Effects of virtual acoustics on dynamic auditory distance perception

Atul Rungta [a] [1], Nicholas Rewkowski [a], Roberta Klatzky [b], Ming Lin [a], and Dinesh Manocha [a]
[a] *Department of Computer Science, University of North Carolina, Chapel Hill*
201 S. Columbia Street , Chapel Hill
rungta@cs.unc.edu, nrewkows@live.unc.edu, lin@cs.unc.edu, dm@cs.unc.edu

[b] *Department of Psychology, Carnegie Mellon University*
5000 Forbes Avenue Pittsburgh
klatzky@cmu.edu

URL: `http://gamma.cs.unc.edu/AcousticDistancePerception`

Auditory distance compression in virtual acoustics

Abstract: Sound propagation encompasses various acoustic phenomena including reverberation. Current virtual acoustic methods ranging from parametric filters to physically accurate solvers can simulate reverberation with varying degrees of fidelity. The effects of reverberant sounds generated using different propagation algorithms on acoustic distance perception are investigated. In particular, two classes of methods for real time sound propagation in dynamic scenes based on parametric filters and ray tracing are evaluated. The study shows that ray tracing enables more distance accuracy as compared to the approximate, filter-based method. This suggests that accurate reverberation in VR results in better reproduction of acoustic distances.

Keywords: Psychoacoustics; Virtual acoustics; Architectural acoustics

PACS numbers: 43.55,43.66B

1. Introduction

Realistic sound effects can increase the sense of presence or immersion in virtual environments. One of the most important of these effects is reverberation, with considerable research going into simulating reverberation in virtual environments. Reverberation is known to have multiple perceptual effects on humans, enhancing auditory distance perception and environment size estimation, but degrading localization and speech clarity Valimaki (2012). In this paper, we consider auditory distance perception in virtual environments with audio rendered using the two sound propagation methods most suitable for all-frequency, broadband signals in real-time systems - an approximate, automatically-calibrated filter-based reverberation method and a state-of-the-art geometric ray tracer - in order to assess whether the increase in the physical accuracy of sound propagation leads to an increase in perceptual differentiation.

Auditory distance perception is known to be compressive Loomis (2002), and distance estimation in virtual environments follows the same trend. Some earlier work on auditory distance perception was based on setting up physical environments, using appropriate sound sources, and evaluating the accuracy and precision of distance estimates. However, such studies confront issues with controlling and quantifying the actual sound stimulus to the ears of the users in a physical scene. Recent developments in interactive and physically-accurate sound propagation methods provide unique capabilities for performing experiments using virtual acoustics. Not only can virtual acoustic techniques provide more flexibility and less expensive solutions in terms of evaluating auditory distance perception but they also allow us to assess the additional compression that might be caused by virtual environments (VE).

Our choice of sound propagation methods used to test compression in VE with reverberant sounds is guided by (1) real-time performance requirement in virtual environments and games Valimaki (2016); (2) broadband frequency range needed in complex VR systems similar in characteristics and complexity of the physical world we live in.

The study was conducted in a large, virtual rectangular-shaped room with sources that were far away from the listener, i.e., between $10m - 40m$. The sound was rendered as if the listener was moving in the environment. We observed that a calibrated, parametric reverberation filter shows similar compression characteristics to those of the accurate ray tracing method. However, the perceived distance is consistently higher for the ray tracing method. This result suggests that accurate reverberation methods can lead to better reproduction of acoustic distances.

2. Related work

2.1. Sound propagation

Sound propagation deals with modeling how sound waves reflect, scatter, and diffract around obstacles as they travel through an environment. Some of the simplest algorithms for modeling reverberation are based on artificial reverberation filters Jot (1991) that capture the statistics of reverberant decay using a small set of parameters Valimaki (2012). These are widely used in games and virtual environments because of their low runtime overhead. Other sound propagation algorithms are based on ray tracing and its variants (e.g., beam tracing or frustum) and assume that the sound travels along linear rays Allen (1979). They work well for high-frequency sources, though some approximate techniques have also been proposed to approximate low-frequency effects such as edge diffraction Antani (2009). The third, and most accurate, way of simulating sound propagation is the wave-based algorithms that numerically solve the acoustic wave equation and compute the sound pressure field and impulse responses Raghuvanshi (2010). As compared to geometric methods, these algorithms are able to accurately model low-frequency effects, but the time complexity increases as the fourth power of the frequency. As a result, they are only practical for low-to-medium-range frequencies (< 2 KHz). Furthermore, they are limited to static scenes due to the high precomputation overhead.

2.2. Acoustic distance perception

Over the last few decades, there has been considerable work on auditory distance perception Zahorik (2005); Kolarik (2016). Studies have found that auditory distance perception depends on the intensity Zahorik (2001), spectral cues Coleman (1968), reverberance (or more specifically direct-to-reverberant ratio (DRR) Mershon (1975)), and binaural cues Kopco (2011). Most of these studies suggest that listeners systematically underestimate the distance to a faraway sound source. In general, subjects tend to overestimate distances $<$ 1m and underestimate distances $>$ 1 m, with reverberation increasing the perceived distance. Kuusinen (2015) did experiments using virtual acoustics in four concert halls for distances ranging from 10m - 26m and found that in such large environments, overestimation may occur up to 10m and beyond, with the results being highly dependent on the acoustics of the halls. Locomotion also affects acoustic distance perception, but distance cues arising from active locomotion, parallax and acoustic tau, have been found to be relatively weak Speigle (1993) and we ignore them in our discussion. Further, our environments were very reverberant, which made distance cues from sound changes relative to the moving listener barely noticeable.

3. Experiment: moving listener in a reverberant environment

3.1. Participants

Seventeen participants took part in the study with informed consent. Their ages ranged from 19 to 47 (mean $= 25.9$ and SD $= 7.4$ - Four females, thirteen males). The participants were recruited from the students and staff at the university. All participants reported normal hearing.

3.2. Apparatus

The set up consisted of a Dell T7600 workstation and the sound was delivered via a pair of Beyerdynamic DT990 PRO headphones. The subjects were blindfolded for the study. The software to compute the RT_{60} and DRR was based on open-source MATLAB code. The calibration and auralization were done using in-house software, also written in MATLAB.

3.3. Sound propagation methods

We take a brief look at the sound propagation methods used. We use a state-of-the-art geometric sound propagation system Schissler (2014) capable of computing higher order specular and diffuse reflections using ray tracing. The system is able to compute 50 - 100 orders of reflections in dynamic scenes at interactive rates on a multi-core desktop PC by exploiting the coherence of the sound field and performing backward ray tracing.

We also use an artificial reverberator based on Schroeder filter design Schroeder (1962). This approach use a parallel bank of comb filters connected to a series of all-pass filters. The comb filters generate a repeated version of the input signals, while the all-pass filters keep the frequency gain of the input at constant value.

For our experiments, we used the Schroeder filter only to compute the late reverberation part of the impulse response, while the early part (direct + early reflections) was computed using the geometric propagation system described above.

3.4. Stimuli

The source was a pre-recorded sound of human clapping. Since clapping is somewhat similar to an impulse, it tends to have a broad frequency content making wave-based methods impractical for virtual environments with such stimuli. The virtual environment consisted of a rectangular room $45m \times 10m \times 3m$ with highly reflective walls to create a highly-reverberant environment with an $8m$ walking path as shown in Figure 1. Seven omnidirectional sound sources were kept at increasing distances from the center of the path starting from $10m$ up to $40m$ in increments of $5m$. The sources were all kept at the same height of $1.7m$ from the floor. This value was chosen assuming a standard listener height of $1.7m$ in the virtual environment. The source sound power was 78dB.

3.5. Filter calibration

The filter was calibrated to match the reverberation characteristics of the geometric sound propagation system by appropriately scaling and splicing the early part of the impulse response ($\sim 80ms$), starting at the approximate onset time of reverberation to match the RT_{60} and DRR

of the ray-traced impulse response. Since for each position the early part of the impulse response was similar in both cases, coupled with the fact that the RT_{60} and DRR values were matched, the loudness for the two methods was equivalent. The RT_{60} ranged from 2.6s to 2.8s while the DRR values ranged from -11.8 dB to -7.2 dB

3.6. Design and procedure

A rectangular room was chosen as the environment with highly reverberant environment similar to a painted, concrete room with no windows. In order to make sure we were comparing the *underlying methods* and not the specific parameters, we matched the RT_{60} and the DRR for both reverberation methods.

3.7. Training

Before the participants started the experiments, they completed a training task in a real-world setting. An $8m$ long walking path was constructed and the sound sources were placed at $3m$ and $6m$ from the center of the walking path, starting with $3m$. The participants were blindfolded before being led into the room so as to not give them an idea of the room dimensions. The dry (without reverberation) sound clip was played from a Harmon/Kardon HK 195 desktop speaker. The participants were asked to point at the sound source with their right hand and keep pointing at the source as they walked along the $8m$ path. Since the participants were blindfolded, they were helped by the test administrators as they walked down the path. Once they reached the end of the path, the participants were asked to give their best evaluation (in meters) as to how far from them they thought the sound source to be when it seemed closest to them. The training task was then repeated with the source moved to $6m$. The subjects were told the actual distances at the end of the training. The training exercise was not meant to be an exact replica of the experiment, as it was not possible to construct a physical room with the same kind of reverberance as the one in the virtual environment. Instead, the training was meant to give the participants a feeling for what to expect and how to make judgments.Please refer to the supplementary video on how the training was performed.

The supplementary video file is named MM1. This is a file of type “mp4” (9.6 MB).

3.8. Method

This was a within-subject study. The walking in the virtual environment was not controlled by the participants; instead, the 81 impulse responses per source were first convolved with sound source and then sampled such that each one of them contributed to 0.1m of the total 8m for a human traveling at an average speed of 1.39 m/s. The contributions from each of these 81

convolved impulse responses were spliced together (with interpolation) to create a sound file for each source. This sound file was played to the participants and they were asked to give the same estimate as they performed in the training, i.e., the perceived distance (in meters) of the sound when it seemed to be the closest to them. The impulse responses were spatialized using a generic HRTF-filter being applied to the direct sound and the early reflections. The participant's head orientation was fixed and they were always looking straight ahead. Each participant rated the complete set of 7 source positions $\times$ 2 reverberation methods with the order of the sources randomized for each block, giving a total of 42 (7 source positions $\times$ 2 methods $\times$ 3 blocks) judgments. The total time for the experiment, including the training, took around 15 minutes. The participants were allowed to take breaks between blocks, as required. No fatigue was reported. The supplementary video shows the environment along with the methodology used. The geometry of our environment coupled with the high reflection coefficients result in flutter echoes making it a special case for large, rectangular, and highly-reverberant virtual rooms.

3.9. Results

A 3-way ANOVA on block, distance, and reverberation method found a significant effect of method ($F(1, 16) = 15.29$, p < 0.01) and distance ($F(6, 96) = 29.12$, p < 0.01). All two-way interactions (block-distance, block-method, distance-method) failed to show significance. This finding indicates that the shape of distance compression is statistically the same for both reverberation methods, and the ray tracing algorithm exhibits an overall tendency to give longer distances. The null effect of block indicates that the experiment showed no learning or training effects and no trends are obscured by averaging over this factor.

4. Analysis

Zahorik Zahorik (2002) performed a comprehensive study of acoustic distance perception in virtual environments that assesses the weights assigned to the principal cues. To analyze the data, he fitted a power function of the form: $D_r = kd_r^a$ where D_r is the perceived distance, k is a constant, a is the power-function exponent that determines the function's rate of growth or decay, and d_r is the actual source distance.

Zahorik's data found that perceived distance was related to the simulated distance with the power parameter averaging 0.39. A value less than 1.0 meant that perceived distance was highly compressive. The multiplicative parameter was 1.3, which would result in over-estimation of very low distance values. There was also substantial variability in the reported judgments across the individuals, particularly for distances $> 1m$. The tendency to compress perceived distance was consistent across the source signal type or direction, but relative weighting of cues did vary with these factors. That being said, few studies have tried distances on the same order as ours and much less so in virtual environments. So our comparisons to previous research serve primarily as a means to establish baseline data.

In our present data, the power function fit on the data generated the following functions: $D_r^{accurate} = 1.56 d_r^{0.58}$ and $D_r^{filter} = 1.08 d_r^{0.66}$. The $R^2_{accurate}$ is 0.94 while the R^2_{filter} is 0.99; thus both functions account well for the observed variability in the mean perceived distance. The exponents exceed by $\sim$50% the average value found by Zahorik for a stationary listener.

The similarity of the power-function exponent for the two reverberation methods confirms the lack of interaction between physical distance and method in the ANOVA, which indicates that the compression of perceived distance relative to simulated distance was comparable across both methods of generating reverberation. The effect of the greater multiplicative parameter for the accurate method is to move the responses closer to the true values for all distances measured in our study and, by projection, any measured beyond this range of 10m – 40m. Any over-estimation resulting from a multiplicative parameter $>$ 1.0 would be expected for much smaller distances. The statistically confirmed result is that across the range of values examined here, the distance perceived with the accurate ray tracing algorithm exceeds that obtained with the filter method by an essentially constant amount.

The degree of compression observed here with virtual sound must be evaluated relative to the compressive perception found in reverberatory environments with real sound. If we take the data from Klatzky (2003) to provide a standard, a linear compression of 0.7 is to be expected with verbal report. The linear fits to the present data were reasonable for the filter and accurate ray tracing (values of R^2_{filter} and $R^2_{accurate}$, respectively), and the corresponding slopes were 0.24 and 0.23. In this context, we can estimate the additional compression due to simulation as a multiplicative factor on the order of $\frac{1}{3}$ giving us a good idea what to expect in terms of perceived distance when using virtual acoustics.

5. Conclusions, limitations, and future Work

In this paper, we compared the performance of approximate techniques based on parametric filters with accurate techniques based on interactive ray tracing in a dynamic scene. Our study shows that although the compression characteristics of the two methods are similar, the more accurate propagation method results in less distance compression in VR in a dynamic scene with a moving listener. This finding suggests that accurate reverberation effects in a VR system can be perceptually useful for different applications. The main limitations of our work include the use of a simple Schroeder filter only despite many other, more sophisticated reverberation filters available. Further, our environment was very simple and the study could have benefited from more complex geometries. In view of this, we would like to extend our evaluation to non-rectangular environments with moving sound sources, dynamic obstacles, and various methods of reverberation, especially methods based on feedback delay networks. It would be useful to combine our results with other cues (e.g., visual perception). Ultimately, we hope to develop VR systems with multi-modal capabilities (including sound), where researchers from other fields (e.g., psychology) can evaluate different hypotheses.

Acknowledgments

The authors would like to thank the subjects who took part in the user studies. The work was supported in part by Oculus Research and NSF grants 1320644, 1456299, and 1518630.

References and links

[1] Author to whom correspondence should be addressed.

Allen, J.B. and Berkley D.A. (**1979**). “Image method for efficiently simulating small-room acoustics”, J. Acoust. Soc. Am. **65**(4), 943–950.

Antani L. , Chandak A. , Taylor M. and Manocha D. (**2009**). “Fast geometric sound propagation with finite-edge diffraction.”, Technical report, University of North Carolina at Chapel Hill, 2009.

Coleman P. (**1968**). “Dual role of frequency spectrum in determination of auditory distance. ”, J. Acoust. Soc. Am. **44**(2), 631–632.

Jot J.C. and Chaigne A. (**1991**). “Digital delay networks for designing artificial reverberators. ”, AES Convention.

Klatzky R., Lippa Y., Loomis J., and Golledge G. (**2003**). “Encoding, learning, and spatial updating of multiple object locations specified by 3-d sound, spatial language, and vision. ”, Experimental Brain Research **149**(1), 48–61.

Kolarik A., Moore B., Zahorik P., Cirstea S., and Pardhan S. J (**2016**). “Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss”, Attention, Perception, & Psychophysics,**78**(2), 373–395.

Kopco N. and Shinn-Cunnigham B.G. (**2011**). “ Effect of stimulus spectrum on distance perception for nearby sources. ”, J. Acoust. Soc. Am. **130**(2), 1530–1541.

Kuusinen A. and Lokki T (**2015**). “ Investigation of auditory distance perception and preferences in concert halls by using virtual acoustics. ”, J. Acoust. Soc. Am. **138**(5), 3148–3159.

Loomis J., Lippa Y., Klatzky R., and Golledge G. (**2002**). “Spatial updating of locations specified by 3-d sound and spatial language ”, Journal of Experimental Psychology: Learning,Memory, and Cognition **28**(2),

Mershon D. and King L.E. (**1975**). “Intensity and reverberation as factors in the auditory perception of egocentric distance. ”, Perception and Psychophysics **18**(6), 409–415.

Raghuvanshi N. and Snyder J., Mehra R., Lin M., and Govindraju N. (**2010**). “Precomputed wave simualtion for real-time sound propagation of dynamic sources in complex scenes. ”, ACM Trans. on Graphics (Proc. ofACM SIGGRAPH) **29**(3), 1530–1541.

Schissler C., Mehra R., and Manocha D. (**2014**). “High-order diffraction and diffuse reflections for interactive sound propagation in large environments.”, ACM Trans. on Graphics (Proc. ofACM SIGGRAPH) **33**(4),

Schroeder M.R (**1962**). “Natural sounding artificial reverberation. ”, Audio Engineering Society **13**

Speigle J.M. and Loomis J. (**1993**). “Auditory distance perception by translating observers. ”, Proceedings of the IEEE Symposium in Research Frontiers in Virtual Reality, 25–26.

Valimaki V., Parker J.D., Savioja L., Smith J.O., and Abel J.S. (**2012**). “Fifty years of artificial reverberation. ”, Trans. Audio, Speech and Lang. Proc. **20**(5), 1421–1448.

Valimaki V., Parker J.D., Savioja L., Smith J.O., and Abel J.S. (**2016**). “More than 50 years of artificial reverberation. ”, Audio Engineering Society Conference: 60th International Conference

Zahorik P. (**1996**). “Auditory distance perception: A literature review. ”, Phd Preliminary Examination, University of West Maddison, Deptartment of Psychology, 19.

Zahorik P. (**2002**). “Auditory distance perception: A literature review. ”, J. Acoust. Soc. Am.,**111**(4), 1832–1846.

Zahorik P. and Wightman F.L. (**2001**). “Loudness constancy with varying sound source distance. ”, Nature neuroscience,**4**(1), 78–83.

Zahorik P., Brungart D., and Bronkhorst A. (**2005**). “Auditory distance perception in humans: A summary of past and present research. ”, Acta Acustica united with Acustica,**91**(3), 409–420.

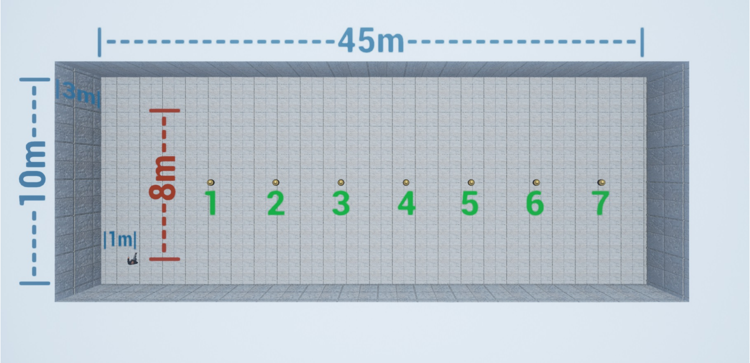

Fig. 1. The room used for the experiment. The path marked in red is the walking path along which the subject walks. The sound sources are perpendicular to the walking path and kept at increasing distances from it. The labels 1-7 show the different source distances sampled uniformly from the range 10-40m

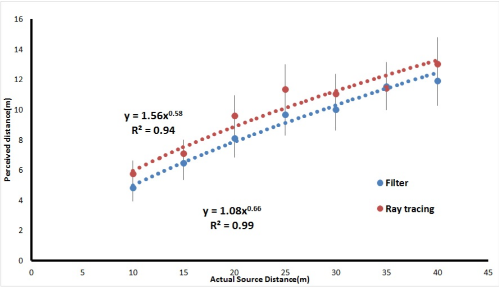

Fig. 2. The power function fit of the distance data for the Schroeder filter (blue) and interactive ray tracing (red) algorithms. This plot suggests that the compression of perceived distance relative to simulated distance was comparable across both methods. The distance perceived with the accurate ray tracing algorithm exceeds that obtained with the filter method by essentially a constant amount.