
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Saarelma, Jukka; Savioja, Lauri

Audibility of dispersion error in room acoustic finite-difference time-domain simulation in the presence of a single early reflection

Published in:
Journal of the Acoustical Society of America

DOI:
[10.1121/1.5095874](https://doi.org/10.1121/1.5095874)

Published: 01/04/2019

Document Version
Publisher's PDF, also known as Version of record

Published under the following license:
CC BY

Please cite the original version:
Saarelma, J., & Savioja, L. (2019). Audibility of dispersion error in room acoustic finite-difference time-domain simulation in the presence of a single early reflection. *Journal of the Acoustical Society of America*, 145(4), 2761-2769. <https://doi.org/10.1121/1.5095874>

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Audibility of dispersion error in room acoustic finite-difference time-domain simulation in the presence of a single early reflection

Jukka Saarelma, and Lauri Savioja

Citation: *The Journal of the Acoustical Society of America* **145**, 2761 (2019); doi: 10.1121/1.5095874

View online: <https://doi.org/10.1121/1.5095874>

View Table of Contents: <https://asa.scitation.org/toc/jas/145/4>

Published by the *Acoustical Society of America*

ARTICLES YOU MAY BE INTERESTED IN

[A round robin on room acoustical simulation and auralization](#)

The Journal of the Acoustical Society of America **145**, 2746 (2019); <https://doi.org/10.1121/1.5096178>

[Room acoustics modelling in the time-domain with the nodal discontinuous Galerkin method](#)

The Journal of the Acoustical Society of America **145**, 2650 (2019); <https://doi.org/10.1121/1.5096154>

[Passive volumetric time domain simulation for room acoustics applications](#)

The Journal of the Acoustical Society of America **145**, 2613 (2019); <https://doi.org/10.1121/1.5095876>

[Introduction to the Special Issue on Room Acoustic Modeling and Auralization](#)

The Journal of the Acoustical Society of America **145**, 2597 (2019); <https://doi.org/10.1121/1.5099017>

[Source directivity approximation for finite-difference time-domain simulation by estimating initial value](#)

The Journal of the Acoustical Society of America **145**, 2638 (2019); <https://doi.org/10.1121/1.5096163>

[A framework for auralization of boundary element method simulations including source and receiver directivity](#)

The Journal of the Acoustical Society of America **145**, 2625 (2019); <https://doi.org/10.1121/1.5096171>



CAPTURE WHAT'S POSSIBLE
WITH OUR NEW PUBLISHING ACADEMY RESOURCES

Learn more 



Audibility of dispersion error in room acoustic finite-difference time-domain simulation in the presence of a single early reflection

Jukka Saarelma^{a)} and Lauri Savioja

Department of Computer Science, School of Science, Aalto University, Otakaari 5, FI-02150, Espoo, Finland

(Received 23 July 2018; accepted 12 October 2018; published online 30 April 2019)

Finite-difference time-domain method has gained increasing interest for room acoustic prediction use. A well-known limitation of the method is a frequency- and direction-dependent dispersion error. In this study, the audibility of dispersion error in the presence of a single surface reflection is measured. The threshold is measured for three different distance conditions with a fixed reflection arrival azimuth angle of 54.7° . The error is placed either in the direct path, or in the reflection path. Additionally a qualitative follow-up experiment to evaluate how the measured thresholds reflect the audibility of error in short room responses is carried out. The results indicate that the threshold varies depending whether the error is in the direct path or in the reflection path. For transient signals the threshold is higher when the error is located in the direct path, whereas for speech signal, the threshold is higher when it is located in the reflection path. Evidence is found that the error is detectable in rendered room responses at the measured threshold levels. © 2019 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1121/1.5095874>

[JFL]

Pages: 2761–2769

I. INTRODUCTION

Wave-based simulation methods have gained interest for large scale room acoustic prediction. *Finite-difference time-domain* (FDTD) method is one potential candidate for such a task due to its scalability properties for parallel computing architectures. For room acoustic applications, compact explicit FDTD schemes and *finite-volume time-domain* (FVTD) methods² have had the most attention.

A specific source of error for explicit FDTD schemes is dispersion; due to the discretization, different frequencies in the simulation domain travel with different velocities. In addition to the frequency dependence, the dispersion is also direction dependent. Depending on discretization scheme, at certain directions a propagating plane wave is error-free, and for the remaining directions propagating plane waves will contain a varying degree of dispersion. For example, one of the simplest FDTD schemes, the standard rectilinear scheme, an error-free propagation direction is towards diagonal of a rectilinear grid, and the worst-case propagation error appears in the direction of each Cartesian coordinate axis. The anisotropy of the error makes the possible correction difficult for full room responses as different propagation paths contain different amounts of dispersion. Therefore it is inevitable that some dispersion error will remain in the simulation result.

The audibility of dispersion error has been previously studied with different experiment designs. Cobos *et al.*³ used simulated room responses to compare several different oversampling factors. Used sampling frequencies for digital waveguide mesh (DWM) simulation were 20, 30, and

40 kHz, and the responses were low-pass filtered from 5000 Hz. Two different speech samples were used in the test. It was reported that the participants were not able to discriminate between the sampling frequencies of 30 and 40 kHz. Twenty percent of the participants were able to discriminate between the conditions 20 and 30 kHz. Southern *et al.*⁴ compared simulated free-field responses of two different propagation directions in a FDTD simulation using sampling frequency of 5000 Hz. Independent variables in the test were low-pass filter cutoff frequency and the distance between the source and receiver. The authors suggest that below a normalized cutoff frequency of 0.18 (900 Hz), the participants could not discriminate between two propagation directions, which both contained some degree of dispersion error. Saarelma *et al.* measured the audibility of the dispersion error of FDTD schemes using free-field responses of several different schemes as a function of distance,⁵ and as a function of the phase velocity error percentage in the presence of air absorption.^{5,20} The responses were generated using the analytic formulation of the dispersion relation of the schemes. The results indicated that with a fixed 2% phase velocity error percentage at 20 kHz, 9.8 m of propagation in the simulation domain introduced an audible error. If air absorption was included in the response, a phase velocity percentage of 0.29% was needed to make the dispersion error inaudible for propagation distance conditions less than 344 m. In both studies, evidence was found that maximum group delay close to 2 ms was present in the signal at the threshold when air absorption was not present (averaged over mean threshold observations with different distances without air absorption in Refs. 5 and 20, $\mu = 2.13$ ms, $\sigma = 0.36$ ms).

The previous work addresses either band-limited responses, or solely the error in a free-field response.

^{a)}Electronic mail: jukka.saarelma@aalto.fi

Experiments where the audibility of the dispersion error in full audible range in the presence of any room-acoustic phenomenon have not been conducted to the present authors' knowledge. In this study, the detectability of dispersion error of a FDTD scheme is measured for a condition where single surface reflection is present. The experiment tests whether the dispersion error is less detectable when introduced in a single surface reflection, opposed to when it is introduced in the direct sound. Additionally, a follow-up study is conducted to study how the measured thresholds reflect the audibility of the dispersion error in an auralized room response in the case of a small room with a short reverberation time.

II. RELEVANT PSYCHOACOUSTIC EFFECTS

The case of a single reflection can be thought as two sources radiating simultaneously from two different directions. For the perception of a single, coherent, reflection, the most relevant effect is referred to as “the law of the first wavefront,” or “precedence effect.”⁶ The effect occurs when two coherent wavefronts arrive at a listening position from different directions with a time difference more than 1 ms. In such cases the position of the auditory event is determined by the wavefront that arrives first. In a case when the time difference exceeds a certain limit, two auditory events occur instead of one. A common measure of the separation of the audio event is the *echo threshold*. The level of the echo threshold varies depending on the stimuli; for short transient, click-like signals, the threshold is in the range of 5–10 ms. For continuous speech, the value is in the range of 30–50 ms.⁷

The influence of coherence⁸ (p. 241) affects the lateralization so, that when fully coherent signals are played to each of the subjects ears, the source is localized as a single event so that the “center of gravity” is on the median plane. When the coherence is reduced, the auditory event widens, and after enough reduction, two separate auditory events occur.

In the case of this study, the effects that may assist the discrimination and evaluation of the error level are difference in the width of the auditory event due to the reduced coherence of the signals, and changes in timbre due to the dispersion error.

III. NUMERICAL DISPERSION

Here the scalar wave equation is used as a model for sound propagation in air, given

$$\frac{\partial^2 p}{\partial t^2} = c^2 \left(\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} \right), \quad (1)$$

where p is the acoustical pressure, and c the speed of sound. Following the notation of Kowalczyk and van Walstijn,¹ Eq. (1) can be discretized using finite differences. The scheme used in this study is *standard rectilinear* (SRL) scheme which has the form

$$\delta_t^2 p_{k,l,m}^n = \lambda^2 (\delta_x^2 + \delta_y^2 + \delta_z^2) p_{k,l,m}^n, \quad (2)$$

where δ_t , δ_x , δ_y , and δ_z are difference operators:

$$\begin{aligned} \delta_t^2 p_{k,l,m}^n &= p_{k,l,m}^{n+1} - 2p_{k,l,m}^n + p_{k,l,m}^{n-1}, \\ \delta_x^2 p_{k,l,m}^n &= p_{k+1,l,m}^n - 2p_{k,l,m}^n + p_{k-1,l,m}^n, \\ \delta_y^2 p_{k,l,m}^n &= p_{k,l+1,m}^n - 2p_{k,l,m}^n + p_{k,l-1,m}^n, \\ \delta_z^2 p_{k,l,m}^n &= p_{k,l,m+1}^n - 2p_{k,l,m}^n + p_{k,l,m-1}^n, \end{aligned} \quad (3)$$

with $p_{k,l,m}^n = p(x, y, z, t)$, with $x = k\Delta x, y = l\Delta x, z = m\Delta x$, and $t = n\Delta t$. Δx is the spatial discretization step size, and λ is the Courant number determining the ratio of the temporal and spatial discretization. The Courant number at the stability limit of the studied scheme is $\lambda = 1/\sqrt{3}$. A dispersion relation of the scheme can be derived directly from Eq. (2):

$$\begin{aligned} \sin^2 \left(\omega \frac{\Delta t}{2} \right) &= \lambda^2 \left[\sin^2 \left(\hat{k}_x \frac{\Delta x}{2} \right) + \sin^2 \left(\hat{k}_y \frac{\Delta x}{2} \right) \right. \\ &\quad \left. + \sin^2 \left(\hat{k}_z \frac{\Delta x}{2} \right) \right], \end{aligned} \quad (4)$$

where ω is angular frequency, and \hat{k}_x , \hat{k}_y , and \hat{k}_z wavenumber components in the directions of the Cartesian coordinate axes. By setting the wavenumber components $\hat{k}_y = 0$ and $\hat{k}_z = 0$, and solving \hat{k}_x as a function of angular frequency, an expression for numerical wavenumber in the direction of the x axis is achieved (same for all axial directions):

$$\hat{k}_x = \frac{2}{\Delta x} \arcsin \left[\frac{1}{\lambda} \sin \left(\omega \frac{\Delta t}{2} \right) \right], \quad (5)$$

which is the worst-case dispersion error direction for the SRL scheme.

From the numerical wavenumber \hat{k}_x the dispersion waveform can be evaluated using a plane wave solution:

$$\hat{F} = e^{-j\omega t} e^{-j\hat{k}_x(\omega)d}, \quad (6)$$

for distance d , for this case in the axial direction of the SRL scheme. Such a waveform corresponds to wave introduced by a planar *hard*⁹ source. More elaborate derivations of the *dispersion filter* may be found in Ref. 5.

IV. EXPERIMENTAL SETUP

The experimental setup in this work is to compare the audibility of dispersion error in a signal that is a combination of the direct sound and a single early reflection. The error levels are chosen so that one of the two, the direct sound or the reflection path, contains the worst case error, and the other one is error free. The surface is assumed to have a reflection coefficient of 1. The independent variables of the experiment are the location of the dispersion error (*dispersion in the direct sound, dispersion in the reflection*), difference between the arrival time of the direct sound and the reflection (5, 10, 20 ms), and sound sample (*click, speech*). The direction of the arrival of the reflected wave is chosen to be 54.7°. The dependent variable in the experiment is the amount of dispersion error that is controlled with the phase velocity error percentage at the frequency of 10 kHz.

The rationale for the decision of the direction of arrival of the reflection is, as illustrated in Fig. 1(b), this direction of

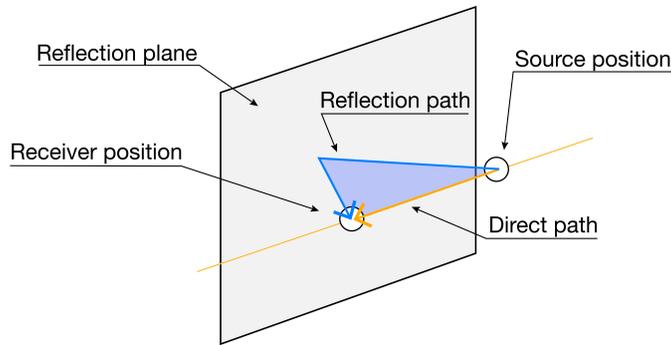
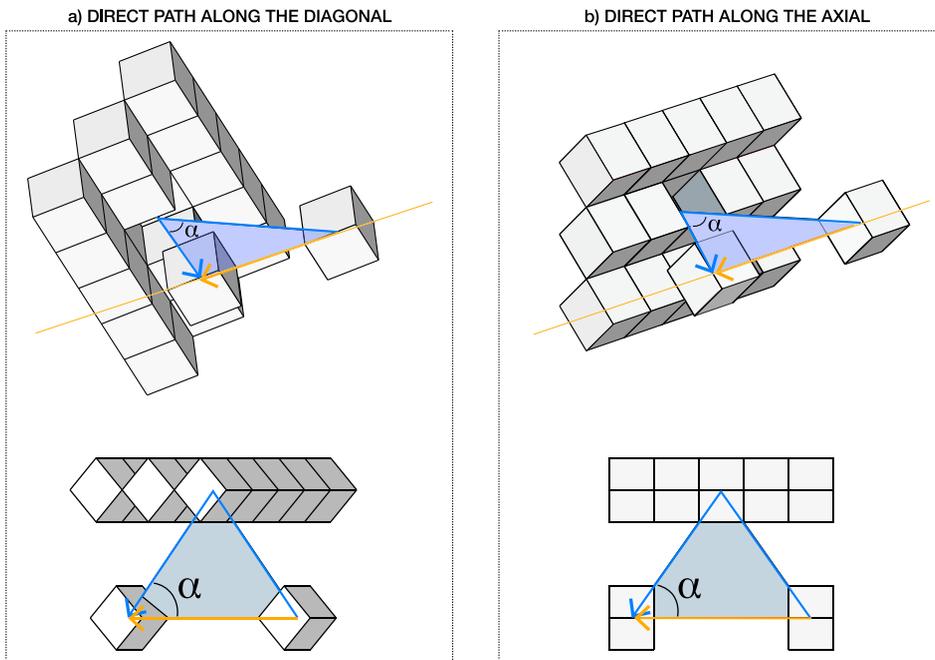


FIG. 1. (Color online) The different reflection conditions used in this study. Condition (a) represents the case where the direct sound propagates in the diagonal direction, and the first reflection arrives close to the axial direction. Condition (b) represents the case where the direct sound propagates across the axial direction, and the reflection arrives at the receiver from the diagonal direction. The azimuth angle (α) of the reflection in condition (a) is 50.7° when the propagation direction is closest to the axial direction. In condition (b) the azimuth angle is 54.7° . The illustrations at the bottom are a top view of each condition. The projection is adjusted so that camera direction is normal to the plane defined by the source, receiver, and reflection point.



arrival (54.7° azimuth, 0° elevation) corresponds to the condition where the direct sound propagates along the axial direction of the grid, and a single reflection then propagates along the diagonal direction. The condition where the propagation directions are nearly opposite is presented in Fig. 1(a). In this condition, the direct sound propagates along the diagonal, and the reflection close to the axial direction. A condition where the reflection propagates solely in the axial direction is not achievable with a single surface, but may occur if two reflecting surfaces are present. Therefore, a fixed angle of arrival is used, so that it corresponds to the condition (b). In order to achieve symmetric conditions for the experiment, the dispersion filter is evaluated for the worst case direction, regardless of the slight misalignment (4°) of the reflection direction in condition (a). The values of reflection delays and angle of arrival correspond to source–receiver distances of 2.34, 4.70, and 9.39 m.

The test hypothesis in this experiment is that the threshold for the level of dispersion is higher in a case where the dispersion error is in a reflection path instead of the direct path from the source. If evidence for this hypothesis is found, it can be argued that schemes with low error in the vicinity of the direction of Cartesian coordinate axes may be perceptually more optimal if it is assumed that source receiver pairs are commonly at similar heights in typical listening conditions.

Additionally, the possibility of orienting the room model so that less error is in the direct path may be considered.

Additionally a qualitative test is carried out using the results achieved from the threshold experiment. A pair-wise comparison of room responses simulated with a model of a small room is used to qualitatively assess how the threshold measurements reflect the audibility of dispersion error in short room responses. Independent variables in the second experiment are the sampling frequency of the simulation (59 583, 119 165, 238 330 Hz, rationale of the decision is given in Sec. VA 1), receiver distance (2.9, 4.0 m), and sound sample (*click, speech*).

A. Stimuli

Two experiments are conducted. First a threshold measurement procedure for the audibility of dispersion error in a signal that contains the direct sound and a single early reflection. The results of the first experiments are then used in a second experiment where the audibility of dispersion error in a full room response is quantified at the measured threshold values.

1. Experiment 1

Two sound samples are used in this study: a click-like signal, and a short speech excerpt. The samples are processed

using generated responses of a direct sound and a single early reflection, resulting in a signal with two channels. The generated responses contain the time delay and distance attenuation according to the distance condition. For the reference signal, neither of the two paths contain dispersion error, and for the stimuli, dispersion error that is evaluated using the dispersion filter is introduced in the signal according to the condition (*dispersion in the direct sound, dispersion in the reflection*).

The two channels of the processed sample, direct sound and the reflection are played back to the participant in an anechoic room from two loudspeakers, one located directly in the front of the participant (azimuth 0°), and one 54.7° on the right from the participants frontal direction (azimuth 54.7°).

2. Experiment 2

The room responses compared in the second experiment are simulated using an existing room acoustic FDTD library¹⁰ using SRL scheme and first-order accurate boundaries.¹¹ The room model that is used for the simulation is a model of a small rectangular room with dimension $7\text{ m} \times 5\text{ m} \times 2.8\text{ m}$. Several diffusor elements are added to the geometry to achieve diffuse reverberation at the receiver positions. The diffusors are designed using quadratic residue diffuser (QRD) design equations¹² (p. 291) ($N = 13$, depth = 20 cm, well width = 21 cm, well height = 40 cm) with a Chinese folding¹² (p. 317) (total wells height = 5, total wells width = 7). Diffusors are placed in the back half of the room so that two elements are on both side walls, three elements on the back wall, and six on the ceiling. The surface materials of the room model are chosen so, that the reverberation time is approximately 0.6 s. Absorption coefficients for the different surfaces are listed in Table I. The receiver and source positions for the simulation are selected so that specular early reflections will reach the receiver positions in the constructed geometry. The simulation setup is illustrated in Fig. 2.

The simulation is run using a soft source and a source signal consisting of a single impulse at the time instance at 0 s. Impulse responses are simulated for a rectilinear receiver point grid with constraints that the points are inside a sphere with a radius of 0.12 m centered at the chosen receiver location, and that the grid spacing is 0.01 m. The achieved impulse responses are resampled, low-pass filtered with a cutoff frequency of 20 kHz, and processed with an air absorption filter (see Appendix A).

TABLE I. The absorption coefficients of the surfaces in the simulated room in experiment 2. Each absorption coefficient value is directly converted to a normal incidence admittance in the simulation. Therefore the values are not to be interpreted as random incidence absorption coefficients.

Layer	α
Ceiling	0.99
Floor	0.1
Walls front ($x < 0.35$)	0.05
Walls back ($x \geq 0.35$)	0.4
Diffusors	0.01

The processed impulse responses are then used to generate a *plane wave decomposition* (PWD) of the sound field around the receiver position.¹³ The PWD, that is represented by a collection of weights that describe the sound field in terms of arriving plane waves in the spherical harmonic domain at the center of the spherical receiver array. Interested reader may refer to Ref. 13 for more thorough derivation of the PWD for volumetric array. A total number of 7122 receiver positions are used for the PWD, and order of the spherical harmonic transform used for the PWD is 24. The implementation of the PWD is adapted from Ref. 14.

A loudspeaker array consisting of 37 loudspeakers, setup in an anechoic room is used for the reproduction of the stimuli. The loudspeaker configuration is described in Appendix B.

The stimulus is evaluated as follows: PWD is calculated from the simulated impulse responses. The PWD is then used to evaluate time-domain impulse responses at azimuth and elevation angles at the range $[-180, 180]$ and $[90, -90]$ deg, respectively, with 3 degree increments, referred to as *dense grid* here. These time-domain responses are then used to evaluate responses for each loudspeaker of the loudspeaker array by summing up the responses of the dense grid that are closest to the direction of the particular loudspeaker direction measured with the central angle. The impulse response for each loudspeaker is then applied to the sound sample via convolution and played back from the respective loudspeaker synchronously.

B. Subjects

Eight test subjects including the present author participated in the experiments. The present author was excluded from experiment 2 due to the explicit knowledge of the experiment conditions. All the participants work in the field of acoustics and have experience in listening tests.

C. Procedure

1. Experiment 1

Experiment 1 consisted of an adaptive staircase routine for measuring the threshold of the dispersion error. A *three alternative forced choice* (3AFC) procedure and QUEST¹⁵ method for determining the trial levels was utilized in the test. A probability level of $P = 0.82$ for correct discrimination was measured similarly to previous work.^{5,20} The trial procedure is as follows: three samples are presented to the subject in random order with 1 s interval between samples. Two of the played samples are the reference sample, and one is the stimulus. The subject is asked to specify which of the samples contained dispersion error, which in this case reduces to an “odd-one-out” discrimination. The experiment consisted of six separate staircase routines for each condition group {distance \times dispersion location}. The number of trials in each routine was limited to 30. The staircase routines were interleaved and split into two sessions. Subsequent trials were randomly picked from the routines selected for the session. An open source library for psychophysics experiments was used to implement the test.¹⁶

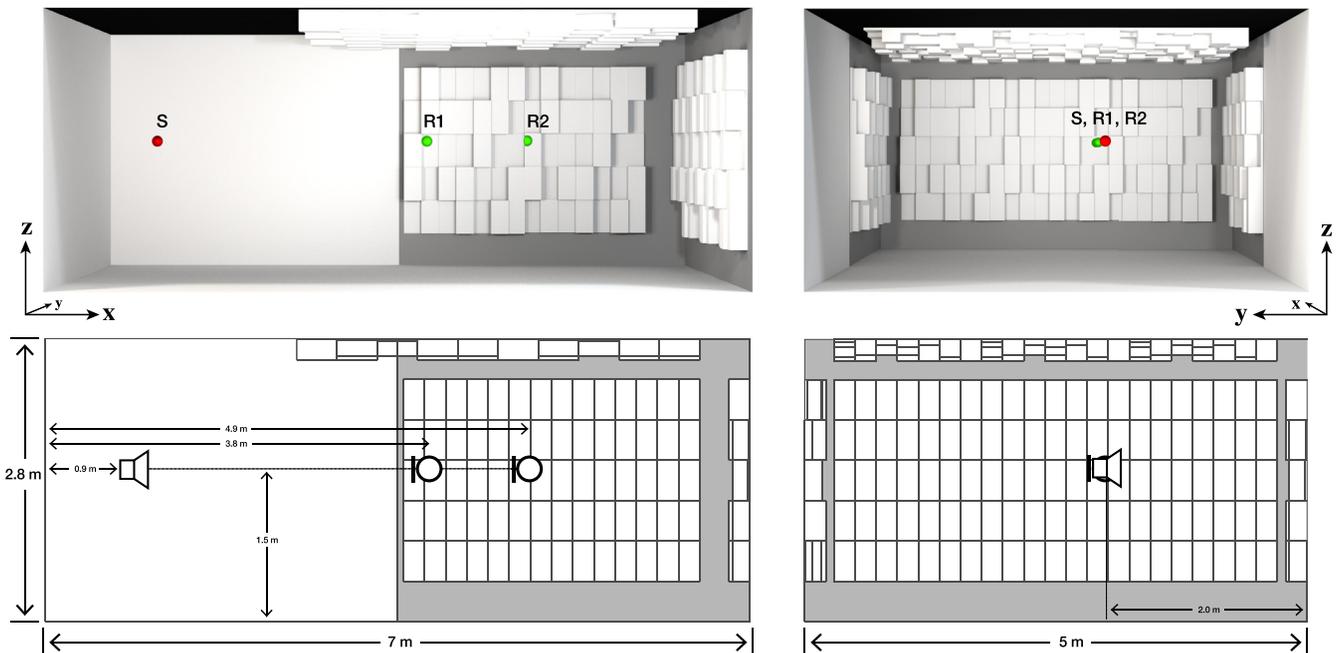


FIG. 2. (Color online) The simulation setup used in experiment 2. The top image is a render of the room with perspective projection. The bottom image is an illustration with orthographic projection from the same point of view as in the top image. The source position is indicated with “S” and the receiver positions with “R1” and “R2.”

Each participant carried out a training routine before the experiments. The training routine contained 10 similar 3AFC tasks than in the experiment with varying error levels with feedback for correct discrimination. No feedback was given to the participants in the actual experiment.

After the experiment, each participant was asked to write down what kind of artifacts did he/she concentrated on when discriminating the stimuli from the references, and report any changes in the artifacts during the experiment.

2. Experiment 2

A paired-comparisons routine was used in the second experiment. The conditions in the sampling frequency category (59 583, 119 165, 238 330 Hz) were compared pair-wise (three unique pairs) within the remaining condition groups (2 receiver distances \times 2 sound samples = 4 condition groups). Each pair was repeated five times. The experiment contained $3 \times 2 \times 2 \times 5 = 60$ comparisons in total. The condition and pair orders were randomized.

The participant was able to play back the two presented sampling frequency conditions in the trial as many times as needed to make the decision. The instruction for each participant was to choose the condition from the two options which contains more audible error. The criterion for the error was left for the participant to decide. This means that even if an audible difference between the conditions is present, it may not be evident to the participant which of the conditions contains more error as “odd-one-out” discrimination is not possible.

An introductory session where the participant was familiarized with the different error levels preceded the experiment. In the session, the participants could freely play back three different sound samples (a percussion excerpt, a

piano chord, and a click-like sample), and switch between three different room responses with the same sampling frequency conditions as in the experiment. The sampling frequency conditions were labeled according to the level of error (59 583 Hz: “Most error,” 119 165 Hz: “More error,” 238 330 Hz: “Error”). The room geometry used in the simulation of the responses in the introductory session was different to the actual experiment. The room model used in the introductory session was a small apartment, with reverberation time of 0.4 s. The change in the room condition for the introductory session was made to avoid the participants to memorize any specific dispersion artifacts related to the room condition.

After the experiment each participant was asked to write down a description of the differences he/she heard between the samples, and elaborate on what was perceived as an artifact in the stimuli while evaluating the pairs. Three categories were given in which to describe the differences and artifacts: *timbre*, *spatial image*, and *reverberation*.

V. EXPERIMENTS

A. Experiment 1

1. Results

The descriptive statistics of the measured phase velocity error percentage thresholds are presented in Table II, and visualized by condition group in Fig. 3. The phase velocity error percentage indicates the phase velocity error at 10 kHz.

The threshold measurement for the sound sample condition *speech* of one of the participants was removed from the results as threshold measurement did not converge in the condition group {*dispersion in the direct sound*, *speech*} with any distance condition (error level stayed at 99%).

TABLE II. Phase velocity error percentage thresholds for different conditions groups. The phase velocity error percentage is measured at 10 kHz. The last column shows the *group delay error* (Gd. err.) that occurs at 20 kHz with the measured threshold and distance condition. This value is comparable to previous work and therefore reported (Refs. 5 and 20). Dash (-) indicates that the error level does not have a real phase velocity at the frequency of 20 kHz.

Dist. (m)	Dispersion in	Sample	μ (%)	σ (%)	Min (%)	Max (%)	Gd. err. (ms)
2.34	Direct	Click	3.42	0.97	2.14	5.00	37.23
2.34	Reflection	Click	3.09	0.75	1.96	3.92	20.0
4.70	Direct	Click	1.93	0.49	1.15	2.52	5.87
4.70	Reflection	Click	1.44	0.31	1.00	1.84	6.18
9.39	Direct	Click	1.30	0.43	0.72	1.80	6.13
9.39	Reflection	Click	0.89	0.26	0.46	1.13	6.37
2.34	Direct	Speech	19.54	35.33	3.38	99.9	—
2.34	Reflection	Speech	15.37	8.9	6.52	33.00	—
4.70	Direct	Speech	9.07	15.93	2.36	45.19	—
4.70	Reflection	Speech	9.44	4.78	5.20	18.13	—
9.39	Direct	Speech	2.16	1.91	1.02	6.44	14.75
9.39	Reflection	Speech	4.50	1.94	3.13	8.36	—

Results of statistical tests conducted for the observations are presented in Table III. Shapiro–Wilk test for normality indicates that the null hypothesis for normality should be rejected for the category groups: {*click-dispersion in reflection—9.39 m*}, {*speech-dispersion in direct sound*}, and {*speech-dispersion in reflection—9.39 m*}.

The main effect to be measured in the observations was between the dispersion location condition (*dispersion in the direct sound, dispersion in the reflection*). Paired t-test was done for the observations made with the sound sample *click* within the dispersion location condition separately with each distance condition. The test indicates that the means of the observations are statistically significantly different for the distance conditions 4.7 and 9.39 m. The means of condition group *Dispersion in reflection* are lower than the condition *Dispersion in direct sound*, meaning the threshold is lower, and therefore the error is harder to discriminate in the condition group *Dispersion in direct sound*.

As the test of normality was rejected for all distances in category group {*speech-dispersion in direct sound*}, a non-parametric Wilcoxon signed rank test was performed for the observations made with sound sample *Speech* within the dispersion location condition separately for each distance condition. The results indicate that the null hypothesis that medians of the observations are the same cannot be rejected with any distance condition. The result is most likely dominated by the threshold measurements of one subject that are outside the interquartile range (IQR) in each distance condition with dispersion location condition *dispersion in direct sound*. Therefore it can be speculated that the main effect may exist regardless of the results of the signed rank test.

The results of the questionnaire indicated that most of the participants used high frequency artifacts for the discrimination with the sample condition *click*. Participants reported chirps, sweeps, and for more difficult trials change in timbre. For the sample condition *speech* the audible effects used for

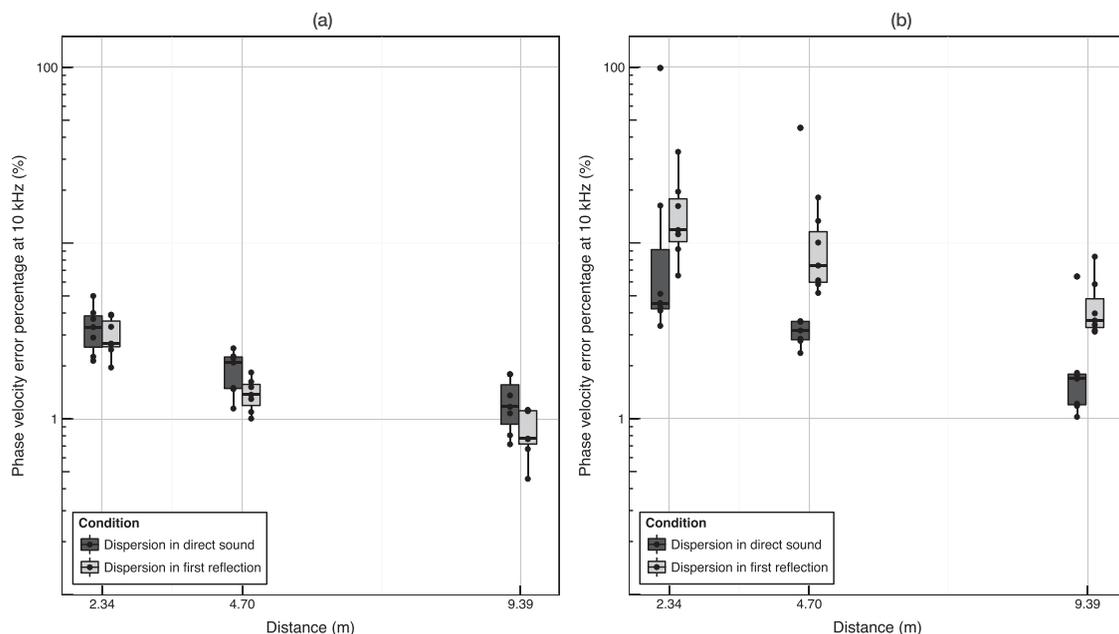


FIG. 3. The threshold observations of the two different sound sample conditions, (a) *click*, and (b) *speech*.

TABLE III. Results of statistical tests performed on the observation. *Within* indicates that the observations are compared within-subjects in the category. S–W: Shapiro–Wilk test for normality.

Dist. (m)	Dispersion in	Sample	Test	Results
2.34	Direct	<i>Click</i>	S–W	W = 0.956, p = 0.774
4.70	Direct	<i>Click</i>	S–W	W = 0.889, p = 0.230
9.39	Direct	<i>Click</i>	S–W	W = 0.905, p = 0.319
2.34	Reflection	<i>Click</i>	S–W	W = 0.890, p = 0.238
4.70	Reflection	<i>Click</i>	S–W	W = 0.948, p = 0.691
9.39	Reflection	<i>Click</i>	S–W	W = 0.837, p = 0.071 ^a
2.34	Direct	<i>Speech</i>	S–W	W = 0.536, p = 0.473e-4 ^a
4.70	Direct	<i>Speech</i>	S–W	W = 0.480, p = 0.916e-5 ^a
9.39	Direct	<i>Speech</i>	S–W	W = 0.605, p = 0.315e-3 ^a
2.34	Reflection	<i>Speech</i>	S–W	W = 0.871, p = 0.190
4.70	Reflection	<i>Speech</i>	S–W	W = 0.865, p = 0.166
9.39	Reflection	<i>Speech</i>	S–W	W = 0.762, p = 0.0167 ^a
2.34	<i>Within</i>	<i>Click</i>	t-test	t(7) = 1.232, p = 0.258
4.70	<i>Within</i>	<i>Click</i>	t-test	t(7) = 3.806, p = 0.00665 ^b
9.39	<i>Within</i>	<i>Click</i>	t-test	t(7) = 3.736, p = 0.00730 ^b
2.34	<i>Within</i>	<i>Speech</i>	W	V = 7, p = 0.297
4.70	<i>Within</i>	<i>Speech</i>	W	V = 7, p = 0.297
9.39	<i>Within</i>	<i>Speech</i>	W	V = 5, p = 0.156

^ap < 0.1.

^bp < 0.01.

discrimination were more varying. Many participants reported hearing high frequency artifacts, such as sweeps, distortion, “weird phase effects,” or “robotic” sound. Some participants reported that for some cases the distortion was more audible in a certain direction. One participant reported that the artifact was more audible on the right side. Discussions with several participants confirmed that the artifact seemed to be localized in a different direction than the speech itself in some cases.

The sampling frequency conditions for the experiment 2 are motivated by the measured threshold values. The lowest threshold 0.89% of phase velocity error at the frequency of 10 kHz with condition group {*click-dispersion in reflection—9.39 m*} is achieved with a sampling frequency of 194 750 Hz. The highest threshold of 19.54% of phase velocity error at 10 kHz is measured with condition group {*speech-dispersion in direct—2.34 m*}, and results in a sampling frequency of 54 250 Hz. The third condition is then picked between these two as an approximate average between the lowest threshold in the sample category *speech* (2.16%), and the highest threshold in the sample category *click* (3.42%), the average being 2.79%. The sampling frequency that is needed to achieve this error level is 113 050 Hz. The sampling frequencies are truncated so that they result in an element size that is a fraction of centimeter so that the geometry is presented accurately between conditions. The element sizes then are 1, 0.5, and 0.25 cm, that result in the used sampling frequency conditions 59 583, 119 165, 238 330 Hz, respectively.

2. Discussion

The test hypothesis in the experiment was that the threshold for dispersion error level is higher in a case where

the worst-case direction of the dispersion error is a reflection path instead of the direct path from the source. The results indicate that the hypothesis should be rejected with the sound sample *click* and distance conditions 4.7 and 9.39 m. The observed effect is actually the opposite: the threshold for dispersion error level is lower when the worst-case direction is a reflection path. This can be explained with the fact that the click-like signal is very short, and subsequently two auditory events occur especially with the longer delays. The reflection path contains more dispersion as it is always longer in distance, and therefore more group delay is accumulated in this path. The result may therefore reflect the total amount of group delay.

For the sound sample *speech*, an effect supporting the hypothesis seems to exist when visually inspecting the observations in Fig. 3. The box plot shows that the 50% IQR are not overlapping between the dispersion location condition in any distance conditions. Regardless of this, the statistical tests indicate that the medians of the groups are not significantly different. Therefore the results may be considered inconclusive regarding the test hypothesis.

The threshold values measured with the distance condition 9.39 m are higher (3.90% phase velocity error at 20 kHz) than the measured threshold for only the direct sound⁵ (2.0% phase velocity error at 20 kHz at the distance of 9.1 m). Therefore there is evidence that clean version of the signal in either, direct sound or reflection, does partially mask the group delay error.

B. Experiment 2

1. Results

The binomial data of the paired comparison task is transferred into a ratio scale with the *Bradley–Terry–Luce* (BTL) model¹⁷ using the implementation of Wickelmaier and Schmid¹⁸ The analysis was done to the transpose of the comparison matrix to attain rating on which condition would contain least error (task in the paired-comparison was to choose the stimuli which contained more error). The results of the comparison task are presented in Fig. 4 for the different sampling frequency conditions. The results indicate that the lowest sampling frequency (59 583 Hz) is perceived to have most error almost unanimously with all conditions. The highest sampling frequency condition (239 330 Hz) is the most probable to be chosen as having least error. The condition (119 165 Hz) has relatively low probability to be chosen to have least error. The condition group {*Speech, receiver 1*} has slightly higher probability to be chosen as the condition with the lowest error in comparison to other groups.

The results of the questionnaire indicated that most participants used change in timbre as the main attribute to evaluate the amount of error in the stimuli. For the sample condition *click*, participants reported high frequency chirps and sweeps. For the sample condition *speech*, participants reported that they evaluated sharper sibilants as having more error. Additionally, increased high frequency content in the stimuli was reported as indicating more error. Some participants reported that the artifacts for certain pairs were perceived more diffuse, and the spatial image was not as

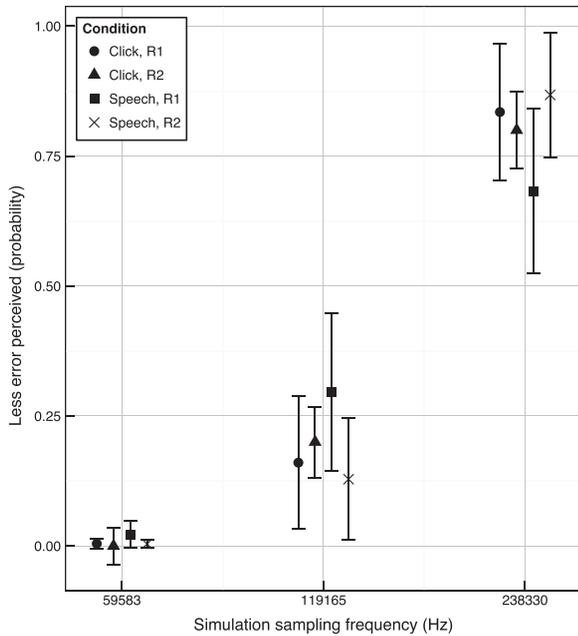


FIG. 4. The results of the BTL analysis of the paired comparison data.

“narrow” as with the stimulus that was evaluated as having less error. Few participants reported that the stimuli evaluated as having more error had more reverberation, and that the reverberation had a color change. On the contrary one participant reported that the stimuli evaluated as having more error was less reverberant.

2. Discussion

The results of the experiment 2 give evidence that the audibility of the error with a simulated room response with prominent early reflections, diffuse reverberation, and reverberation time of approximately 0.6 s, reduces as the sampling rate conditions is increased. There is no evidence that that error difference would be unnoticeable between any of the sampling rate conditions. If the error would be unnoticeable between all of the conditions, the probabilities should be equal at 0.3. The lowest sampling frequency is practically always chosen to have most error. If the condition 59 583 Hz is excluded and the two remaining conditions are compared, rating should be 0.5 in the case that the error is unnoticeable between them. The probability of choosing the sampling frequency condition 119 165 Hz as having least error is fairly low, having maximum value with the condition group {speech, R1} at $p = 0.27$. Additional experiments are needed to quantify whether the perceived error would further decrease by increasing the sampling frequency above the highest condition value of 238 330 Hz.

Several notions should be made when comparing the results of experiment 1 and experiment 2. The sampling frequency condition 119 165 Hz was scaled close to the lowest measured threshold made with the sample condition speech. From the results of experiment 2, it is evident that the decrease of error is audible when the sampling frequency condition is changed to 238 330 Hz with all conditions. Therefore it may be speculated that the threshold with the

distance condition 9.39 m does not reflect the audibility of the error in the room response with the sound sample speech.

In the case of the sound sample click, such conclusion may not be speculated, as the highest sampling frequency condition 238 330 Hz corresponds to the lowest measured threshold in experiment 1. Therefore an effect whether or not a higher sampling frequency makes the error harder to perceive cannot be observed, and a possible saturation point may lie higher, or possibly at the condition 238 330 Hz.

VI. CONCLUSIONS

A test for the audibility of dispersion error in the presence of single early reflection was conducted. The results of the experiment indicated that for short transient sample the dispersion error is easier to perceive when the error is located in the reflection in comparison to the condition where the dispersion is in the direct sound. The results give evidence that in such a case the total amount of group delay error in the signal defines the audibility, as more group delay error is accumulated in the reflection path. For a speech signal, some evidence was found that the effect is the opposite; the error is harder to perceive when it is located in the reflection path.

A follow-up experiment from the results of the first experiment was made to quantify whether the measured error levels reflect the audibility of the error in a short room response. The results indicate that the participants could evaluate the error levels correctly with high confidence. It may be concluded that the measured error levels thresholds are not low enough to make the error imperceptible in a short room response for speech signal. For a click-like sample, the results indicate that the least error is perceived in the highest sampling frequency condition that corresponds to the lowest measured threshold, but a conclusion is this high-enough sampling frequency cannot be drawn.

ACKNOWLEDGMENTS

Sebastian Prepelita is acknowledged for commenting the manuscript. This work has received funding from Academy of Finland, Project No. 265824.

APPENDIX A: AIR ABSORPTION FILTER

Following the formulation of Bass *et al.*¹⁹ absorption of air can be approximated with

$$\hat{\alpha}(F) = F^2 \frac{p_s}{p_{s0}} \left\{ 1.84 \times 10^{-11} \left(\frac{T}{T_0} \right)^{1/2} + \left(\frac{T}{T_0} \right)^{-5/2} \times \left[0.01278 \frac{e^{-2239.1/T}}{F_{r,O} + F^2/F_{r,O}} + 0.1068 \frac{e^{-3352/T}}{F_{r,N} + F^2/F_{r,N}} \right] \right\}, \quad (A1)$$

where $F = f/p_s$, $F_{r,O} = f_{r,O}/p_s$, $F_{r,N} = f_{r,N}/p_s$. $f_{r,O}$ and $f_{r,N}$ are relaxation frequencies for oxygen and nitrogen, respectively, and f is frequency in Hz. The absorption $\hat{\alpha}$ has unit nepers/m. Formulations for the relaxation frequencies may be found from Ref. 19 and are omitted here.

TABLE IV. Loudspeaker directions for the setup in the anechoic chamber.

Elevation	Azimuth
90°	0°
45°	0°, ±90°, 180°
22°	0°, ±30°, ±55°
0°	0°, ±15°, ±30°, ±45°, ±60°, ±75°, ±90°, ±105°, ±120°, ±135°, ±150°, 180°
-22°	0°, ±30°
-45°	0°, ±90°, 180°

The absorption for a given distance d can be evaluated from Eq. (A1) with

$$H(f) = e^{-d\hat{\alpha}(f)} \times e^{-j2\pi f t}, \quad (\text{A2})$$

where $t=0$, and f is frequency in Hz. The absorption is applied to the simulated impulse response in overlapping time windows. Window length of 128 samples, 50% overlap and Hanning windowing is used. The distance d is evaluated from the beginning of the window with $sample\ index/ fs \times c$, where c is the speed of sound (taken 344 m/s here), and fs the sampling frequency.

APPENDIX B: LOUDSPEAKER CONFIGURATION

The reproduction system consisted of 37 loudspeakers (Genelec 8030B) located at 2.2 m distance from the listening position. The loudspeaker positions are listed in Table IV. Elevation angles are $[-90^\circ, 90^\circ]$ starting from below the listening position and ending directly above the listening position, 0° being parallel to the horizontal plane. Azimuth angles are $[-180^\circ, 180^\circ]$, measured from left, behind the listener to right behind the listener, 0° being directly in front of the listening position.

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

²B. Hamilton, "Finite difference and finite volume methods for wave-based modelling of room acoustics," Ph.D. thesis, The University of Edinburgh, 2016.

³M. Cobos, J. Escolano, J. J. López, and B. Pueo, "Subjective effects of dispersion in the simulation of room acoustics using digital waveguide mesh," in *124th Convention of Audio Engineering Society*, Amsterdam, Netherlands (2008), Paper No. 7471.

⁴A. Southern, D. Murphy, T. Lokki, and L. Savioja, "The perceptual effects of dispersion error on room acoustic model auralization," in *Proc. Forum Acusticum*, Aalborg, Denmark (2011), pp. 1553–1558.

⁵J. Saarelma, J. Botts, B. Hamilton, and L. Savioja, "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 (2016).

⁶H. Wallach, E. Newman, and J. Rosenzweig, "The precedence effect in sound localization," *Am. J. Psychol.* **62**, 315–336 (1949).

⁷R. Y. Litovsky, H. S. Colburn, W. A. Yost, and S. J. Guzman, "The precedence effect," *J. Acoust. Soc. Am.* **106**(4) 1633–1654 (1999).

⁸J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, 1st ed. (MIT, Cambridge, MA, 1983).

⁹A. Taflov, *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, 2nd ed. (Artech House, Boston, 2000).

¹⁰J. Saarelma and L. Savioja, "An open source finite-difference time-domain solver for room acoustics using graphics processing units," in *Proc. Forum Acusticum*, Krakow, Poland (2014).

¹¹C. Webb and S. Bilbao, "Computing room acoustics with CUDA-3D FDTD schemes with boundary losses and viscosity," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process* (2011), pp. 317–320.

¹²T. Cox and P. d'Antonio, *Acoustic Absorbers and Diffusers: Theory, Design and Application*, 2nd ed. (Taylor & Francis, New York, 2009).

¹³J. Sheaffer, M. Van Walstijn, B. Rafaely, and K. Kowalczyk, "Binaural reproduction of finite difference simulations using spherical array processing," *IEEE Trans. Audio, Speech, Lang. Process.* **23**(12), 2125–2135 (2015).

¹⁴J. Sheaffer and B. Fazenda, "Wavecloud: An open source room acoustics simulator using the finite difference time domain method," in *Proc. Forum Acusticum*, Krakow, Poland (2014).

¹⁵A. B. Watson and D. G. Pelli, "Quest: A Bayesian adaptive psychometric method," *Percept. Psychophys.* **33**(2), 113–120 (1983).

¹⁶J. W. Peirce, "PsychoPy Psychophysics software in Python," *J. Neurosci. Methods* **162**(1), 8–13 (2007).

¹⁷R. A. Bradley and M. E. Terry, "Rank analysis of incomplete block designs: I. The method of paired comparisons," *Biometrika* **39**(3/4), 324–345 (1952).

¹⁸F. Wickelmaier and C. Schmid, "A Matlab function to estimate choice model parameters from paired-comparison data," *Behav. Res. Methods, Instrum., Comput.* **36**(1), 29–40 (2004).

¹⁹H. 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).

²⁰J. Saarelma and L. Savioja, "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 (2016).