

Psychoacoustic Characterization of Propagation Effects in Virtual Environments

ATUL RUNGTA, SARAH RUST, and NICOLAS MORALES, University of North Carolina at Chapel Hill
ROBERTA KLATZKY, Carnegie Mellon University
MING LIN and DINESH MANOCHA, University of North Carolina at Chapel Hill

As sound propagation algorithms become faster and more accurate, the question arises as to whether the additional efforts to improve fidelity actually offer perceptual benefits over existing techniques. Could environmental sound effects go the way of music, where lower-fidelity compressed versions are actually favored by listeners? Here we address this issue with two acoustic phenomena that are known to have perceptual effects on humans and that, accordingly, might be expected to heighten their experience with simulated environments. We present two studies comparing listeners' perceptual response to both accurate and approximate algorithms simulating two key acoustic effects: diffraction and reverberation. For each effect, we evaluate whether increased numerical accuracy of a propagation algorithm translates into increased perceptual differentiation in interactive virtual environments. Our results suggest that auditory perception does benefit from the increased accuracy, with subjects showing better perceptual differentiation when experiencing the more accurate rendering method: the diffraction experiment shows a more linearly decaying sound field (with respect to the diffraction angle) for the accurate diffraction method, whereas the reverberation experiment shows that more accurate reverberation, after modest user experience, results in near-logarithmic response to increasing room volume.

CCS Concepts: • **Computing methodologies** → **Simulation evaluation**;

Additional Key Words and Phrases: Auditory perception, virtual environments/reality

ACM Reference Format:

Atul Rungta, Sarah Rust, Nicolas Morales, Roberta Klatzky, Ming Lin, and Dinesh Manocha. 2016. Psychoacoustic characterization of propagation effects in virtual environments. *ACM Trans. Appl. Percept.* 13, 4, Article 21 (July 2016), 18 pages.
DOI: <http://dx.doi.org/10.1145/2947508>

1. INTRODUCTION

Sound plays a vital role in increasing the degree of realism in virtual environment (VE) systems [Begault 1994; Larsson et al. 2002] and other interactive applications, such as video games. This observation has motivated the development of different sound propagation methods that are used to simulate how sound waves, emitted from a source, travel through an environment and interact with the objects before reaching a listener. These methods are used to model well-known acoustic phenomena, such as diffraction, reverberation (comprising early reflections and late reverberation), and scattering.

This work was supported in part by NSF grants 1320644, 1456299, and 1518630.

Authors' addresses: A. Rungta; email: rungta@cs.unc.edu; S. Rust; email: serust@live.unc.edu; N. Morales; email: nmorales@cs.unc.edu; R. Klatzky; email: klatzky@cmu.edu; M. Lin; email: lin@cs.unc.edu; D. Manocha; email: dm@cs.unc.edu.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.

© 2016 ACM 1544-3558/2016/07-ART21 \$15.00

DOI: <http://dx.doi.org/10.1145/2947508>

At a broad level, sound propagation methods are categorized into geometric [Krokstad et al. 1968; Borish 1984; Allen and Berkley 1979] and wave-based [Zienkiewicz 2005; Yee 1966; Cheng and Cheng 2005] algorithms. Although these computational techniques have been studied for many decades in different fields, only recent advancements in terms of new algorithms and fast hardware have enabled the development of interactive propagation systems that are useful for VE. These include interactive geometric methods based on ray tracing and beam tracing [Schissler et al. 2014; Tsingos et al. 2001; Taylor et al. 2009; Funkhouser et al. 2004] that can simulate approximate diffraction, early reflections, and high-order late reverberation. Furthermore, advancements in computational acoustics solvers [Mehra et al. 2013; Raghuvanshi et al. 2009; Webb and Gray 2013; Mehra et al. 2015] have made it possible to compute highly accurate solutions to the wave equation for large domains, and thereby perform interactive sound propagation.

Given the recent developments in interactive sound propagation algorithms, it is imperative to evaluate their perceptual effectiveness. Psychoacoustics researchers have focused on evaluating the perceptual effects of many of these acoustic phenomena, and many important results have been published on how different propagation phenomena affect our perception of the environment [Fastl and Zwicker 2007]. Most of these studies were conducted in either real-world environments or in very simple VEs that could simulate only limited acoustic effects. The advent of interactive and accurate sound propagation techniques makes the task of perceptually evaluating these phenomena simpler and less expensive.

In this article, we mainly focus on two of the aforementioned phenomena: diffraction and late reverberation (hereafter referred to as reverberation). Early reflections were not considered here, as they are easy to simulate and have been widely studied in the literature [Haas 1951; Djelani and Blauert 2001]. Reverberation is also a well-studied phenomenon in psychoacoustics and enhances immersion [Kuttruff 2007]. Despite this, one of its fundamental effects, namely to convey the size of the environment, remains relatively unexamined. Reverberation is typically approximated using artificial filters, and such filters are widely used in computer games and VE [Jot and Chaigne 1991]. Therefore, we evaluate the reverberation computed in a physically accurate manner using an interactive, state-of-the-art geometric propagation system to a precomputed (Schroeder-type) filter and evaluate their relative effectiveness in telling us the size of an environment. Perceptual effects of diffraction, although important [Torres et al. 2001b], are seldom evaluated in VEs primarily due to the complexity of modeling diffraction. Most interactive geometric sound propagation systems approximate edge diffraction based on the uniform theory of diffraction (UTD) [Tsingos et al. 2001]. Therefore, we evaluate the perceptual performance of diffraction effects computed using UTD with a numerically accurate solver that directly solves the wave equation. We report two separate comparative studies to evaluate whether increased numerical accuracy of sound propagation translates to perceptual differentiation.

In our diffraction study, we construct a virtual test scene similar to Kawai [1981] and perform a psychoacoustical evaluation. We evaluate the diffracted sound field around an obstacle by placing subjects along a semicircle for the two methods: UTD and wave based. The subjects are asked to rate the perceived loudness for different positions along the semicircle. Our results show that wave-based diffraction results in a diffracted field that decays nearly linearly with an increasing diffraction angle compared to UTD-based diffraction, which shows erratic behavior.

Prior psychoacoustics studies in reverberation normally focus on evaluating how the human response to the environment varies with the changing reverberation parameters (e.g., RT_{60}). Our study builds on these evaluations and seeks to compare the effectiveness of two competing methods to model reverberation: Schroeder parametric filter and physically accurate path tracing. We use free-magnitude estimation to compute the magnitude of the internal responsiveness of the subjects to the change in the physical dimensions of the environment, and thereby its reverberation characteristics.

Our results show poor size discrimination for both methods but also show increasing discrimination ability for the physically accurate one as the subjects become more familiar with the task.

The rest of the article is organized as follows. In Section 2, we give a brief overview of the two acoustic effects and the corresponding techniques and algorithms used in the study to model them and their evaluation. Sections 3 and 4 describe the two studies, including their designs and procedures. Section 5 offers a discussion on our observed results for the two studies. In Section 6, we discuss our conclusion, limitations, and the future work.

2. BACKGROUND

In this section, we give an overview of the two sound phenomena, the methods used to model them, and the related work in psychoacoustic evaluation of these phenomena.

2.1 Diffraction

Diffraction refers to the bending of a wave around an obstacle when the obstacle's dimensions are comparable to the wavelength of the wave. Diffraction helps a listener hear sounds not in the line of sight and is observed in everyday life [Tsingos et al. 2001]. Diffraction is explained through the wave equation; hence, wave-based methods emulate it automatically. Geometric methods, on the other hand, assume rectilinear propagation of sound. As these methods do not account for the bending of sound rays around an obstacle, this effect must be incorporated separately. Doing so is difficult and computationally expensive, which is why most VEs avoid incorporating diffraction even though the theories for approximating its effects have existed for decades [Biot and Tolstoy 1957; Kouyoumjian and Pathak 1974]. With the advent of better algorithms and fast hardware, real-time diffraction is now used for interactive geometric propagation [Tsingos et al. 2001; Schissler et al. 2014].

2.2 Reverberation

Reverberation forms the later part of the impulse response (IR) within closed environments. It is caused by successive reflections or “echoes” as they diminish in intensity. Reverberation forms a critical part of the acoustics of an environment and directly correlates with the size and the clarity of sound in the environment. For these reasons, reverberation plays an important role in architectural acoustics, especially while constructing auditoriums and concert halls. This has led to considerable research in characterizing reverberation, and several parameters, such as the reverberation time (RT_{60}) and clarity index (C_{50} and C_{80}), have been formulated [Kuttruff 2007]. We concern ourselves with RT_{60} in this study. RT_{60} is defined as the time it takes for the sound to decay by 60dB.

Given the importance of reverberation to architectural/room acoustics, and complexity of simulating it from physical principles, empirical methodologies have been developed to artificially simulate reverberation in virtual acoustics such as digital and convolution reverberators. These reverberators are parameterized using several values, the most important of which is the RT_{60} . We use a digital Schroeder filter for our evaluation, as these filters are the most common form of digital reverberators in use today. A well-known empirical formula used to estimate RT_{60} is Sabine's equation, which gives the relationship between the RT_{60} of a room, its volume, and the total absorption by

$$RT_{60} \approx 0.1611sm^{-1} \frac{V}{Sa}, \quad (1)$$

where V is the total volume of the room in m^3 , S is total surface area in m^2 , a is the average absorption coefficient of the room surfaces, and Sa is the total absorption in sabins. Another more accurate way to estimate the RT_{60} of a room is to look at the room's impulse response (RIR) and directly compute the time it took for the sound to decay by 60dB as specified in ISO 3382-1:2009. This method involves

a reverse cumulative trapezoidal integration to estimate the decay of the IR and using a linear least-squares fit to estimate the slope between 0dB and -60dB. This is the method we used to compute the RT_{60} for our digital reverberator.

2.3 Geometric Acoustics

Geometric acoustics methods work under the assumption that the wavelength of sound is much smaller (i.e., higher frequency) than the objects in the scene. This assumption allows these methods to assume that sound waves travel in straight lines and thereby use ray tracing [Vorländer 1989; Krokstad et al. 1968; Taylor et al. 2012; Schissler and Manocha 2016] and its variants, such as beam tracing [Funkhouser et al. 2004] and frustum tracing [Lauterbach et al. 2007; Chandak et al. 2008]. Other methods such as image sources [Allen and Berkley 1979; Borish 1984; Antani et al. 2012] have also been employed. To approximate diffraction, geometric methods use formulations based on UTD or Biot-Tolstoy-Medwin (BTM). We focus on UTD-based diffraction, as that has been used for interactive applications.

2.3.1 Uniform Theory of Diffraction. UTD [Kouyoumjian and Pathak 1974] is a high-frequency approximation of the phenomenon of diffraction. These methods were initially developed for the propagation of light but later used for sound [Tsingos et al. 2001]. UTD assumes that a diffracting edge is of infinite length and acts as a secondary sound source. Another assumption made by UTD is that the source and listener are far away compared to the wavelength of sound. According to UTD, an incoming sound ray hitting an infinite wedge results in a cone of diffracted rays, and a single ray with the shortest distance to the listener (in a homogeneous medium) forms the diffracted field.

2.4 Numerical or Wave-Based Acoustics

Sound propagation is governed by the acoustic wave equation (in time domain):

$$\nabla^2 p - \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} = F(\mathbf{x}, t) \quad \mathbf{x} \in \Omega, \quad (2)$$

where ∇^2 is the Laplacian, p is pressure, $c = 343\text{ms}^{-1}$ is the speed of sound, $F(\mathbf{x}, t)$ is forcing term corresponding to the source, and Ω is the domain of interest.

Solving Equation (2) gives us the sound pressure P at any point in the domain. Unfortunately, closed-form solutions to the wave equation only exist for the simplest of domains, and most solvers use numerical techniques. However, the complexity of these methods increases as a fourth power of the frequency.

2.4.1 Adaptive Rectangular Decomposition. This technique was developed by Raghuvanshi et al. [2009] and constitutes a domain-decomposition technique to solve the wave equation (Equation (2)) in homogeneous media. The underlying principle of this technique is based on the observation that the wave equation can be solved analytically inside a rectangular domain. Therefore, this technique first decomposes the domain (scene) into a set of connected rectangles and computes the pressure in each of those rectangles analytically. The pressure is then transferred between rectangles using a finite-difference stencil. For more details on the technique, refer to Raghuvanshi et al. [2009], Mehra et al. [2012], and Morales et al. [2015].

2.5 Related Work

Reverberation and its effects have been widely studied in psychoacoustics. Zannini et al. [2011] explored the effect of source localization in reverberant conditions and found that the localization

deteriorates with changing reverberation times. Hartmann [1983] found that localization accuracy decreases in a reflecting room compared to the same absorbing room. Rakerd and Hartmann [1985] showed that reverberation has a considerable effect on localization even at low frequencies. Giguère and Abel [1993] found that localization was poorer in reverberant environments compared to absorbing ones. Other studies involving reverberation have analyzed the impact of reverberation on speech clarity. Knudsen [1932] showed that the presence of reverberation reduces the number of sounds that are heard correctly. Galster [2007] revealed the relation between the speech transmission index (STI) as a function of signal-to-noise ratio and the reverberant conditions and found that speech intelligibility decreases significantly in small, highly reflective rooms. Although reverberation decreases localization accuracy, it is known to have positive effects with respect to externalization. Externalization is defined as the perception of the sound source emanating from a point in the world, in contrast to internalization, where the sound appears to be emanating from within one's own head. Begault [1992] conducted studies showing that spatial reverberation increased externalization. Our reverberation experiment borrows from this work in that we use synthetic reverberation and construct IRs as they did in their work. Another important perceptual effect of reverberation is that of size estimation. Hameed et al. [2004] conducted studies to assess the effect of two reverberation parameters—the reverberation time RT_{60} and the direct-to-reverberant energy ratio (D/R) ratio—and found that RT_{60} is the most important parameter in room size estimation. Cabrera et al. [2005] performed experiments using real and virtual rooms and investigated the role of various parameters, including the reverberation time. They found that reverberation strongly affects the room size perception. Pop and Cabrera [2005] tested the relationship between acoustical characteristics of a room and its perceived size for three real rooms. They found a negative relation between sound level and room size.

Most studies on diffraction have been numerical in nature [Kawai 1981; Chu et al. 2007]. Few psychoacoustical studies have been performed on diffraction. Torres et al. [2001b] conducted a study investigating the audibility of diffracted sound for a simple open geometry and found that diffracted sound is audible in nonshadow regions. Torres et al. [2001a] concluded that reflected-diffracted combinations are significant and audible even for minute spectral changes. Torres and Kleiner [1998] performed listening tests to compute the audibility of edge diffraction in a stage house and found that first-order diffraction is significantly more audible than second-order diffraction. Mehra et al. [2015] conducted experiments to evaluate the subject localization performance using wave-based and geometric propagation and found that diffraction played a part in helping the subjects localize faster. However, this study did not consider diffraction as a separate phenomenon from other acoustic effects.

In this article, we consider the psychoacoustical evaluation of diffraction and reverberation in a VE. In the case of diffraction, we consider how the diffracted sound field changes with respect to the diffraction angle for two methods of diffraction: UTD based and wave based. In the case of reverberation, we consider how two competing methods of rendering late reverberation perform with respect to the size estimation of various rooms of increasing volumes. Although reverberation and its associated perceptual effects have been studied, size estimation remains relatively unexplored, motivating us to consider this particular phenomenon in our experiments.

3. EXPERIMENT 1: DIFFRACTION STUDY

3.1 Participants

Sixteen subjects participated in this study with informed consent. The ages ranged from 22 to 28 years (mean = 24.6 with SD = 1.4, 2 females and 14 males). The participants were recruited from the staff and students at a university campus. All participants reported normal hearing.

3.2 Apparatus

The setup consisted of a Dell T7600 workstation with the sound delivered through a pair of Beyerdynamic DT990 PRO headphones. The subjects wore a blindfold. The software consisted of in-house code to auralize precomputed IRs. All code was written in C++.

3.3 Stimuli

The source consisted of a prerecorded sound of a ringing bell. Since diffraction is a frequency-dependent phenomenon and more prominent at low frequencies, the bell clip was low-pass filtered with a cut-off frequency of 300Hz. The choice of frequency range was motivated by two considerations:

- The edge-diffraction phenomenon is most prominent in the 20 to 350 Hz range (Torres et al. [2001b]),
- The loudness characteristic of sound is least complex (near linear) with respect to frequency in the frequency range chosen by us (see Figure 2).

The subjects were placed at a radius of 5m from the sound source along a semicircle in our virtual scene covering two orders of diffraction as shown in Figure 1. The scene was an open scene making sure that no reflected sound reached the subjects. The semicircle was sampled at 10 degrees giving 18 positions, plus 1 for the direct sound region, for a total of 19 positions. Each trial consisted of the subject being placed at one these 19 positions randomly with the audio either auralized using the UTD-based or the wave-based diffraction, the order of which was also randomized. The IR for each of these 19 positions and the two methods of diffraction were precomputed and stored on file. The stimuli consisted of 38 combinations (19 positions \times 2 diffraction types) of convolved source and IR signals.

3.4 Design and Procedure

We designed a virtual environment to ensure that only the diffracted sound reached the subject. The environment was effectively open, the ground was perfectly absorptive, and the source was not kept in the line of sight (see Figure 1). Ideally, we would need to have an infinitely high obstacle and an infinitely long barrier as shown in Figure 1 to make sure that the sound reaching the listener is the one that diffracts around the lateral plane of the obstacle. Otherwise, the sound would diffract around the top of the obstacle and the edge of the barrier. These dimensions can obviously not be achieved in practice, and to *simulate* an open environment with seemingly infinite features, we enclosed the scene with a perfectly absorptive box. The box was 100m \times 12m \times 100m.

Another advantage of having an enclosure on the scene is that it acts as an infinitely far away region for wave-based methods. Wave-based methods inherently cannot handle open scenes and need an artificial boundary to *act* as infinity. To ensure that this artificial infinity does not reflect back into the scene, an absorbing boundary condition called the *perfectly matched layer* (PML) [Berenger 1994] is used at the boundaries. To add the PML boundary, the aforementioned bounding box is enclosed within a slightly bigger bounding box. The space between these boxes acts as the absorbing layer.

This makes the environment effectively open with respect to sound propagation and would give the same result as in a truly open, similar environment. Moreover, the environment also ensured that computing the propagated sound field with a wave-based propagation system was computationally tractable.

The barrier was added to one side of the cuboid to ensure that no sound could reach the subject from the other side (see Figure 1). In addition, the barrier was perfectly absorbing to avoid spurious reflections in the scene. We tested up to two orders of diffraction. It should be noted that the concept of diffraction “order” does not apply to wave-based methods but does in the case of geometric methods. An order of diffraction is defined as the number of edges that sound has to diffract around to reach the

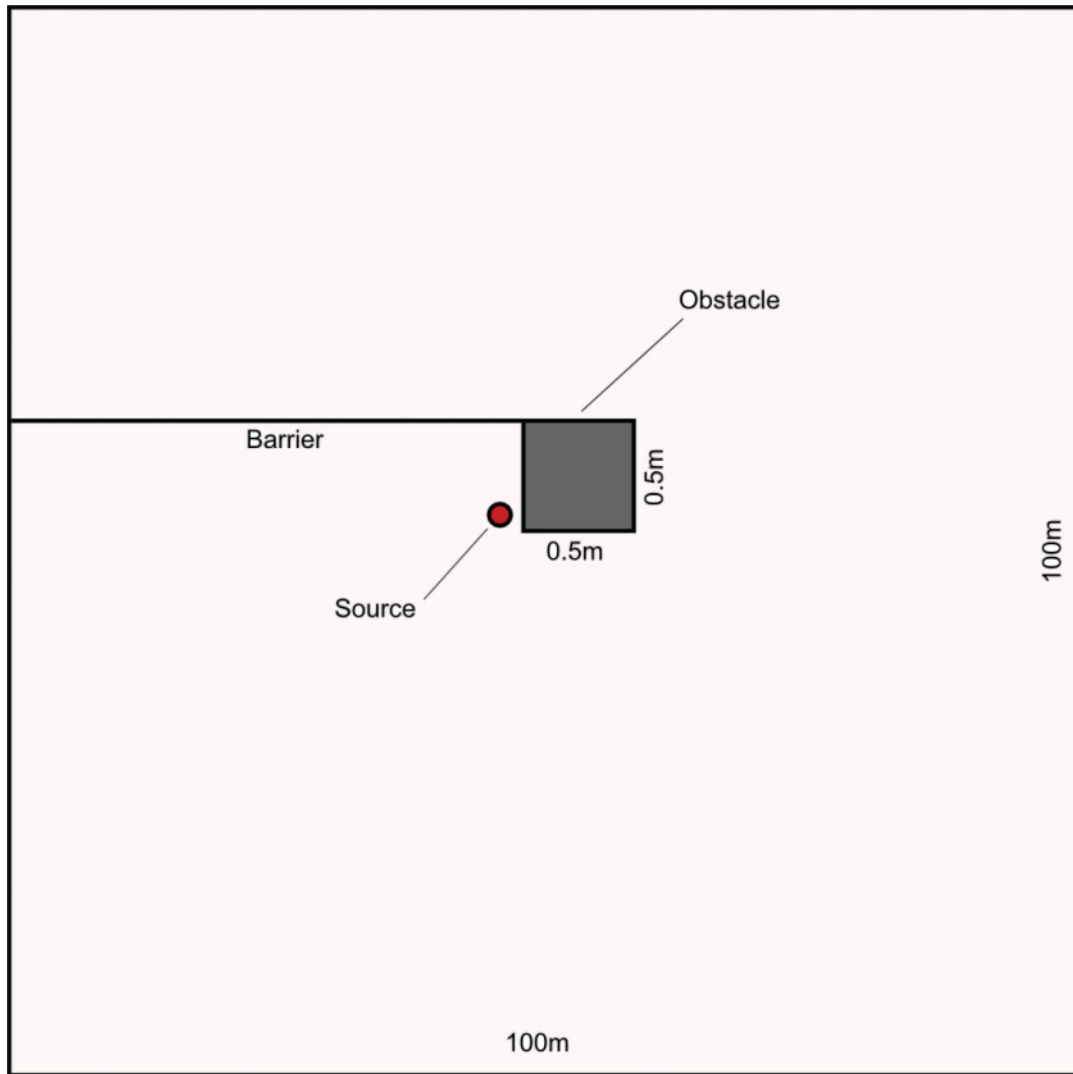


Fig. 1. The figure shows the setting used for characterizing diffraction. The obstacle was $0.5\text{m} \times 0.5\text{m}$. The solid barrier prevents the sound from reaching the listener from behind the obstacle. The barrier and the obstacle are fully absorptive to prevent any reflections and make sure that only the diffracted sound reached the listener. The figure is not drawn to scale for better viewing of the experimental setup.

listener. Typically, the cost of geometric diffraction methods such as UTD increases exponentially with increasing order.

This was a within-subject study with the same participants for each method of diffraction. The subjects wore a blindfold, and the audio was monaural sound delivered using headphones. The head orientation was kept fixed with the subjects facing the source, as we were only interested in the perceived loudness measure. The absence of variable head orientation meant that we could disregard the spatial dimensions of the head, and hence no spatial or binaural rendering was done. Before starting

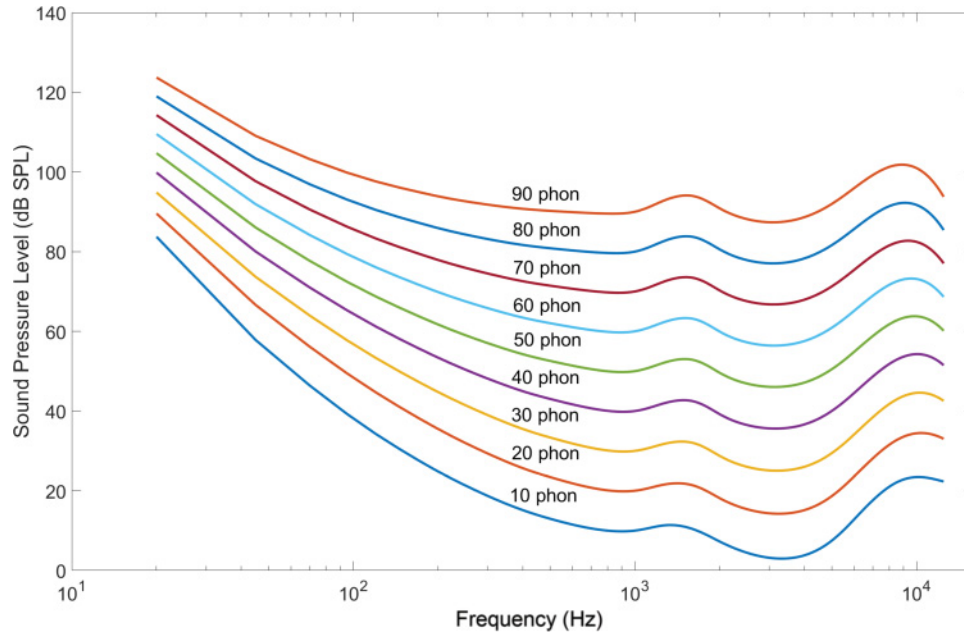


Fig. 2. ISO equal-loudness contours with frequency in hertz (from ISO 226:2003).

the experiments, subjects were played a sample of the sound source to ensure that they were familiar with it.

A total of 16 participants took part in each group. For each of the 19 positions, the subjects were asked to rate the loudness of the sound heard on an arbitrary, nonphysical scale ranging from 1 to 20. The loudness scale was explained before the start of the experiment. The extrema of the scale were compared to a verbal standard: 1 was the loudness of a falling leaf, whereas 20 was the loudness of someone shouting nearby. The sounds for the two methods were level matched by matching the root mean square (RMS) of the signals in the direct sound region.

A block consisted of 38 trials (19 positions \times 2 diffraction methods). There were three blocks per subject, giving a total of 114 readings. The virtual placement of the subjects was controlled by the experimenters, who pressed a key that placed the subjects at a random position along the semicircle and randomly chose one of two diffraction methods. The subjects were allowed to take as many breaks as needed. Subjects took an average of 25 to 30 minutes for the entire experiment.

Please refer our uploaded video (https://www.youtube.com/watch?v=_9Ye2hrDxz0) for details on the experimental design and setup.

3.5 Results

To equate use of the scale for analysis, each subject's ratings were averaged over the three blocks, normalized by the subject's mean score (over all listener positions and diffraction methods), and then scaled by the grand mean. The results are shown in Figure 3, which shows the average values with the standard errors for different listener positions around the obstacle for the two methods of diffraction. As can be seen in the figure, the decay trend for the wave-based diffraction is more uniform than the UTD-based one, which shows substantial variation in the perceived loudness.

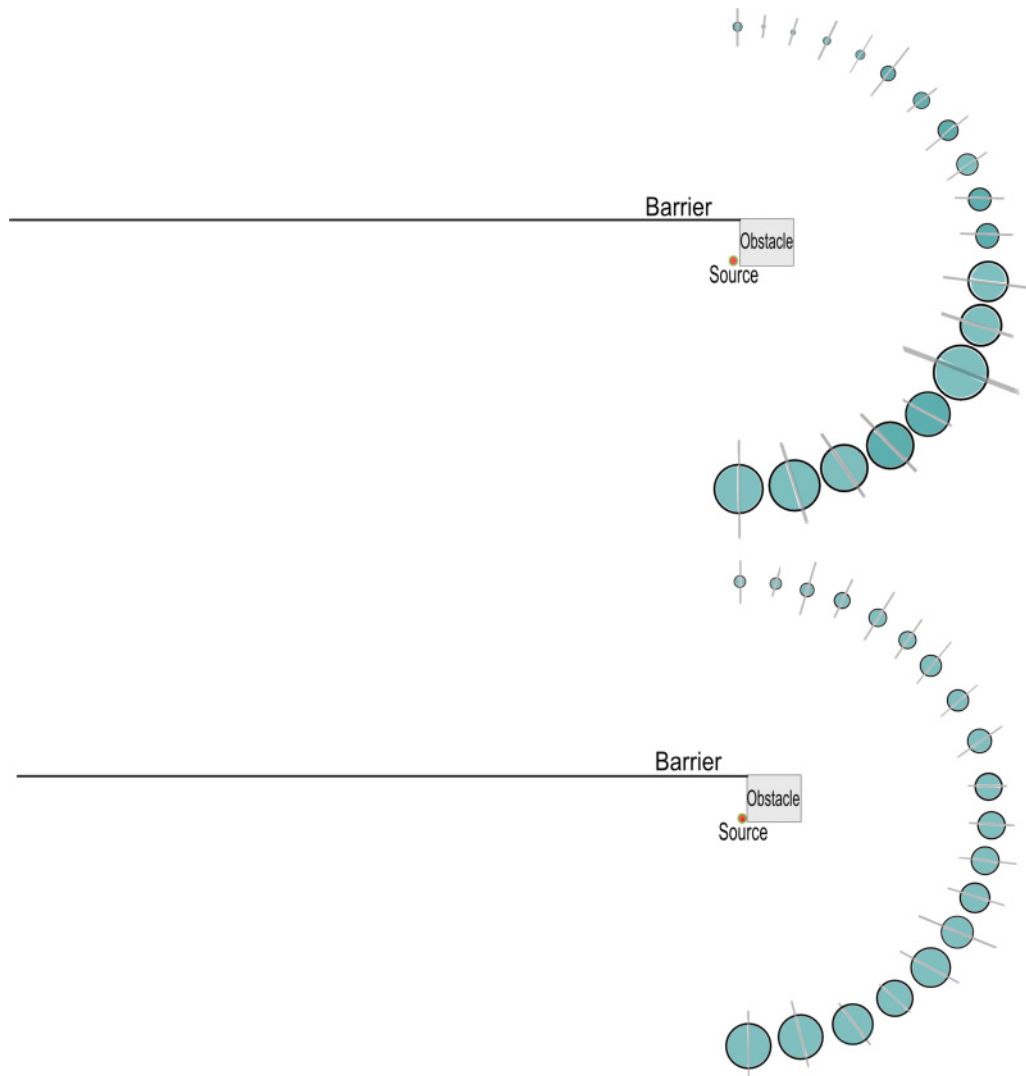


Fig. 3. Mean subject ratings for different positions along a semi-circle for two methods of diffraction, plus one direct sound region. The cross-hatches show the between-subjects standard deviation of the responses. The upper figure shows the results for the UTD-based diffraction while the lower one shows results for wave-based diffraction. Wave-based diffraction shows a strongly linear decay while the UTD-based diffraction shows considerable deviation from linear decay.

The significance of these trends was confirmed by a two-way, repeated measures ANOVA with factors diffraction method and position. It showed a significant two-way interaction: $F(18,270) = 48.64$, $p < .001$, indicating variations in the positional response across the two methods.

To analyze the additional effect of diffraction order within a factorial design, the direct sound condition was eliminated, and a subsequent three-way ANOVA with factors diffraction method, diffraction angle (the 18 nondirect angles), and order was performed. It showed a significant three-way interaction, $F(8, 120) = 34.52$, $p < .001$, indicating that the relation between diffraction order and diffraction angle changes with the diffraction method. All two-way interactions were also significant: diffraction

angle by diffraction method, $F(8, 120) = 15.83, p < .001$; diffraction angle by order, $F(8, 120) = 20.10, p < 0.001$; and diffraction method by order, $F(1, 15) = 370.00, p < .001$. The ANOVA also yielded significant main effects for diffraction angle $F(8, 120) = 288.99, p < .001$ and diffraction order $F(1, 15) = 672.54, p < .001$.

4. EXPERIMENT 2: REVERBERATION STUDY

4.1 Participants

Twelve subjects took part in this study, all males. The ages ranged from 18 to 23 years (mean = 21 with SD = 1.5). All reported normal hearing. The participants were recruited from the staff and students at a university campus.

4.2 Apparatus and Stimuli

The setup consisted of a MacBook Pro laptop with the sound delivered through a pair of Beyerdynamic DT990 PRO headphones. The subjects were blindfolded.

Since reverberation forms the later part of the IR of a room, the initial part of the IR has to remain consistent across the two cases to isolate reverberation as the only variable. To ensure that *only* the reverberation part of the IR is compared, we use the same system based on Schissler et al. [2014] to perform our task. This allows us to ensure that the early part of the IR is constant for both cases. In the first case, the IR is computed as is, without any modification corresponding to computing the reverberation using full IR. In the second case, we disable the physically computed reverberation and use a Schroeder-type filter [Schroeder 1962] to compute the reverberation and add it to the final IR. Gypsum board was chosen for all surfaces, including the ceiling and the floor. This was done because gypsum boards are commonly used in wall paneling and hence would retain the subjects' familiarity with everyday environments. The sound source was a male voice.

Reverberation filters must be hand tuned by setting the RT_{60} and D/R ratio to get the reverberation for an environment. Typically, these parameters are set by the audio designer based on his or her perception of what the environment would sound like. To remove this subjectivity, we decided to compute the parameters from the actual room parameters. The impulse response for each of the rooms was obtained, and the parameters were estimated from the response as explained in Section 2. Subjective ratings of room size were obtained using the method of free-magnitude estimation.

The stimuli consisted of seven cubes of increasing dimensions (100m^3 to 1600m^3 in increments of 250m^3) for each of the two reverberation conditions (Figure 4). The sound source was placed near the center of the room, slightly to the right. The subjects were allowed to move along a fixed path on the floor to hear the sound at different positions and get an average estimate of the reverberation in the room.

4.3 Design and Procedure

This was a within-subject study with the same participants for both methods of reverberation. The subjects wore a blindfold and headphones. The sound was rendered in stereo using the standard stereo panning technique. Even in the case of this experiment, the subjects' head orientation was fixed and they were always facing straight ahead. And although HRTF-based binaural rendering is known to increase the elevation localization accuracy, in the absence of variable head orientation and fixed source position, binaural auralization was considered redundant and simple stereo-based rendering was deemed sufficient. Hence, panning-based stereo rendering was employed.

Before starting the study, examples of sound originating in a small-size room and a large-size room were played for the blindfolded subjects. This was done so that the subjects could get an idea about

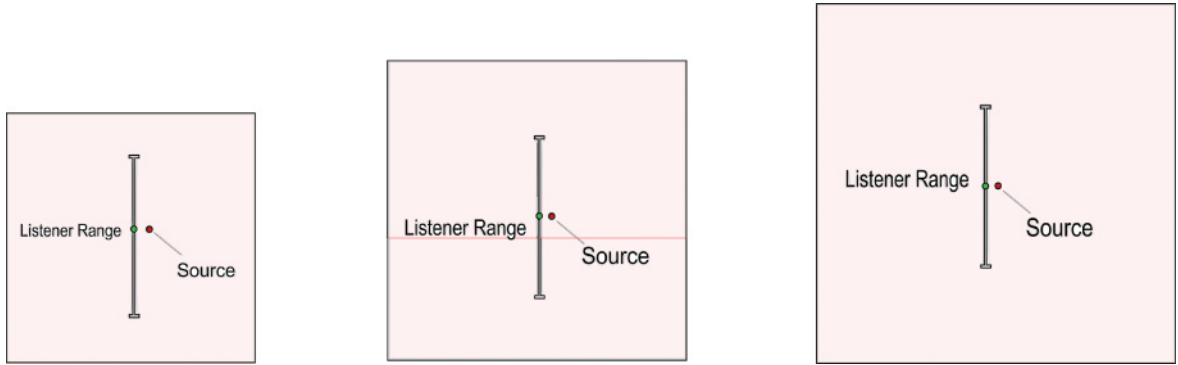


Fig. 4. The reverberation experiment setup consisted of seven cubes of increasing dimension, three of which are shown here. The listener path length in all cubes was the same.

the reverberation characteristics of the two ends of the room size spectrum and scale their internal response accordingly. They then underwent a training round in which each of the rooms were played in random order with one of the two reverberation methods. This was done to let the subjects come up with their own scale for scaling the room size. It is important to clarify that the subjects were in cubic meters; instead, they were asked to give a dimensionless number representing how large they thought the rooms were.

Each subject rated the complete set of 7 rooms \times 2 reverberation conditions with the order of the rooms randomized, on each of three separate blocks, giving a total of 42 measurements per subject (7 rooms \times 2 reverberation conditions \times 3 blocks). The subjects always started at the same initial position (relative to the room) and were asked to use one key on the keyboard to move forward and another to move backward. The subjects' head orientation was fixed, and they always looked straight ahead. On reaching the end of the translating distance in the virtual room on either side, they were notified by the experimenter and asked to move in the opposite direction. The walking distance was kept constant irrespective of the size of the room so that the subjects could not get a sense of the room size by the distance walked. The subjects were allowed to take as many breaks as needed. The average duration of the experiment was 1 hour. No fatigue was reported.

Please refer to our uploaded video (<https://www.youtube.com/watch?v=.9Ye2hrDxz0>) for details on the experimental design and setup.

4.4 Results

To analyze the data, we first normalized it for each subject to account for the different scales adopted by the different subjects [Zwislocki and Goodman 1980]. This was done by first taking the mean of the subjects' scores for the three trial rounds. Then, each subject's mean score was computed over all seven rooms and two reverberation conditions. This mean was used to normalize the scores. Finally, the normalized scores were scaled by the grand mean over all subjects to give a sense of the scale used.

Figure 5 shows the average magnitudes for each of the seven rooms for the two reverberation conditions across different blocks.

The three-way ANOVA with reverberation method, room size, and block showed three significant effects. First, there was a main effect of room size, $F(6,66) = 2.63$, $p = .025$, indicating that overall, participants did differentiate reverberant quality. Second, the interaction between room size with reverberation method was significant, $F(6,66) = 4.01$, $p = .003$, indicating that room size had different effects on the two reverberation methods. Finally, room size interacted with block, $F(2,22) = 4.029$,

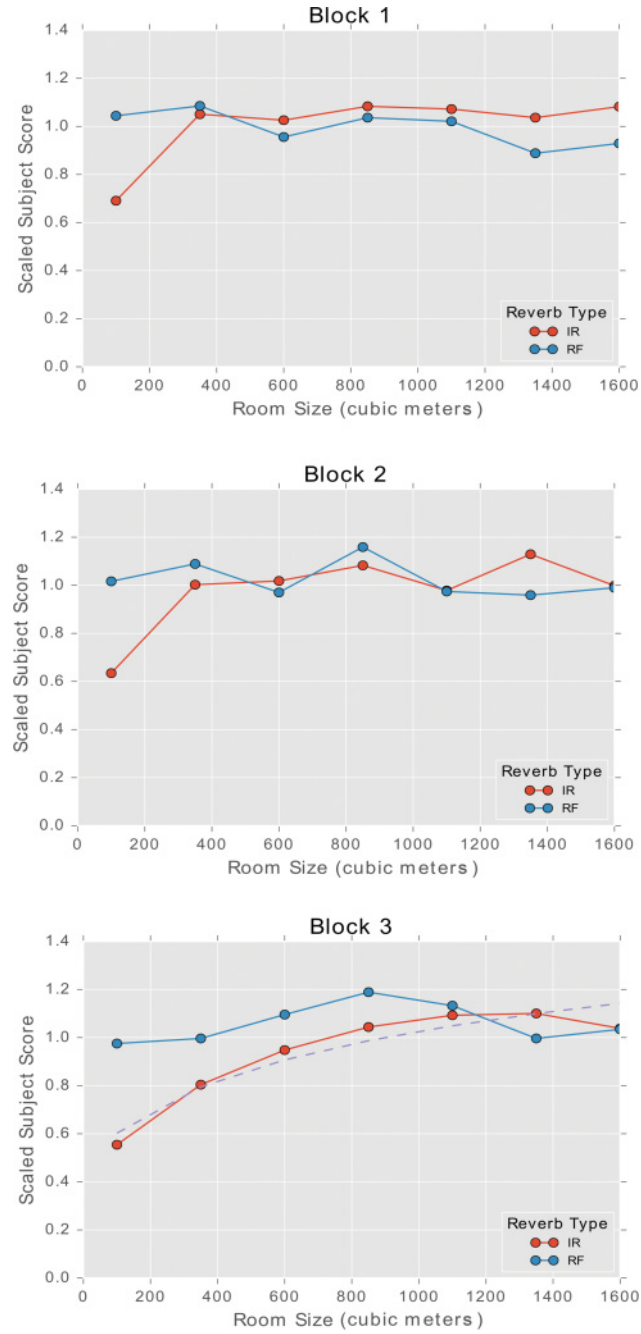


Fig. 5. The average subjective magnitude given by the subjects for the two reverberation conditions across the three blocks of trials. The IR-based reverberation starts doing better as the experiment progresses and starts showing a logarithmic relation between the volume of the room and perceived reverberant intensity (the purple dotted line in block 3 is the log fit). The reverberation filter shows no such learning effect.

$p = .034$, indicating a change in room size perception with experience making the judgments. The interaction-involving method led us to examine the two reverberation methods separately.

A two-way ANOVA for the reverberation filter with trial block and room size failed to show any significance for block, $F(2,22) = 0.794$, $p = .466$, room size, $F(6,66) = 0.759$, $p = .605$, or room size by block, $F(12,132) = 0.77$, $p = .678$, thus indicating that neither the room size nor the trial block mattered. Essentially, the subjects were unable to discriminate different room sizes and did not show any improvement even when the experiment was repeated across blocks (see Figure 5).

In contrast, the same two-way ANOVA for the reverberation computed using full IR showed significance for block, $F(2,22) = 5.08$, $p = .016$, and room size, $F(6,66) = 5.855$, $p < .001$. Although room size did not significantly interact with block, $F(12,132) = 0.709$, $p = .741$ (see Figure 5), it is clear that the changes with experience were confined to the smaller room sizes. A log fit to the block 3 mean data accounts for a substantial amount of the variance in the means across room size (91.5%). Thus, after two blocks of practice with the full IR method, subjects were differentiating the intensity of reverberation in an approximately logarithmic relation to room volume. Such compressive scaling of perceptual dimensions is commonly found, and the logarithmic relation is known as Fechner's law [Wolfe et al. 2014].

5. ANALYSIS

5.1 Diffraction Study

In any VE, the sound field should not appear discontinuous to the listeners, as that could break presence. Figure 3 shows how the sound field was perceived by the subjects in the two cases. A simple linear regression fit further solidifies the notion: the coefficient of determination R^2_{UTD} evaluates to 0.91, whereas $R^2_{Wave} = 0.98$ for the same set of diffraction angles relative to the source.

The high R^2 values for both diffraction methods indicate that the subject responses for both methods at least approximate a linearly decreasing trend in terms of perceived loudness. Although this is the expected trend for perceived loudness in such a setting, as shown by our experiment, the UTD-based diffraction shows substantially more deviation from linear. In contrast, the wave-based diffraction is highly linear in its trend. Figure 6 shows how amenable each of the methods is to a linear model fitting.

Based on these results, it can be concluded that UTD-based diffraction is a reasonable diffraction model perceptually and for applications where some discontinuity in the sound field is acceptable (e.g., games). In addition, UTD serves as a relatively inexpensive means to get reasonably good diffracted sound. On the other hand, applications where presence should not be affected by abruptly changing sound fields should use wave-based methods.

5.2 Reverberation Study

It is well known that volume estimation is a complex phenomenon, and the trends in the data for size scaling from reverberant cues reflect this in showing that people are relatively insensitive to these cues. We would expect to see a monotonically increasing trend in both cases. The full IR shows this characteristic in that the subjective magnitude for the full IR shows an increasing trend over the smaller room sizes, but this trend quickly saturates. However, the subjects seem to learn to scale with this technique over the course of the experiment. The scaling by reverberation filter, on the other hand, tends to vary in a nonsystematic manner with increasing room size and shows high intersubject variability as well.

Based on the log-linear regression model as a prediction of listener responses, the reverberation filter consistently performs poorly in rendering a sense of room volume, whereas the physically accurate reverberation method improves with user experience. Figure 7 shows how the value of R^2 varies

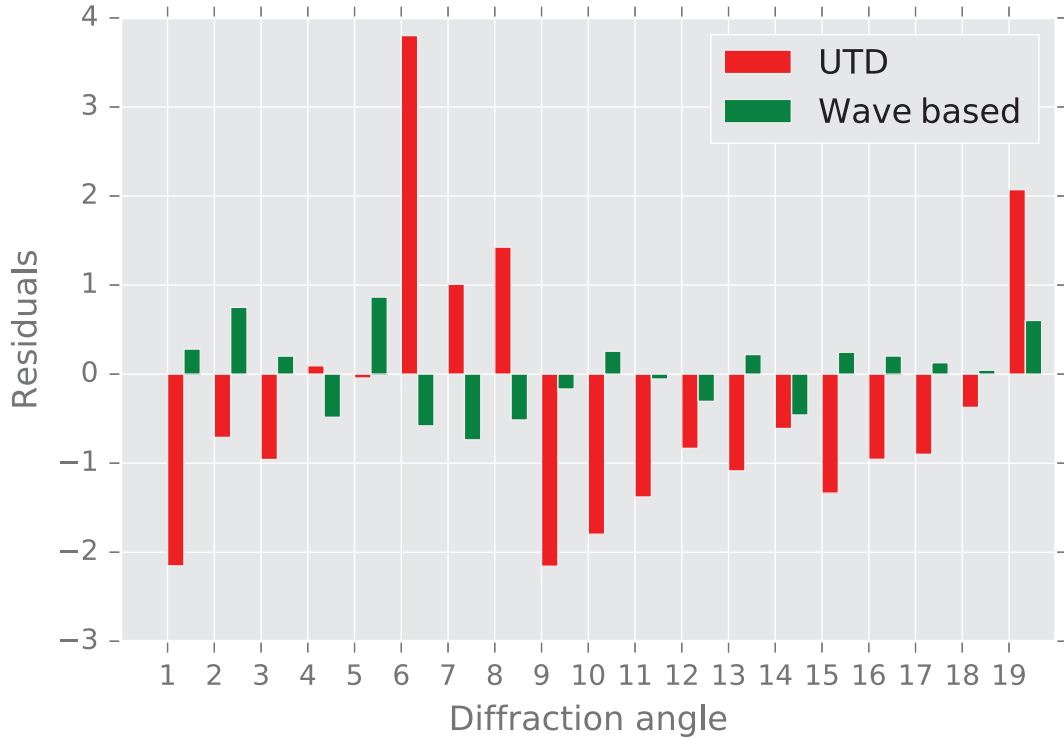


Fig. 6. The residual error from the simple linear regression for both diffraction methods. As is clearly evident, UTD-based diffraction shows high error when trying to fit its response to a straight line compared to the wave-based method. The x -axis represents the 19 equispaced diffraction angles ranging from 0 to 180 degrees.

with the block: the reverberation filter performs poorly across the three blocks, indicating that it does not lend itself to any meaningful relation between the perceived reverberation intensity and the volume of the environment. Accurate reverberation, on the other hand, starts showing improvement with $R^2_{(IR)_{B3}} = 0.92$ for the third block, indicating that the logarithmic relation between volume and reverberation intensity becomes more prominent as subjects start to learn to judge the volume based on sound cues alone.

Although the data indicate that untrained listeners are not particularly good at judging the size of an environment, it appears that using full IR offers some benefit in the way of consistency and seems to offer much superior discriminatory ability in room size perception as the subjects become trained. Hameed et al. [2004] pointed out that subjects tend to only use the RT_{60} in room size perception, and if fidelity with respect to this phenomenon is desired, the absorption coefficient of the room can be adjusted to get the proper relationship to the volume of the space.

We would like to point out that the reason we used a restricted scale for the diffraction experiment while using a free-magnitude estimation for the reverberation was because, in case of the diffraction experiment, we were interested in very specific points of discrepancy in the two methods, whereas in case of reverberation, we wanted to trace out the magnitude estimation function. We find that in the latter case, people are insensitive to the manipulations once we get away from the extreme low end.

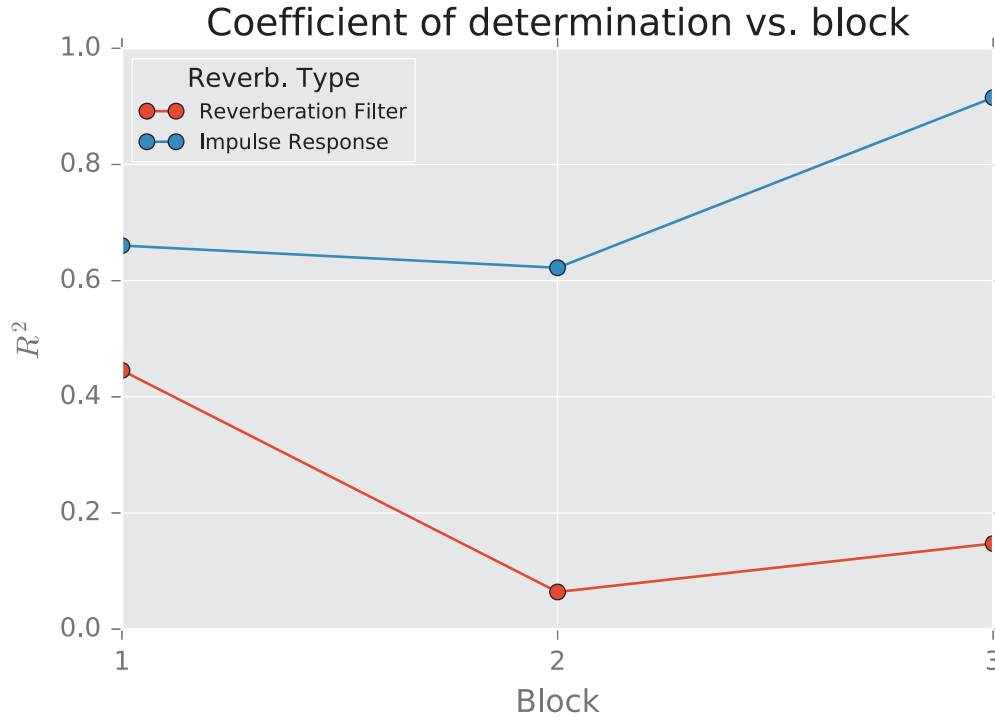


Fig. 7. The value of R^2 for the reverberation types across all three trial blocks.

6. CONCLUSIONS, LIMITATIONS, AND FUTURE WORK

In this article, we have presented two user studies to evaluate the perceptual merits of accurate and approximate interactive sound propagation algorithms. To the best of our knowledge, the diffraction study is the first of its kind in evaluating the perceived smoothness of the diffracted field for different methods, whereas the reverberation study evaluates the ability to provide accurate space cues for different methods. The study results show that accurate sound propagation algorithms offer an increased perceptual differentiation capability, exemplified by the reverberation study. Although the subjects were not able to differentiate the volumes of the rooms very well, the accurate reverberation method resulted in a significantly better discriminatory capability compared to the reverberation filter.

The regression fits offer a concise and numerical value that encapsulate the perceptual differentiation capabilities of the algorithms and serve as simple metrics to evaluate the performance of these methods. We expect that these metrics will serve as a means to let users decide which algorithms to use specific to their requirements. Currently, the R^2 values offer a very high level view of the algorithms and may not present an accurate picture for all scenarios. This can easily be the case if the wave-based diffraction had an unacceptably high peak at particular listener position(s) but still lent itself to a better linear fit than UTD. We believe that our choice of these metrics can be made more robust to such anomalies by using more sophisticated statistical models as part of future work.

We would also like to point out that the environments in both experiments were kept relatively simple so that the approximate algorithms could work to their full potential. Normally, reverberation filters are tuned with respect to simple cubical environments, and it is known that the performance of UTD-based diffraction degrades substantially with an increasing number of diffraction edges. Thus, in a way, these experiments also represent the best-case scenario for the performance of the

approximate algorithms. Since most VEs are going to be significantly more complicated, these approximate algorithms should be expected to perform below the level established in our experiments. Further analyses are needed to quantify the effects of room shape and interior features on performance of the algorithms. However, the number of potential geometric variations, combined with the range of variables that might be used to measure performance, makes determining the ecological validity of these methods a much more challenging research problem.

In addition, the studies can be extended in multiple ways. It would be interesting to vary the spectral content of the source in the diffraction experiment, as diffraction is a frequency-dependent phenomenon, and evaluate the subjects' responses. The subject's distance from the source can also be varied and evaluated. The reverberation experiment could be verified by constructing real-world rooms and using an actual sound source to verify the logarithmic relation of subjects' responses to changing room size.

ACKNOWLEDGMENTS

The authors would like to thank Alok Meshram, Ravish Mehra, and Jennifer Baulier for valuable insights and help at various stages of the project. The authors would also like to thank the subjects who took part in the user studies.

REFERENCES

- Jont B. Allen and David A. Berkley. 1979. Image method for efficiently simulating small-room acoustics. *Journal of the Acoustical Society of America* 65, 4, 943–950.
- Lakulish Antani, Anish Chandak, Micah Taylor, and Dinesh Manocha. 2012. Efficient finite-edge diffraction using conservative from-region visibility. *Applied Acoustics* 73, 3, 218–233.
- Durand R. Begault. 1992. Perceptual effects of synthetic reverberation on three-dimensional audio systems. *Journal of the Audio Engineering Society* 40, 11, 895–904.
- Durand R. Begault. 1994. *3-D Sound for Virtual Reality and Multimedia*. Academic Press, Cambridge, MA.
- Jean-Pierre Berenger. 1994. A perfectly matched layer for the absorption of electromagnetic waves. *Journal of Computational Physics* 114, 2, 185–200.
- Maurice Anthony Biot and Ivan Tolstoy. 1957. Formulation of wave propagation in infinite media by normal coordinates with an application to diffraction. *Journal of the Acoustical Society of America* 29, 3, 381–391.
- Jeffrey Borish. 1984. Extension of the image model to arbitrary polyhedra. *Journal of the Acoustical Society of America* 75, 6, 1827–1836.
- Densil Cabrera, Daeup Jeong, Hyun Jeong Kwak, Ji-Young Kim, and Jeonju Duckjin-Gu. 2005. Auditory room size perception for modeled and measured rooms. In *Proceedings of the 2005 Congress and Exposition on Noise Control Engineering (Inter-Noise'05)*.
- Anish Chandak, Christian Lauterbach, Micah Taylor, Zhimin Ren, and Dinesh Manocha. 2008. Ad-frustum: Adaptive frustum tracing for interactive sound propagation. *IEEE Transactions on Visualization and Computer Graphics* 14, 6, 1707–1722.
- Alexander H.-D. Cheng and Daisy T. Cheng. 2005. Heritage and early history of the boundary element method. *Engineering Analysis with Boundary Elements* 29, 3, 268–302.
- Dezhang Chu, Timothy K. Stanton, and Allan D. Pierce. 2007. Higher-order acoustic diffraction by edges of finite thickness. *Journal of the Acoustical Society of America* 122, 6, 3177–3194.
- Thomas Djelani and Jens Blauert. 2001. Investigations into the build-up and breakdown of the precedence effect. *Acta Acustica United with Acustica* 87, 2, 253–261.
- Hugo Fastl and Eberhard Zwicker. 2007. *Psychoacoustics: Facts and Models*. Vol. 22. Springer Science and Business Media.
- Thomas Funkhouser, Nicolas Tsingos, Ingrid Carlbom, Gary Elko, Mohan Sondhi, James E. West, Gopal Pingali, Patrick Min, and Addy Ngan. 2004. A beam tracing method for interactive architectural acoustics. *Journal of the Acoustical Society of America* 115, 2, 739–756.
- Jason A. Galster. 2007. *The Effect of Room Volume on Speech Recognition in Enclosures with Similar Mean Reverberation Time*. Ph.D. Dissertation. Vanderbilt University, Nashville, TN.
- Christian Giguère and Sharon M. Abel. 1993. Sound localization: Effects of reverberation time, speaker array, stimulus frequency, and stimulus rise/decay. *Journal of the Acoustical Society of America* 94, 2, 769–776.
- ACM Transactions on Applied Perception, Vol. 13, No. 4, Article 21, Publication date: July 2016.

- Helmut Haas. 1951. Über den Einfluß eines Einfachechos auf die Hörsamkeit von Sprache. *Acta Acustica United with Acustica* 1, 2, 49–58.
- Sharaf Hameed, Jyri Pakarinen, Kari Valde, and Ville Pulkki. 2004. Psychoacoustic cues in room size perception. In *Proceedings of the 116th Audio Engineering Society Convention*.
- William M. Hartmann. 1983. Localization of sound in rooms. *Journal of the Acoustical Society of America* 74, 5, 1380–1391.
- Jean-Marc Jot and Antoine Chaigne. 1991. Digital delay networks for designing artificial reverberators. In *Proceedings of the 90th Audio Engineering Society Convention*.
- T. Kawai. 1981. Sound diffraction by a many-sided barrier or pillar. *Journal of Sound and Vibration* 79, 2, 229–242.
- Vern Oliver Knudsen. 1932. *Architectural Acoustics*. John Wiley & Sons.
- Robert G. Kouyoumjian and Prabhakar H. Pathak. 1974. A uniform geometrical theory of diffraction for an edge in a perfectly conducting surface. *Proceedings of the IEEE* 62, 11, 1448–1461.
- A. Krokstad, S. Strom, and S. Sørsdal. 1968. Calculating the acoustical room response by the use of a ray tracing technique. *Journal of Sound and Vibration* 8, 1, 118–125.
- Heinrich Kuttruff. 2007. *Acoustics: An Introduction*. CRC Press, Boca Raton, FL.
- Pontus Larsson, Daniel Vastfjall, and Mendel Kleiner. 2002. Better presence and performance in virtual environments by improved binaural sound rendering. In *Proceedings of the 22nd International Audio Engineering Society Conference: Virtual, Synthetic, and Entertainment Audio*.
- Christian Lauterbach, Anish Chandak, and Dinesh Manocha. 2007. Interactive sound rendering in complex and dynamic scenes using frustum tracing. *IEEE Transactions on Visualization and Computer Graphics* 13, 6, 1672–1679.
- Ravish Mehra, Nikunj Raghuvanshi, Lakulish Antani, Anish Chandak, Sean Curtis, and Dinesh Manocha. 2013. Wave-based sound propagation in large open scenes using an equivalent source formulation. *ACM Transactions on Graphics* 32, 2, 19.
- Ravish Mehra, Nikunj Raghuvanshi, Lauri Savioja, Ming C. Lin, and Dinesh Manocha. 2012. An efficient GPU-based time domain solver for the acoustic wave equation. *Applied Acoustics* 73, 2, 83–94.
- Ravish Mehra, Atul Rungta, Abhinav Golas, Ming Lin, and Dinesh Manocha. 2015. WAVE: Interactive wave-based sound propagation for virtual environments. *IEEE Transactions on Visualization and Computer Graphics* 21, 4, 434–442.
- Nicolas Morales, Ravish Mehra, and Dinesh Manocha. 2015. A parallel time-domain wave simulator based on rectangular decomposition for distributed memory architectures. *Applied Acoustics* 97, 104–114.
- Claudio B. Pop and Densil Cabrera. 2005. Auditory room size perception for real rooms. In *Proceedings of the Australian Acoustical Society Conference*.
- Nikunj Raghuvanshi, Rahul Narain, and Ming C. Lin. 2009. Efficient and accurate sound propagation using adaptive rectangular decomposition. *IEEE Transactions on Visualization and Computer Graphics* 15, 5, 789–801.
- B. Rakerd and W. M. Hartmann. 1985. Localization of sound in rooms, II: The effects of a single reflecting surface. *Journal of the Acoustical Society of America* 78, 2, 524–533.
- Carl Schissler and Dinesh Manocha. 2016. Adaptive impulse response modeling for interactive sound propagation. In *Proceedings of the 20th ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*. ACM, New York, NY, 71–78.
- Carl Schissler, Ravish Mehra, and Dinesh Manocha. 2014. High-order diffraction and diffuse reflections for interactive sound propagation in large environments. *ACM Transactions on Graphics* 33, 4, 39.
- Manfred R. Schroeder. 1962. Natural sounding artificial reverberation. *Journal of the Audio Engineering Society* 10, 3, 219–223.
- Micah Taylor, Anish Chandak, Qi Mo, Christian Lauterbach, Carl Schissler, and Dinesh Manocha. 2012. Guided multiview ray tracing for fast auralization. *IEEE Transactions on Visualization and Computer Graphics* 18, 11, 1797–1810.
- Micah T. Taylor, Anish Chandak, Lakulish Antani, and Dinesh Manocha. 2009. RESound: Interactive sound rendering for dynamic virtual environments. In *Proceedings of the 17th ACM International Conference on Multimedia*. ACM, New York, NY, 271–280.
- R. R. Torres, M. Kleiner, U. P. Svensson, and B. I. Dalenbäck. 2001a. Edge diffraction and surface scattering in auralization. In *Proceedings of SIGGRAPH Campfire*.
- Rendell R. Torres and Mendel Kleiner. 1998. Audibility of edge diffraction in auralization of a stage house. *Journal of the Acoustical Society of America* 103, 5, 2789–2789.
- Rendell R. Torres, U. Peter Svensson, and Mendel Kleiner. 2001b. Computation of edge diffraction for more accurate room acoustics auralization. *Journal of the Acoustical Society of America* 109, 2, 600–610.
- Nicolas Tsingos, Thomas Funkhouser, Addy Ngan, and Ingrid Carlbom. 2001. Modeling acoustics in virtual environments using the uniform theory of diffraction. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*. ACM, New York, NY, 545–552.

- Michael Vorländer. 1989. Simulation of the transient and steady-state sound propagation in rooms using a new combined ray-tracing/image-source algorithm. *Journal of the Acoustical Society of America* 86, 1, 172–178.
- Craig Webb and Alan Gray. 2013. Large-scale virtual acoustics simulation at audio rates using three dimensional finite difference time domain and multiple graphics processing units. In *Proceedings of Meetings on Acoustics*, Vol. 19.
- J. Wolfe, D. Levi, K. Kluender, L. Bartoshuk, R. Herz, R. Klatzky, S. Lederman, and D. M. Merfeld. 2014. *Sensation and Perception* (4th ed.). Sinauer Associates, Sunderland, MA.
- Kane S. Yee. 1966. Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media. *IEEE Transactions on Antennas and Propagation* 14, 3, 302–307.
- Cecilia Maria Zannini, Raffaele Parisi, and Aurelio Uncini. 2011. Binaural sound source localization in the presence of reverberation. In *Proceedings of the 2011 17th International Conference on Digital Signal Processing (DSP'11)*. IEEE, Los Alamitos, CA, 1–6.
- Olek C. Zienkiewicz. 2005. *The Finite Element Method for Fluid Dynamics*. Ph.D. Dissertation. University of Wales.
- J. J. Zwislöcki and D. A. Goodman. 1980. Absolute scaling of sensory magnitudes: A validation. *Perception and Psychophysics* 28, 1, 28–38.

Received May 2016; accepted May 2016