THE EFFECTS OF ATTENUATION MODELING ON SPATIAL SOUND SEARCH

Kyla A. McMullen
The University of Michigan
Electrical Engineering and Computer Science
Ann Arbor, MI 48109
kyla@clemson.edu

Gregory H. Wakefield
The University of Michigan
Electrical Engineering and Computer Science
Ann Arbor, MI 48109
ghw@umich.edu

ABSTRACT

Virtual spatial audio often utilizes the inverse-square law to model the relationship between intensity and distance for sources in the far-field. The present study explores the potential advantages of an "inverse-Nth" law, where N is greater than two, for dense, noisy environments where sources are distributed over a wide range of distances, or potentially sparse environments where the distance varies little. The findings of the study show significantly improved listener search and recall performance, without affecting sound search time, when using an inverse-8th law.

1. INTRODUCTION

Virtual spatial audio typically utilizes an inverse-square law to model the relationship between distance and intensity of a virtual source (e.g.: [1], [2], [3], [4]). While the inverse-square law is physically accurate, its use does not address the challenges of real system integration. For example, when used by a system operator to monitor a large number of sound sources at varying distances and levels, both the number of perceived sources and masking among sources increases, such that accuracy and search time for finding a particular source will degrade. A listener interacting with a VAE in a small environment space where source distance is more uniform could face the same challenges, as the change in level with distance as the listener moves through the environment will not be as salient as cue.

To address all of these challenges, it is necessary to investigate the effects of changing the attenuation model and its impact on search performance. One method to consider is to increase the sensitivity between intensity and distance by an "inverse-N" law, where $N >> 2$. Such an attenuation model could lower the signal to noise ratio in dense environments, by decreasing the effective density of detectable sources at the listener's ear. Additionally, for $N >> 2$, incremental changes in relative distance between a source and the listener will produce a much greater change in intensity, which is likely to facilitate auditory search, given that intensity is a prominent spatial cue [4].

On the other hand, too large an N could hinder spatial sound search by effectively limiting the number of sources that are detectable from any one position in the virtual space. In our previous studies ([3]) on virtual search, source position and level were controlled such that every sound was detectable at any point within the interface using an inverse-square law. For $N >> 2$, it is possible for some sounds to potentially be hidden within the environment, as the likelihood that stronger sounds at farther distances are more likely to be masked by nearby softer sounds. This could increase the time it takes to search for distant sources and would require far different search strategies to do so.

The present study examined the effects of two alternative attenuation models on listener accuracy and exploration time. Listeners explored a five-source VAE by moving their location in the environment to the position of each source. The locations of the sources was fixed for each trial, and randomly varied across trials. An inverse-eighth model of attenuation was used to study the case when $N >> 2$. For comparison, an "inverse-zero" law was also explored, in which level doesn't change with position. Positioning accuracy, angular accuracy, labeling accuracy, and exploration time were measured and compared to an identical search task, performed in a previous session which used inverse-square attenuation.

2. METHODS

2.1. Participants, Stimuli and Apparatus

Five undergraduate students at the University of Michigan (2 women and 3 men) participated in the current experiment and were paid \$10 per hour for their services. Each listener was experienced in the spatial-audio listening task, having participated in experiments on our laboratory's HRTF-customization [5] and training procedures. They had also completed a version of the present experiment using an inverse-square law attenuation model.

2.2. Stimuli

Table 1: Environmental sounds and their labels

Sound	Labels
Drumsticks striking a drum at regular intervals	Drums
Computer generated electronic noises	Electronic
A river flowing rapidly	River
Crickets chirping	Crickets
Typewriter keys being pressed	Typewriter

The experimental stimuli consisted of five distinct environmental sounds (see Table 1). The sounds were chosen from [6] based upon their ability to be perceptually segregated as five individual sounds. These specific sounds were chosen because of their distinct spectro-temporal patterns. Each signal also contained

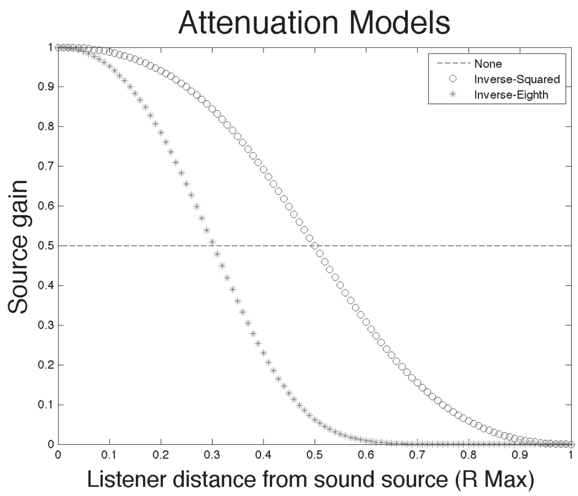

Figure 1: Attenuation Models used in each experimental condition.

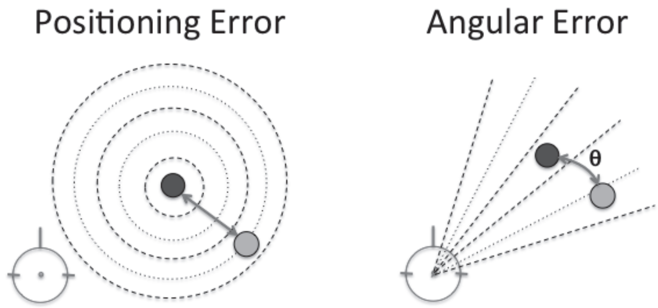

Figure 2: Positioning error (left) and angular error (right). In both figures, the blue circle represents the true location of the sound source and the green circle represents the location of the source as marked by the user.

broadband energy distribution and transients, to support binaural localization ([7], [8]). Furthermore, these signals are mutually inconsistent (unlikely to occur together in a natural acoustic environment) to mitigate the potential effects of a listener's long-term association of one source's location relative to that of another. Each sound was between 23 and 80 seconds long, and repeated continuously. Stimuli were presented at an audible level, as adjusted by the participant.

2.3. Apparatus

Experiments were conducted in a Tracoustics soundproof booth in the Computer Science and Engineering Building at the University of Michigan. Each participant was seated at a table in front of an iMac desktop computer. An Apogee Duet audio interface was used to generate the left and right channels of the audio signal, which was delivered to the listener over Beyerdynamic headphones. As in [3], a real-time spatial auditory system programmed in MATLAB was used. The system spatialized the auditory stimuli using each participant's customized HRTF, which was created using the procedure described in [5]. The controller sampled the participant's position and orientation at a rate of 10 Hz. To facilitate a real-time double buffering audio scheme, a timer was used to generate a new frame of audio corresponding to the participant's position within the VAE. The timer called routines to read the sound file from disk and convolve the audio input using the orientation-adjusted interpolated HRTFs. Audio was controlled using the PsychToolbox extension of OpenAL by double buffering. The timer also queried OpenAL sound source to determine if one of the buffers had finished playing, so that the next frame of audio could be loaded into the buffer, to be played after the current buffer. Interpolation of the HRTFs was implemented by constructing the minimum-phase impulse response of a system whose magnitude spectrum is determined from a log mixture of the adjacent measured HRTFs (sampled every 10 degrees) and convolving the result with an all-phase system using a fractional-delay method. An "inverse-N" law was used to attenuate the sounds as the listener moved through the VAE. Standard mouse to screen cursor mapping was used. Clicking or dragging the mouse in a new location moved the position of the onscreen avatar. The keyboard's right and left arrows were used to control the yaw of the avatar's head by rotating their orientation in steps of two degrees.

2.4. Procedure

Before beginning each condition, participants completed training to orient themselves to the auditory environment as realized through the particular "inverse-N" law. The training procedure followed that used originally to train the listeners in the use of the spatial audio system under an inverse-square law model.

On each trial, the participant "walks through the auditory environment" by moving the cursor on the computer screen using a free-search procedure with the goal of finding and memorizing the locations of the five sound sources. When the participant believes they have acquired the spatial configuration of the sources, they are aurally cued by monaural sample of each sound source and asked to indicate the position of that source by clicking on its screen location using the mouse. The experiment uses a balanced design. Figure 1 shows the attenuation conditions. The participant performs each trial 20 times for both experimental conditions.

3. RESULTS

Results of the current experiment were compared with each subject's own results from a previous experiment in which a standard inverse-squared attenuation model was used.

3.1. Accuracy

The effects of attenuation model on positioning accuracy were examined through a comparison of error during the aforementioned condition and the error during the a previous experimental session using inverse-squared gain. Two accuracy measures were used: positioning, angular (Figure 2). Positioning error was defined as the straight-line distance between the true and marked sound location. Position error was calculated in term of R_{Max}, which represents the maximum distance from which a sound source could be located with respect to the listener's position in the center of the interface. Angular error was defined as the unsigned angular difference between the true and perceived sound. The sound's angle was defined as the angle between the interaural axis and a line

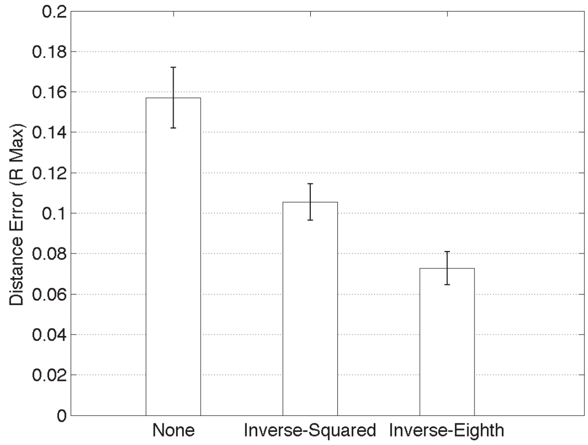

Figure 3: Effects of attenuation modeling on positioning accuracy. Along the abscissa are the attenuation models and along the ordinate is the positioning error.

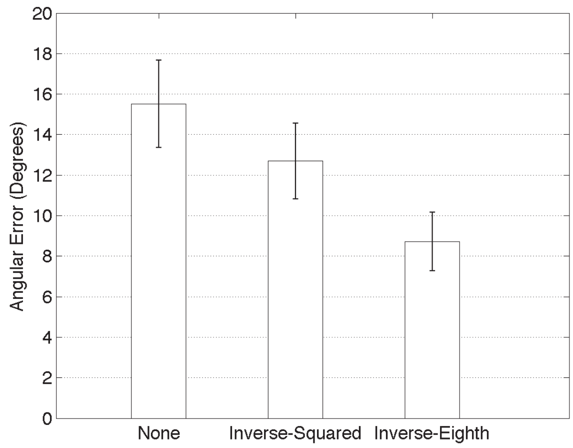

Figure 4: Effects of attenuation modeling on angular accuracy. Along the abscissa are the attenuation models and along the ordinate is the angular error.

originating at the center of the interface that intersects the sound's location.

Figure 3 shows that the choice of attenuation model has a significant effect on positioning error [$F_{2,1497}$=56.18, $p<0.05$]. A Tukey LSD multiple comparison test showed that positioning error was the lowest when an inverse-eighth attenuation model was used. Additionally, positioning error was lower using inverse-squared attenuation than the inverse-zeroth (no attenuation) model.

Figure 4 shows the effect of attenuation model on angular error. Similar to positioning error results, a significant difference was observed between conditions [$F_{2,1497}$=13.12, $p<0.05$]. A Tukey LSD multiple comparison test showed that angular error was the lowest when an inverse-eighth attenuation model was used. Additionally, angular error was lower using inverse-squared attenuation

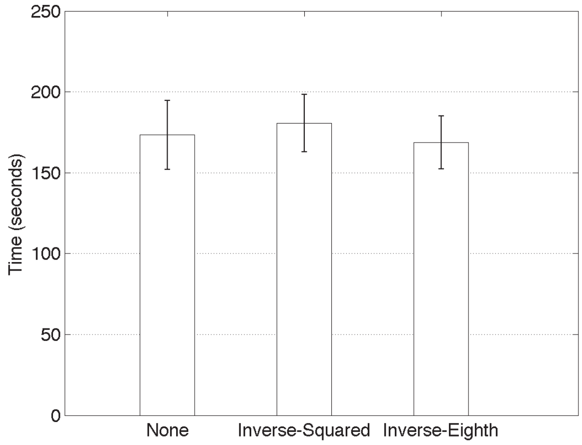

Figure 5: Effects of attenuation modeling on exploration time. Along the abscissa are the attenuation models and along the ordinate is the total time listeners explored the environment while localizing the five sound sources.

than no attenuation.

3.2. Exploration Time

Figure 5 shows the effect of attenuation model on search time. The analysis shows that the choice of attenuation model does not affect search time [$F_{2,297}$=0.42, p=0.66].

4. CONCLUSION

The results suggest the potential utility of non-physically realizable attenuation models for auditory search and recall in VAEs. On average, listeners were most accurate under an inverse-eighth law and were least accurate when there was no change in source level with distance. However, there may be a point at which increasing the value of N is no longer beneficial as when strong, but distant, sources are masked below detection threshold by weak, but close, sources.

Using inverse-eighth attenuation resulted in higher positioning and angular accuracy. This may have occurred because the environment was quieter during search and there was less interference from competing sources in the environment. Perhaps inverse-eighth attenuation narrows the search space of the possible locations of a sound source since the sound source is heard when the listener is within a very close proximity of the source. Therefore, if the listener can hear the source, they can conclude that they must be fairly close to it particularly if the levels of sources are roughly equal. In the inverse-squared and flat attenuation conditions, sound sources can be heard over a wider range of locations and relative level is a weaker (or nonexistent) cue for proximity.

The interface with no attenuation modeling had the lowest positioning and angular accuracy. This was also observed to be the loudest interface. This suggests that some attenuation modeling is necessary, due to the observation that in the cases of no attenuation, performance was severely degraded.

It was also observed that listeners need similar amounts of exploration time for each attenuation model. This is particularly surprising, given that there were significant differences in accuracy across conditions. One would expect a significant increase in search time as well. The results suggest that it may take just as long to find a drastically attenuated source as to narrow down the exact position of an easily detected sound source.

Nevertheless, our findings suggest that VAE designers can improve accuracy and recall, without affecting exploration time, by adopting an inverse-eighth law as opposed to the more physically accurate inverse-square law of attenuation. Further study is needed to determine the potential effects that such an attenuation model imposes on performance under environments with a large number of sources that vary substantially in both distance and level.

5. REFERENCES

[1] D. R. Begault, "Perceptual effects of synthetic reverberation on three-dimensional audio systems," *Perceptual effects of synthetic reverberation on three-dimensional audio systems*, vol. 40, 1992.

[2] R. L. Martin, K. I. McAnally, and M. A. Senova, "Free-field equivalent localization of virtual audio," *Journal of the Audio Engineering Society*, 2001.

[3] A. Roginska, G. H. Wakefield, T. S. Santoro, and K. A. McMullen, "Effects of interface type on navigation in a virtual spatial auditory environment," in *Proceedings of the International Conference on Auditory Displays*, 2010.

[4] P. Zahorik, "Auditory display of sound source distance," *Source*, pp. 1–7, 2002. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.137.3154&rep=rep1&type=pdf

[5] K. McMullen, A. Roginska, and G. Wakefield, "Subjective selection of head-related transfer functions (hrtf) based on spectral coloration and interaural time differences (itd) cues," in *Audio Engineering Society Convention 133*, 2012.

[6] BBC, *The BBC Sound Effects Library*. Princeton, NJ: Princeton, NJ: Films for the Humanities and Social Sciences, 1991, vol. 1-40.

[7] D. Begault, "3-d sound for virtual reality and multimedia," 2000.

[8] J. Blauert, *Spatial hearing: the psychophysics of human sound localization*. MIT Press, 1983.