forum acusticum 2023

COMPARISON OF SIMULATED HEAD-RELATED TRANSFER FUNCTIONS ACCURACY FOR DIFFERENT MODEL COMPLEXITIES USING THE FINITE-DIFFERENCE TIME-DOMAIN METHOD

Julie Meyer, Lorenzo Picinali
Dyson School of Design Engineering, Imperial College London, England

ABSTRACT

The use of finite-difference time-domain (FDTD) simulations is relevant for several applications in virtual acoustics. One of these is the numerical calculation of head-related transfer functions (HRTFs). This study investigates the effect of varying the geometrical complexity (shape, level of details) of a human head/torso model on the calculation of its HRTFs using an FDTD solver. In particular, the interest is on the accuracy of the obtained simulation results with respect to the human head/torso model complexity. For that aim, a solution verification process is undertaken, and a single sphere, a two-sphere and a human head and torso models are considered. The results indicate that relatively small 95% confidence intervals on the solution verification results are achieved, indicating relatively good accuracy for the prediction of HRTFs up to relatively high frequencies for the single and two-sphere models considered. However, for the simplified human head and torso model, a similar accuracy is achieved only up to a lower frequency.

Keywords: *head-related transfer function, finite-difference time-domain simulation, accuracy, solution verification*

1. INTRODUCTION

A head-related transfer function (HRTF) is the Fourier transform of a head-related impulse response (HRIR) which essentially encodes the transfer path between a source and a subject's ears in the free-field. The HRTF plays a predominant role in spatial hearing as it characterises binaural cues, and thus represents a key function to integrate in augmented/virtual reality, immersive or spatial audio applications. HRTFs are specific for every individual, as they depend on morphological features of the external ear, the head and the shoulders/torso. In order to achieve a high level of realism in the simulation, it is important to use individually measured, or at least personalised, HRTFs [1].

Since acoustically measuring HRTFs is a long and tedious task, several numerical methods have been employed in the past, such as the boundary element method (BEM) [2] and finite-difference time-domain (FDTD) [3–5]. Though BEM is more commonly used for the task of simulating HRTFs, FDTD has several advantages which makes it a relevant method for the calculation of such functions. One of these advantages is the compatibility of the method with parallel processing (due to its explicit nature) to improve its computational performance. Thus, FDTD simulations are relatively fast, considering that the whole 3D domain is discretised, which can be useful to generate large data sets, e.g., for training machine learning models (e.g., [6]) or for large-scale simulations for acoustic applications (e.g., [7–10]).

Using wave-based modelling techniques, it is possible to simulate HRTFs by embedding the listener's geometry in the simulation domain and "recording" the sound field at specific locations in space, taken to be the entrance of the ear canal for the calculation of HRTFs. Concretely, incorporating a listener's geometry into the simulation domain first consists in taking a 3D scan of the listener's head and torso and generating a mesh from the resulting point cloud. Note that for simplified models of a human head and torso such as the snowman model, the mesh can be directly created from closed surfaces using most common 3D modelling software.

While using numerical methods comes with several advantages (e.g., the simulation domain, including the source type and position, is highly controllable; there is no misalignment of the listener's mesh position with respect to the source position), there seems to be surprisingly few

formal studies on the numerical errors arising from the employed modeling methods for acoustical tasks. However, in order to be sufficiently confident that the simulated results are reliable, any prediction tools should typically first undergo two mathematical processes called code verification and solution verification [11], p. 249. The code verification process aims at providing evidence that the implemented code does not contain any coding mistake, and the solution verification process aims at assessing the accuracy of the computed results.

This paper, which utilises an FDTD solver that previously underwent a code verification process [12], focuses on the solution verification process in the context of HRTF modelling. More specifically, the aim of this paper is to evaluate the change in accuracy in the FDTD-computed HRTFs when the model complexity is increased in shape and level of details. It is noted that the present interest is in comparing different levels of model complexity rather than focusing on accurate human head/torso modelling. As such, the two first models considered are rather crude approximations of a human shape.

2. 3D MODELS

Three models were considered for the prediction of their HRTFs: (1) a single sphere model solely representing a human head; (2) a two-sphere model, also known as the snowman model [13], which further includes a simplified model of the torso in comparison to the single sphere model; (3) a model of an adult male which, in contrary to both previous models, includes the ears as well as other important anthropometric data (e.g., by considering the presence of the shoulders).

For the single sphere and the snowman models, the meshes were manually created with solid spheres using the 3D modelling software Rhinoceros [14]. The smaller sphere of the snowman model, hereafter referred to as the *head*, was the same as the single sphere model whose radius is 8.25 cm. The larger sphere of the snowman model representing the torso was related to the radius of the *head* denoted r_{head} by $r_{\text{head}} \times$ 1.3253, as in [15]. As for the human head and torso model, the original mesh of the adult male, which was taken from [16], was obtained via a 3D scan of the person. However, this latter mesh was too detailed to be voxelised and, as a consequence, a simplified version of the mesh was created for the purpose of this study. This simplification was done by using the *ReduceMesh* function of Rhinoceros and reducing the number of polygons from 89394 to 13385. Fig. 1 shows the three models considered and the details of the generated meshes are listed in Tab. 1.

The coordinate system adopted, whose origin was set as indicated in Tab. 1, was the spherical coordinate one (r, θ, ϕ) where r is the radial distance or radius, $0° \leq \theta < 360°$ is the azimuthal angle or azimuth and $-90° \leq \phi \leq 90°$ is the elevation angle. Four HRTF directions were selected at a fixed elevation $\phi = 0°$ and azimuth varying from 0° to 90° with 30° increments. It is worth mentioning that only the magnitude of the HRTFs is considered in the present study. Further calculation details on the HRTFs are given in Sec. 3.2.

Table 1. Details on the meshes.

Model	Origin	Vertices	Polygons
Sphere	center	1986	3968
Snowman	center of the *head*	2×1986	2×3968
Human	center of the interaural axis	17113	13385

3. METHOD

3.1 Solution verification

As previously mentioned, a solution verification process is concerned with the assessment of the accuracy of the computed results. Typically, such an assessment is done by either estimating the discretisation error or computing asymptotic predictions from a series of simulations, as here will be done. Essentially, an asymptotic prediction is the outcome of a linear regression model that is devised either from the knowledge of numerical error behaviour or hypothesis testing on model parameters. Here, the use of knowledge of error behaviour is relied upon to choose the correct linear regression model, which will provide the asymptotic predictions.

As the spatial grid spacing X and the time step T (which is related to the sampling rate f_s by $T = 1/f_s$) approach zero, the expected order of accuracy is determined by the source of error with the lowest order. This herein implies that the expected order of accuracy of the current problem at hand will be determined by the voxelisation error which is first-order accurate in space [17]. As a result, a first-order asymptotic model is considered due to the first-order accuracy behaviour of the voxelisation error. The first-order asymptotic model can be expressed as

$$H_i = H_{\text{asymptotic}} + \beta X_i \,, \tag{1}$$

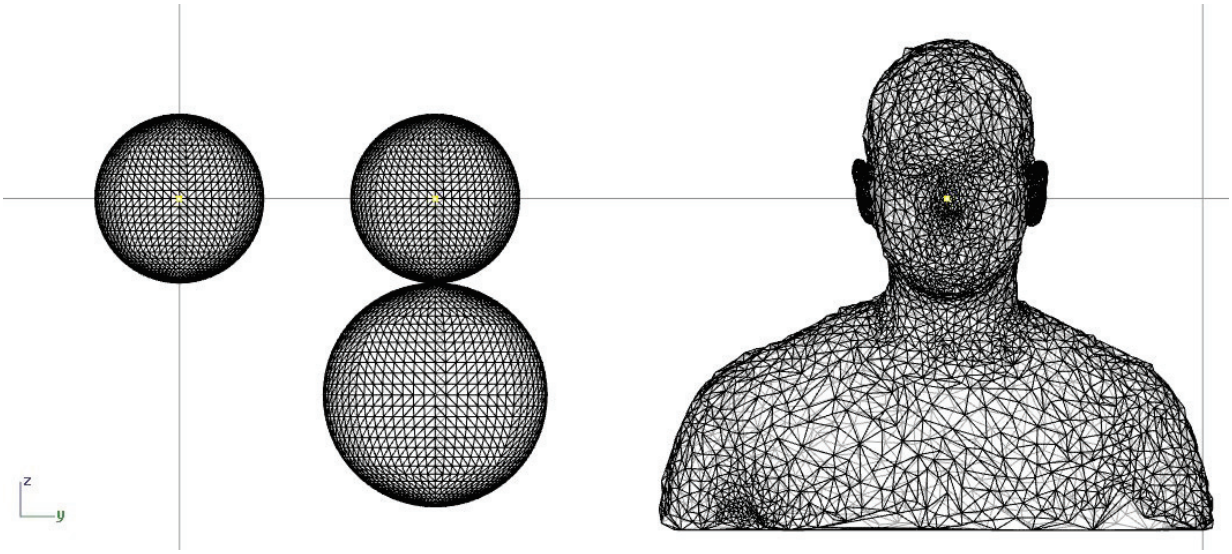


Figure 1. Right view of the meshes of the three models considered. The yellow dots indicate the origin of the models.

where H_i is the simulated HRTF magnitude using a spatial grid spacing of X_i, β represents the coefficient of the principal error term, and $H_{\text{asymptotic}}$ denotes the sought asymptotic prediction. In practice, $H_{\text{asymptotic}}$ is found by evaluating the intercept of the least-squares fit of the first-order asymptotic model depicted in Eqn. (1). It is worth mentioning that Eqn. (1) assumes that the spatial grid spacing X is in the asymptotic range (defined in, e.g., [18]), which if failure is observed may indicate that the grids used were not. The failure of the first-order asymptotic model was assumed when the predicted $H_{\text{asymptotic}}$ was below machine epsilon in single precision (i.e. approximately 1.2×10^{-7}) in which case $H_{\text{asymptotic}}$ was fixed to this latter value to appear visible in the results.

3.2 FDTD simulations

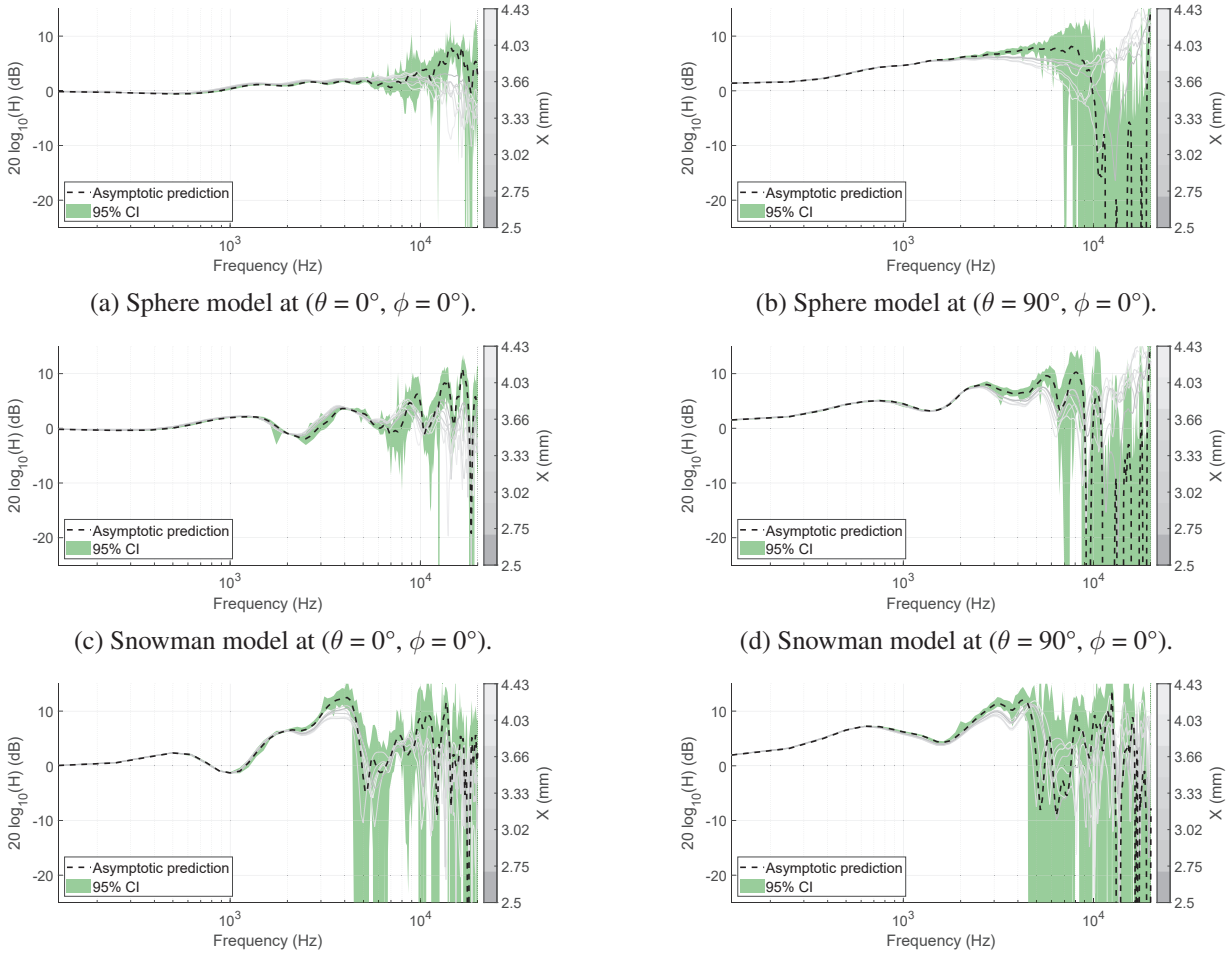
An open source FDTD solver for room acoustics [19] implementing the standard rectilinear scheme [20] for the acoustic wave equation was utilised to run the simulations with single precision floating-point numbers. The solver utilised two graphics processing units (Tesla A100) for improved performance.

The sound source was a Gaussian pulse, $G(t) = e^{-(t-1.8\times10^{-4})^2/2\times0.034\times10^{-8}}$, located at 82.5 cm from the origin of the coordinate system (see also Tab. 1) and at the four azimuthal angles θ and elevation ϕ set to 0°

Table 2. Parameters of the series of simulations for each HRTF direction. The grid spacing X are rounded to 5 decimal places. f_s denotes the temporal sampling frequency, N is the number of time steps.

#	f_s (Hz)	X (mm)	N
1	238375	2.50	1907
2	216750	2.75	1734
3	197000	3.02	1576
4	179125	3.33	1433
5	162875	3.66	1303
6	148000	4.03	1184
7	134625	4.43	1077

as described in Sec.2. Since the solver does not include the implementation of a perfectly matched layer, a rigid box with dimensions 4 m × 4 m × 4 m was designed around the origin of the coordinate system to define the boundaries of the simulation domain. The box dimensions were chosen such that the simulation time would be long enough to "record" the interactions of the sound waves with the models without capturing the reflected waves from the box. For each source position, a single receiver (i.e. the left-ear only was calculated) was positioned at the entrance of the ear canal. For the sphere and the snow-



(a) Sphere model at ($\theta = 0°$, $\phi = 0°$).

(b) Sphere model at ($\theta = 90°$, $\phi = 0°$).

(c) Snowman model at ($\theta = 0°$, $\phi = 0°$).

(d) Snowman model at ($\theta = 90°$, $\phi = 0°$).

(e) Human model at ($\theta = 0°$, $\phi = 0°$).

(f) Human model at ($\theta = 90°$, $\phi = 0°$).

Figure 2. Log-magnitude of the computed asymptotic predictions using Eq. (1). The seven individual FDTD-simulated HRTF magnitudes used to compute the asymptotic predictions are also included as a grey colour gradient. It is also reminded that only the left-ear signals are herein considered.

man models, the ear canal was assumed to be on the *head*. However, to ensure that the receiver did not end up inside the models, its position was fixed slightly outside of the *head* as follows

$$r = r_{\text{head}} + X\sqrt{3} + \varepsilon\,, \tag{2}$$

where r_{head} = 8.25 cm is the radius of the *head*, X is the spatial grid spacing from the FDTD grid, ε is machine epsilon in single precision.

After obtaining the time-domain signals from the FDTD solver, a discrete Fourier transform (DFT) was calculated for each simulated response. The FDTD solver was also run separately using the four source positions with a receiver positioned at the origin of the models in the free field (i.e. in the absence of the models). A DFT was calculated for the time-domain responses of the captured sound field at the origin simulated in the free-field as well. Finally, the HRTF was obtained by dividing the two resulting DFTs for each direction. Tab. 2 summarises the parameters used in the series of simulations.

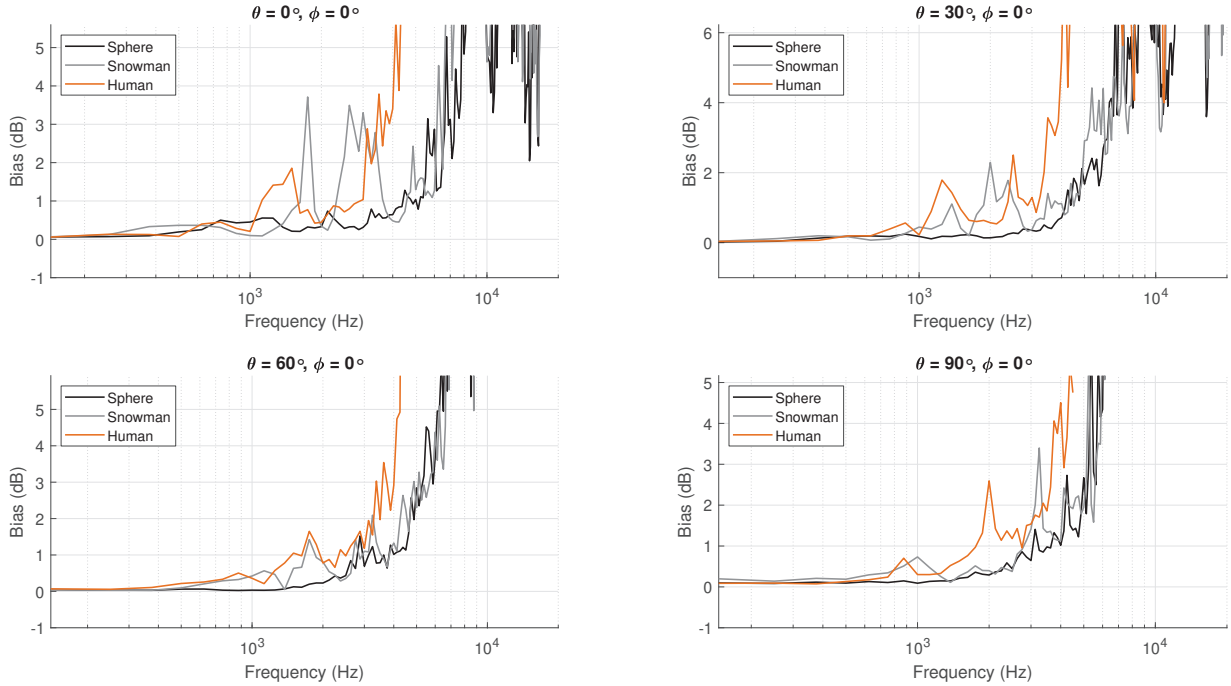


Figure 3. Bias as a function of frequency for the three models per HRTF direction.

4. RESULTS

The results on the asymptotic predictions are presented for the three models and two HRTF directions in Fig. 2. Overall, the results obtained are in line with previous studies [21,22], showing that using a series of simulations to compute asymptotic predictions give better results than using a single FDTD simulation ran over a small grid. This is seen by the smooth behaviour of the asymptotic predictions in comparison to some of the individual FDTD-simulated HRTF magnitudes, despite these latter ran over small grid spacings. For example, it can be seen from 2 that when the CIs are relatively small (indicating satisfactory results for the asymptotic predictions) the asymptotic predictions exhibit less deep notches and peaks than for FDTD simulations using the larger grid spacings of the simulation series. However, the accuracy of the asymptotic prediction for the single sphere presented here is smaller than in [21]. The main first difference between the two studies which could explain this result is that smaller grids were included in the calculation of the asymptotic prediction in [21] in comparison to the present study. This demonstrates that decreasing the number of small grids in the first-order asymptotic model is detrimental to the accuracy of the predictions. The second difference, which is tightly related to the first, is the total number of grids included in the first-order asymptotic model, since increasing the number of data points in a linear regression model increases its reliability.

In order to evaluate the uncertainty of the asymptotic predictions, the 95% confidence intervals (CIs) were estimated using the bias-corrected and accelerated ($\mathrm{BC_a}$) bootstrap method with 100 replicates [23]. The bias, defined as the width of the CIs for each frequency bin and shown in Fig. 3, was also computed per HRTF direction to compare the results obtained as a function of model complexity. It is clear that the bias gets larger as a function of frequency, which highlights the frequency-dependency of the discretisation error. As a reminder, the discretisation error is commonly assumed to be the dominant error in a partial differential equation-based scientific simulation [24]. For example, relatively small uncertainty on the first-order asymptotic model (e.g., bias < 2 dB) is achieved up to 4125 Hz for the sphere model across the four HRTF directions, while the same uncertainty is attained at 1625 Hz and 1875 Hz for the snowman and human models, respectively.

Failure of the first-order asymptotic model was removed from Fig. 3, hence the absence of data for some of the frequency bins. From Fig. 3, it can also be seen that failure of the first-order asymptotic model increases as the model complexity increases, indicating that the accuracy of the HRTFs decreases with model complexity. This can be due to the small details that are not properly captured by the FDTD grid size.

5. CONCLUSION

Three 3D models of a human head/torso with increased complexity were simulated using an FDTD solver. For each model, the accuracy of the FDTD-computed HRTFs was assessed through a solution verification process. The results show that as the model complexity increases, the upper limit of the frequency bandwidth, for which the asymptotic predictions are relatively accurate, reduces. These results demonstrate the difficulties of the FDTD solver to predict HRTFs for complex models using a small number of grids to compute asymptotic predictions. In addition to using more grids to compute the asymptotic predictions, the results also suggest that smaller grids should be included in the first-order asymptotic model to achieve better accuracy, especially for models with a high level of details representing complex shapes.

Future studies in this direction could include the use of computational models to quantify the relevance of the spectral variations in terms of human perception, specifically looking at sound sources localisation and externalisation (e.g., as done in [25]). Furthermore, comparisons with other methods such as BEM will definitely be of interest.

6. ACKNOWLEDGEMENT

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.

7. REFERENCES

[1] L. Picinali and B. F. Katz, "System-to-user and user-to-system adaptations in binaural audio," in *Sonic Interactions in Virtual Environments*, pp. 115–143, Springer International Publishing Cham, 2022.

[2] B. F. Katz, "Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation," *The Journal of the Acoustical Society of America*, vol. 110, no. 5, pp. 2440–2448, 2001.

[3] T. Xiao and Q. Huo Liu, "Finite difference computation of head-related transfer function for human hearing," *The Journal of the Acoustical Society of America*, vol. 113, no. 5, pp. 2434–2441, 2003.

[4] C. J. Webb and S. Bilbao, "Binaural simulations using audio rate FDTD schemes and CUDA," in *Proc. of the 15th Int. Conference on Digital Audio Effects (DAFx-12), York, United Kingdom*, 2012.

[5] S. Prepeliţă, M. Geronazzo, F. Avanzini, and L. Savioja, "Influence of voxelization on finite difference time domain simulations of head-related transfer functions," *The Journal of the Acoustical Society of America*, vol. 139, no. 5, pp. 2489–2504, 2016.

[6] P. Siripornpitak, I. Engel, I. Squires, S. J. Cooper, and L. Picinali, "Spatial up-sampling of HRTF sets using generative adversarial networks: A pilot study," *Frontiers in Signal Processing*, p. 54, 2022.

[7] C. Webb and A. Gray, "Large-scale virtual acoustics simulation at audio rates using three dimensional finite difference time domain and multiple graphics processing units," in *Proceedings of Meetings on Acoustics ICA2013*, vol. 19, p. 070092, Acoustical Society of America, 2013.

[8] F. Pind, A. P. Engsig-Karup, C.-H. Jeong, J. S. Hesthaven, M. S. Mejling, and J. Strømann-Andersen, "Time domain room acoustic simulations using the spectral element method," *The Journal of the Acoustical Society of America*, vol. 145, no. 6, pp. 3299–3310, 2019.

[9] G. Fratoni, B. Hamilton, and D. D'Orazio, "Rediscovering the Acoustics of a XII-Century Rotunda through FDTD Simulation," in *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, pp. 1–8, IEEE, 2021.

[10] G. Fratoni, B. Hamilton, and D. D'Orazio, "Feasibility of a finite-difference time-domain model in large-scale acoustic simulations," *The Journal of the Acoustical Society of America*, vol. 152, no. 1, pp. 330–341, 2022.

[11] W. L. Oberkampf and C. J. Roy, *Verification and validation in scientific computing*. Cambridge University Press, 2010.

[12] J. Saarelma, "Finite-difference time-domain solver for room acoustics using graphics processing units," *Master's thesis, Aalto University, Finland*, 2013.

[13] N. A. Gumerov, R. Duraiswami, and Z. Tang, "Numerical study of the influence of the torso on the HRTF," in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. II–1965, IEEE, 2002.

[14] Robert McNeel & Associates, "Rhino3D," 2018-02-06. Version 6.6.18323.1.

[15] N. A. Gumerov and R. Duraiswami, "Computation of scattering from n spheres using multipole reexpansion," *The Journal of the Acoustical Society of America*, vol. 112, no. 6, pp. 2688–2701, 2002.

[16] H. Braren and J. Fels, "A high-resolution individual 3D adult head and torso model for HRTF simulation and validation: 3D data," *Medical Acoustics Group, Institute of Technical Acoustics, RWTH Aachen*, 2020.

[17] S. T. Prepeliţă, J. Gómez Bolaños, M. Geronazzo, R. Mehra, and L. Savioja, "Pinna-related transfer functions and lossless wave equation using finite-difference methods: Verification and asymptotic solution," *The Journal of the Acoustical Society of America*, vol. 146, no. 5, pp. 3629–3645, 2019.

[18] C. J. Roy, "Review of code and solution verification procedures for computational simulation," *Journal of Computational Physics*, vol. 205, no. 1, pp. 131–156, 2005.

[19] J. Saarelma and L. Savioja, "An open source finite-difference time-domain solver for room acoustics using graphics processing units," in *Forum Acusticum, Krakow, Poland*, p. SS11.8, 2014.

[20] S. D. Bilbao, *Wave and scattering methods for the numerical integration of partial differential equations*. PhD thesis, Stanford University, 2001.

[21] J. Meyer, M. Smirnov, A. Khajeh-Saeed, P. F. Hoffmann, and S. T. Prepeliţă, "Finite-difference time-domain simulations: Verification on head-related transfer functions of a rigid sphere model," *JASA Express Letters*, vol. 2, no. 6, p. 062401, 2022.

[22] J. Meyer, S. T. Prepeliţă, A. Khajeh-Saeed, M. Smirnov, and P. F. Hoffmann, "Verification on Head-Related Transfer Functions of a Snowman Model Simulated Using the Finite-Difference Time-Domain Method," *IEEE Transactions on Audio, Speech & Language Processing*, 2023. Accepted for publication.

[23] B. Efron, "Better bootstrap confidence intervals," *Journal of the American statistical Association*, vol. 82, no. 397, pp. 171–185, 1987.

[24] C. J. Roy and W. L. Oberkampf, "A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing," *Computer methods in applied mechanics and engineering*, vol. 200, no. 25-28, pp. 2131–2144, 2011.

[25] I. Engel, D. F. Goodman, and L. Picinali, "Assessing HRTF preprocessing methods for Ambisonics rendering through perceptual models," *Acta Acustica*, vol. 6, p. 4, 2022.