CHALMERS
UNIVERSITY OF TECHNOLOGY

Perceptual detection thresholds for numerical dispersion in binaural auralizations of two acoustically different rooms

JASA ARTICLE

Perceptual detection thresholds for numerical dispersion in binaural auralizations of two acoustically different rooms

Julie Meyer,[1,a] Tapio Lokki,[1] and Jens Ahrens[2]
[1]Acoustics Laboratory, Department of Signal Processing and Acoustics, Aalto University, Espoo, 02150, Finland
[2]Division of Applied Acoustics, Chalmers University of Technology, Gothenburg, 412 96, Sweden

ABSTRACT:
Room acoustic simulations using the finite-difference time-domain method on a wide frequency range can be computationally expensive and typically contain numerical dispersion. Numerical dispersion can be audible and, thus, constitutes an artifact in auralizations. There is a need to measure perceptual thresholds for numerical dispersion to achieve an optimal balance between computational complexity and audibility of dispersion. This work measures the perceptual detection thresholds for numerical dispersion in binaural auralizations of two acoustically different rooms. Numerical dispersion is incorporated into measured binaural room impulse responses (BRIRs) by the means of filters that represent the dispersion that plane waves experience, which propagate in the simulation in the direction of the worst-case dispersion error. The results show that the perceptual detection threshold is generally lower for the most reverberant room and greatly depends on the source signal independently of the room in which the threshold is measured. It is the most noticeable in the pure BRIRs, i.e., with an impulse as source signal, and almost unnoticeable with speech. The results also show that there was no statistical evidence that the perceptual thresholds for the conditions where numerical dispersion was present or absent in the direct path of the BRIRs be different.
 https://doi.org/10.1121/10.0014830
(Received 12 May 2022; revised 17 September 2022; accepted 26 September 2022; published online 20 October 2022)
[Editor: Brian F. G. Katz] Pages: 2266–2276

I. INTRODUCTION

The discretization error, which is due to the approximation of continuous differential operators by discrete operators, is one of the numerical errors intrinsically present in finite-difference time-domain (FDTD) simulation results. In fact, this numerical error is usually assumed to be the largest of the numerical errors in a partial differential equation-based scientific simulation (Oberkampf and Roy, 2010; Roy and Oberkampf, 2011). In the context of room acoustic modelling of this article, the discretization error depends on the Courant number, the propagation direction, the formal solution of the partial differential equation (which, in this context, is the inhomogeneous scalar wave equation), and increases with frequency and simulated time (Prepeliţă *et al.*, 2019). To reliably reduce the error, "oversampling," i.e., refining the spatial grid, can be employed (Hamilton, 2016).

However, oversampling largely increases the computational cost. For example, for the simplest FDTD scheme, the computational density (Kowalczyk and Van Walstijn, 2011; Van Walstijn and Kowalczyk, 2008) increases by a factor of 16 when the spatial grid spacing is halved. This also implies that the computation time increases with the same rate. Note that the temporal sampling frequency of the simulation is inversely proportional to the grid spacing. In addition, simulating the full audible frequency range is already computationally expensive due to the existence of a cutoff frequency (Kowalczyk and Van Walstijn, 2011), determining the highest frequency of the wave that can propagate in all directions, and at this limit, the simulation is not accurate and the waves are strongly affected by the discretization error.

In virtual acoustics applications where auralization is commonly employed, another downside is that this error can lead to audible phase artefacts (Hamilton and Bilbao, 2017). Considering that the "operational validity" (Sargent, 2010) of a simulation method is context dependent, perceptual metrics that represent the audibility of the error can be useful. A first step in defining the operational validity from a perceptual point of view is measuring the perceptual threshold for the error in various acoustic scenarios so as to provide guidelines for generating room acoustic simulations that are free of audible numerical artefacts.

A convenient and commonly used approach to analyze the discretization error in the context of room acoustic modelling is through dispersion analysis in which a plane wave solution to the wave equation is assumed (Schneider and Wagner, 1999). As this approach is herein adopted, the terms numerical dispersion or dispersion error will be employed in the remainder of the article.

II. PREVIOUS RELATED STUDIES

Several previous studies investigated the perception of numerical dispersion in FDTD simulations. A summary of these studies is presented in Table I. Amongst them, the

[a]Also at Department of Computer Science, Aalto University, Espoo, 02150, Finland. Electronic mail: julie.meyer@aalto.fi

https://doi.org/10.1121/10.0014830

TABLE I. A summary of previous studies focusing on perceptual evaluations of numerical dispersion in FDTD simulations and related parameters investigated. The symbol "—" means that the information was not explicitly given in the referenced article.

Reference	Cobos *et al.* (2008)	Southern *et al.* (2011)	Saarelma *et al.* (2016)	Saarelma and Savioja (2016)	Saarelma and Savioja (2019) Experiment 1	Saarelma and Savioja (2019) Experiment 2	Meyer *et al.* (2020)
Scheme	SRL[a]	SRL	SRL, CCP, IWB	CCP	SRL	SRL	SRL
Propagation direction[b]	—	Axial/diagonal	Axial, diagonal, Diagonal	Diagonal	Axial	Axial	Axial
Sampling frequency (Hz)	20 000, 30 000, 40 000	5000	264 030, 113 860, 107 780	—	—	59 583, 119 165, 238 330	156 796
Distance (m)	—	0.8/0.6, 4.1/2.9, 6.9/4.9, 9.8/6.9, 11.8/8.3	≈ up to 60	10, 50, 100, 344	2.34, 4.70, 9.39	2.9, 4	2.9
Simulated environment	Reflective	Free field	Free field	Free field with/without air absorption	Direct path + one reflection	Reflective with air absorption	Free field
Stimulus playback	—	—	Diotic (headphones)	Diotic (headphones)	2-Loudspeaker Setup	3D spatial (37 loudspeakers)	Binaural (headphones)
Sound sample	Speech (female and male)	Trombone, Violin/cello	Click, speech (male)	Click	Click, speech (male[c])	Click, speech (male[c])	Castanet

[a]The equivalent explicit finite-difference scheme of the K-DWM method is indicated, which was employed with a rectilinear mesh topology.
[b]For the direct path.
[c]Information not provided in the manuscript but known by one of the authors of this article who participated in the experiment.

Kirchhoff digital waveguide mesh (K-DWM; López *et al.*, 2007) and commonly used FDTD schemes were employed, although only Saarelma *et al.* (2016) considered comparing the standard rectilinear (SRL), the close-cubic packed (CCP), and the interpolated wideband (IWB) schemes with each other. Note that the FDTD schemes and waveguides based on the same stencil are mathematically equivalent (Karjalainen and Erkut, 2004; Kowalczyk and Van Walstijn, 2011).

The earlier studies presented in Table I employed diotic playback over headphones to auralize the FDTD-simulated or emulated responses, e.g., as in Saarelma *et al.* (2016). In the more recent years, studies like Saarelma and Savioja (2019) and Meyer *et al.* (2020) considered spatial reproduction systems, although only experiment two from Saarelma and Savioja (2019) considers full room responses as well.

In Cobos *et al.* (2008), the source was located in the right part of a medium sized room while the receiver was placed in its left part. In Southern *et al.* (2011), five different distances that were dependent on axial and diagonal directions were simulated. In Saarelma *et al.* (2016), the numerical dispersion was evaluated for the worst-case propagation direction respective to each scheme, and the audibility of the dispersion error was measured as a function of simulated distance. Four different distances (10, 50, 100, and 344 m) were evaluated in Saarelma and Savioja (2016). As for the simulated environment, the numerical dispersion was evaluated in the free field in all of these previously mentioned studies except Cobos *et al.* (2008) and Saarelma and Savioja (2019) in which room reflections were included. In addition to free-field conditions, the audibility of the dispersion error was measured in the presence of absorption of air in Saarelma and Savioja (2016).

The sampling frequency used in the simulations differed between these studies, which makes them difficult to compare. Cobos *et al.* (2008) used three different sampling frequencies, 20, 30 and 40 kHz, while a single sampling frequency (per scheme) of 5 kHz was used in Southern *et al.* (2011) and Saarelma *et al.* (2016) (264 030 Hz for SRL, 113 860 Hz for CCP, and 107 780 Hz for IWB). Finally, the audibility of dispersion error was measured as a function of the phase velocity error level at 20 kHz in Saarelma and Savioja (2016), thus, for various sampling frequencies.

As for the outcomes of these studies, Cobos *et al.* (2008) reported that the results were not fully conclusive because in the performed listening tests, only 20% and 30% of the listeners managed to discriminate between simulations with sampling frequencies of 20 and 30 kHz for the male and female speech phrases that were used, respectively. None of the participants was able to discriminate between simulations with sampling frequencies of 30 and 40 kHz. In Southern *et al.* (2011), it was found that under the chosen modeled conditions, dispersion became perceivable when the normalized frequency was in the 0.12–0.15 range. However, the authors highlight that a change of the sampling frequency does not necessarily lead to a similar change of the normalized frequency range that is free of audible artefacts (Southern *et al.*, 2011). The audibility of the dispersion error measured as a function of simulation distance in Saarelma *et al.* (2016) led to significantly different thresholds, depending on the sound samples (click-like signal and male speech) but similar thresholds across the SRL, CCP, and IWB schemes. The propagation distance for which numerical dispersion becomes audible for the modeled free-field conditions was found to be 9.1 m. In Saarelma and Savioja (2016), the lowest mean detection

threshold for the dispersion error was found to be at a phase velocity error percentage of 0.28% at 20 kHz (achieved for a source-receiver distance of 100 m) when air absorption was included.

In an experiment from a more recent study (Saarelma and Savioja, 2019), the audibility of the dispersion error in FDTD simulations was measured in the presence of a single early reflection. The threshold was found to vary depending on whether the error was in the direct path or reflection path. Recall that the spatial grid can be rotated to achieve either. The authors also found that for transient signals, the threshold was higher when the error was in the direct path, whereas for speech, the threshold was higher when it was in the reflection path (Saarelma and Savioja, 2019). In a second experiment, the authors simulated a full room response to quantify the audibility of the dispersion error at the measured threshold values (lowest, highest, and average between the lowest and highest) from the first experiment described above. As the three sampling frequencies (59 583, 119 165, and 238 330 Hz) that were chosen for the second experiment corresponded to previous thresholds measured at different distances than those that were evaluated in experiment 2, the audibility of dispersion evaluated in the full room response did not correspond to the threshold values for the simulated conditions. The authors highlight that a conclusion about what sampling frequency is required to make the error inaudible could not be drawn from experiment 2 for the click-like signal that was used.

As an extension to most of the studies reported in Table I, this work employs binaural signals and full room responses. Additionally, the sampling frequency, which directly controls the extent of numerical dispersion, is varied incrementally until the perceptual threshold for audibility of numerical dispersion is reached. Although the results from perceptual studies on numerical dispersion seem to be context dependent, by taking two acoustically different rooms and three representative sound samples, this work covers relevant cases. Moreover, by considering the worst-case propagation direction of the dispersion error, the lowest perceptual detection thresholds for the error are herein measured. The present study, therefore, provides safe guidelines on how to choose the sampling frequency to simulate room responses such that auralizations are free of audible dispersion artefacts over the entire audible frequency range.

III. BINAURAL AURALIZATIONS

Binaural synthesis using the FDTD method has been addressed by several authors using mainly two approaches. One approach, employed in Mokhtari *et al.* (2008), Sheaffer *et al.* (2013), and Webb and Bilbao (2012), consists of embedding the listener's head/head and torso morphology directly in the FDTD grid while the second employs free-field head-related transfer functions and virtual spherical receiver arrays whose pressure signals are decomposed using spherical harmonic processing (Meyer *et al.*, 2020; Saarelma and Savioja, 2019; Sheaffer *et al.*, 2014, 2015). Earlier work investigating differential microphone array modelling and spherical harmonic encoding in FDTD grids was also explored in Southern *et al.* (2012).

The first approach results in computationally demanding FDTD simulations because of the necessity to use high sampling frequencies to faithfully represent the listener's morphology, which is usually captured by highly detailed scans. As for the second approach, it is bandwidth constrained by the parameters defining the spherical receiver array, such as the array configuration and spatial sampling scheme. Guidelines for how to choose the parameters of the virtual receiver array such that the audibility of dispersion is not affected are provided in Meyer *et al.* (2020). A third approach that integrates a spherical harmonic spatial encoding process directly in the FDTD scheme was proposed in Bilbao *et al.* (2019). To the best of the authors' knowledge, this latter approach has not yet been used to generate binaural auralizations.

While these proposed methods for binaural synthesis differ, they all require running FDTD simulations. It is difficult to control the amount of dispersion on a continuous scale while keeping all of the other parameters constant. This work, therefore, proposes a novel approach based on measured binaural room impulse responses (BRIRs) and digital filters to create binaural signals that contain an emulation of numerical dispersion.

A. BRIRs

The control room 1 (CR1) and small broadcast studio (SBS) from the WDR (Westdeutscher Rundfunk) radio broadcast studios were selected as the two acoustically different rooms for the experiment (Stade *et al.*, 2012). CR1 is an acoustically dry control room used primarily for the production and recording of classical music (volume = 92.79 m^3) with a reverberation time of approximately 0.2 s (Ahrens and Andersson, 2019). SBS is a concert room for chamber music and small ensembles (volume = 1246.77 m^3) with a reverberation time of approximately 1 s in the 200–3000 Hz frequency range and above 0.5 s outside of this range; see also Fig. 7 in Stade *et al.* (2012).

The BRIR of SBS was directly taken from the data provided in Stade *et al.* (2012), which was obtained using a Neumann KU100 dummy head (Berlin, Germany) that faces the center loudspeaker (a horn loaded public address rig from AD-Systems, Wesel, Germany). In CR1, the same dummy head was placed in the center of the left and right main monitor loudspeakers (Genelec 8260A, Iisalmi, Finland) facing the mixing consoles. To create a (virtual) source that would face the dummy head as measured in SBS, the BRIR of CR1 was created by adding the two respective left- and right-ear signals measured from the left and right main monitor loudspeakers. The two BRIRs were then resampled to 44 100 Hz. The resulting lengths of the BRIR were 0.35 s for CR1 and 1.56 s for SBS. Note that the direct sound of the BRIRs was aligned in time. The direct sound appeared at around 2.5 ms in the BRIRs, and there

https://doi.org/10.1121/10.0014830

was an approximate 2 ms time delay between the direct sound and the first reflections in both BRIRs.

B. Numerical dispersion

Numerical dispersion was introduced to the BRIRs by convolution with the impulse response of the parametric dispersion error filter for the SRL scheme that was presented in Saarelma *et al.* (2016). The latter was designed to correspond to a plane wave propagating in the direction of the worst-case error, which is the axial direction. The dispersion error filter was also designed by considering the scheme at its stability limit (i.e., with a Courant number set to $\lambda = 1/\sqrt{3}$). The filter was validated against FDTD simulations in Saarelma *et al.* (2016). A more detailed description of the time-domain impulse responses used in this work is given below.

The dispersion relation, which relates temporal frequencies to spatial frequencies and whose derivation from the homogeneous acoustic wave equation can be found, e.g., in Schneider and Wagner (1999) and Van Mourik and Murphy (2014), can be expressed for the SRL scheme as

$$\sin^2\left(\omega\frac{T}{2}\right) = \lambda^2 \sin^2\left(\hat{k}_x\frac{X}{2}\right), \tag{1}$$

where T is the time step, which is equivalent to $1/f_s$, where f_s denotes the temporal sampling frequency of the simulation. X is the spatial grid spacing, $\lambda = 1/\sqrt{3}$ is the Courant number, and $\hat{k}_x$ is the only nonzero numerical wavenumber component corresponding to an axial propagation direction in the simulation domain. c denotes the speed of sound, which was set to 344 m/s.

The time-domain impulse response representing the FDTD-simulated pressure recorded by the receiver located at a distance d in the axial direction from the source can then be expressed as

$$h(t) = F^{-1}\left[e^{i(\omega t - \hat{k}_x d)}\right], \tag{2}$$

where $t = 0$, and $\hat{k}_x = (2/X)\arcsin((1/\lambda)\sin(\omega T/2))$ is obtained by solving Eq. (1) for the numerical wavenumber component, $\hat{k}_x$, as a function of the angular frequency, ω. $F^{-1}[\cdot]$ denotes the inverse Fourier transform.

For brevity, the time-domain impulse responses, $h(t)$, from Eq. (2) will be referred to as dispersion filters in the remainder of the article.

1. Simulated distance

As mentioned in Sec. I, numerical dispersion increases with simulated time and, thus, with simulated distance. To introduce an increasing amount of numerical dispersion as a function of time in the BRIRs, several dispersion filters representing different propagation distances in the simulation domain were created. To reduce the computation time of the successive convolutions between the dispersion filters and corresponding time steps of the BRIRs, segmentation of the BRIRs was performed instead of a sample-by-sample application and update of the filter. The segment size was set to 100 samples after adjusting it until a satisfying trade-off between computation time and absence of perceptual artefacts was attained. A Hann window was applied to each BRIR segment. Finally, the dispersion filters were applied (via convolution) to the corresponding windowed-BRIR time segments with 50% overlap. It is also worth mentioning that to further reduce the computation time of the successive convolutions, the length of the dispersion filters was shortened to between 2000 and 50 000 samples.

The dispersion filters were further low-pass filtered, as in Saarelma *et al.* (2016), with a cutoff frequency of 20 kHz using a 20-sample long finite impulse response (FIR) filter and smoothed with a 21-point Chebyshev window with 80 dB of sidelobe attenuation. The delay introduced by the FIR filter and propagation distance was removed. An example of two dispersion filters representing two different propagation distances is illustrated in Fig. 1(top).

2. Sampling frequency

As mentioned in Sec. I, numerical dispersion also increases with frequency. Also, recall from Sec. I that in the present context of room acoustic modelling, the simulation bandwidth is limited by the cutoff frequency of the numerical FDTD scheme and is $0.196 \times f_s$ for the SRL scheme (as a reminder, f_s denotes the sampling frequency of the simulation). One useful quantity to quantify numerical dispersion is the phase velocity error, which is defined as the ratio of the numerical wave speed over the real wave speed (Kowalczyk and Van Walstijn, 2011). The phase velocity error profile for the SRL scheme in the direction of the worst-case dispersion error is shown in Fig. 2. As can be seen from Fig. 2, for a fixed frequency, f, the maximal phase velocity error percentage occurs at the cutoff frequency and is approximately 32%. In this study, 12 different phase velocity error percentages, varying from 0.5% to 32% at a frequency f chosen to be 20 kHz, were employed. An example of two dispersion filters using two different sampling frequencies is illustrated in Fig. 1(bottom). Informal listening suggested that the audibility of dispersion can be low such that the two conditions, *nonfys 20* and *nonfys 200*, were additionally created in which two extra dispersion filters were applied to the BRIRs to produce exaggerated dispersion. The two additional dispersion filters corresponded to a maximum phase velocity error of ≈32% at 20 kHz and a propagation distance, d, set to 20 m and 200 m at the start of the BRIRs, respectively.

3. Direct path

Whereas there is a propagation direction for which numerical dispersion is maximum, there also exists a direction for which the error does not exist. For the SRL scheme considered, this occurs in the diagonal direction when considering the diagonal of a cube that is in the direction ($\theta = 45°$, $\phi \approx 35°$) if the cube's edges are parallel to the main

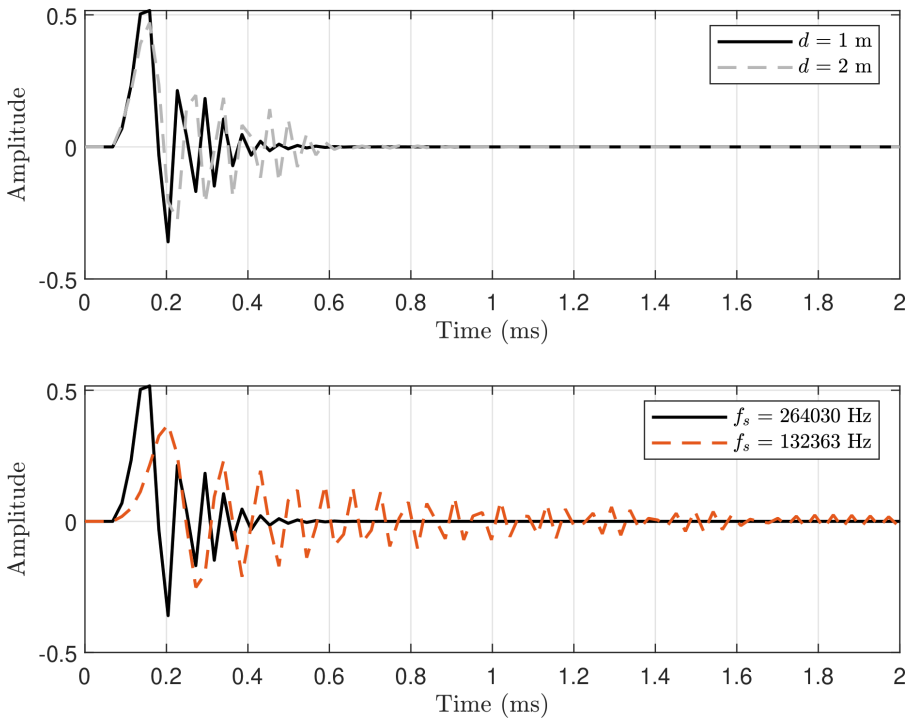

FIG. 1. (Color online) (Top) Dispersion filters, $h(t)$, corresponding to propagation distances $d = 1$ m and 2 m with a sampling frequency of $f_s = 264\,030$ Hz (fixing the phase velocity error to 2% at 20 kHz). (Bottom) Dispersion filters, $h(t)$, corresponding to a propagation distance $d = 1$ m with sampling frequencies $f_s = 264\,030$ Hz and $f_s = 132\,363$ Hz (with the latter fixing the phase velocity error to 10% at 20 kHz). Note that the propagation delay has been removed.

axes of the spatial grid, where θ and ϕ denote the azimuth and elevation, respectively. This implies that in an FDTD simulation domain, the orientation of the three-dimensional (3-D) geometry (e.g., a room) can be chosen such that the direct path of the simulated response is free of numerical dispersion.

To test if the perceptual thresholds for numerical dispersion depend on whether numerical dispersion is present or absent in the direct path, conditions were created in which the dispersion filters were applied only to those parts of the BRIRs that occurred after the direct sound. The conditions where numerical dispersion was present and absent in the direct path will be hereafter referred to as *full* and *partial*, respectively.

IV. EXPERIMENT

A. Sound samples

All of the BRIRs (with and without numerical dispersion) were further convolved with two dry recordings: a castanet excerpt (duration 2.1 s) and a male voice pronouncing the first sentence of "The Rainbow Passage" (duration 4.7 s; Fairbanks, 1960). In addition, a third signal in the experiment was the pure BRIRs, i.e., with an impulse as source signal, which will be hereafter referred to as impulse. The castanet sound sample was chosen because of its transient nature and spectrum, which exhibits significant energy up to 12 kHz. On the other hand, the speech signal whose spectrum contained most of its energy below 5 kHz was chosen because it represents a more familiar acoustic scenario and, therefore, is relevant in the context of room acoustic auralization. Note that several musical excerpts were also considered as sound sample conditions for the experiment. Informal listening sessions revealed that the audibility of numerical dispersion in such sound samples was similar to when using speech. As such, no musical excerpt was included in the experiment to limit the number of measurement conditions and, thus, the duration of the experiment, which, in turn, also limits the fatigue of the participants.

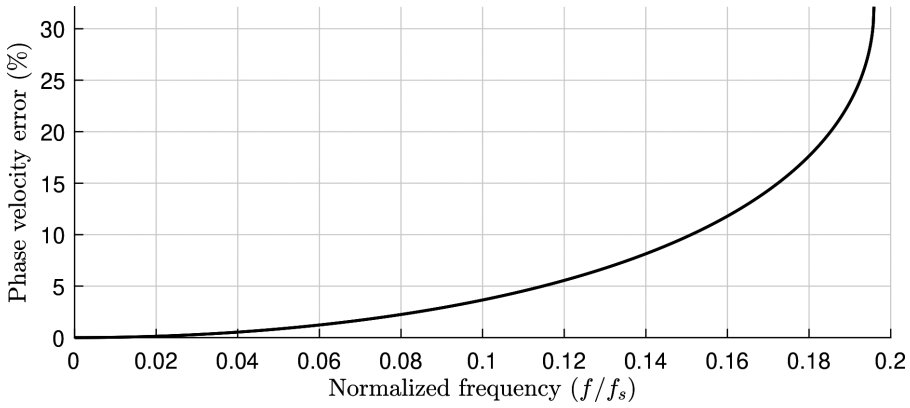

FIG. 2. The phase velocity error profile for the worst-case propagation direction of the SRL scheme, which is the axial direction.

https://doi.org/10.1121/10.0014830

B. Experimental setup

The experiment took place in two different acoustics laboratories. At Aalto University, the experiment ran simultaneously in two separate listening booths specifically designed for running listening experiments in which the measured background noise levels were $L_{\mathrm{Aeq,30s}} = 22$ dB. At Chalmers University, the experiment was conducted in an acoustically treated laboratory room with a measured background noise level of $L_{\mathrm{Aeq}} = 19$ dB.

The user interface was designed and launched with MATLAB (2019) running on a computer (Mac mini, Apple, Cupertino, CA) connected to a headphone amplifier (Objective 2 + ODAC, distributed by JDS Labs, Collinsville, IL) and a pair of open-back headphones (Sennheiser HD 650, Wedemark, Germany) at Aalto University. At Chalmers University, the interface ran on a computer (iMac Pro, Apple) connected to an audio interface (Scarlett 2i2, Focuswrite, High Wycombe, England) to which a headphone amplifier (Lake People phone-amp G109, Konstanz, Germany) and a pair of headphones (Sennheiser HD 650) were connected. Headtracking was not applied in the experiment.

C. Experimental design

An adaptive procedure was adopted to measure the perceptual thresholds for numerical dispersion. More specifically, a transformed up-down staircase procedure following a two-down/one-up algorithm, estimating 70.7% correct (Levitt, 1971), was used with a triangular procedure. The experimental task consisted of selecting the odd sample out amongst the three presented samples. The reference was always the sample that did not contain numerical dispersion and appeared twice amongst the stimuli triplet. The step size between each up/down error level was fixed and differed across the conditions impulse/castanet/speech, as can be seen in Fig. 3. Note that prior to running the experiment, an informal listening session was conducted in which the same step size was applied across the three impulse/castanet/speech conditions. However, from the informal listening session, large differences were observed in the measured thresholds between these conditions, especially between the speech and other two conditions. As a result, the step size for the experiment was adjusted for each condition to include error levels that would be as close as possible to the true (unknown) thresholds. In addition, the number of error levels was limited to avoid fatigue, thus, making the intervals between the step sizes slightly unequal.

The staircase procedure stopped either when six reversal points were counted or one of the two extreme error levels of the error scale (i.e., the minimum or maximum error level) was attained for five successive trials. In the former case, the perceptual threshold was estimated as the average phase velocity error percentage corresponding to these six reversal points, and in the latter case, the threshold was

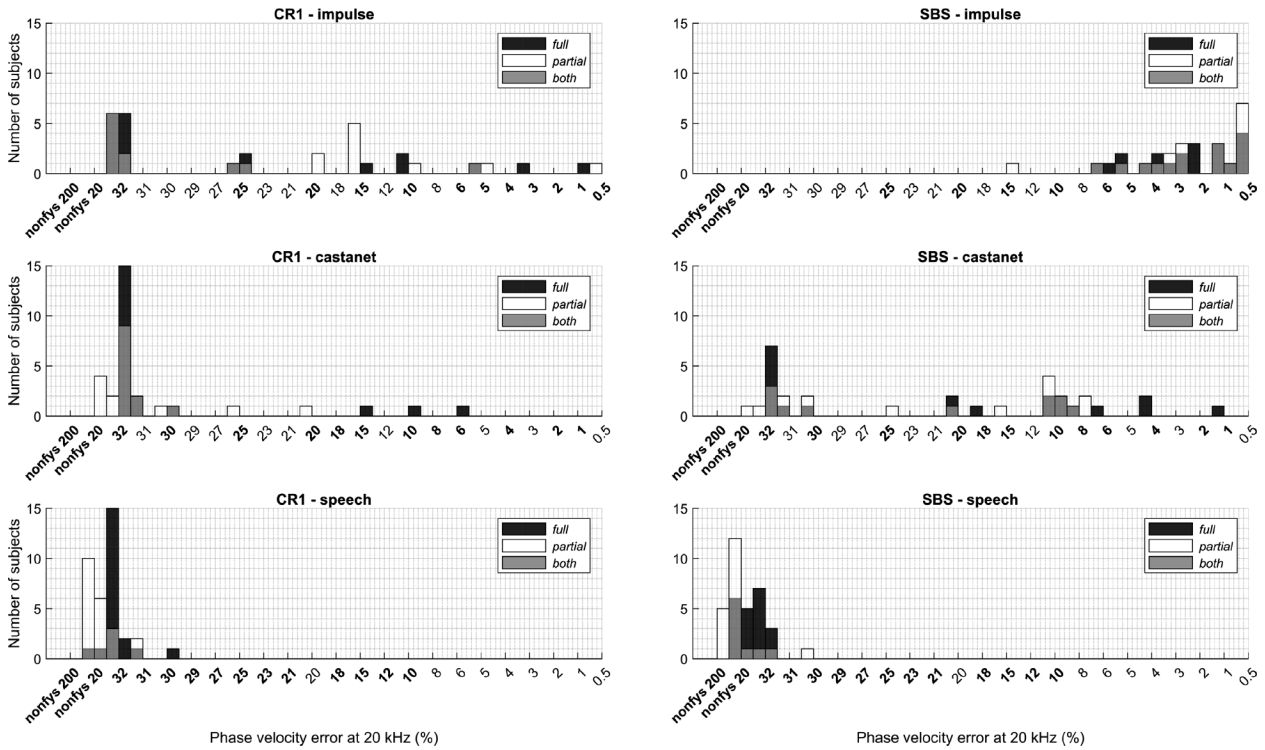

FIG. 3. Histograms of the individual measured perceptual detection thresholds for all of the conditions. It is noted that *full* refers to the condition where the dispersion error filters were applied to the entire BRIRs, whereas *partial* refers to the condition where the direct sound in the BRIRs was free of numerical dispersion. *Both* denotes overlapping measured thresholds between the conditions *full* and *partial*. The bars, whose widths are set to 0.5, are stacked except when *full* and *partial* conditions are overlapping. The *x* axes in each column are unified such that the subfigures can be directly compared, but recall that the error levels used in the experiment varied across the sound sample (see Sec. IV C). The error levels used for each condition in the experiment are indicated in bold.

estimated as the phase velocity error percentage attained in the five successive successful trials. It is worth mentioning that if *nonfys* error levels were attained at the reversal points or the *nonfys 200* error level was attained in the five successive trials, the phase velocity error percentage considered in the estimation of the individual threshold for the statistical analysis (see also Sec. V A) was the maximum phase velocity error percentage.

Every eight triplets, the highest error level (which was the same across the impulse/castanet/speech conditions) was inserted in place of the odd sample out to remind the listeners of the audible artefacts to detect as well as to provide confidence to the participants for correct detection. These additional "easy-to-discriminate" triplets were excluded from the perceptual threshold estimation measures.

Prior to starting the experiment, subjects were provided with written instructions describing the task and they completed a training phase to familiarize themselves with the task and user interface. The training phase consisted of five trials with a decreasing degree of error if a correct answer was provided and an increasing degree of error otherwise. Feedback for correct/incorrect answers was provided at the end of each trial during the training phase. The subjects were allowed to adjust the reproduction level during the training phase and asked to not modify it afterward. The presentation order of the conditions CR1/SBS and *full*/*partial* was randomized across participants. The conditions impulse/castanet/speech were always presented in the same order such that the difficulty level of the task was progressively increased. Each room was evaluated in two separate sessions spanned over two different days to limit fatigue. This means that in one session, the perceptual threshold for numerical dispersion was measured for 2 (*full*/*partial*) × 3 (impulse/castanet/speech) = 6 conditions.

D. Subjects

The number of participants in the experiment was 14 (of which 3 were female) at Aalto University and 7 (of which 2 were female) at Chalmers University, thus, resulting in a total of 21 subjects. All of the participants had self-reported normal hearing (no hearing test was performed) and ranged from 25 to 43 years of age (average = 31 years old, standard deviation = 5 yr). They all had previous experience with listening tests and an educational background in acoustics. On average across all of the participants, the experiment lasted 28 min (standard deviation = 10 min) for CR1 and 33 min (standard deviation = 7 min) for SBS. Participation was voluntary and not compensated for.

V. RESULTS AND DISCUSSION

A. Overview

To determine whether the measured perceptual detection thresholds were statistically different between the two independent groups of participants from the two different acoustics laboratories, a nonparametric Wilcoxon rank sum test (Wilcoxon, 1945) was conducted for each of the 12 conditions.

TABLE II. The mean and standard deviation (indicated in parentheses) of the individual measured perceptual thresholds for each condition. All of the values are reported as phase velocity error percentages measured at 20 kHz. Note that the thresholds which were above the maximum phase velocity error percentage were clipped to 32%.

		Sound sample		
Room	Direct path	Impulse	Castanet	Speech
CR1	*Full*	23.8 (9.8)	28.6 (7.6)	31.7 (0.7)
	Partial	21.1 (9.3)	30.7 (2.5)	31.9 (0.3)
SBS	*Full*	3.0 (2.3)	19.8 (11.3)	31.9 (0.2)
	Partial	2.9 (3.5)	20.8 (9.7)	31.8 (0.5)

For all of the conditions, the results of the tests ($-1.4699 \leq Z \leq 0.8720$, and $0.1416 \leq p\text{-value} \leq 0.9511$ across all of the conditions) indicated that there was not enough evidence to reject the null hypothesis of equal medians between the two groups of participants at the 5% significance level.

The individual measured thresholds for each participant and averaged threshold values across participants are reported for all of the 12 conditions in Fig. 3 and Table II, respectively. As can be seen in Fig. 3 and Table II, the thresholds (stated as phase velocity error percentages) are generally lower for the room SBS, which exhibited the longer reverberation time in comparison to CR1. This difference observed between the two rooms is less pronounced for the condition castanet. As for the speech, the thresholds are similar for the two rooms. Another observation concerns the sound sample: it seems that the thresholds for both rooms are generally the lowest for the condition impulse, i.e., for an impulse as source signal, and the highest with speech. It can also be seen that there is no clear threshold difference between the conditions *full* and *partial* across the other measured conditions (i.e., rooms and sound samples). Finally, small individual differences were observed for some conditions. For example, the lowest measured perceptual detection thresholds for the conditions CR1-*full* with the impulse and castanet was attained by the same participant.

Recall that the error levels *nonfys 20* and *nonfys 200* do not represent valid cases as they comprise exaggerated dispersion (see Sec. III B 2). These levels were included in the experiment to provide high dispersion error levels for which the audible artefacts would be easily perceived in all of the conditions. Note that some detection thresholds are in the range of this exaggerated dispersion, for example, for both rooms when speech is used, as can be seen in Fig. 3 (bottom row). In the statistics presented in Table II and the following statistical analysis, these threshold values were clipped to the maximum phase velocity error percentage at 20 kHz (i.e., 32%) that was achieved in the employed stimuli when choosing practically viable parameters. In Fig. 3, these threshold values are not clipped to distinguish the instances when the individual thresholds reached these high error levels from when the maximum phase velocity error percentage was attained. The displayed histograms from Fig. 3 were obtained by first transforming the scale of the tested error levels expressed as the percentage of phase velocity error

https://doi.org/10.1121/10.0014830

into a series of uniformly spaced integers ranging from 1 to 14, where 1 represents the maximum error level (i.e., the *nonfys 200* condition) and 14 represents the minimum error level tested (which differed across sound sample conditions). Second, after calculating the measured individual thresholds based on the integer scale of the error levels, the individual thresholds were displayed as the corresponding error levels expressed as percentage of phase velocity error.

B. Statistical analysis

A Shapiro-Wilk test (Shapiro and Wilk, 1965) for normality was conducted for each of the 12 conditions. For all of the conditions except for the triplet CR1-impulse-*partial*, the null hypothesis stating that the data are normally distributed was rejected at the 5% significance level, denoted α.

Because evidence was provided that the data were non-normally distributed for almost all of the conditions, a non-parametric Wilcoxon signed-rank test (Wilcoxon, 1992) was considered to test whether there was a statistical difference between the thresholds measured for the two rooms, separately for each of the three sound samples and *full/partial* conditions. However, the assumption that the distribution of the differences between the two "paired" room groups (by paired, it is meant that the same subjects evaluated both rooms) is symmetrical was violated for the *full/partial*, together with the impulse and speech conditions. As a result, a sign test was performed instead of the Wilcoxon signed-rank test for these latter conditions. Violation of the assumption that the distribution of differences between paired observations is symmetrical was considered when the skewness of the distribution of differences between the two room groups was outside of the interval [-0.5,0.5]. The results of the statistical tests are reported in Table III.

The null hypothesis of the Wilcoxon signed-rank test stating that there was a median difference in measured thresholds between the two rooms was rejected at $\alpha = 5\%$ for all of the tested conditions. Similarly, the null hypothesis of the sign test was rejected at the same significance level for the conditions, including the impulse. However, for the speech sample, the null hypothesis of the sign test could not be rejected at $\alpha = 5\%$. These results suggest that the measured thresholds in the two rooms were statistically different except when the sound sample speech was included.

An intuitive explanation for the difference in the measured thresholds between the two rooms is given by the fact that the drier the room is, the shorter the BRIR is. Since numerical dispersion increases as a function of simulated distance, which is directly related to the duration of the BRIR, it is evident that there will be more numerical dispersion in the BRIR that is the most reverberant, and, thus, will become easier to detect.

To test whether the thresholds were different between the conditions where numerical dispersion was present and absent in the direct path, a similar statistical analysis as for testing differences between rooms was conducted. First, the skewness of the distribution of differences between the two *full/partial* groups was calculated. When the skewness was within the interval [-0.5,0.5], indicating that the data are fairly symmetrical, a Wilcoxon signed-rank test was conducted. When the skewness was outside of the same interval, a sign test was performed on the data. The results from these statistical tests are reported in Table III, which shows that the null hypothesis of the two test types, stating that the difference between the medians of the paired samples is zero, could not be rejected at $\alpha = 0.05$ for all of the sound sample and room conditions. These results provide evidence that the thresholds were not statistically different between the conditions where numerical dispersion was present and absent in the direct path. This result suggests that there is no need to orient the spatial grid such that numerical dispersion will be absent in the direct path of the simulated response.

TABLE III. The results of the statistical tests. *Between* indicates that the thresholds from the corresponding group columns were compared.

Test	Room	Direct path	Sound sample	Test statistic	p-value	Decision[a]
Wilcoxon signed-rank	*Between*	*Full*	Castanet	$Z = 2.7879$	0.0053	Reject
Wilcoxon signed-rank	*Between*	*Partial*	Castanet	$Z = 3.2888$	0.0010	Reject
Sign test	*Between*	*Full*	Impulse	$Z = 3.9269$	8.5683×10^{-5}	Reject
Sign test	*Between*	*Partial*	Impulse	$Z = 3.9269$	8.5683×10^{-5}	Reject
Sign test	*Between*	*Full*	Speech	$Z = -0.2673$	0.7893	Fails to reject
Sign test	*Between*	*Partial*	Speech	$Z = 0$	1.0000	Fails to reject
Wilcoxon signed-rank	CR1	*Between*	Impulse	$Z = 1.2320$	0.2180	Fails to reject
Sign test	CR1	*Between*	Castanet	$Z = -0.7500$	0.4533	Fails to reject
Sign test	CR1	*Between*	Speech	$Z = -1.4434$	0.1489	Fails to reject
Sign test	SBS	*Between*	Impulse	$Z = 0.2357$	0.8137	Fails to reject
Wilcoxon signed-rank	SBS	*Between*	Castanet	$Z = -0.8720$	0.3832	Fails to reject
Sign test	SBS	*Between*	Speech	$Z = -0.9487$	0.3428	Fails to reject
Friedman	CR1	*Full*	*Between*	$\chi^2(2) = 12.37$	0.0021	Reject
Friedman	CR1	*Partial*	*Between*	$\chi^2(2) = 27.91$	8.6904×10^{-7}	Reject
Friedman	SBS	*Full*	*Between*	$\chi^2(2) = 40.67$	1.4769×10^{-9}	Reject
Friedman	SBS	*Partial*	*Between*	$\chi^2(2) = 40.67$	1.4769×10^{-9}	Reject

[a]The decision on the null hypothesis of the test.

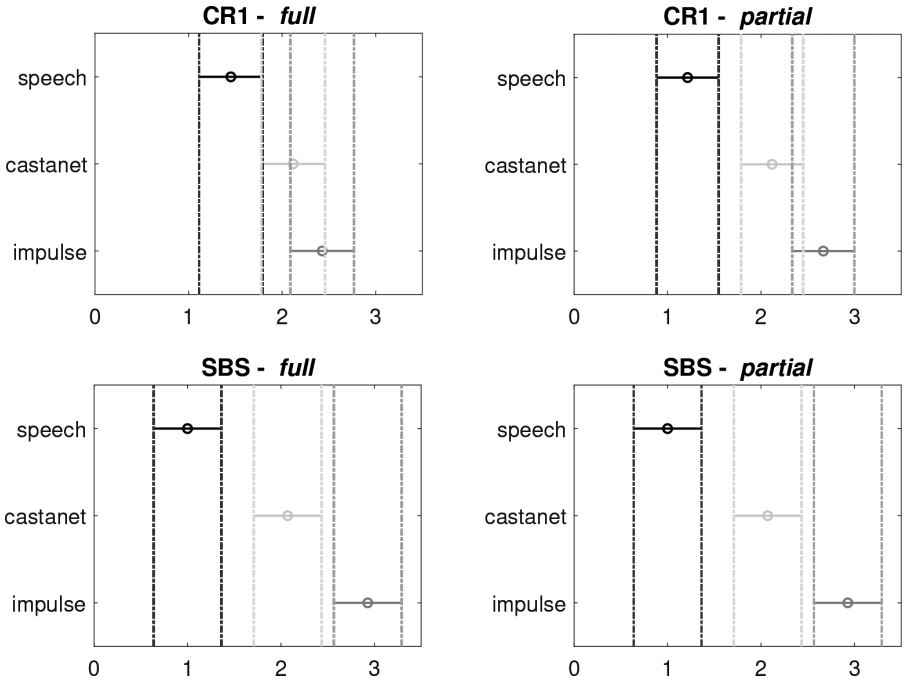

FIG. 4. The results of the multiple comparison tests. The symbol "○" and the horizontal line extending out from the symbol represent the group mean and the 95% comparison interval, respectively. Two group means are significantly different if their intervals are disjoint. Alternatively, if their intervals overlap, the two group means are not significantly different. Notice the similarity of the results between the conditions small broadcast studio (SBS)-*full* and SBS-*partial*.

Separately for each combination of room CR1/SBS and *full/partial* conditions, a nonparametric Friedman test (Friedman, 1937) on the sound sample groups was performed. The test result, reported in Table III, indicated that the null hypothesis stating that all of the samples are drawn from the same population should be rejected for all of the combinations of conditions at the α level 5%, suggesting that at least one sample mean is significantly different than the other sample means.

To determine which pairs of means significantly differed, a multiple comparison *post hoc* test using the Bonferroni method was performed. The results of the multiple comparison test, shown in Fig. 4, revealed that in all of the cases except for the condition CR1-*full*, the threshold means between impulse and speech, as well as between castanet and speech, are significantly different at the 5% significance level. These results also demonstrated that for CR1, the threshold means were not statistically different between the castanet and impulse sound sample groups. This is contrary to the results for SBS, where the multiple comparison test indicated that the threshold means were statistically different between the three possible combinations of two of the sound sample groups (i.e., between the impulse and castanet groups, between the impulse and speech groups, and between the castanet and speech groups). Compared to findings from the literature on the perception of numerical dispersion, the present results confirm the dependency of the measured perceptual detection thresholds on the sound sample, as in Saarelma *et al.* (2016) and Saarelma and Savioja (2019).

It is also reminded that the present perceptual detection thresholds for numerical dispersion were measured as percentages of the phase velocity error at 20 kHz. Other previous related studies focusing on perceptual evaluations of numerical dispersion, such as Saarelma *et al.* (2016) and Saarelma and Savioja (2016), have also employed the group delay as a measure of perceptual thresholds for the error. However, since the binaural auralizations evaluated in this work resulted in complex and frequency-dependent signals, the group delay was not considered. Overall, it is still not evident what error metric(s) should be employed as a measure of the accumulated error in the auralizations, and this could be the object of future work. Nevertheless, although it is difficult to translate these results if the amount of phase velocity error was redefined for each condition (e.g., based on the phase velocity error percentage calculated at the effective source bandwidth instead of 20 kHz), the present results still provide guidelines for generating FDTD simulations that would be free of audible artefacts for different scenarios.

C. Perceptual properties of dispersion

To better understand the perceptual changes due to numerical dispersion, interviews with the participants were conducted after they completed the experiment. As in previous related studies, most participants reported hearing frequency chirps or sweeps in the signals for high error levels. The majority of the participants also described the stimuli containing numerical dispersion to be smoother, softer, or less impulsive than the reference in the attack and onsets. The focus was mostly on the fricative sounds of the speech sample, like the "s," suggesting that the error could be more noticeable in a word whose spectrum has high frequency peaks/content. Less than half of the subjects also reported coloration/timbre and loudness/level differences. Fewer subjects additionally described the odd one out to be more stretched/wider in time than the reference. This last remark is particularly interesting because time-domain responses

polluted with numerical dispersion are, in fact, smeared in time compared to their dispersion error-free counterparts.

VI. CONCLUSION

The lowest perceptual detection thresholds for numerical dispersion were measured in binaural auralizations of two acoustically different rooms. The BRIRs were taken from measurements, and numerical dispersion was introduced in the BRIRs by the means of filters that represent the dispersion that plane waves experience, which propagate in the direction of the worst-case dispersion error. The BRIRs were further convolved with a dry recording of a speech signal and castanet excerpt. A third test signal consisted of the pure BRIRs, i.e., with an impulse as source signal. The results from the experiment revealed that keeping the direct sound free of numerical dispersion did not lead to statistically different perceptual thresholds than when the dispersion error filters were applied to all of the BRIRs. For the auralizations produced with the impulse and castanet sample, smaller perceptual thresholds were attained for SBS compared to CR1, suggesting that a given amount of dispersion is more audible with longer reverberation. The results also showed that the mean perceptual thresholds for numerical dispersion in speech were almost reaching the maximum phase velocity error percentage that is attainable with the simulations, independently of the room.

Given the dependence of the measured thresholds on the sound sample used in the auralization, it is recommended that the application be determined prior to simulating a room using the FDTD method such that an optimal balance between computational complexity and audibility of dispersion can be achieved. More specifically, auralizations including impulsive/transient or high-pitched sounds will lead to the most conservative perceptual detection thresholds for numerical dispersion. To obtain a dispersion error-free auralization of a room with a reverberation time of approximately 1 s with such a sound sample, the FDTD simulation should be run using a sampling frequency set such that the phase velocity error measured at 20 kHz is around 0.5%. For the simulation of drier rooms, the sampling frequency can be set such that the phase velocity error is around 2% at 20 kHz. Finally, for rooms with a larger reverberation time than approximately 1 s, running an experiment similar to the one presented in this study (i.e., measuring the perceptual detection thresholds for numerical dispersion) is recommended as it was shown that the measured thresholds were lower for the most reverberant room.

ACKNOWLEDGMENTS

The authors would like to thank Nils Meyer-Kahlen from Aalto University for sharing his valuable insights on binaural synthesis, as well as the participants of the experiment.

Ahrens, J., and Andersson, C. (**2019**). "Perceptual evaluation of headphone auralization of rooms captured with spherical microphone arrays with respect to spaciousness and timbre," J. Acoust. Soc. Am. **145**(4), 2783–2794.

Bilbao, S., Politis, A., and Hamilton, B. (**2019**). "Local time-domain spherical harmonic spatial encoding for wave-based acoustic simulation," IEEE Signal Process. Lett. **26**(4), 617–621.

Cobos, M., Escolano, J., López, J. J., and Pueo, B. (**2008**). "Subjective effects of dispersion in the simulation of room acoustics using digital waveguide mesh," in *Audio Engineering Society Convention 124* (Audio Engineering Society, Amsterdam, The Netherlands).

Fairbanks, G. (**1960**). *Voice and Articulation Drillbook*, 2nd ed. (Harper and Row, New York), p. 127.

Friedman, M. (**1937**). "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," J. Am. Stat. Assoc. **32**(200), 675–701.

Hamilton, B. (**2016**). "Finite difference and finite volume methods for wave-based modelling of room acoustics," Ph.D. thesis, The University of Edinburgh, Edinburgh, Scotland.

Hamilton, B., and Bilbao, S. (**2017**). "FDTD methods for 3-D room acoustics simulation with high-order accuracy in space and time," IEEE/ACM Trans. Audio. Speech. Lang. Process. **25**(11), 2112–2124.

Karjalainen, M., and Erkut, C. (**2004**). "Digital waveguides versus finite difference structures: Equivalence and mixed modeling," EURASIP J. Appl. Signal Process. **2004**, 978–989.

Kowalczyk, K., and Van Walstijn, M. (**2011**). "Room acoustics simulation using 3-D compact explicit FDTD schemes," IEEE Trans. Audio. Speech. Lang. Process. **19**(1), 34–46.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**(2B), 467–477.

López, J. J., Escolano, J., and Pueo, B. (**2007**). "Simulation of complex and large rooms using a digital waveguide mesh," in *Audio Engineering Society Convention 123* (Audio Engineering Society, New York).

MATLAB (**2019**). *Version 9.6.0.1307630 (R2019a)* (The Mathworks, Inc., Natick, MA).

Meyer, J., Lokki, T., and Ahrens, J. (**2020**). "Identification of virtual receiver array geometries that minimize audibility of numerical dispersion in binaural auralizations of finite difference time domain simulations," in *Audio Engineering Society Convention 149* (Audio Engineering Society), online.

Mokhtari, P., Takemoto, H., Nishimura, R., and Kato, H. (**2008**). "Computer simulation of HRTFs for personalization of 3D audio," in *Second International Symposium on Universal Communication, 2008 (ISUC'08)* (IEEE, Osaka, Japan), pp. 435–440.

Oberkampf, W. L., and Roy, C. J. (**2010**). *Verification and Validation in Scientific Computing* (Cambridge University Press, Cambridge).

Prepeliţă, S. T., Gómez Bolaños, J., Geronazzo, M., Mehra, R., and Savioja, L. (**2019**). "Pinna-related transfer functions and lossless wave equation using finite-difference methods: Verification and asymptotic solution," J. Acoust. Soc. Am. **146**(5), 3629–3645.

Roy, C. J., and Oberkampf, W. L. (**2011**). "A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing," Comput. Methods Appl. Mech. Eng. **200**(25–28), 2131–2144.

Saarelma, J., Botts, J., Hamilton, B., and Savioja, L. (**2016**). "Audibility of dispersion error in room acoustic finite-difference time-domain simulation as a function of simulation distance," J. Acoust. Soc. Am. **139**(4), 1822–1832.

Saarelma, J., and Savioja, L. (**2016**). "Audibility of dispersion error in room acoustic finite-difference time-domain simulation in the presence of absorption of air," J. Acoust. Soc. Am. **140**(6), EL545–EL550.

Saarelma, J., and Savioja, L. (**2019**). "Audibility of dispersion error in room acoustic finite-difference time-domain simulation in the presence of a single early reflection," J. Acoust. Soc. Am. **145**(4), 2761–2769.

Sargent, R. G. (**2010**). "Verification and validation of simulation models," in *Proceedings of the 2010 Winter Simulation Conference* (IEEE, Baltimore, MD), pp. 166–183.

Schneider, J. B., and Wagner, C. L. (**1999**). "FDTD dispersion revisited: Faster-than-light propagation," IEEE Microwave Guided Wave Lett. **9**(2), 54–56.

Shapiro, S. S., and Wilk, M. B. (**1965**). "An analysis of variance test for normality (complete samples)," Biometrika **52**(3/4), 591–611.

Sheaffer, J., Van Walstijn, M., Rafaely, B., and Kowalczyk, K. (**2014**). "A spherical array approach for simulation of binaural impulse responses

using the finite difference time domain method," in *Proceedings of Forum Acusticum*, Kraków, Poland.
Sheaffer, J., Van Walstijn, M., Rafaely, B., and Kowalczyk, K. (**2015**). "Binaural reproduction of finite difference simulations using spherical array processing," IEEE/ACM Trans. Audio. Speech. Lang. Process. **23**(12), 2125–2135.
Sheaffer, J., Webb, C., and Fazenda, B. M. (**2013**). "Modelling binaural receivers in finite difference simulation of room acoustics," Proc. Mtgs. Acoust. **19**(1), 015098.
Southern, A., Murphy, D., Lokki, T., and Savioja, L. (**2011**). "The perceptual effects of dispersion error on room acoustic model auralization," in *Proceedings of Forum Acusticum*, Aalborg, Denmark, pp. 1553–1558.
Southern, A., Murphy, D. T., and Savioja, L. (**2012**). "Spatial encoding of finite difference time domain acoustic models for auralization," IEEE Trans. Audio. Speech. Lang. Process. **20**(9), 2420–2432.
Stade, P., Bernschütz, B., and Rühl, M. (**2012**). "A spatial audio impulse response compilation captured at the WDR broadcast studios," in *27th Tonmeistertagung-VDT International Convention*, pp. 551–567.
Van Mourik, J., and Murphy, D. (**2014**). "Explicit higher-order FDTD schemes for 3D room acoustic simulation," IEEE/ACM Trans. Audio, Speech Lang. Process. **22**(12), 2003–2011.
Van Walstijn, M., and Kowalczyk, K. (**2008**). "On the numerical solution of the 2d wave equation with compact fdtd schemes," in *Proceedings of Digital Audio Effects (DAFx)*, Espoo, Finland, pp. 205–212.
Webb, C. J., and Bilbao, S. (**2012**). "Binaural simulations using audio rate FDTD schemes and CUDA," in *Proceedings of the 15th International Conference on Digital Audio Effects (DAFx-12)*, York, England.
Wilcoxon, F. (**1945**). "Individual comparisons by ranking methods," Biom. Bull. **1**, 80–83.
Wilcoxon, F. (**1992**). "Individual comparisons by ranking methods," in *Breakthroughs in Statistics* (Springer, New York), pp. 196–202.