Pinna-related transfer functions and lossless wave equation using finite-difference methods: Validation with measurements

Sebastian T. Prepelit,ă, Javier Gómez Bolaños, Michele Geronazzo, Ravish Mehra, and Lauri Savioja

Citation: The Journal of the Acoustical Society of America **147**, 3631 (2020); doi: 10.1121/10.0001230
View online: https://doi.org/10.1121/10.0001230
View Table of Contents: https://asa.scitation.org/toc/jas/147/5
Published by the Acoustical Society of America

JASA ARTICLE

Pinna-related transfer functions and lossless wave equation using finite-difference methods: Validation with measurements

Sebastian T. Prepeliță,[1,a)] Javier Gómez Bolaños,[2,b)] Michele Geronazzo,[3,c)] Ravish Mehra,[4] and Lauri Savioja[1,d)]
[1]*Department of Computer Science, Aalto University, Otaniementie 17, P.O. Box 15500, FI-00076 AALTO, Finland*
[2]*Department of Signal Processing and Acoustics, Aalto University, P.O. Box 13000, FI-00076 Aalto Espoo, Finland*
[3]*Department of Architecture, Design, and Media Technology, Aalborg University, A. C. Meyers Vænge 15, 2450 København SV, Denmark*
[4]*Facebook Reality Labs, 8747 Willows Road, Redmond, Washington 98052, USA*

ABSTRACT:
Nowadays, wave-based simulations of head-related transfer functions (HRTFs) lack strong justifications to replace HRTF measurements. The main cause is the complex interactions between uncertainties and biases in both simulated and measured HRTFs. This paper deals with the validation of pinna-related high-frequency information in the ipsilateral directions-of-arrival, computed by lossless wave-based simulations with finite-difference models. A simpler yet related problem is given by the pinna-related transfer function (PRTF), which encodes the acoustical effects of only the external ear. Results stress that PRTF measurements are generally highly repeatable but not necessarily easily reproducible, leading to critical issues in terms of reliability for any ground truth condition. On the other hand, PRTF simulations exhibit an increasing uncertainty with frequency and grid-dependent frequency changes, which are here quantified analyzing the benefits in the use of a unique asymptotic solution. In this validation study, the employed finite-difference model accurately and reliably predict the PRTF magnitude mostly within ±1 dB up to ≈8 kHz and a space- and frequency-averaged spectral distortion within about 2 dB up to ≈ 18 kHz.
 https://doi.org/10.1121/10.0001230
(Received 16 October 2019; revised 20 April 2020; accepted 21 April 2020; published online 26 May 2020)
[Editor: Steffen Marburg] Pages: 3631–3645

I. INTRODUCTION

Human perception relies on acoustic information included in the head-related transfer function (HRTF), which accounts for the linear acoustic transformations produced by the listener's head, pinna, torso, and shoulders.[1] Unless adaptation and learning occurs,[2] HRTFs are usually not perceptually transferable[3] mainly due to the uniqueness of the human pinna.[4,5]

Individualized HRTFs are mainly estimated through acoustic measurements and numerical simulations. Although measurements are fundamentally considered a better indicator of reality,[6] HRTF measurements are still impractical and physically un-validated at any standardized level, despite their proven perceptual validity[7–9] and their potentially short acquisition times.[10–12] Due to various limitations[13,14] in accuracy, scalability, reproducibility, and ground-truth definition(s), current HRTF validation studies result in cross-validating[14,15] HRTF measurements with simulations. Although wave-based numerical simulations could offer greater flexibility, HRTF simulations are also limited by, e.g., topological inconsistencies,[16,17] numerical errors,[18] or boundary modelling errors.[19,20] Moreover, their validity in the full audible frequency range has not been yet established using rigorous Verification&Validation (V&V) studies.[21,22] The boundary element method[23–27] (BEM) and the finite difference time domain[16,28,29] (FDTD) methods are the most studied HRTF simulation methods.

Broadly, simulation V&V studies aim to measure the magnitude of the involved errors relative to the working definitions and premises/assumptions. Verification studies aim to quantify the numerical errors contained in a simulated result, while validation aims to measure the adequacy of using the simulated model in predicting the modelled real-world processes.

Although HRTF validation studies based on frequency-smoothed magnitude responses generally show a very good agreement between wave-based simulations and measurements above 200 Hz,[15,30] direct HRTF magnitude validation generally shows an acceptable agreement with measurements below 3 kHz.[16,24,26] Such high frequency mismatch could be caused by measurement errors, simulation errors, or both. In fact, any previous HRTF validation result could be criticized since none independently quantified either the numerical errors or the measurement errors—two critical prerequisites[6] in evaluating the quality of any model prediction. The present study aims to advance the current state in the validation of wave-based HRTF simulations by assessing and investigating the involved measurement and numerical errors.

[a)]Current address: Facebook Reality Labs, 8747 Willows Road, Redmond, WA 98052, USA. Electronic mail: sebastian.prepelita@aalto.fi, ORCID: 0000-0001-9577-660X.
[b)]Current address: Hefio Ltd., Otakaari 5 A, 02150 Espoo, Finland.
[c)]ORCID: 0000-0002-0621-2704.
[d)]ORCID: 0000-0002-8261-4596.

Since the pinna is mainly responsible for HRTF features above 3 kHz,[31] the HRTF validation problem can be simplified for an acoustically-rigid scatterer: the sole influence of the pinna can be studied through the pinna-related transfer function (PRTF). Although there is no simple linear relation between the effects of various anatomical structures in the final HRTF spectra,[32,33] validating PRTFs is one important aspect of validating HRTFs since (i) for a rigid boundary, the most challenging phenomena to be described by numerical solutions are the high-frequency[31] pinna effects, and (ii) at lower frequencies, HRTFs are similar to the acoustic scattering of a sphere,[34] which has been previously verified[28] and validated.[35,36] Nevertheless, HRTFs are generally more difficult to validate since they involve posture changes,[37] or high-frequency impedance effects from hair or clothing.[38,39]

The present validation work extends and completes a companion verification study[18] where formal PRTF solutions together with their precision were estimated. The previously-obtained predictions and grid-specific PRTF computations are compared here to PRTF measurements to further investigate the effects of the numerical errors caused by the complex pinna geometry. The ultimate objective of the current study is to answer whether the inhomogeneous wave equation without any losses is adequate in predicting the magnitude of blocked-meatus far-field PRTFs for a stiff scatterer.

The remainder of the paper is organized as follows: Sec. II briefly describes the PRTF problem; Sec. III A presents a detailed account of some common errors for an HRTF/PRTF validation study and the employed means to minimize/assess such errors; next, the data used in the validation study is characterized: the numerical simulations in Sec. III B and acoustical measurements in Sec. III C; validation results and a validation metric are presented in Sec. IV; Sec. V highlights the limitations of the employed methods and experimental design; Sec. VI contextualizes the work and the results; finally, the findings are summarized and concluded in Sec. VII.

II. PRTFS

The PRTFs are formally defined in the frequency domain as the complex division of a sound pressure signal captured at a fixed location of interest inside the pinna cavity/ear canal, $P_{\text{ear}}(\omega)$, and a reference pressure, $P_{\text{ref}}(\omega)$, captured at $\mathbf{r}_{\text{ref}} \in \mathbb{R}^3$ with the pinna absent. ω denotes angular frequency.

Both signals are measured in the free-field and in a quiescent medium, while the $P_{\text{ear}}(\omega)$ measurement could be done with the ear placed in a finite baffle.[40] Both measurements assume an ideal point source placed at a location $\mathbf{r}_1 \in \mathbb{R}^3$ relative to $\mathbf{r}_{\text{ref}}$, and ideal point receivers. Using linear, time-invariant, and identical instrumentation for both measurements, the biases induced by the measurement chain would cancel out, yielding the "true" free-field corrected PRTF,[41]

$$\text{PRTF}(\mathbf{r}_1, \omega) = \frac{P_{\text{ear}}(\mathbf{r}_1, \omega)}{P_{\text{ref}}(\mathbf{r}_1, \omega)}. \quad (1)$$

III. METHODS

This section describes the simulations and measurements used in the validation study together with the steps taken to minimize the validation errors.

A. Sources of error

A validation study has specific sources of error.[44] Table I summarizes the relevant sources of error acknowledged for the current study. The larger the number of sources of errors, the more likely that error cancellation occurs, which could increase the chance of type I/II errors in the validation results. Thus, for reliable conclusions, all the acknowledged errors must be addressed.

1. Measurement errors

Any reliable validation study requires some form of quantification of the measurement error.[6] Linear HRTF/PRTF measurements in controlled environments are generally highly *repeatable*[15,45] (i.e., repeated measurements of the same measurand under identical conditions,[46] like set-up, procedure, inputs, observer, etc.), but could suffer from low *reproducibility*[47–50] (i.e., replicated measurements of the same measurand under changed conditions[46]). Reproducibility is mainly affected by measurement biases like orientation mismatches[51] or unwanted scattering, which are difficult to quantify.

HRTF/PRTF measurement errors include: limited dynamic range caused by acoustical,[52] vibrational, and electrical noise; inappropriate head/body posture;[34] issues with direct current (DC) voltages[53] and gain mismatches; non-ideal sound source such as a directive or non-linear source; non-ideal receivers such as directive microphones; drifts in environmental conditions such as disturbances in humidity, temperature,[54] equilibrium pressure, or airflow; variations within measurement apparatus caused by, e.g., thermal effects or hysteresis; time-invariance violations caused by, e.g., living subjects;[51,55] errors related to the measurement point such as pressure leakages, physical distortions of the outer ear,[34] difficulty in positioning a sensor inside the meatus;[47,56] data acquisition errors caused by, e.g., electro-magnetic interference, errors in digital/analog converters, quantization errors, buffering inconsistencies; errors and limitations induced by the excitation signal,[57,58] especially for stochastic sequences;[55] potential non-linear vibro-acoustic coupling/losses for scatterers of large mechanical compliance;[59] reduction and post-processing errors/interactions; or other unacknowledged errors mostly driven by the skill and expertise of the experimenter.

2. Input errors

Geometry errors: Acoustically, the topology of the pinna is of paramount significance for PRTF simulations.[17,60]

https://doi.org/10.1121/10.0001230

TABLE I. A summary of the relevant sources of error for the current validation study.

Validation error	Source	Effects on the validation process
FDTD simulation	Discretization errors	Could bias and scatter the computation relative to the formal solutions (see Sec. II D in the companion study[18]).
	Round-off errors	
	Pollution errors	
	Voxelization errors	
Measurement	Electrical and acoustical noise	Decreased dynamic range in the measured HRTFs. Higher impact for the contralateral ear.
	Unwanted scattering	Biases the measurement which effectively changes the validation problem.
	Source directivity	Biases in the measurements causing changes in the PRTF features. The errors introduced by non-ideal sources are larger in the near-field.
	Source size	
	Non-ideal receivers/sensors	
	Source non-linearity	Level-dependent biases in the PRTF features.
	Environmental inconsistencies	Peak/notch shifts or even unwanted scattering.
	Excitation signal errors	Can reduce the dynamic range and SNR or could[42] bias the resulting HRTFs.
	Data acquisition errors	Can induce both random error and bias.
Input	Geometrical/topological inconsistencies	Differences in PRTF spectra (both fabrication errors and 3D mesh representation).
	Boundary surface impedance values	Could affect the PRTF features.[43]
Alignment	Real objects or 3D mesh	Measured and simulated PRTFs at incorrect locations.
Modeling	Approximations or poorly known phenomena/definitions	Quantified in the validation process if other unacknowledged errors are insignificant.
Post-processing	Inconsistent interaction with the errors in the two domains	Usually induce biases in the validation result.

It is generally difficult to have complete and exact access to the scatterer of interest. To avoid such ground-truth issues, the scanned pinna mesh was considered the "true" geometry of the pinna and then three-dimensional (3D) printed (see Fig. 1).

Impedance errors: Another source of input error is the boundary impedance. To minimize such source of error and restrict the validation domain, the pinna was 3D printed in a rather stiff material. See Sec. III B 3.

Source/receiver errors: Directional sources will cause deviations in the scattered field compared to the available point source in a simulation. The off-axis amplitude deviation of the loudspeaker used in the measurements (see Fig. 1) at $r = 1$ m was coarsely quantified: for a fixed microphone orientation, the deviation in dB from the on-axis response was calculated when the loudspeaker was manually rotated approximately $\pm 5^\circ$ and $\pm 10^\circ$ in both azimuth and elevation. The analysis was carried in an anechoic chamber based on measured impulse responses (IRs) obtained using a logarithmic sweep.[61] For $\pm 5^\circ$ loudspeaker rotations, measured results showed a maximal deviation of about ± 0.5 dB within 0.5–20 kHz and about ± 1 dB within 20–24 kHz. Larger deviations are seen for $\pm 10^\circ$ loudspeaker rotations: ± 2 dB within 0.5–20 kHz and up to –5 dB within 20–24 kHz. Therefore, the pinna was printed and mounted with a very compact baffle such that the (on-axis) solid angle seen by the loudspeaker is small enough to assume constant-amplitude incoming wavefronts. The bounding box of the pinna is roughly $5.2 \times 4.3 \times 7.7\,\text{cm}^3$ yielding a maximal angular coverage at 1 m from the center of the loudspeaker of about $\pm 2.92^\circ$. For such angular span, amplitude deviations less than 1 dB in the loudspeaker directivity up to 20 kHz are expected.

One microphone model was used: a miniature electret condenser microphone FG-23329 (Knowles, Itasca, IL). Its directivity, together with sensitivity to small misalignments,

FIG. 1. (Color online) The used loudspeaker and the mounting of the pinna (lower right). The marked angle was verified to be 90° (angle bracket used during first measurement session).

was crudely assessed for one specimen using the same setup used in the P_{ref} measurements. Six sweep measurements were taken in the free-field with the microphone fixed to a rigid wire at 1 m. In each measurement, the microphone was rotated, which also caused about ± 2 mm location misalignments. Such uncertainty was quantified using a student-t distribution based on the measured magnitude levels in dB: the 95% confidence standard error of the mean is generally within ± 0.3 dB up to about 13.5 kHz with the exception of a ± 0.8 dB peak around 8.9 kHz. The uncertainty increases to around ± 0.9 dB up to 19.3 kHz, after which it reaches ± 2 dB at 24 kHz.

3. Alignment errors and coordinate system

Since the pinna was printed without any reference planes or points, an M8 nut was used for orientation matching. First, the location of the nut relative to the pinna was assessed by measuring three distances.

Next, three Cartesian Euler angles are matched. To match rotations in the x and z axes, the back of the nut can be used: four reference points on the surface of the external ear were chosen (see Fig. 3) and the distances to the reference plane were each repeatedly measured five times with the Digimatic 500–311 caliper in random order by one of the authors. The same distances were created in the simulation domain and the ear manually rotated until the four points on the mesh surface were within the 95% confidence intervals (CIs), assuming a Gaussian error. To match the y axis, the pinna was mounted and rotated such that a vertical plane passed through three easily defined points (intertragal notch, center of the microphone, and top of the ear) and such that the inter-point distances could easily be measured and cross-checked on the mesh (see Fig. 2). The alignment errors for the x, y, and z axes are estimated to be within $\pm 0.3^\circ$, $\pm 1^\circ$, and $\pm 0.5^\circ$, respectively. Finally, the arm onto which the bolt was mounted for the PRTF measurements was checked to be properly aligned within the $\pm 0.2^\circ$ tolerance of a level tool (rotations in x axis).

B. Wave-based simulations

To limit the difficulty of the validation study, the present validation study will address only the magnitude of the ipsilateral blocked-meatus PRTFs, computed in the far-field with a lossless wave-based model. Since the continuous problem is linear and *well-posed*, the estimated FDTD asymptotic solutions (i.e., free of numerical errors) will theoretically converge to the same solution, independent of the continuous formulation and simulation method (e.g., FDTD, BEM).

1. Models used in validation

The same models and simulations as in the verification study[18] are employed: the inhomogeneous 3D wave equation coupled with the standard rectilinear update ran at maximal stable Courant number $\lambda_C = 1/\sqrt{3}$. The FDTD update is run on uniform Cartesian grids characterized by the grid

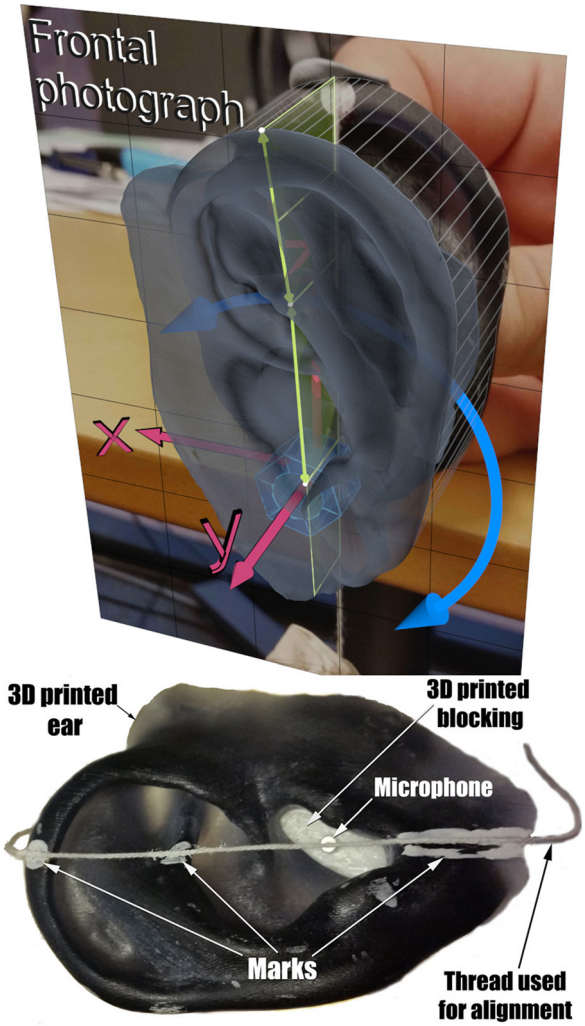


FIG. 2. (Color online) Vertical ear alignment. The picture shows the white thread and the markers on the 3D-printed ear replica. The upper picture shows the extra primitives used to minimize the parallax effects from the used reference picture.

spacing ΔX. The pinna surface is voxelized on the same grid resulting in stairstepped boundaries.

Based on a weighted linear regression model,[62] a reliable estimate of the formal solution and its associated uncertainty are obtained by asymptotically extrapolating the response on multiple grids. For more details on the employed models and asymptotic prediction, see the companion study.[18]

In the present work, the principle of reciprocity is used in the simulations. Only acoustically rigid boundaries $\beta = 0$ are considered in this study, where β is the specific acoustic admittance at the boundary (see Ref. 63, p. 261).

2. Simulations used in validation

To avoid extra difficulties in the validation such as correspondence with reality, the PRTF domain was chosen without an absorbing bounding layer (such as a perfectly matched layer).

For the P_{ear} simulation, the pinna mesh (see Sec. III B 3) was placed at the center of a domain bounding-box such that

https://doi.org/10.1121/10.0001230

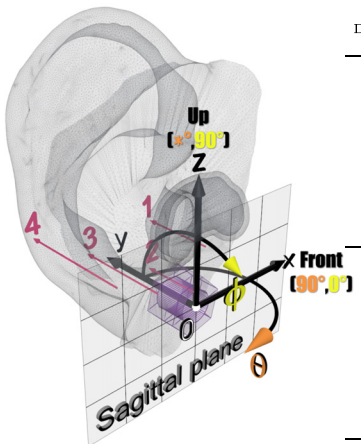
no reflections from the domain boundary would reach the receiver locations described in Fig. 3 during the considered time window of 6.25 ms. The reference simulations to estimate P_{ref} in Eq. (1) only contained the bounding-box and the source was placed and interpolated at a location close to the origin of the pinna coordinate system (see Fig. 3).

To estimate the asymptotic (here, as $\Delta X \to 0$) PRTFs, the same grid ensemble as in the companion study[18] are used (see Table II). A discrete-time impulse sequence $\delta[n]$ ($\delta[0] = 1$, $\delta[i \in \mathbb{Z}^+] = 0$) was used to drive both simulations in Eq. (1). A sound speed of 344.34 ms^{-1} was employed which corresponds (see Ref. 64, p. 121) to a temperature of 294.8 K.

3. The human ear replica

The input mesh used in the PRTF simulations is a laser scan of the cast of an otologically normal human pinna. To minimize the discrepancy in the used geometry in the model versus reality, additive manufacturing (AM) is employed: the pinna mesh is 3D printed at 100% infill using ProJet SD 3000 (3D Systems, Rock Hill, SC) having a claimed accuracy of 0.025–0.05 mm per inch. The surface of the printed pinna was slightly sanded to smoothen the surface and minimize the resulting stair-stepping[65,66] form-error in the AM process. The pinna was printed in a rather stiff material: gray Acrylic Plastic VisiJet® SR200 of claimed (ASTM D638) tensile modulus of 866MPa.[67]

To confirm the printing accuracy, physical measurements were compared with the corresponding distances in the 3D mesh. Such distances were found to be similar within about 0.5 mm. Real-world lengths were measured using the Digimatic 500–311 (Mitutoyo Corp., Kawasaki, Japan) caliper of accuracy 1 μm, while the lengths in the simulation domain were obtained with the tape tool in 3ds max® (Autodesk Inc., San Rafael, CA).



The points 1-4 were used to match the orientation in x and z rotations.

Direction	Azim. θ [°]	Elev. ϕ [°]	Radius [m]
back	−90	0	1
	−45	0	1
left	0	0	1
	45	0	1
front	90	0	1
	−90	40.6	1.01
	−45	40.6	1.01
	0	40.6	1.01
	45	40.6	1.01
	90	40.6	1.01

FIG. 3. (Color online) Used Cartesian and spherical coordinate systems (left) together with the directions used in the validation study (right). The origin O is at the center of the nut and in the back plane (here, a sagittal plane). $\theta \in [-90°, 90°]$ represents azimuth angles and $\phi \in [-90°, 90°]$ represents elevation angles while directions are given as (θ, ϕ) pairs. The directions in the first column are for the left ear of a hypothetical head.

TABLE II. Grids used in the PRTF convergence study. N_{voxels} represents the total number of voxels used in the full computational domain [without any message passing interface (MPI) halos]. N_{freqs} represents the number of positive frequencies, Δf represents the frequency resolution used in the analysis, f_s the sampling frequency, while N_{dirs} the total number of PRTF directions. $\Delta X, N_{\text{voxels}}$ values are rounded to two decimal places.

Grid	f_s	ΔX	$\text{MPI}_{\text{Nodes}}$	N_{voxels}
	(Hz)	(mm)		($\times 10^9$)
1	1 014 240	0.59	7	100.47
2	922 080	0.65	5	75.33
3	838 240	0.71	4	56.72
4	762 080	0.78	3	42.62
5	692 800	0.86	3	31.98
6	629 760	0.95	3	23.98
$\Delta f = 160$ Hz		$N_{\text{freqs}} = 151$		$N_{\text{dirs}} = 10$

To measure at the blocked-meatus location, an ear blocking was designed to fit the measurement microphone and rigidly fit inside the entrance of the ear canal. The ear blocking was 3D printed in white polylactic acid at 100% infill with a 3 Dual Extruder (MiniFactory Oy LTD, Seinäjoki, Finland) printer.

In the simulations, the microphone is replaced by a cylinder and the mesh of an M8 hexagonal nut is added for orientation matching. No measurement rigs or mounting equipment were included in the simulation domain.

4. Considered directions

A metal M8 nut was glued to the printed pinna. It was used to match the orientation in the simulation domain: the origin was chosen in the back plane of the nut, in the center of the corresponding bolt. The back plane of the nut is considered a sagittal plane (see Fig. 3) while the central axis of the bolt is defined to be parallel to the interaural y axis. The *up* axis was chosen as the Cartesian z-axis yielding the x axis as the *front* direction.

From the defined origin, a vertical-polar spherical coordinate system is defined such that the azimuth angle θ varies from 90° in the frontal direction to −90° in the back while the elevation angle ϕ varies from 90° (top direction) to −90°. Ten directions are considered sufficient for the present validation study: five in the horizontal plane and five at about $\phi = 40°$ elevation for a radius of $r = 1$ m. A $\Delta\theta = 45°$ azimuthal separation is considered (see Fig. 3). Two measurement sessions were conducted: due to physical limitations, the elevation in the second setup was slightly larger and measured to be approximately $\phi = 41.18°$. To avoid running two independent convergence studies, an average elevation angle of $\phi = 40.6°$ is considered in the simulations which will introduce approximately $\pm 0.6°$ error in elevation angle between measurements and simulations outside the horizontal plane.

5. Simulation post-processing

To reduce the uncertainty in the predicted solutions, the simulations were trillinearly interpolated in space for both the source and receiver at a consistent 3D continuous location.[18] To further improve the precision of the asymptotic predictions, prior to the frequency division in Eq. (1), the interpolated FDTD solutions for each grid are processed with a sampled Gaussian window,

$$w(t) = e^{-\alpha^2(t-t')^2}, \tag{2}$$

where $t' = t_{\text{obs}}/2$ and α was chosen such that $w(t_{\text{obs}}) \approx 1 \times 10^{-7}$. Here, t_{obs} represents the observation interval, which is equal to the length of the measurement temporal window or the considered simulation interval. The window in Eq. (2) reduces (i) the edge effects in discrete-Fourier domain for the FDTD simulations and (ii) any potential unwanted reflections in the PRTF measurements.

C. PRTF measurements

Two measurement sessions were conducted in a large chamber fitted with 80 cm long wedges that created an anechoic environment above 100 Hz and a noise floor below 5 dB sound pressure level (SPL) within 0.1–16 kHz. All IRs were obtained using 10 s long logarithmic sweeps[61] designed between 0.1 Hz and 24 kHz and sampled at $f_{s,m} = 48$ kHz. The separable[42] nonlinear parts were excluded from the measured IRs.

Only the ambient temperature was monitored prior to each measurement. It was found to vary about $\pm 0.5^\circ$ for p_{ear}, with an average of 294.83 and 295.61 K for the two measurement sessions.

All the measurements used one loudspeaker (shown in Fig. 1), which consisted of a wooden enclosure and a 2-inch Peerless (Tymphany, Taipei City, Taiwan) audio driver having a lower 3 dB cutoff at 150 Hz and a magnitude response within ± 3 dB between 0.15–20 kHz. The loudspeaker was designed[68] to efficiently radiate sound energy in the 100 Hz–20 kHz range and was fed through an MX-70 (Yamaha, Hamamatsu, Japan) amplifier from an RME Fireface 400 (Audio AG, Haimhausen, Germany) audio interface.

The P_{ear} measurements received more attention than those for P_{ref} since it was assumed that P_{ref} embeds a negligible amount of measurement errors.

P_{ref} **measurements**: Only one P_{ref} measurement was conducted for each measurement session. The microphone was placed at 90° incidence at the center of a turntable at $r \in \{1, 1.01\}$ m.

P_{ear} **measurements**: Two measurement sessions were conducted. For each, every direction was repeatedly measured three times in randomized order.

To minimize acoustic leakage, the used electret condenser microphone FG-23329 was tightly secured in the designed blocking while the hole at the end of the ear-canal location through which the microphone cable passed was covered with polymer clay and tape. No special vibration-isolation treatment was employed around the microphone to address vibrational leakage.

Measurement session #1: An automatic rotating table ET250–3D (Outline s.r.l., Brescia, Italy) of claimed accuracy of 0.5° was employed for the azimuth positioning. Its surface was verified to be within the $\pm 0.2^\circ$ accuracy of a level tool. On top of the turntable, a thin structure made of 2×2 cm^2 aluminum profiles (Aluflex AB, Hesingborg, Sweden) was constructed [see Fig. 4(a)]. The pinna was then mounted to the structure such that the backplane of the nut was vertically within about 1 mm from the center of the dish through which the rotation axis of the turntable should pass. The vertical alignment of the pinna was also adjusted

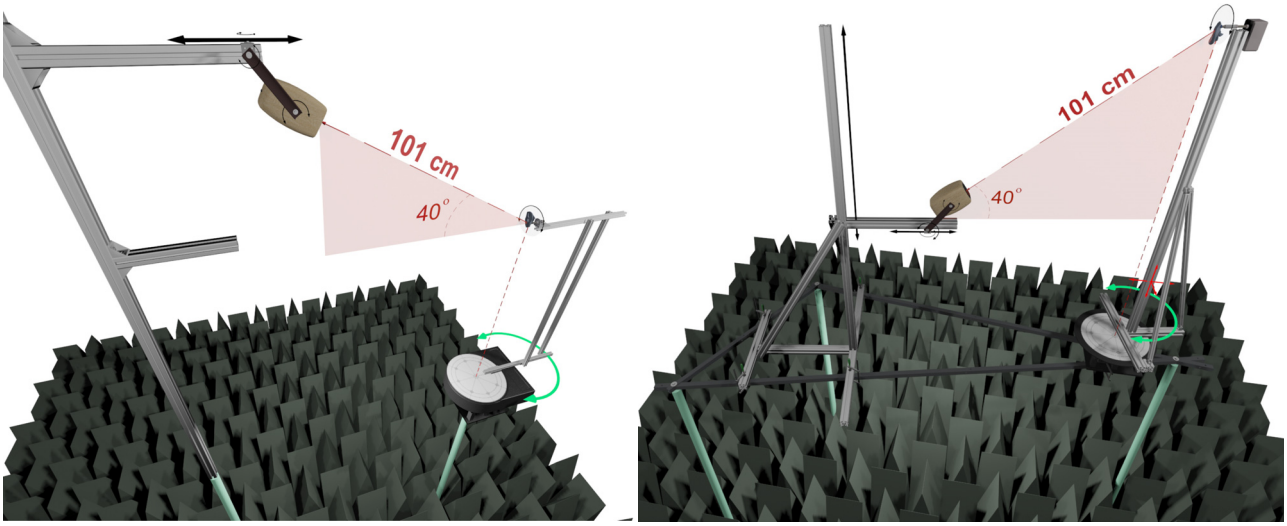


FIG. 4. (Color online) 3D sketches of the two measurement systems. In the 2nd measurement session, the pinna is mounted upside down. Arrows indicate allowable rigid transformations of movable objects in the setups (black arrows indicate fixed transformations for a given elevation; i.e., stationary objects). (a) First measurement session ($\phi = 40^\circ$). (b) Second measurement session ($\phi = 40^\circ$).

https://doi.org/10.1121/10.0001230

as mentioned in Sec. III A 3. No severe vertical deviations of the structure were observed. The pinna structure was mechanically isolated from the structure on which the loudspeaker was mounted.

Measurement session #2: To estimate the measurement bias, a second measurement session was conducted. The pinna was attached to a more rigid structure that was built with Aluflex 4×4 cm^2 aluminum profiles and which were expected to increase the amount of unwanted scattering. The loudspeaker structure was similar as in the first session (see Fig. 4). In this measurement session, the pinna and loudspeaker structures were mechanically connected through a steel triangular frame which was calibrated to be almost horizontal and was mounted on three vertical poles in the anechoic chamber. A small tilt in the triangular frame of about 0.5° was introduced by the weight of the other structures.

In order to reduce the amount of acoustical scattering from the heavier structures, acoustically absorptive material was used while the pinna was mounted up-side down for $\phi = 40°$. The pinna was aligned as described in Sec. III A 3.

A manual "turntable" was used which consisted of a planar round support surface with azimuth angle grading and a marked rotation axis. Such surface was fixed to the triangular frame and calibrated to be horizontal with the level tool. Two of the authors alternatively hand-operated and positioned the pinna structure on top of the turntable for each measurement. As in the first measurement session, three measurements per direction were acquired in a random order.

With the help of a Bosch GLL 3–80 Professional (Robert Bosch GmbH, Gerlingen-Schillerhöhe, Germany) laser level (set on "free"), the back of the nut was quantified to be within ±1 mm from the rotation axis, while the error in the azimuth angle is expected to be within ±1°. The source was at $r = 100.25$ cm ± 2 mm in the horizontal plane and $r = 101.4$ cm ± 2 mm for $\phi = 40°$.

1. Post-processing

In order not to introduce post-processing errors in the validation, both measurements and simulations must be identically transformed. To begin with, the analysis time-window is set to $t_{\text{obs}} = 6.25$ ms, equivalent to 300 measurement samples, yielding a frequency resolution of $\Delta f = 160$ Hz.

For the measured IRs, the temporal origin is determined by subtracting the 1 m travel time (based on the recorded temperature) from the average group delay in the 1–5 kHz frequency range for p_{ref}. The subsequent 300 samples are extracted and the Gaussian window in Eq. (2) is sampled at and applied to each measured p_{ref} and p_{ear}. It is assumed that such a window has negligible effects on the cancellation of the microphone and loudspeaker magnitudes in Eq. (1): no significant gain bias was observed in the measured PRTFs when a rectangular window was applied.

IV. VALIDATION RESULTS

In the following, for consistency[18] and analysis of the asymptotic predictions for more noisy numerical data, the analysis is done up to 24 kHz despite the irrelevance above 20 kHz for sound perception.

A first qualitative comparison is shown in Figs. 5–8. Note the uncertainty embedded in the CIs is not propagated through the logarithm function in the plots in Fig. 5. For ease of readability, the 95% estimated two-sided CIs of the asymptotic PRTFs from Fig. 5 are presented in Fig. 7.

Since quantitative comparisons are usually favored in computational physics[69] and to better support the qualitative data in Figs. 5–8, a frequency-dependent[16] spectral distortion (SD) is first used as an HRTF comparison metric. Due to the magnitude-based FDTD predictions, the unsigned average difference of the SD metric reads

$$\mathrm{SD}[\omega_i] = \frac{1}{N_{\text{freqs}}} \sum_{i=1}^{N_{\text{freqs}}} \frac{1}{N_{\text{dirs}}} \times \sum_{j=1}^{N_{\text{dirs}}} 20 \log_{10} \left| \frac{\widetilde{\mathrm{PRTF}}\left[\omega_i, \theta_j, \phi_j, \Delta X\right]}{\mathbb{E}[\mathrm{PRTF}_m]\left[\omega_i, \theta_j, \phi_j\right]} \right|, \tag{3}$$

where $\mathbb{E}[\mathrm{PRTF}_m]$ is calculated for different measurements taken at the same direction and frequency bin, and $\widetilde{\mathrm{PRTF}}[\omega_i, \theta_j, \phi_j, \Delta X]$ is either the asymptotic $\Delta X \to 0$ prediction or an individual computation on one of the grids in Table II.

A. The two measurement sessions

Regarding the two measurement sessions, measurements generally agree (see Fig. 5). Moreover, each measurement session was quite repeatable: a maximal deviation from the average for all directions and frequencies below 20 kHz was within 0.6 dB for the first measurements and 1.0 dB for the second measurements. This indicates that potential drifts in the environment or apparatus for each measurement session were insignificant.

However, there are some noticeable qualitative differences. First of all, the second measurement session generally shows an increased PRTF amplitude especially for spectral peaks: qualitatively in Fig. 5; quantitatively, signed SD metric [i.e., without the outer absolute in Eq. (3)] shows that the levels in the second measurement session are, on average, about 0.5–1 dB higher between 6.5–9.3 kHz (mostly statistically insignificant) and 12–16 kHz and even 3 dB higher above 20 kHz (plot not shown). The most likely cause is the potential vibration in the structure and the pinna in the first measurement session. If true, the second measurement session could be viewed as having an increased signal-to-noise ratio (SNR) relative to a rigid scatterer. This could be the reason why the second measurements usually show more pronounced PRTF features: some features are smoother in the first measurements, e.g., the extra dip around 14 kHz for the (0, 0) direction. Note no accurate dB SPL levels were measured.

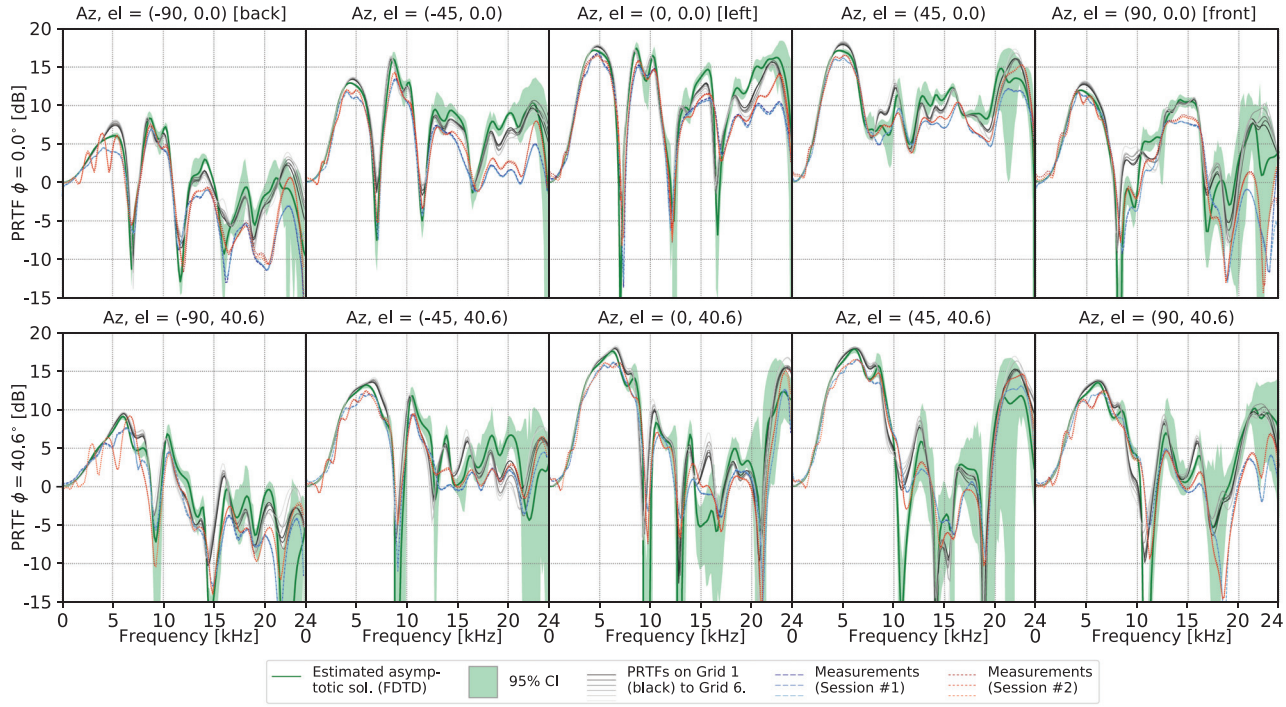


FIG. 5. Measurements for both sessions and first order weighted least squares asymptotic solution estimate for the FDTD simulation (thick, solid line). Top row of images shows the direction in the horizontal plane ($\phi = 0°$), while bottom row shows the elevated directions ($\phi = 40°$). All three measurements per direction are shown: dashed lines for first session and dotted lines for the second session. The transparent filling represents bias-corrected and accelerated bootstrapped-pairs CIs. The 6-grid pool of computed solutions from the convergence study is also displayed for each direction in shades of grey.

Another major difference can be seen for directions in the back $[(-90, 0)$ and $(-90, 40.6)]$: the second measurements show an additional modal pattern below 7 kHz. A subsequent measurement with a damping material placed at the back of the pinna showed that such pattern decreased in magnitude (plot not shown). Thus, the most likely cause of such pattern is due to measurement error caused by the unwanted scattering of the larger pinna mounting system.

Finally, there are sporadic mismatches of some PRTF features between the two measurements, e.g., the last notch for the (90, 0) direction. The most likely cause was orientation mismatches—both measurements were done with some alignment errors.

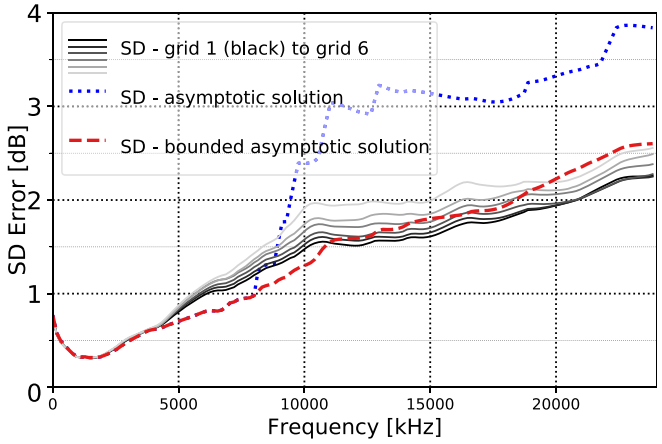


FIG. 6. (Color online) SD metric in Eq. (3) (averaged in the $[0, \omega]$ frequency band and across all the $N_{\text{dirs}} = 10$ directions) for the asymptotic prediction (dotted line), the bounded asymptotic solution (dashed line), and individual computations on the six grids.

B. General magnitude differences

Figure 6 shows the SD metric from Eq. (3): each point represents the SD calculated from DC across all directions. Considering Fig. 6, the SD for the asymptotic solution increases sharply at the predicted PRTF notches. Due to limited dynamic range, the measured notches are bounded by the measurement noise floor. Moreover, the asymptotic predictions could pass below the computation noise floor imposed by the round-off error. Finally, the uncertainty in the asymptotic solution is high at deep notches. As such, the

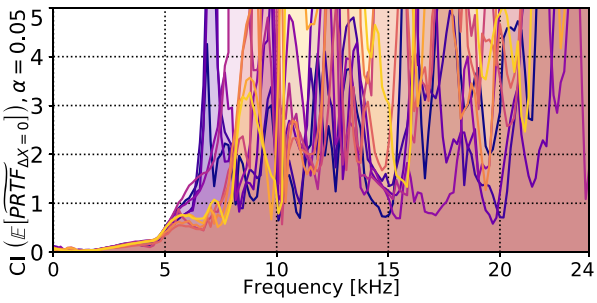
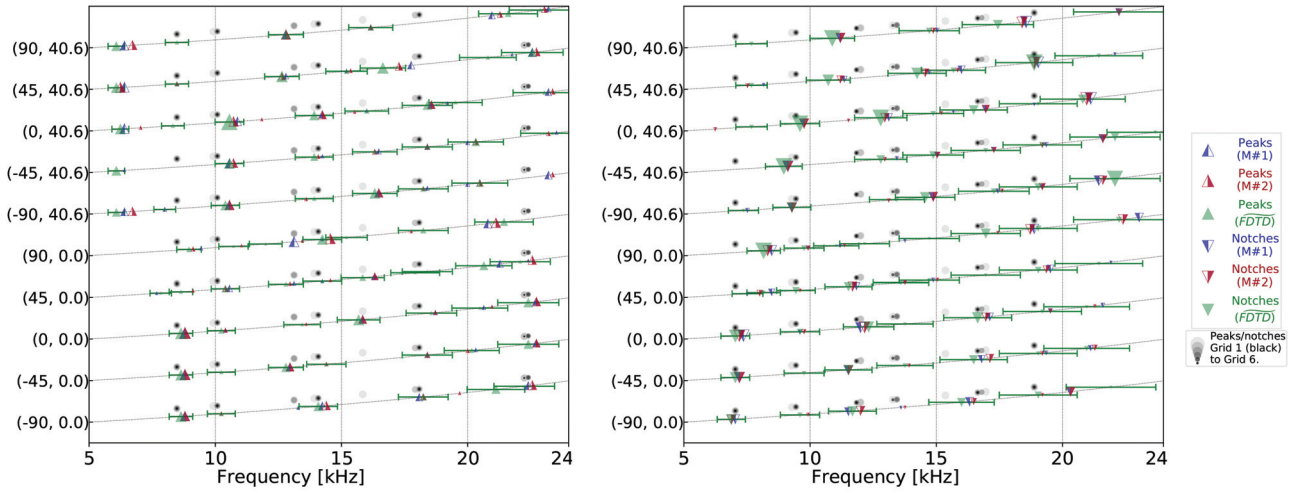


FIG. 7. (Color online) Estimated CIs for the asymptotic solution (same as in Fig. 5). CI estimation was done with bias-corrected and accelerated bootstrapped-pairs method. Same color coding as in Fig. 9. The ordinate axis is truncated at 5 dB.

https://doi.org/10.1121/10.0001230



(a) peaks (b) notches

FIG. 8. PRTF feature extraction. For the asymptotic solution and measurements, the marker sizes are proportional to $\log_{1.4}(2+\wp)$, where $\wp$ is the prominence of the peaks/notches in dB as extracted by the *find_peaks*() function (see text). The grid-axes for each direction are curved to discriminate overlapping error bars that are the fixed-level static frequency difference limens at 8 kHz according to Moore *et al.* (Ref. 70). *M*#1 and *M*#2 represent the first and second measurement session, respectively. $\widetilde{\text{FDTD}}$ is the predicted asymptotic solution. PRTF features extracted from computations on each grid in Table II are also shown vertically displaced to minimize data clutter—for these, the marker size is proportional to the grid index, not the prominence.

SD analysis on the asymptotic solution could be considered misfocused for the present study.

A more appropriate comparison is to bound the notch depth in the asymptotic predictions. Figure 6 also shows the SD results with the asymptotic predictions bounded to –25 dB (value chosen based on the minimum of the measured PRTF across all directions). Thus, considering the "corrected" SD in Fig. 6, such a metric shows an improved asymptotic prediction compared to individual computations up to around 11 kHz, after which it degrades faster than individual computations.

Figures 5 and 6 show that, in contrast to single-grid computations, the asymptotic solution shows improved qualitative results relative to the acquired measurement levels within 1–11 kHz magnitude corrections are seen compared to the individual computations.

On the other hand, the magnitudes of the asymptotic solutions usually depart further from the measurements compared to individual computations at higher frequencies as shown by the averages in Fig. 6 (above 8 kHz for some individual computations or worse than any grid computation above 20 kHz) or by many individual examples in Fig. 5 such as for the 12–15 kHz bandwidth for the $(-90, 0)$ direction. This could indicate inaccurate modeling at higher frequencies (e.g., higher losses occur in reality).

Similarly to the companion study,[18] Fig. 6 exemplifies once more the caveats in using a single-grid computation: validation conclusions drawn on computations lacking rigorous error analysis could be misleading due to, e.g., error cancellation. Ignoring the inherent uncertainties, one can wrongly consider the FDTD computations as accurate at higher frequencies based on some coarse-grid simulation as the SD metric suggests in Fig. 6; modeling errors (e.g., high-frequency losses) or measurement errors could cancel out with the numerical errors. Note that the CIs also increase with frequency as shown in Fig. 7; thus, such behavior is not certain.

Notice in Fig. 7 the low reliability of the asymptotic prediction (i.e., CIs are usually larger than 2 dB) above about 10 kHz even for the sub-millimeter voxel sizes used. For some directions, the precision of the asymptotic prediction starts degrading at even lower frequencies—at the location of the first notch. At higher frequencies, smaller grids are needed for reasonable confidence of the PRTF predictions in magnitude.

Finally, the actual SD values in Fig. 6 are generally below reported values on unsmoothed HRTFs in the literature[16,71–73] up to about 20 kHz for both the bounded asymptotic solution and individual computations. One main reason is the sub-millimeter grids used in the present study. The SD is below 1 dB up to around 8 kHz for the asymptotic solution. As such, the SD metric shows that the present work improves on the general problem of wave-based HRTF/PRTF validation, but it also suggests the need to improve the modeling and/or the measurements.

C. Differences in spectral features

Peaks and notches were extracted from the magnitude levels in dB within the 6–24 kHz bandwidth for the asymptotic solution $\widetilde{\text{PRTF}}$ and the two measurement sessions. In particular, the Python function *find_peaks*[74] from the *scipy.signal* package was used for such feature extraction without any interpolation of the magnitude. Notches were analyzed

as the peaks of $20\log_{10}(|\widetilde{\text{PRTF}}|^{-1})$. Figure 8 shows the results.

To begin with, most of the actual peaks and dips of the asymptotic solution get displaced in frequency towards the features in the measurements: for instance, the peak at 10 kHz for the (45, 0) direction in Fig. 5. When compared to the measurements, only about 10.7% of the single-grid peaks are closer in frequency to the measurements compared to the asymptotic peaks, while 9.9% of the notches for the single-grid computation are closer to the measurement notches (percentage calculated from the total number of identified features; features identified manually in Fig. 8). Generally, there is still room for improvement for the features of the asymptotic solution: higher frequency shifts of such features are needed to accurately match the measurements. Similar to the companion study,[18] Fig. 8 shows that there could be more spurious features in the single-grid solutions compared to the asymptotic PRTFs; similarly, such single-grid computations miss more PRTF features present in measurements (for instance, the notches above 20 kHz and most peaks/notches around 15–16 kHz).

To have a rough idea about the discriminability of individual peaks, the $\pm 5.5\% f_{\text{peak}}$ and $\pm 8.0\% f_{\text{notch}}$ just-noticeable differences (JNDs) from the work by Moore *et al.*[70] are also shown as error bars. Although there are very good matches such as the peak for $(90°, 40.6°)$ at 12.8 kHz or the notch for $(-90°, 40.6°)$ at 9.3 kHz, poorly or missing predicted features of large prominence also exist: for instance, the peak for $(90°, 0°)$ around 13–14 kHz. Nevertheless, considering spectral discrimination,[70] the feature prediction appears satisfactory for most cases. In Fig. 8, independent of the prominence, there is at least one measurement feature (i.e., peak or notch) within each assumed JND for 85.62% of the peaks/notches of the asymptotic solution. Of the asymptotic FDTD features above 6 kHz, 10.27% do not correspond to any measurement feature (i.e., no measured feature is within the assumed JND) captured in Fig. 8. Here, measured features which were close in frequency but outside the assumed JNDs of the corresponding simulated features were only counted once as a misprediction (i.e., they were not considered a separate feature).

Thus, the asymptotic solution contains most of the measured PRTF features. This is also some evidence that no significant extra scattering from the measurement apparatus was present in the post-processed measurements.

Still, there are some noticeable differences: the PRTF peaks and notches are rarely perfectly predicted by the asymptotic solution. There are many possible causes for such mismatches: orientation mismatches, surface impedance mismatches, temperature mismatches, mechanical coupling, or source/receiver location mismatches. The first two are the most likely and their quantification would require additional sensitivity analyses.

Finally, Figs. 5 and 8 show qualitatively that the asymptotic solution exhibits more pronounced notches compared to measurements, likely due to an increased SNR, e.g., the first notch around 9 kHz for the $(-45, 40.6)$ direction. Quantitatively, the more pronounced notches in the asymptotic solution can be quantified through the notch prominences of the magnitudes levels: the estimated average notch prominence above 6 kHz for the asymptotic solution is about 17 dB, while it is 6.83 and 7.11 dB for the first and second measurement session, respectively. The average notch prominence differences were found to be significantly different using a Welch's t-test ($p_{\text{M1vsFDTD}} = 0.015$ and $p_{\text{M2vsFDTD}} = 0.017, N_{\text{M1}} = 57, N_{\text{M2}} = 59, N_{\text{FDTD}} = 69$)—note the notches appeared to be following an exponential distribution. Such differences are mostly due to the first pinna notch, which is much deeper in the asymptotic predictions: a two-tailed nonparametric Mann-Whitney test was not significant ($p_{\text{M1vsFDTD}} = 0.5$ and $p_{\text{M2vsFDTD}} = 0.55$), while the median notch prominence is smaller in the simulations [med(M1) = 4.89 dB, med(M2) = 3.64 dB, med(FDTD) = 3.21 dB].

D. The ratio-scale measure of the dissimilarity

There are at least two main issues with the SD metric in Eq. (3): averaging across frequencies losses information,[16] which in turn could render it a poor perceptual predictor (see, e.g., the related work by Andéol *et al.*[75] on Middlebrooks' *spectral difference/strength* metric), and it does not include the inherent uncertainties in a validation study. An appropriately designed validation metric could yield useful information about the employed continuous mathematical model.

Validation metrics are designed to both capture the error of interest and to infer about the confidence in such error (see Ref. 22, pp. 486–490). The error of interest is a ratio-scale measure of the dissimilarity between the measurements and simulations. For the HRTF problem, it is relevant to measure the magnitude level error in dB for direction (θ, ϕ) at angular frequency ω. Following Oberkampf and Roy[22] (p. 495), the validation metric is defined as

$$E_{\text{dB}}[\omega,\theta,\phi] = \left(\mathbb{E}\left[|\widetilde{\text{PRTF}}[\omega,\theta,\phi,\Delta X]|\middle|\Delta X = 0\right]\right)_{\text{dB}} - \left(\mathbb{E}[|\text{PRTF}_m[\omega,\theta,\phi]|]\right)_{\text{dB}} \pm U_{\text{dB},95\%}[\omega], \tag{4}$$

where $(\cdot)_{\text{dB}}$ is a notation for $20\log_{10}(\cdot)$, and $U_{\text{dB},95\%}[\omega]$ represents the total validation uncertainty at ω for the level in dB at a confidence level of $\alpha = 0.05$. The validation uncertainty $U_{\text{dB},95\%}$ is composed of measurement uncertainty and simulation uncertainty $\widetilde{U}_{95\%}$. The former is composed of the uncertainty $U_{\text{ear},95\%}$ in the P_{ear} measurement and uncertainty $U_{\text{ref},95\%}$ in the P_{ref} measurement.

Since the asymptotic estimate is based on magnitude, uncertainty propagation laws are used to calculate $U_{\text{dB},95\%}$. Propagating first the uncertainty in the $\log_{10}$ function (see Ref. 76, p. 19) and using the relative error summation rule (Ref. 77, p. 52) twice (once for the $|\widetilde{\text{PRTF}}|/|\text{PRTF}_m|$

https://doi.org/10.1121/10.0001230

division, and once for the $|P_{m,\text{ear}}|/|P_{m,\text{ref}}|$ magnitude-division in the measured $|\text{PRTF}_m|$) yields

$$U_{\text{dB},95\%}[\omega] = \frac{20}{\ln(10)}\left[\frac{\widetilde{U}_{95\%}[\omega]}{\mathbb{E}|\widetilde{\text{PRTF}}[\omega]|} + \frac{U_{\text{ear},95\%}[\omega]}{\mathbb{E}|P_{m,\text{ear}}[\omega]|} + \frac{U_{\text{ref},95\%}[\omega]}{\mathbb{E}|P_{m,\text{ref}}[\omega]|}\right], \quad (5)$$

where $\widetilde{U}_{95\%}$ is estimated from the convergence analysis,[18] $\mathbb{E}|\widetilde{\text{PRTF}}[\omega]|$ is the asymptotic solution, while $U_{\text{ear},95\%}$ is roughly estimated based on the standard error of the mean and the appropriate two-tailed t values as $U_{\text{ear},95\%} = t_{0.025,N_m-1}\hat{\sigma}$, where $\hat{\sigma} = \hat{\sigma}[\omega]$ is the sample-based standard error of the mean estimator of the considered N_m measurements. Since no repeated measurements were done for the reference measurement, the standard error variability in the magnitude response for microphone positioning and orientation reported in Sec. III A 2 is used for $U_{\text{ref},95\%}$. Note in Sec. III A 2, the reported uncertainty is calculated on the magnitude level in dB.

The validation metric in Eq. (4) can be calculated for one measurement session ($N_m = 3$) or for both sessions ($N_m = 6$). A gain mismatch in the measurement chain between the two measurement sessions was found which unjustifiably inflates $U_{\text{dB},95\%}$. As such, the average gain mismatch between the two measurement sessions (calculated across frequencies, directions, and ear/ref measurements) was removed from the second measurement session when calculating the P_{ear} measurement contribution to uncertainty.

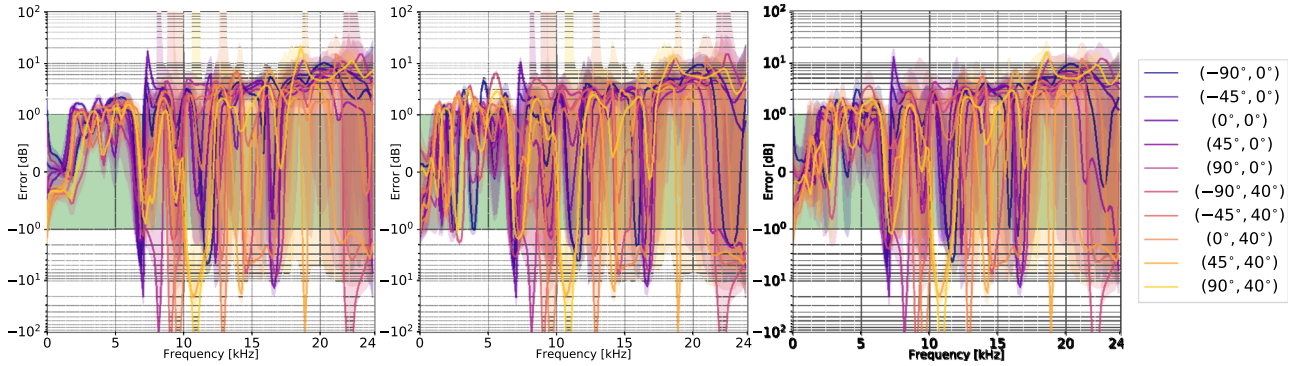
Figure 9 shows the validation metric in Eq. (4) for the first measurement session [Fig. 9(a)], the second measurement session [Fig. 9(b)], and all the pooled measurements [Fig. 9(c)]. The same $U_{\text{ref},95\%}$ is used for all three plots in Fig. 9. To begin with, the validation metric is rather strongly affected by the slight mismatches in frequency of the peaks and notches (see Fig. 5) above about 7 kHz. It is difficult to conclude whether the simulations agree better with one measurement session over the other.

One consistent qualitative result in Fig. 9 is an increased positive amplitude bias with frequency in the simulations as compared to measurements. Quantitatively, correlation analyses are usually employed to establish trends in the error.[78] For example, the Pearson correlation coefficient ρ could be used on the average errors E_{dB} in Fig. 9; results show statistical significance ($p_{\text{two-tailed}} \leq 0.05$) of such a positive bias with frequency only in the horizontal plane, independent on the considered measurement session. For $\phi = 40°$, the back direction ($\theta = -90°$) shows a significant negative correlation of about $\rho = -0.26$, while the other directions show small slopes $|\rho| \leq 0.14$ that are statistically insignificant. Note such analysis ignored the CIs $U_{\text{dB},95\%}$ in Eq. (4). Finally, Fig. 9 suggests that the prediction of the used model(s) have low reliability above 15 kHz where amplitude differences are as high as 10 dB and the CIs become unreasonably large.

The increased magnitude in the simulations could indicate the lossless assumption in the model is not entirely correct even for the used stiff scatterer. Nevertheless, the exact cause is unclear: the almost linear increase on the log-scale could indicate small-amplitude air absorption[79] as a possible cause; nevertheless, sole air-absorption cannot account for 10 dB during $t_{\text{obs}} = 6.25$ ms. Another cause could be viscothermal losses at the boundary (see, e.g., Morse and Ingard,[63] p. 286). Additional losses in the pinna material or vibration of the pinna structure would also add to the discrepancy; the acoustically-rigid assumption is an idealized condition that is very difficult to achieve in reality. Finally, an orientation mismatch is also expected to cause an increase in the validation error with frequency: this could explain errors with alternating sign in Fig. 9 and further aggravate losses-induced mismatches. Further validation studies are needed to pinpoint the dominant source of error.

V. POTENTIAL LIMITATIONS

Since the number of potential errors is unbounded during measurements, the chance of type I/II errors becomes relatively high. This is why validation studies usually require higher levels of characterization (see Ref. 22, p. 464), especially when sensitivity/uncertainty analyses for input parameters are missing.



(a) 1st measurement session (b) 2nd measurement session (c) both measurement sessions

FIG. 9. (Color online) Validation metric defined in Eq. (4). For the simulations, the predicted asymptotic solution is used. The transparent fills around each error curve represent the total validation uncertainty [see Eqs. (4) and (5)]. For (c), the gain mismatch between the two P_{ear} measurement sessions is corrected. The green area is based on an 1dB ILD JND (see text for details).

As such, it is arguably important to document and explore the acknowledged limitations in this study. To begin with, the topology of the printed pinna was not independently validated except for a few linear distances.

A lack of well-defined reference planes is another potential design flaw. This, coupled with the size of the scatterer, could cause orientation mismatch between the simulations and measurements.

Although no significant extra scattering was identified in the measurements, the simulations did not include any mounting apparatus that could further add to the mismatch. Despite being quantified to be small, the mismatch in the directivity of the source is another general issue.

Unverified fabrication errors (e.g., in the nut or mounting bolt) could introduce additional orientation bias. Moreover, a slight mismatch in the receiver placement exists in the simulations: the source was interpolated on G5 (see Table II), $|\Delta \mathbf{r}| \approx 1.3$ mm away from the blocked-meatus surface. Besides the 0.6° elevation mismatches in the simulations outside the horizontal plane, elevation deviations of about 1°–2° were estimated in positioning the loudspeaker during the first measurement session. Despite such acknowledged errors, the errors in Fig. 9 seem to be similar to the horizontal plane. Small radial mismatches around 5 mm were also identified during the second measurement session.

It was subsequently found that the rotating dish of the ETD250–3D in the first measurement session was not rotating in a consistent plane. Consequently, small measurement bias is present: repeatable and hard to quantify θ-dependent alignment and orientation mismatches. Based on the results in Figs. 8 and 9, the captured bias seems to be rather small.

Although the pinna was quite stiff, the used material had a quite low density (tabulated[67] value of 1100 kg/m^3) and was mounted on a structure. As such, potential vibrations in the structure could affect the apparent surface impedance of the pinna surface, especially for the first measurement session.

Finally, although informative, the feature analysis from Fig. 8 and Sec. IV C was kept rather simple. An accurate feature extraction should consider PRTF-magnitude interpolation, the frequency-dependence of the JNDs,[70] and the uncertainty in the exact frequency of the peaks and notches in the asymptotic solution.

VI. GENERAL DISCUSSION

The magnitude-only results in Fig. 9 cannot clearly find the continuous model "correct." For example, additional losses unaccounted in the model could be significant enough. Studying the phase would likely bring additional complications to the results.

Zero dissimilarity with zero validation uncertainty is virtually unattainable in practice. As such, the problem of degree of adequacy of the model arises which is generally difficult due to incomplete coverage of the application and validation domain (Ref. 21, p. 300),[69] or subjective expert opinion.[80] In acoustics, one can rely on meaningful perceptual metrics such as JNDs.[81] Nevertheless, such an approach has limitations due to the incomplete coverage of auditory perception/dimensions for any relevant JND. For instance, some perceptual dimensions could integrate multiple sound attributes or combine multiple physical quantities.[82]

From a physical point of view, validation at the most difficult to satisfy physically-based JND would be sufficient for a uni-dimensional percept. Nevertheless, even for directional localization, such a worst-case JND is presently unknown. For the present work, a relevant JND would be a worst-case pure-tone binaural interaural level difference (ILD). The JND for such a localization cue for 500 Hz tone-bursts is as low as 1 dB or even lower for higher SPLs.[83] Using, for example, a prudent ±1 dB threshold for the magnitude (green background in Fig. 9) would result in the model being marginally validated against the measured data up to the first concha mode (below 8 kHz). Notice the ILD is based on binaural signals; extrapolation to two monaural signals might be unfounded.

The 1 dB threshold is also at the upper end of reported loudness JNDs for wideband noise. (Ref. 84, p. 144). Nevertheless, such JND is smaller for pure tones at more common dB SPL levels,[85] while a broadband (say, 0–8 kHz) 1 dB JND implies lower per-frequency level differences. As such, it is considered that the predicted PRTFs are not validated when the loudness JND is considered. However, considering Fig. 8 and the inherent errors in the measurements, the prediction is likely validated when the broadband discrimination of each prominent PRTF feature is considered.[70] Note there could be caveats in employing purely spectral JNDs such as the 1 dB ILD due to the complexity of the nervous and auditory systems.[86]

Above about 8 kHz, the validation results show poor reliability. Although high accuracy was targeted, it appears that physics-based HRTF validation studies require:

(1) More tight measurement error control.
(2) A high number of reasonably-accurate reproduced measurements on the same scatterer under multiple conditions (such that an average can be obtained as a reliable approximation of a "true" HRTF/PRTF).

For the first aspect, some deficiencies in experimental design were identified in Sec. V. Nevertheless, increasing accuracy would quickly become highly expensive and difficult.

For the second aspect, HRTF reproducibility studies[13,14,48,49] show that present replicated HRTF measurements are hardly useful for validating physics-based simulations due to high measurement inconsistencies. The present study also shows the difficulties in the replicability of the measured HRTF. Moreover, a lack of well-defined measurement protocols[13] currently prevents high reproducibility.

To partly mitigate such stringent requirements, physics-based HRTF validation studies could be further complemented with HRTF perceptual validation results.

Usually, extrapolation of validation results to similar/neighboring problems needs to be done with care.[69] The

results in the present work cannot be fully extrapolated to HRTFs for all directions and distances (e.g., near-field HRTFs). The current analysis can be viewed as a study of the ipsilateral ear and both the simulation and measurement errors might behave differently for the acoustically-shadowed contralateral ear. The results should also not be extrapolated without evidence to *in vivo* HRTF due to extra modeling (see Ref. 22, p. 575) and/or input uncertainty; different temperature gradients in the pinna cavities and additional surface impedance uncertainty are expected. Finally, inference even to the space of all acoustically-rigid ispsilateral ears, although plausible, should be presently viewed with care. More validation cases are required which sample the ear-space more thoroughly.

VII. CONCLUSION

The present work assesses the validity of using lossless wave-based models to predict the high-frequency spectra of the HRTFs by evaluating PRTF simulations for the ispilateral ear. Accurate estimates of asymptotic (i.e., at $\Delta X \to 0$) PRTFs were found to better match measurements than computations on any grid. Such results indicate that a single computation not only is different than the asymptotic solution,[18] but can also give misleading validation results. Consequently, rigorous computational error analysis and accurate measurements are required for a validation study. The latter were found difficult to achieve, and estimation of measurement bias is recommended.

Results showed that the inhomogeneous wave equation with acoustically-rigid boundaries qualitatively predicts the PRTF features for a stiff scatterer. Nevertheless, slight frequency mismatches could still be present which can be attributed to computational and measurement uncertainty. Some potential modeling errors were observed. An increased mismatch with frequency was found across directions, possibly due to unaccounted losses in the model. Still, the predicted PRTF magnitudes agree with the measurements within ± 1 dB up to around 8 kHz.

It was also argued that validation at the physics level cannot be practically achieved without perceptual considerations. This was exemplified by considering three JNDs: an ILD JND (model marginally validated up to 8 kHz), a peak/notch discrimination JND (model validated for prominent features up to 20 kHz), and a loudness JND (model appears invalid).

Finally, this study suggests the need to lower the validation uncertainty, while the modeling needs to improve before wave-based HRTF/PRTF predictions could be considered accurate.

ACKNOWLEDGEMENTS

This research has received funding from the Academy of Finland, Project No. 265 824 (L.S. and S.T.P.) and Facebook Reality Labs. The computational resources provided by the Aalto Science-IT project and CSC-IT Center for Science, Espoo, Finland, are also acknowledged. M.G. was supported by the 2016-2021 strategic program "Knowledge for the World" of Aalborg University with a personal grant. S.T.P. would like to thank all the persons that shared their time and patience in various related conversations (Vesa Välimäki from Aalto University; Terry Cho, Pablo Hoffmann, Dave Ali, and others from Oculus Research). Jukka Saarelma, Terry Cho, Pablo Hoffmann, Antti Kuusinen, and Julie Meyer are thanked for commenting and giving useful suggestions for previous drafts of the paper. Ville Pulkki is thanked for providing the 3D-printed pinna and its mesh. Marko Hiipakka and Kari Metso are thanked for their help in digging up the necessary details of the existing 3D-printed pinna. Ilkka Huhtakallio is thanked for the help with the ear-blocking 3D-printing. The two anonymous reviewers are acknowledged for the useful text suggestions.

[1]B. Xie, *Head-Related Transfer Function and Virtual Auditory Display*, 2nd ed. (J. Ross Publishing, Plantation, FL, 2013), 505 pp.
[2]R. Trapeau, V. Aubrais, and M. Schönwiesner, "Fast and persistent adaptation to new spectral cues for sound localization suggests a many-to-one mapping mechanism," J. Acoust. Soc. Am. **140**(2), 879–890 (2016).
[3]E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization using nonindividualized head-related transfer functions," J. Acoust. Soc. Am. **94**(1), 111–123 (1993).
[4]A. J. Hoogstrate, H. Van Den Heuvel, and E. Huyben, "Ear identification based on surveillance camera images," Sci. Justice **41**(3), 167–172 (2001).
[5]M. Geronazzo, S. Spagnol, and F. Avanzini, "Do we need individual head-related transfer functions for vertical localization? The case study of a spectral notch distance metric," IEEE/ACM Trans. Audio Speech Lang. Process. **26**(7), 1243–1256 (2018).
[6]W. L. Oberkampf and T. G. Trucano, "Verification and validation in computational fluid dynamics," Prog. Aerosp. Sci. **38**(3), 209–272 (2002).
[7]M. Morimoto and Y. Ando, "On the simulation of sound localization," J. Acoust. Soc. Jpn. **1**(3), 167–174 (1980).
[8]F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. II: Psychophysical validation," J. Acoust. Soc. Am. **85**(2), 868–878 (1989).
[9]E. H. A. Langendijk and A. W. Bronkhorst, "Fidelity of three-dimensional-sound reproduction using a virtual auditory display," J. Acoust. Soc. Am. **107**(1), 528–537 (1999).
[10]P. Majdak, P. Balazs, and B. Laback, "Multiple exponential sweep method for fast measurement of head-related transfer functions," J. Audio Eng. Soc. **55**(7/8), 623–637 (2007).
[11]G. Enzner, "3D-continuous-azimuth acquisition of head-related impulse responses using multi-channel adaptive filtering," in *2009 IEEE Workshop Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY (October 18–21, 2009), pp. 325–328.
[12]P. Dietrich, B. Masiero, and M. Vorländer, "On the optimization of the multiple exponential sweep method," J. Audio Eng. Soc. **61**(3), 113–124 (2013).
[13]A. Andreopoulou, D. Begault, and B. Katz, "Inter-laboratory round robin HRTF measurement comparison," IEEE J. Sel. Top. Signal Process. **9**(5), 895–906 (2015).
[14]F. Brinkmann, M. Dinakaran, R. Pelzer, P. Grosche, D. Voss, and S. Weinzierl, "A cross-evaluated database of measured and simulated HRTFs including 3D head meshes, anthropometric features, and headphone impulse responses," J. Audio Eng. Soc. **67**(9), 705–718 (2019).
[15]F. Brinkmann, A. Lindau, S. Weinzierl, S. van der Par, M. Müller-Trapet, R. Opdam, and M. Vorländer, "A high resolution and full-spherical head-related transfer function database for different head-above-torso orientations," J. Audio Eng. Soc. **65**(10), 841–848 (2017).
[16]S. Prepeliță, M. Geronazzo, F. Avanzini, and L. Savioja, "Influence of voxelization on finite difference time domain simulations of head-related transfer functions," J. Acoust. Soc. Am. **139**(5), 2489–2504 (2016).

[17]M. Dinakaran, F. Brinkmann, S. Harder, R. Pelzer, P. Grosche, R. R. Paulsen, and S. Weinzierl, "Perceptually motivated analysis of numerically simulated head-related transfer functions generated by various 3D surface scanning systems," in *Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, Canada (April 15–20, 2018), pp. 551–555.
[18]S. T. Prepeliță, 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," J. Acoust. Soc. Am. **146**(5), 3629–3645 (2019).
[19]B. E. Treeby, J. Pan, and R. M. Paurobally, "An experimental study of the acoustic impedance characteristics of human hair," J. Acoust. Soc. Am. **122**(4), 2107–2117 (2007).
[20]M. Vorländer, "Computer simulations in room acoustics: Concepts and uncertainties," J. Acoust. Soc. Am. **133**(3), 1203–1213 (2013).
[21]P. J. Roache, *Verification and Validation in Computational Science and Engineering* (Hermosa, Albuquerque, NM, 1998), 464 pp.
[22]W. L. Oberkampf and C. J. Roy, *Verification and Validation in Scientific Computing*, 1st ed. (Cambridge University Press, New York, 2010), 790 pp.
[23]B. F. G. Katz, "Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation," J. Acoust. Soc. Am. **110**(5), 2440–2448 (2001).
[24]Y. 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).
[25]W. 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).
[26]C. Jin, P. Guillon, N. Epain, R. Zolfaghari, A. van Schaik, A. Tew, C. Hetherington, and J. Thorpe, "Creating the Sydney York Morphological and Acoustic Recordings of Ears Database," IEEE Trans. Multimed. **16**(1), 37–46 (2014).
[27]H. Ziegelwanger, P. Majdak, and W. Kreuzer, "Numerical calculation of listener-specific head-related transfer functions and sound localization: Microphone model and mesh discretization," J. Acoust. Soc. Am. **138**(1), 208–222 (2015).
[28]T. 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).
[29]P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, "Frequency and amplitude estimation of the first peak of head-related transfer functions from individual pinna anthropometry," J. Acoust. Soc. Am. **137**(2), 690–701 (2015).
[30]J. Fels, P. Buthmann, and M. Vorländer, "Head-related transfer functions of children," Acta Acust. united Ac. **90**(5), 918–927 (2004).
[31]V. 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).
[32]V. Algazi, R. Duda, R. Morrison, and D. Thompson, "Structural composition and decomposition of HRTFs," in *2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY (October 21–24, 2001), pp. 103–106.
[33]M. Geronazzo, S. Spagnol, and F. Avanzini, "Mixed structural modeling of head-related transfer functions for customized binaural audio delivery," in *Proceedings of the 18th International Conference on Digital Signal Process (DSP)*, Santorini, Greece (July 1–3, 2013), pp. 1–8.
[34]E. A. G. Shaw, "Transformation of sound pressure level from the free field to the eardrum in the horizontal plane," J. Acoust. Soc. Am. **56**(6), 1848–1861 (1974).
[35]R. O. Duda and W. L. Martens, "Range dependence of the response of a spherical head model," J. Acoust. Soc. Am. **104**(5), 3048–3058 (1998).
[36]B. E. Treeby, J. Pan, and R. M. Paurobally, "The effect of hair on auditory localization cues," J. Acoust. Soc. Am. **122**(6), 3586–3597 (2007).
[37]F. Brinkmann, R. Roden, A. Lindau, and S. Weinzierl, "Audibility and interpolation of head-above-torso orientation in binaural technology," IEEE J. Sel. Top. Signal Process. **9**(5), 931–942 (2015).
[38]G. F. Kuhn, "Model for the interaural time differences in the azimuthal plane," J. Acoust. Soc. Am. **62**(1), 157–167 (1977).
[39]B. E. Treeby, R. M. Paurobally, and J. Pan, "The effect of impedance on interaural azimuth cues derived from a spherical head model," J. Acoust. Soc. Am. **121**(4), 2217–2226 (2007).
[40]E. A. G. Shaw and R. Teranishi, "Sound pressure generated in an external-ear replica and real human ears by a nearby point source," J. Acoust. Soc. Am. **44**(1), 240–249 (1968).
[41]H. Møller, "Fundamentals of binaural technology," Appl. Acoust. **36**(3–4), 171–218 (1992).
[42]A. Torras-Rosell and F. Jacobsen, "A new interpretation of distortion artifacts in sweep measurements," J. Audio Eng. Soc. **59**(5), 283–289 (2011).
[43]P. M. Morse, "Some aspects of the theory of room acoustics," J. Acoust. Soc. Am. **11**(1), 56–66 (1939).
[44]Note that in reviewing the potential issues with an HRTF/PRTF validation study, the commonly-used term *error* is chosen in favor of *uncertainty* despite some inherent practical drawbacks. When quantifying the validation results in Sec. IV, the term *uncertainty* is found to better suit the discussion, which is assumed to encapsulate the total amount of errors it represents. A further discussion and analysis of the two terms is out of scope; see, e.g., section 2.4 from Oberkampf and Roy (Ref. 22).
[45]K. A. J. Riederer, "Repeatability analysis of head-related transfer function measurements," in *Proceedings of the 105th Convention of the Audio Engineering Society*, San Francisco, CA (September 26–29, 1998), pp. 4846:1–4846:62.
[46]JCGM, *The International Vocabulary of Metrology-Basic and General Concepts and Associated Terms (VIM), 3rd edition JCGM 200: 2012* (Joint Committee for Guides in Metrology, France, 2012).
[47]V. R. Algazi, C. Avendano, and D. Thompson, "Dependence of subject and measurement position in binaural signal acquisition," J. Audio Eng. Soc. **47**(11), 937–947 (1999).
[48]B. F. G. Katz and D. R. Begault, "Round robin comparison of HRTF measurement systems: Preliminary results," in *Proceedings of the 19th International Congress on Acoustics (ICA 2007)*, Madrid, Spain (September 2–7, 2007), pp. 1–6.
[49]X.-L. Zhong and B.-S. Xie, "Consistency among the head-related transfer functions from different measurements," Proc. Mtgs. Acoust. **19**, 050014 (2013).
[50]J.-G. Richter, G. Behler, and J. Fels, "Evaluation of a fast HRTF measurement system," in *Proceedings of the Audio Engineering Society Convention 140*, Paris, France (June 4–7, 2016).
[51]J. Gómez Bolaños and V. Pulkki, "HRIR Database with Measured Actual Source Direction Data," in *Proceedings of the Audio Engineering Society Convention 133*, San Francisco, CA (October 26–29, 2012), pp. 8759:1–8759:8.
[52]B. Gardner and K. Martin, "HRTF measurements of a KEMAR dummy-head microphone," Technical Report No. 280 (Massachusetts Institute of Technology, Cambridge, MA, 1994).
[53]D. Hammershøi and H. Møller, "Binaural technique—Basic methods for recording, synthesis, and reproduction," in *Communication Acoustics*, edited by J. Blauert (Springer, Berlin-Heidelberg, 2005), pp. 223–254.
[54]B. P. Bovbjerg, F. Christensen, P. Minnaar, and X. Chen, "Measuring the head-related transfer functions of an artificial head with a high-directional resolution," in *Proceedings of 109th Convention of the Audio Engineering Society*, Los Angeles, CA (September 22–25, 2000), pp. 5264:1–5264:17.
[55]P. Zahorik, "Limitations in using Golay codes for head-related transfer function measurement," J. Acoust. Soc. Am. **107**(3), 1793–1796 (2000).
[56]D. Hammershøi and H. Møller, "Sound transmission to and within the human ear canal," J. Acoust. Soc. Am. **100**(1), 408–427 (1996).
[57]M. Rothbucher, K. Veprek, P. Paukner, T. Habigt, and K. Diepold, "Comparison of head-related impulse response measurement approaches," J. Acoust. Soc. Am. **134**(2), EL223–EL229 (2013).
[58]P. M. S. Müller, "Transfer-function measurements with sweeps—Director's cut including previously unreleased material and some corrections," J. Audio Eng. Soc. **49**(6), 443–471 (2001).
[59]A. Jost and D. R. Begault, "Observed effects of HRTF measurement signal level," in *Proceedings of the AES 21st International Conference on Architectural Acoustics & Sound Reinforcement*, St. Petersburg, Russia (June 1–3, 2002), pp. 1–5.
[60]J. Fels and M. Vorländer, "Anthropometric parameters influencing head-related transfer functions," Acta Acust. united Ac. **95**(2), 331–342 (2009).
[61]A. Farina, "Advancements in impulse response measurements by sine sweeps," in *Proceedings of the 122nd Convention of the Audio Engineering Society*, Vienna, Austria (May 5–8, 2007), pp. 7121:1–7121:21.
[62]L. Eça and M. Hoekstra, "A procedure for the estimation of the numerical uncertainty of CFD calculations based on grid refinement studies," J. Comput. Phys. **262**, 104–130 (2014).
[63]P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (Princeton University Press, Princeton, NJ, 1986).

https://doi.org/10.1121/10.0001230

[64]L. E. Kinsler, F. R. Austin, C. B. Alan, and S. V. James, *Fundamentals of Acoustics*, 4th ed. (Wiley, New York, 2000).
[65]A. Boschetto and L. Bottini, "Accuracy prediction in fused deposition modeling," Int. J. Adv. Manuf. Technol. **73**(5), 913–928 (2014).
[66]A. Armillotta, S. Bianchi, M. Cavallaro, and S. Minnella, "Edge quality in fused deposition modeling: II. experimental verification," Rapid Prototyp. J. **23**(4), 686–695 (2017).
[67]VisiJet "VisiJet SR200 documentation," http://infocenter.3dsystems.com/materials/mjp/visijet-sr200 (Last viewed March 20, 2019).
[68]R. H. Small, "Direct radiator loudspeaker system analysis," J. Audio Eng. Soc. **20**(5), 383–395 (1972).
[69]W. L. Oberkampf, T. G. Trucano, and C. Hirsch, "Verification, validation, and predictive capability in computational engineering and physics," Appl. Mech. Rev. **57**(5), 345–384 (2004).
[70]B. C. J. Moore, S. R. Oldfield, and G. J. Dooley, "Detection and discrimination of spectral peaks and notches at 1 and 8 kHz," J. Acoust. Soc. Am. **85**(2), 820–836 (1989).
[71]P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, "Comparison of simulated and measured HRTFs: FDTD simulation using MRI head data," in *Proceedings of the Audio Engineering Society Convention 123*, New York (October 5–8, 2007), pp. 7240:1–7240:12.
[72]P. 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 *e-Proceedings of the International Workshop on the Principles and Applications of Spatial Hearing (IWPASH 2009)*, Zao, Miyagi, Japan (November 11–13, 2009) (World Scientific, Singapore, 2009), pp. 1–4.
[73]K. Young, G. Kearney, and A. I. Tew, "Acoustic validation of a BEM-suitable 3D mesh model of KEMAR," in *Proceedings of the 2018 AES International Conference on Spatial Reproduction - Aesthetics and Science*, Tokyo, Japan (August 7–9, 2018), EB2-9:1–EB2-9:5.
[74]"scipy.signal.find_peaks - SciPy v1.1.0 Reference Guide," https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.find_peaks.html (Last viewed March 20, 2019).
[75]G. Andéol, E. A. Macpherson, and A. T. Sabin, "Sound localization in noise and sensitivity to spectral shape," Hear. Res. **304**, 20–27 (2013).
[76]D. C. Baird, *Experimentation: An Introduction to Measurement Theory and Experiment Design*, 3rd ed. (Addison-Wesley, Englewood Cliffs, NJ, 1994).
[77]J. R. Taylor, *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*, 2nd ed. (University Science Books, Sausalito, CA, 1996).
[78]G. Bellocchi, M. Rivington, M. Donatelli, and K. Matthews, "Validation of biophysical models: Issues and methodologies," in *Sustainable Agriculture Volume 2*, edited by E. Lichtfouse, M. Hamelin, M. Navarrete, and P. Debaeke (Springer, Dordrecht, 2011), pp. 577–603.
[79]H. E. Bass, L. C. Sutherland, A. J. Zuckerwar, D. T. Blackstock, and D. M. Hester, "Atmospheric absorption of sound: Further developments," J. Acoust. Soc. Am. **97**(1), 680–683 (1995).
[80]I. Babuska and J. T. Oden, "Verification and validation in computational engineering and science: Basic concepts," Comput. Methods Appl. Mech. Eng. **193**(36), 4057–4066 (2004).
[81]M. Vorländer, "Prediction tools in acoustics-Can we trust the PC," in *Proceedings of the Baltic-Nordic Acoustic Meeting (BNAM)*, Bergen, Norway (May 10–12, 2010), pp. 1–8.
[82]S. McAdams and A. Bregman, "Hearing musical streams," Comput. Music J. **3**(4), 26–60 (1979).
[83]R. M. Hershkowitz and N. I. Durlach, "Interaural time and amplitude JNDs for a 500-Hz tone," J. Acoust. Soc. Am. **46**(6B), 1464–1467 (1969).
[84]B. C. J. Moore, *An Introduction to the Psychology of Hearing*, 6th ed. (BRILL, Leiden, Boston, MA, 2012), 458 pp.
[85]N. F. Viemeister and S. P. Bacon, "Intensity discrimination, increment detection, and magnitude estimation for 1-kHz tones," J. Acoust. Soc. Am. **84**(1), 172–178 (1988).
[86]V. R. Algazi and R. O. Duda, "Headphone-based spatial sound," IEEE Signal Process. Mag. **28**(1), 33–42 (2011).