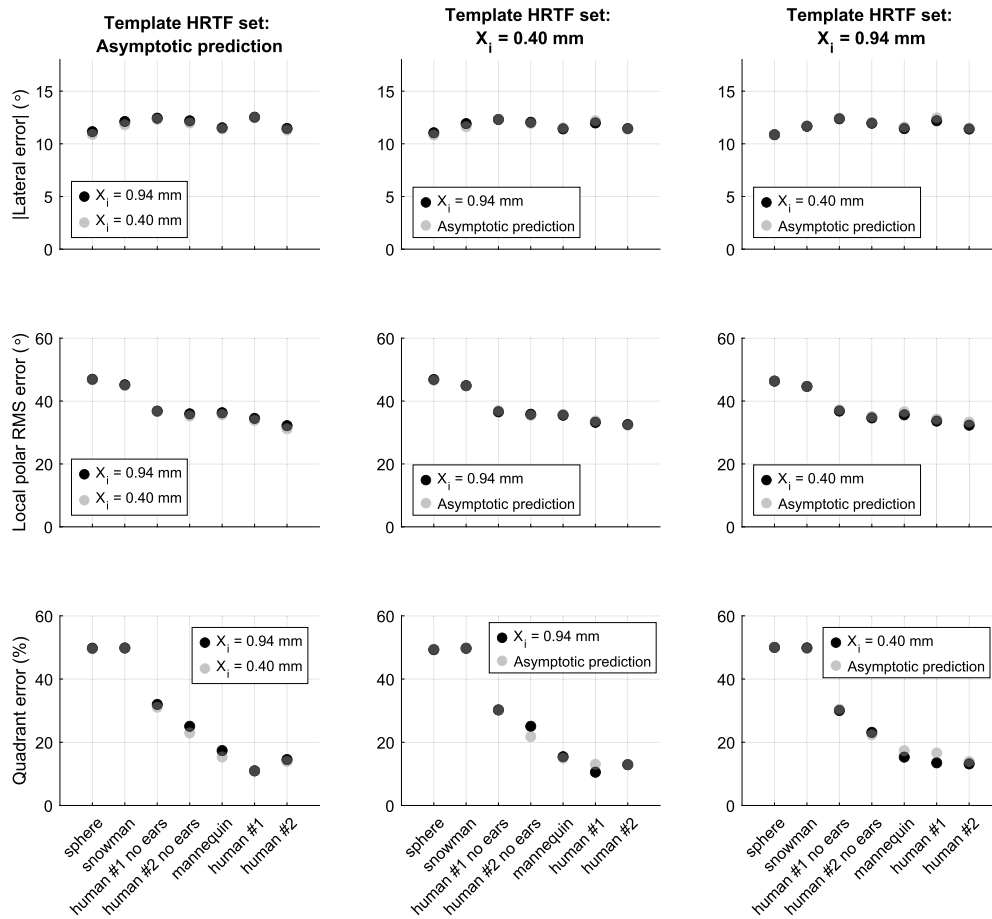


Fig. 7. Error metrics resulting from localization performance prediction of the computational model described in [41] fed with all combinations of template - targets triplets.

sults revealed that there was a significant two-way interaction between target and template HRTF sets on the mean absolute lateral angle error ($F(2, 18) = 7.810$, $p = 0.004$) and on the local polar RMS error ($F(2, 18) = 13.501$, $p < 0.001$). Following on these significant results, a one-way ANOVA test was performed to investigate the effect of the between-subjects factor (i.e. template HRTF set) on the error metrics for each target HRTF set. The results (not shown) were not significant for both error metrics, thus a one-way repeated measures ANOVA was run for the within-subjects variable (i.e. target HRTF set) for each template HRTF set. The results from this within-subjects ANOVA showed that there was a significant difference between the two target HRTFs for the mean absolute lateral angle error ($F(1, 6) = 22.6$, $p = 0.003$) and for the local polar RMS error ($F(1, 6) = 15.4$, $p = 0.008$) when the template HRTF was the asymptotic prediction. The effect of target HRTF was also significant on the mean absolute lateral angle error ($F(1, 6) = 8.52$, $p = 0.027$) and on the local polar RMS error ($F(1, 6) = 11.4$, $p = 0.015$) when the template was the simulated-HRTF from grid #10 ($X_i = 94$ mm). Interestingly, when the template HRTF was from grid #1 ($X_i = 40$ mm), the tests revealed that there was no significant difference between the localization performance of the two target HRTFs.

5. Conclusion

The HRTF magnitudes of seven different 3D models of head and/or torso with increased complexity (i.e. varying the level of detail and the shape approximation of a human head/torso) were simulated using the FDTD method. A solution verification process was applied to the FDTD-simulated HRTFs to quantify the numerical errors present in the computed results. The solution verification results showed that more

accurate HRTF magnitudes can be obtained by considering a series of simulations and applying a first-order asymptotic model instead of a running a single FDTD simulation, even using a sub-millimeter grid. The mean bias representing the accuracy of the asymptotic predictions resulting from the first-order asymptotic model increased as a function of frequency for all the 3D models considered. The mean bias also increased as a function of model complexity, and was especially high for the 3D models including ears at frequencies above 10 kHz.

The results from the Bayesian computational model predicting localization performance showed that the largest localization errors were obtained with the sphere and the snowman model. Interestingly, when taking the asymptotic prediction and the FDTD-simulated HRTFs on the coarsest grid as the template HRTFs, significant differences between the two target HRTFs were found in the data for two of the three error metrics considered. However, when the template HRTF set was the FDTD-simulated HRTFs on the finest grid, no significant difference were found between the two FDTD-simulated HRTFs. This difference in the results obtained between the two target HRTFs show that significant differences in sound source localization performance can occur when comparing FDTD-simulated results ran on a single grid with asymptotically extrapolated results from a series of simulations. Significant differences in sound source localization performance also occurred when comparing FDTD-simulated results ran on different single grids. As such, these results show the importance of not only numerically assessing FDTD-computed results but also perceptually evaluating them. Although the choice of the template HRTF against which the two target HRTFs are evaluated was not a ground truth per se, potential future work could focus on comparisons between measured and FDTD-simulated HRTFs. Similarly, BEM- and FDTD-simulated HRTFs could be also compared in

the future, as it seems that BEM is more commonly used than the FDTD method in the context of HRTF modeling.

CRedit authorship contribution statement

Julie Meyer: Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Sebastian Prepelit :** Writing – review & editing, Software, Resources. **Lorenzo Picinali:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank Roberto Barumerli for sharing his valuable insights on the Bayesian computational model. This project has received funding from the SONICOM project (www.sonicom.eu), a European Union's Horizon 2020 research and innovation programme under grant agreement No. 101017743.

Data availability

The authors do not have permission to share data.

References

- Brinkmann F, Dinakaran M, Pelzer R, Grosche P, Voss D, Weinzierl S. A cross-evaluated database of measured and simulated HRTFs including 3D head meshes, anthropometric features, and headphone impulse responses. *J Audio Eng Soc* 2019;67(9):705–18.
- Engel I, Daugintis R, Vicente T, Hogg AO, Pauwels J, Tournier AJ, et al. The SONICOM HRTF dataset. *J Audio Eng Soc* 2023;71(5):241–53.
- Bronkhorst AW. Localization of real and virtual sound sources. *J Acoust Soc Am* 1995;98(5):2542–53.
- M ller H, S rensen MF, Jensen CB, Hammersh i D. Binaural technique: do we need individual recordings? *J Audio Eng Soc* 1996;44(6):451–69.
- Simon LS, Zacharov N, Katz BF. Perceptual attributes for the comparison of head-related transfer functions. *J Acoust Soc Am* 2016;140(5):3623–32.
- Picinali L, Katz BF. System-to-user and user-to-system adaptations in binaural audio; 2023.
- Kim C, Lim V, Picinali L. Investigation into consistency of subjective and objective perceptual selection of non-individual head-related transfer functions. *J Audio Eng Soc* 2020;68(11):819–31.
- Kahana Y, Nelson PA, Petyt M, Choi S. Numerical modelling of the transfer functions of a dummy-head and of the external ear. In: Audio engineering society conference: 16th international conference: spatial sound reproduction. Audio Engineering Society; 1999.
- Katz BF. Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation. *J Acoust Soc Am* 2001;110(5):2440–8.
- Walsh T, Demkowicz L, Charles R. Boundary element modeling of the external human auditory system. *J Acoust Soc Am* 2004;115(3):1033–43.
- Otani M, Ise S. Fast calculation system specialized for head-related transfer function based on boundary element method. *J Acoust Soc Am* 2006;119(5):2589–98.
- Ziegelwanger H, Majdak P, Kreuzer W. Numerical calculation of listener-specific head-related transfer functions and sound localization: microphone model and mesh discretization. *J Acoust Soc Am* 2015;138(1):208–22.
- 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(1):370–86.
- Jin CT, Guillon P, Epain N, Zolfaghari R, Van Schaik A, Tew AI, et al. Creating the Sydney York morphological and acoustic recordings of ears database. *IEEE Trans Multimed* 2013;16(1):37–46.
- Xiao T, Liu Q, Huo. Finite difference computation of head-related transfer function for human hearing. *J Acoust Soc Am* 2003;113(5):2434–41.
- Mokhtari P, Takemoto H, Nishimura R, Kato H. Comparison of simulated and measured HRTFs: FDTD simulation using MRI head data. Audio engineering society convention, vol. 123. Audio Engineering Society; 2007.
- Webb C, Bilbao S. Binaural simulations using audio rate FDTD schemes and CUDA. In: Proceedings of the 15th international conference on digital audio effects; 2012.
- Ziegelwanger H, Reichinger A, Majdak P. Calculation of listener-specific head-related transfer functions: effect of mesh quality. Proceedings of meetings on acoustics, vol. 19. AIP Publishing; 2013.
- Sargent RG. Verification and validation of simulation models. In: Proceedings of the 2010 winter simulation conference. IEEE; 2010. p. 166–83.
- Oberkampf WL, Roy CJ. Verification and validation in scientific computing. Cambridge University Press; 2010.
- Meyer J, Smirnov M, Khajeh-Saeed A, Hoffmann PF, Prepelit  ST. Finite-difference time-domain simulations: verification on head-related transfer functions of a rigid sphere model. *JASA Exp Lett* 2022;2(6).
- Meyer J, Prepelit  S, Khajeh-Saeed A, Smirnov M, Hoffmann P. Verification on head-related transfer functions of a snowman model simulated using the finite-difference time-domain method. *IEEE/ACM Trans Audio Speech Lang Process* 2023.
- Prepelit  ST, Bola os J G mez, Geronazzo M, Mehra R, Savioja L. Pinna-related transfer functions and lossless wave equation using finite-difference methods: verification and asymptotic solution. *J Acoust Soc Am* 2019;146(5):3629–45.
- Meyer J, Picinali L. Comparison of simulated head-related transfer functions accuracy for different model complexities using the finite-difference time-domain method. In: Proceedings of forum acusticum; 2023.
- Algazi VR, Duda RO, Duraiswami R, Gumerov NA, Tang Z. Approximating the head-related transfer function using simple geometric models of the head and torso. *J Acoust Soc Am* 2002;112(5):2053–64.
- McNeel Robert, Associates. Rhino3D. <https://www.rhino3d.com/>. Version 5.14.522.8390 (2017-05-22).
- Gumerov NA, Duraiswami R. Computation of scattering from N spheres using multipole reexpansion. *J Acoust Soc Am* 2002;112(6):2688–701.
- Cheng Y, Leow WK, Lim TC. Automatic identification of frankfurt plane and mid-sagittal plane of skull. In: 2012 IEEE workshop on the applications of computer vision (WACV). IEEE; 2012. p. 233–8.
- Bilbao SD. Wave and scattering methods for the numerical integration of partial differential equations. Ph.D. thesis. Stanford University; 2001.
- Cohen-Or D, Kaufman A. Fundamentals of surface voxelization. *Graph Models Image Process* 1995;57(6):453–61.
- Saarela J, Botts J, Hamilton B, Savioja L. Audibility of dispersion error in room acoustic finite-difference time-domain simulation as a function of simulation distance. *J Acoust Soc Am* 2016;139(4):1822–32.
- Kowalczyk K, Van Walstijn M. Room acoustics simulation using 3-D compact explicit FDTD schemes. *IEEE Trans Audio Speech Lang Process* 2010;19(1):34–46.
- Roache PJ. Verification and validation in computational science and engineering, vol. 895. Hermosa Albuquerque, NM. 1998.
- Saarela J, Savioja L. Audibility of dispersion error in room acoustic finite-difference time-domain simulation in the presence of a single early reflection. *J Acoust Soc Am* 2019;145(4):2761–9.
- Meyer J, Lokki T, Ahrens J. Perceptual detection thresholds for numerical dispersion in binaural auralizations of two acoustically different rooms. *J Acoust Soc Am* 2022;152(4):2266–76.
- Prepelit  S, Geronazzo M, Avanzini F, Savioja L. Influence of voxelization on finite difference time domain simulations of head-related transfer functions. *J Acoust Soc Am* 2016;139(5):2489–504.
- E a L, Hoekstra M. A procedure for the estimation of the numerical uncertainty of CFD calculations based on grid refinement studies. *J Comput Phys* 2014;262:104–30.
- Efron B. Better bootstrap confidence intervals. *J Am Stat Assoc* 1987;82(397):171–85.
- Duda RO, Martens WL. Range dependence of the response of a spherical head model. *J Acoust Soc Am* 1998;104(5):3048–58.
- Mills AW. Lateralization of high-frequency tones. *J Acoust Soc Am* 1960;32(1):132–4.
- Barumerli R, Majdak P, Geronazzo M, Meijer D, Avanzini F, Baumgartner R. A Bayesian model for human directional localization of broadband static sound sources. *Acta Acust* 2023;7:12.
- Hogg AOT, Jenkins M, Liu H, Squires I, Cooper SJ, Picinali L. HRTF upsampling with a generative adversarial network using a gnomonic equiangular projection. *IEEE/ACM Trans Audio Speech Lang Process* 2024;1–15. <https://doi.org/10.1109/TASLP.2024.3375635>.
- Engel I, Goodman DF, Picinali L. Assessing HRTF preprocessing methods for Ambisonics rendering through perceptual models. *Acta Acust* 2022;6:4.
- Majdak P, Hollomey C, Baumgartner R. AMT 1. x: a toolbox for reproducible research in auditory modeling. *Acta Acust* 2022;6:19.
- Majdak P, Zotter F, Brinkmann F, De Muynke J, Mihocic M, Noisternig M. Spatially oriented format for acoustics 2.1: introduction and recent advances. *J Audio Eng Soc* 2022;70(7/8):565–84. <https://github.com/sofacooustics/SOFAtoolbox>.
- Daugintis R, Barumerli R, Picinali L, Geronazzo M. Classifying non-individual head-related transfer functions with a computational auditory model: calibration and metrics. In: ICASSP 2023-2023 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE; 2023. p. 1–5.
- Middlebrooks JC. Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency. *J Acoust Soc Am* 1999;106(3):1493–510.
- Middlebrooks JC. Sound localization. *Handb Clin Neurol* 2015;129:99–116.