SPATIAL SOUND AND SOUND LOCALIZATION ON A HORIZONTAL SURFACE FOR USE WITH INTERACTIVE SURFACE (TABLETOP) COMPUTERS

by

Jonathan Hak Shing Lam

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of
Master of Science (MSc)
in
Computer Science

Faculty of Science
University of Ontario Institute of Technology
Oshawa, Ontario, Canada
August 2012

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Jonathan Hak Shing Lam

CONTENTS

ABSTRACT

Tabletop computers (also known as surface computers, smart tables, and interactive surface computers) have been growing in popularity for the last decade and are poised to make in-roads into the consumer market, opening up a new market for the games industry. However, before tabletop computers become widely accepted, there are open problems that must be addressed with respect to audio interaction including: "What loudspeaker constellations are appropriate for tabletop computers?" "How does our perception of spatial sound change with these different loudspeaker configurations?" and "What panning methods should be used to maximally use the spatial localization abilities of the user(s)?" Using a custom-built tabletop computer setup, the work presented in this thesis investigated these three questions/problems via a series of experiments. The results of these experiments indicated that accurately localizing a virtual sound source on a horizontal surface is a difficult and error-prone task, for all of the methods that were used.

KEYWORDS

Tabletop computer, interactive surface computer, audio interaction, loudspeaker configuration, amplitude panning, sound localization, spatial sound.

LIST OF PUBLICATIONS DIRECTLY ARISING FROM THIS THESIS

Refereed Journal Publications

[1] J. Lam, B. Kapralos, K. Collins, A. Hogue, K. Kanev, and M. Jenkin. Sound localization on tabletop computers: A comparison of two amplitude panning methods. *ACM Computers in Entertainment*. To appear 2012.

Refereed Conference and Workshop Proceedings

[2] D. Nakano, J. Lam, B. Kapralos, K. Kanev, K. Collins, A. Hogue, and M. Jenkin. A framework for sound localization experiments and automation. In *Proceedings of The ACM Joint International Conference on Human-Centered Computer Environments (HCCE 2012)*. March 8-13, 2012, Hamamatsu, Japan, pp. 137-144.

[3] J. Lam, B. Kapralos, K. Collins, A. Hogue, and K. Kanev. Amplitude panning-based sound system for a horizontal surface computer: A user-based study. In *Proceedings of the International Symposium on Haptic Audio-Visual Environments and Games*. October 16-17, 2010, Phoenix, AZ. USA, pp. 1-5.

[4] J. Lam, C. Collins, B. Kapralos, A. Hogue and M. A. Garcia-Ruiz. Wiimote-controlled stereoscopic MRI visualization with sonic augmentation. In *Proceedings of the ACM FuturePlay 2010 International Conference on the Future of Game Design and Technology*. May 6-8, 2010, Vancouver, British Columbia, Canada, pp. 261-262.

ACKNOWLEDGMENTS

I would like to thank my supervisors Drs. Bill Kapralos and Andrew Hogue for their patience and support with the writing of this thesis. I would also like to thank Drs. Christopher M. Collins, Mikael Eklund, and Lennart Nacke. Dr. Collins (Faculty of Science, University of Ontario Institute of Technology) served on my supervisory committee and provided very useful feedback regarding my thesis. Dr. Mikael Eklund (Faculty of Engineering and Applied Science, University of Ontario Institute of Technology) served as the external thesis examiner and also provided me with very useful feedback and finally, Dr. Nacke served as the Chair during my thesis defense.

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 – INTRODUCTION

1.1 The Importance of Sound in Video Games

Sound plays a vital role in interactive media and video games in particular: it communicates important information to the player; it serves as a sound symbol or leitmotif; it situates the player in a specific location; and it reduces learning curves and creates a greater sense of realism. As such, implementing sounds in games for optimal playback involves the music, dialogue, sound effects and ambient sound-beds being carefully produced and placed in the sound space (the "mix") according to a well-established tradition of audio-visual media.

For many decades now, we have experienced our audio-visual media on a vertical screen; our televisions, movie theaters, and computer screens have all presented information vertically in front of us. As a result, sound (music, dialogue, and sound effects) for television, film, software, and games has been designed accordingly, with the placement of the loudspeakers and the sound mixing all developed based on this format. Recently, smart tabletop touchscreen computers (also known as surface computers, smart tables, interactive surface computers, and tabletop computers), where users position themselves around a horizontal computer screen in a manner similar to sitting around a "traditional'' table, have been introduced. One example is the Microsoft PixelSense, a multi-touch surface computing platform that responds to natural hand gestures and real-world objects.

Tabletop computing has been growing in popularity for the last decade. Most major computer companies have now developed or are working towards a tabletop computer, and we can expect to see tabletop computers make in-roads into the consumer market soon, particularly with the popularity of touch tablets, since they could employ similar APIs. Microsoft's new *LightSpace* technology [Wilson and Benko 2010] allows any table to be repurposed as a tabletop computer, and Microsoft's *SecondLight* system [Izadi et al. 2008] similarly employs surface technology, detecting where the surface of a table is being touched and allows an image to be projected on to a material held above the surface. While these computers remain prototypes or outside the price range of the average consumer, tabletop computing could well become an important market as developers move towards a consumer model. As with much consumer computer technology, entertainment applications will likely drive the success of consumer-model tabletop computing. Until now, audio has been overlooked for tabletop computing yet it remains to be a key component of interactive applications. There are many open questions regarding the generation of effective audio for interactive multi-user tabletop displays.

Video games are a logical application for tabletop computing technology, given that games have for thousands of years been played on table-like surfaces (from ancient games such as Go and Chess to modern board games) rather than the vertical screens of modern video games. One can easily anticipate the translation of traditional games into digital tabletop games (tabletop "cocktail" games were also commonly available in the arcades of 1980s but disappeared along with the arcades). The move to tabletop computers will likely introduce a whole new market for the games industry as the technology encourages multi-player social gaming, whereby many users can crowd around a table quite naturally.

However, the move from vertical-screen digital games to a horizontal tabletop introduces interesting questions with respect to the implementation of images/graphics and sound. Questions of co-operation, orientation and angle will drive innovation in imagery and have been explored elsewhere [Kruger et al. 2004; Scott and Carpendale 2010], but regarding sound and its use in an interactive tabletop setting, as described below, there are many questions that remain unanswered. As previously described, for decades, we have experienced our digital audio-visual media on a vertical screen. When multiple users are present, all of the users sit "in front" of the vertical display surface. As such, content for these media has been designed accordingly, with the assumption that the users are directly in front of the screen with, at minimum, a stereo pair of loudspeakers directed towards them. With tabletop computing these assumptions are no longer valid. Users are now "above" the display surface and surround it, and viewing angles are typically oblique. Given that tabletop computers are intended for multiple users and with an emphasis on collaboration amongst them, headphones are typically not appropriate, since they may limit, and even distract, natural verbal interaction amongst the users. Effective spatialized audio delivery on a tabletop computer will undoubtedly involve one or more external loudspeakers.

The PC and console games industry has until now had to rely on loudspeaker configurations designed for movies and television. This development of audio positioning for a vertical screen has largely come from conventionalization over decades of use, primarily from the movie and home audio listening industry. Few examples of the exploration of surround sound for horizontal surfaces exist despite the potential significant implications this may have for the design of sound in games, many of which rely on spatial information. However, the difficulty with respect to tabletop computers is that the listener configuration changes; two listeners may be faced opposite to each other in front of the table and there are no configurations that plan for optimal reception of sound in this format. A particular difficulty with surround sound and smart tables is the issue of three-dimensional applications (such as games). With loudspeakers (and therefore sound) positioned on a horizontal plane around the users, how can we generate sound such that the user perceives its source is at a particular 3D location? Is it a matter of altering the pitch of sounds, or is it necessary to physically change the loudspeaker location to achieve this effect (essentially, tilting the whole 5.1 surround sound set-up on an axis)? We propose two possible solutions to this problem: i) move the loudspeakers, and ii) move the position of the sound in the mix of the application/game based on the number of users. It is more practical, of course, for users to not have to move the loudspeakers based on the number of people using the smart table, and so the second option offers the most viable alternative, although both solutions should be examined in a variety of ways.

Among the large number of open research questions regarding sound for tabletop computing, we seek to explore the following in the context of sound generation for interactive media:

1. What loudspeaker constellations are appropriate for tabletop computers?
2. How does our perception of spatial sound change with these different loudspeaker configurations?
3. What panning methods should be used to maximize the spatial localization abilities of the user(s)?

The work summarized in this thesis investigates these three questions/problems and the results presented here provide us with a greater understanding of sound localization on a horizontal surface, bringing us closer to providing an optimal solution with minimal errors to these problems. Two spatial sound techniques for tabletop computers were developed and tested (with human participants) with three loudspeaker configurations.

1.2 Sound Localization Experiments

In this thesis, a series of experiments that examined various aspects regarding spatial sound generation and sound localization for tabletop computers were conducted and described. Each of the experiments builds upon the previous one and was generally designed to explicitly address an issue/finding of the previous experiment.

In Experiment One (Chapter 3.3), a simple and computationally efficient bilinear interpolation-based amplitude panning method was designed specifically for horizontal tabletop computers with four loudspeakers, one at each corner of the table facing inwards towards the center of the table (surface). User-based experiments were conducted to test the effectiveness of the method and results showed that virtual sound source positions very close to the user lead to the greatest localization error while the localization error for virtual sound source positions along the border of the surface was less. It was hypothesized that this error was due to the fact that for the positions resulting in the largest error (those closest to the participant), the two loudspeakers were facing away from the participants.

In Experiment Two (Chapter 3.4), the loudspeakers were "flipped" such that they faced (and emitted sound) upwards in order to test whether the errors did in fact result from the fact that the two loudspeakers faced away from the participants. However, the experimental results did not support this claim.

Given the presence of errors, particularly for those positions that were closest to the participants, and the fact that previous work had already determined that a diamond loudspeaker configuration, whereby a loudspeaker was placed at each of the four sides of the tabletop computer was the preferred configuration by participants [Collins et al. 2011], two additional experiments were performed. Experiments Three and Four (Chapter 3.5) examined the application of two amplitude panning techniques to a diamond loudspeaker configuration: the bilinear interpolation method that involves panning of the sound between loudspeaker pairs, and the inverse-distance method where the sound emanating from each loudspeaker is scaled by the distance between the (virtual) sound source and the corresponding loudspeaker. Results from these experiments showed that there was no significant differences between the two methods and that both methods are prone to error.

Although Experiments One through Four measured sound localization of virtual sound sources on a horizontal surface, and compared different panning methods or different loudspeaker configurations, "ground truth" data to compare these results with, was lacking. In other words, just how accurately can we localize a sound on a horizontal surface when the sound is emanating from an actual sound source at the corresponding location? In Experiment Five (Section 3.6), a novel sound verification hardware setup and methodology was used to collect "ground truth" data in order to allow for meaningful comparisons of the previous results to be made. It allowed a single physical sound source to be moved to 36 pre-defined places (positioned on a grid with x- and y-axis separations of 0.15 m) in a simple and efficient manner. The results of Experiment Five indicate that sound localization on a horizontal surface with actual sound sources is erroneous albeit to a lesser degree than virtual sound sources.

1.3 Thesis Organization

The remainder of this thesis is organized as follows. Background information is provided in Chapter 2. More specifically, details regarding amplitude panning and spatial sound as it relates tabletop computers will be presented. In Chapter 3, experimental methods and results are presented while a discussion of these results is provided in Chapter 4. Finally, conclusions and plans for future work are outlined in Chapter 5.

CHAPTER 2 - BACKGROUND

2.1 Overview

The majority of work related to the generation of spatial sound and sound localization has focused primarily on sounds associated with loudspeakers (and screens; our televisions, movie theaters and computer screens) aligned vertically in front of the listener. Very few researchers have examined sound generation and localization on a horizontal surface. Collins et al. [Collins et al. 2011] examined listener preference of the traditional, and diamond loudspeaker configurations. In the context of video games, a touch-table electronic version of tabletop air hockey similar to the two-player Pong game distributed by Atari in the 1970's was developed. In this game, players directed a simulation of a puck into the opponent's "net'' while preventing the puck from entering their own net. The Audio Air Hockey game had two modes: i) standard, and ii) audio-based. The standard mode provided both audio and visual cues to the location of the puck while the audio-based mode required the players to rely only on sound to determine the location of the puck. Distinct sounds were mapped to collisions between the puck and the paddle, the puck and the walls, and the puck and a net. The simulated puck itself emitted a continuous soft white noise sound as it moved. All sounds were spatialized using the inverse-distance amplitude panning method whereby the sound emanating from each loudspeaker is scaled by its distance to the virtual sound source. Participants played the game with a visible puck against a trained opponent for ten minutes before the puck was made invisible and players played by localizing the sound of the puck on the surface. Participants were then asked to complete a short questionnaire regarding their ability to play, and their preference for loudspeaker positioning. Players reported that they preferred the diamond loudspeaker configuration as it allowed them to localize the position of the puck more accurately and therefore "play the game better''. Despite the preference for the diamond loudspeaker configuration, sound localization accuracy was not explicitly examined in that study, and this became the motivation for experiments three and four described in this thesis.

Regardless of the method actually used for sound generation, given that tabletop computers are intended for multiple users, the interaction amongst the users is essential and, as previously described, headphones are typically not an option. Therefore, tabletop computer sounds systems will generally involve multiple loudspeakers. In such scenarios, spatial sound generation with multiple loudspeakers (particularly more than two) will typically involve some form of amplitude panning rather than incorporate head-related transfer functions (HRTFs). An HRTF is a response function comprised of all the interactions between a sound and the listener. As a sound travels from its source to the listener, the various parts of the listener's body (such as the head, the pinna of each ear, and torso) interact with and modify the properties of the sound before it finally reaches the ear drum. These interactions are also dependent on the distance and direction of the sound source, thus to calculate the HRTFs for all possible scenarios is far too complex and computationally expensive (see [Kapralos et al. 2008] for greater details regarding spatial sound generation including HRTF-based spatial sound). Given the importance of amplitude panning, the remainder of this chapter will provide a review of amplitude panning and loudspeaker-based sound generation methods. In Sections 2.2.1 and 2.2.2, the bilinear interpolation and distance-based amplitude panning methods are described. These two methods form the basis for the amplitude panning employed in this work.

2.2 Amplitude Panning

Amplitude panning, or intensity panning, is the most-widely used method of panning for simulating sounds emanating from locations between the two loudspeakers [Pulkki and Karjalainen 2001]. Two or more sound sources (loudspeakers) are placed equidistant from the listener. When the same sound signal is played through each of the source at different amplitudes (or gain factors), a new signal is formed when the original signals reach the listener's ear canals. The signals are summed up and interpreted as a single auditory event by the listener [Pulkki 1997]. Such approaches attempt to mimic (to some degree) the binaural hearing mechanism of humans. As described Kapralos et al. [Kapralos et al. 2008], unless the sound source lies on the median plane (the plane equidistant from the left and right ears) the distance traveled by sound waves emanating from a sound source to the listener's left and right ears differs. This causes the sound to reach the ipsilateral ear (the ear closer to the sound source) prior to reaching the contralateral ear (the ear farther from the sound source). The interaural time delay (ITD) is the difference between the onsets of sounds at the two ears. When the wavelength of the sound wave is small relative to the size of the head, the head acts as an occluder and creates an acoustical shadow which attenuates the sound pressure level of the sound waves reaching the contralateral ear. The difference in sound level at the ipsilateral and contralateral ears is commonly referred to as the interaural level difference (ILD) although it is also referred to as the interaural intensity difference (IID) as well.

The attributes of these signals specify the location of an amplitude-panned virtual sound source [Pulkki 2001], in a process called summing localization [Pulkki and Karjalainen 2001] [Pulkki 2001]. When a large number of loudspeakers are used in a system, pair-wise amplitude panning may be used. This is where a sound signal is panned between and played through only two loudspeakers from the whole system. The pair of loudspeakers actually chosen depends on the location of the virtual sound source. For three-dimensional amplitude panning methods, triplet-wise panning may also be use, working under the same principle as pair-wise panning except utilizing three loudspeakers instead of two [Pulkki 1999].

In two-dimensional amplitude panning, the loudspeakers are placed coplanar to the listener [Pulkki 2001]. Two common two-dimensional panning systems found today are stereophonic and quadraphonic. In stereophonic systems, loudspeakers are placed on a horizontal plane in front of the listener. The setup is symmetrical; the loudspeakers are placed equidistant from the listener [Pulkki and Karjalainen 2001], typically forming the optimum angle of 60° between them [Malham and Myatt 1995]. The accuracy of sound localization diminishes as the angle between loudspeakers increases [Pulkki and Karjalainen 2001], particularly for sounds that are near the midpoint between a pair of loudspeakers [Malham and Myatt 1995].

There are many approaches to form a relationship between the gain factors of each loudspeaker and the perceived direction of the virtual sound source in two-dimensional stereophonic amplitude panning [Pulkki and Karjalainen, 2001]. One such law is the sine law, is given as:

$$\frac{\sin\theta_S}{\sin\theta_O} = \frac{g_1 - g_2}{g_1 + g_2} \tag{1}$$

where θ_O and θ_S are the actual and perceived azimuth angles of the virtual sound source respectively, g_1 and g_2 are the gain factors for each loudspeaker. The sine law assumes the interaural time difference (ITD) for the virtual sound sources are the same as that of the real sound sources [Pulkki and Karjalainen, 2001]. Interaural level differences (ILD) are not taken into account in this estimation model [Pulkki 2001]. Although the ITD for frequencies above the range of 400 [Pulkki 2001] to 600 Hz [Pulkki 1999] for the listener model used is not valid, Pulkki has found that in a horizontal setup such as this one, the laws give a good estimation up to 1100 Hz [Pulkki and Karjalainen 2001] [Pulkki 2001].

A second law, the tangent law, was derived using improved models of how sound interacts with the head of the listener. It is given as:

$$\frac{\tan\theta_T}{\tan\theta_O} = \frac{g_1 - g_2}{g_1 + g_2} \tag{2}$$

with the same assumptions and variables used as the sine law, but with the perceived azimuth angle labeled as θ_T instead of θ_S [Pulkki and Karjalainen 2001]. The tangent law is more correct over the sine law when the listener is facing the virtual source [Pulkki 1999].

2.2.1 Bilinear Interpolation Amplitude Panning

The system consists of four loudspeakers (each at one of the four table corners) facing the computer table's surface (the additional option of a centre channel and LFE will be explored at a later time). This setup is similar to a traditional quadraphonic surround sound system. However, traditional quadraphonic stereo techniques are intended for one listener and therefore, not applicable in this work. A number of amplitude panning methods were experimented with, including a simple distance-based amplitude panning method whereby the sound is output at each of the four loudspeakers but the level of the sound output at each loudspeaker is scaled by the distance between the corresponding loudspeaker and the virtual sound source [Lossius et al. 2009]. Another intuitive and computationally simple technique is based on bilinear interpolation and the sound is panned between loudspeaker pairs. Referring to Figure 1(a), first the "left-horizontal" scalar V_L is determined for the front-left and rear-left loudspeakers (S_{FL} and S_{RL} respectively) by dividing the horizontal distance between them and the virtual sound source D_L (the virtual sound source is denoted by V_S) and the total distance between the left- and the right-hand pair of loudspeakers (D_X). Similarly, the "right-horizontal" scalar V_R for the front-right and rear-right loudspeakers (S_{FR} and S_{RR} respectively), is determined by dividing the x-axis distance between them and the virtual sound source D_R and D_X:

$$V_L = D_L/D_X \tag{3}$$

$$V_R = D_R/D_X \tag{4}$$

Next, in a similar manner and referring to Figure 1(b), the loudspeakers are divided into a front pair (S_{FL} and S_{FR}) and rear pair (S_{RL} and S_{RR}) and the following scalars (V_F and V_B) are determined:

$$V_F = D_F/D_Y \tag{5}$$

$$V_B = D_B/D_Y \tag{6}$$

Finally, the amplitude levels for each of the four loudspeakers are determined as follows [Lam et al. 2012]:

$$S_{FL} = V_F \text{ x } V_L \tag{7}$$

$$S_{FR} = V_F \text{ x } V_R \tag{8}$$

$$S_{RL} = V_B \text{ x } V_L \tag{9}$$

$$S_{RR} = V_B \text{ x } V_R \tag{10}$$

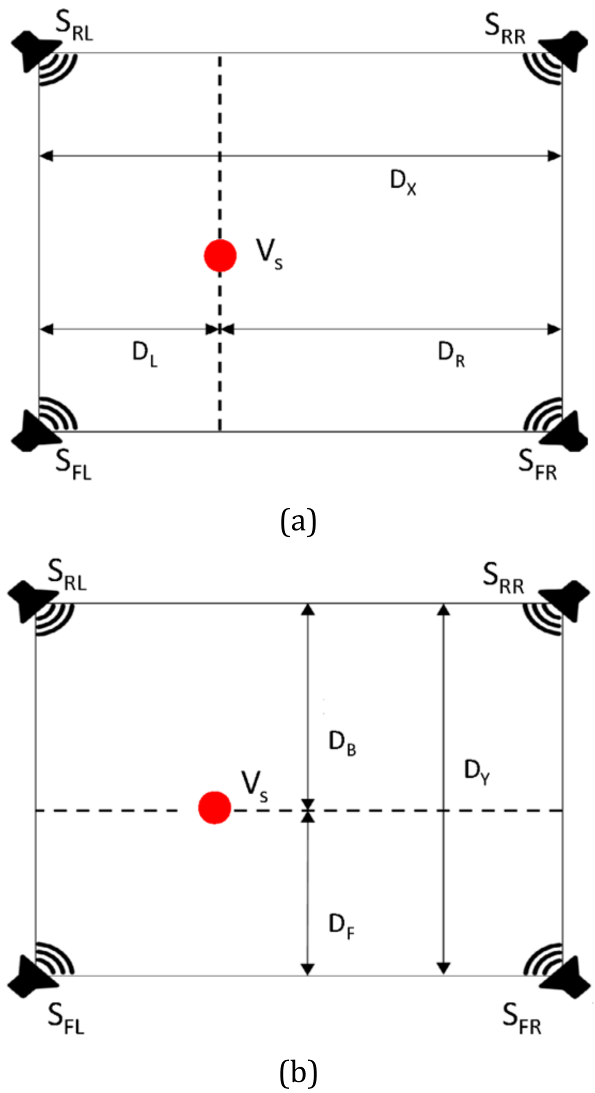


Figure 1. Bilinear interpolation amplitude panning example. (a) Horizontal scalars. (b) Vertical scalars.

2.2.2 Distance-Based Amplitude Panning

Distance-based amplitude panning (DBAP) is an amplitude panning method that does not rely on a particular loudspeaker setup or does it rely on the listener being in a particular location or "sweet spot". Any number of loudspeakers in any arbitrary configuration can be used with DBAP. The sound is panned between these loudspeakers based on the Euclidean distance between each one and the virtual sound source. The listener is free to move to any location around or among the loudspeakers, as DBAP is not reliant on a "sweet spot", where the listener needs to be to order to localize the virtual sound source properly, nor on the loudspeaker being equidistant from the listener. The loudspeaker weight for each loudspeaker is given as:

$$v_i = \frac{k}{2d_i a} \tag{11}$$

where v_i is the weight of a particular loudspeaker i, k is a coefficient based on the relative position of the virtual source and all loudspeakers, d_i is the Euclidean distance used in the calculation for loudspeaker i, and a is a roll-off coefficient. Each loudspeaker is assigned a field or *convex hull,* which determines how d_i is calculated. If the virtual sound source falls within a loudspeaker's hull, then $d_i$is the normal Euclidean distance between the virtual source and the loudspeaker. If the virtual sound source does not fall into a loudspeaker's hull, then a projection of the virtual source's location onto the hull is calculated. This projected point has the minimum Euclidean distance between the virtual sound source and any point inside the loudspeaker's hull. d_i is then calculated from the Euclidean distance between the virtual source and the projected point.

DBAP is enhanced using spatial blur and loudspeaker weights. Spatial blur is done simply by adding a constant value to the normal Euclidean distance formula. This prevents a sound from a virtual source from being played through only one loudspeaker, in cases where the location of the virtual source is the same as that of the loudspeaker. It also prevents division by zero errors in the amplitude calculations. Loudspeaker weights (gain values) can optionally be assigned to each loudspeaker for different virtual sound sources to restrict where a particular sound source can be heard and control the spatial spread for each loudspeaker [Lossius et al. 2009].

With the inverse distance-based amplitude panning method, the weight (gain) of each loudspeaker is scaled by the distance between the corresponding loudspeaker and its distance to the virtual sound source:

$$v_i = \frac{1}{(d_i^r + k)} \tag{12}$$

Here v_i is the weight (gain) of a particular loudspeaker i, v_a is the sound output from the virtual sound source, and d_i is the distance between the i^{th} loudspeaker and the virtual sound source, r is roll-off coefficient (and in the experiments conducted in this work, it was equal to 1.6 and derived empirically through informal testing), and k is a small constant value (= 0.001 in this work) mainly for preventing errors from division by zero. Each s_i is normalized and this normalized signal (sn_i) is applied to the corresponding loudspeaker.

Figure 2 illustrates the four loudspeakers surrounding the simulated environment's playing area. Included in the diagram are two markers, one red and one blue, each representing the position of a virtual sound source. In this particular example, the level of the sound corresponding to the blue virtual sound source (S_1) will be loudest on the rear left loudspeaker, the loudspeaker closest to it. Similarly, the level of the sound corresponding to the red virtual source (S_2) should be greatest at the loudspeaker closest to itself. This method is independent of each listener's physical position; the sounds are being simulated as coming from their position on the table, and if the user were to move farther away from the desk, the volume would naturally get quieter in relation to their distance away from the table [Lam et al. 2012].

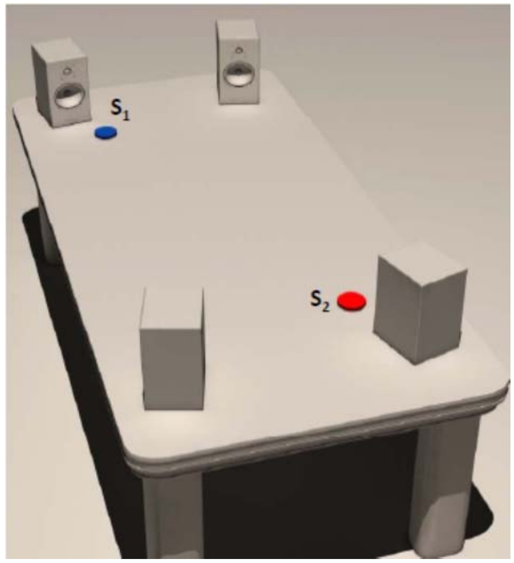


Figure 2. Virtual sound example illustrating the inverse-distance amplitude panning method.

2.2.3 Vector-Based Amplitude Panning

Another method of calculating the gain factors is the *vector-based amplitude panning* (VBAP) technique. This technique can be used with an arbitrary number of loudspeakers and supports both two and three-dimensional loudspeaker configurations. It allows the loudspeakers to be placed in any position given that they are nearly equidistant from the listener and that the listening room is not overly reverberant. In the stereo VBAP configuration the two-channel stereo setup is treated as a two-dimensional vector base defined by two unit length vectors, each vector with its origin at the listener and pointing to one of the two loudspeakers. A third unit vector points to the direction of the virtual sound source and is formulated as a linear combination of the two loudspeaker vectors. The two loudspeaker scaling factors (gains) are calculated using simple linear algebra techniques. The formulation of two-dimensional VBAP can be generalized to handle a three-dimensional loudspeaker configuration where three equidistant loudspeakers are conceptualized as positioned on an imaginary unit radius sphere. Three loudspeaker unit vectors point from the listener's position to each of the three loudspeakers, and a fourth unit vector points to the position of the virtual sound source. The virtual sound source can then be mapped to a location within the active triangle formed by the three loudspeakers. As with the two-dimensional stereo configuration, the vector pointing to the virtual sound source is expressed as a linear combination of the three loudspeaker vectors and the appropriate gain is calculated (using simple linear algebra techniques) and used to scale the signal output to each loudspeaker. The VBAP technique is a relatively simple and computationally efficient method allowing for the maximum virtual sound source localization accuracy possible with amplitude panning. In the three-dimensional configuration, maximum localization accuracy is proportional to the physical dimensions of the active triangle. Although the dimension of the active triangle can be decreased by increasing the number of loudspeakers, increasing the number of loudspeakers is sometimes impossible. As with all pair-wise and triplet-wise amplitude panning techniques, the virtual sound source spreads when it is panned between loudspeakers. Finally, although VBAP allows for accurate virtual sound source localization on the azimuthal plane (particularly near the median plane), the localization of virtual sound sources that do not lie on the azimuthal plane (e.g., non-zero elevation) is unpredictable since it is listener dependent. However, with a large number of loudspeakers, elevation localization becomes acceptable [Pulkki, 1997].

2.3 Surround Sound

Surround sound systems consists of any number of loudspeakers (usually three or more) surrounding a listener in order to provide them a greater sense of realism, giving the sound a greater physical presence and sense of realism whether it is from a musical performance or a movie. Surround sound systems allow the listener to hear sounds coming from all directions, not only in front as with traditional stereo setups. One of the earliest systems was the "Wall of Sound". It used an array of up to 80 microphones placed in a row horizontally across the front of an orchestra. The playback of sound over an equal number of loudspeakers produced very accurate and pleasing results. Such a large number of microphones resulted in a large sweet spot, providing the listener greater freedom to move about in the environment. However, the use of such a large number of microphones and loudspeakers was clearly impractical and so, the number of microphones and loudspeakers were reduced to three [Kapralos et al., 2003]. A complete and detailed discussion of surround sound is beyond the scope of this thesis. Below only a sample of some of the surround sound systems is provided but a more detailed overview is provided by Streicher and Everest [2006].

2.3.1 Quadraphonics

The Quadraphonic (also known as "Quadrisonics"), or "Quad" system was the first surround sound system to be introduced to consumers. Quadraphonic systems were developed to improve the limitations associated with monaural (single channel) and stereo-recorded sound, namely they did not provide the listener with the sense of physical presence of a live performance. In order to accomplish this, sounds would have to reach the listener from any direction in three-dimensional space, something clearly not achievable by monaural and stereo systems. Quadraphonic systems consist of four loudspeakers, two in front of the listener, front left (F_L) and front right (F_R) and two in back of the listener back left (B_L) and back right (B_R). The actual placement of the loudspeakers were not standardized, however they were typically placed at the four corners of a listening area, either facing inwards towards the listening area or the two rear loudspeakers could face the two front loudspeakers. In both cases the angle of separation between each of the loudspeakers is 90°, equally dividing the entire 360° space surrounding a listener. Quadraphonic systems were intended to allow for the perception of sound emanating from any direction on the plane in which the four loudspeakers were placed. Each of the loudspeakers received a signal which was previously recorded from a microphone element, intended to capture sounds emanating from the direction corresponding to the position of the loudspeakers.

Despite the promise of full 360° localization on the azimuthal plane (e.g., the ability to convey 3D sound), Quadraphonic systems were inaccurate and non-realistic in presenting a 3D sound source. Once encoded, the original signals can never be completely reconstructed as information will always be lost in the process, resulting in undesirable effects. As with any loudspeaker auditory display, crosstalk also degrades the performance and effects of the resulting playback sound. In a Quadraphonic setup, the sweet spot is located equidistant from all four loudspeakers (e.g., in the center of the listening area) and is rather narrow. Small head movements by the listener would result in dramatic changes in the desired effect. In addition, Quadraphonics did not find great success with consumers and lasted for a short time only. Given the widespread use of stereo equipment, consumers were reluctant to purchase new and expensive equipment to support Quadraphonics on their existing systems. Furthermore, different record companies and stereo equipment manufacturers each supported different incompatible encoding and decoding schemes, creating much confusion amongst consumers. Despite its shortcomings and lack of interest by consumers, Quadraphonics paved the way for the surround systems currently available [Kapralos et al., 2003].

2.3.2 Ambisonics

The ambisonic surround-sound system is a two-part process that addresses the problems of encoding sound directions and amplitudes and reproducing them over practical loudspeaker systems so that listeners can perceive sounds located in 3D space. This can occur over a 360-degree, horizontal-only soundstage (pantophonic systems), or over a full sphere (periphonic systems). The system encodes signals using a format known as B-format, which contains three channels for pantophonic systems and a further channel for periphonic, which includes information for height reproduction. These signals convey directionally encoded information with a resolution equal to first-order microphones (cardioid, figure-eight, etc.). Accurate reproduction requires at least four loudspeakers for sounds limited to the horizontal plane and eight if height is required. Additional loudspeakers may be needed for larger performance areas. It is not required to consider the actual details of the reproduction system during the original recording or synthesis of a sound field. The only exception to this is that the vertical dimension is essential if a height is required in the replay system. If the B-format specifications are followed, assuming suitable loudspeaker/decoder systems are used, then operation in different venues will be as similar as local acoustics allow. In all other respects the two parts of the system, encoding and decoding, are completely separate [Malham and Myatt 1995].

2.3.3 Wave Field Synthesis

The *wave field synthesis* method involves audio signals fed to a large number of closely-spaced loudspeakers so that a highly natural sound field is produced, including the reproduction of the wave front curvature that would result from real sound sources. Thus, wave field synthesis allows for the simultaneous reproduction of an arbitrary number of virtual sound sources. Wave field synthesis is based on Huygens' principle, which states that at every time instant every point on a primary wavefront can be thought of as a continuous emitter of secondary wavelets combining to produce a new wavefront in the direction of propagation. Given a wave field (that is specified with respect to pressure and normal particle velocity) on a boundary surface S of a closed volume V free of any sources, the sound pressure at any point within V can be determined. Loudspeakers that surround the listening area are driven to produce a volume flux proportional to the normal component of the particle velocity of the original wave field at each corresponding position. For practical purposes (e.g., hardware and computational power requirements) rather than using multiple planes of loudspeakers to enclose the listener, linear loudspeaker arrays are used. This leads to several problems, most notable of which is that sound reproduction is correct for wave field components in the horizontal plane only. Unlike other loudspeaker-based systems whose intended effect is restricted to the listener sweet spot, wave field synthesis systems generate a wave field with natural time and space properties that envelops the listening area.

Multiple listeners are free to move about within this area without fear of losing the correct acoustical impression. This has made wave field synthesis an attractive approach for applications such as sound enhancement in theaters, multipurpose auditoriums, and the reproduction of multichannel recordings. However, wave field synthesis is impractical in many virtual reality settings due to several inherent limitations, most notably the requirement that the distance between loudspeakers be as small as possible in order to avoid spatial aliasing; the highest frequency that can be represented is inversely proportional to the spacing between loudspeakers. This results in the requirement for a large number of loudspeakers and extensive computation [Kapralos et al. 2008].

CHAPTER 3 – EXPERIMENTS AND DEMO

3.1 Experimental Goals

The goals of the experiments conducted in this thesis were two-fold. First, the experiments were conducted to determine the effectiveness of both the bilinear and inverse-distance amplitude panning methods (described in Sections 2.2.1 and 2.2.2 respectively) with respect to their ability to simulate the location of a sound source using four loudspeakers located within an area between them. Second, the experiments tested the effectiveness of different loudspeaker configurations in order to determine which one would be best suited for amplitude panning.

3.2 Experimental Setup

All participants were unpaid volunteers who were either researchers or students from the University of Ontario Institute of Technology. Participants reported no history of auditory disease or disorders. All experiments abided by the University of Ontario Institute of Technology Ethics Review process for experiments involving human participants.

The system used in the experiments is intended to accommodate multiple users (one to four) and consists of a tabletop computer, and four loudspeakers (currently, JVC SX-XSW31 are being used). The multi-touch table is a custom built display system (see Figure 3 for the tabletop computer and experimental setup although for the purposes of the experiments described in this thesis, only sound was required, that is, there were no visuals and there was no touch interface employed). An ultra-short throw projector is used for rear projection (Hitachi CP-A100 which allows great control of the projection size). An Optitrack camera is used to detect/track user touch on the screen as it provides direct illumination with its built-in IR LEDs, operates at 100 fps and provides decent resolution/performance trade-offs. The Optitrack camera has on-board processing that reduces the overall latency of the touch location sensing to high interactive rates. The open source

Touchlib project integrates with the Optitrack camera data to provide centroid determination of the multiple finger touch locations.

The computer used to power the tabletop computer is a Dell Inspiron 560 with an M-Audio Delta 1010LT sound card installed on the system, which provides the outputs to the four loudspeakers. The loudspeakers' outputs were controlled using custom software using the *BASS 2.4.6* audio library. *BASS* is a simple cross-platform library that provides audio playback and recording functionalities in a variety of different formats. [Un4seen Developments, 2012] The software's main function in the experiments was to calculate the appropriate output level for each of the four individual loudspeakers given their positions, the position of the virtual sound source and the interpolation method used and subsequently played a sound sample through the loudspeakers at the calculated levels. The software was also used to actually carry out the experiments; it allowed for the generation of a random sequence of nodes from which a virtual sound is played (described in more details below) and in later revisions being able to automatically run and record participant responses. Figure 4 illustrates the software interface. It provides the experimenter with useful information regarding the state of sound system, including the output levels of each loudspeaker, the panning method currently used, and the location of the virtual sound source.

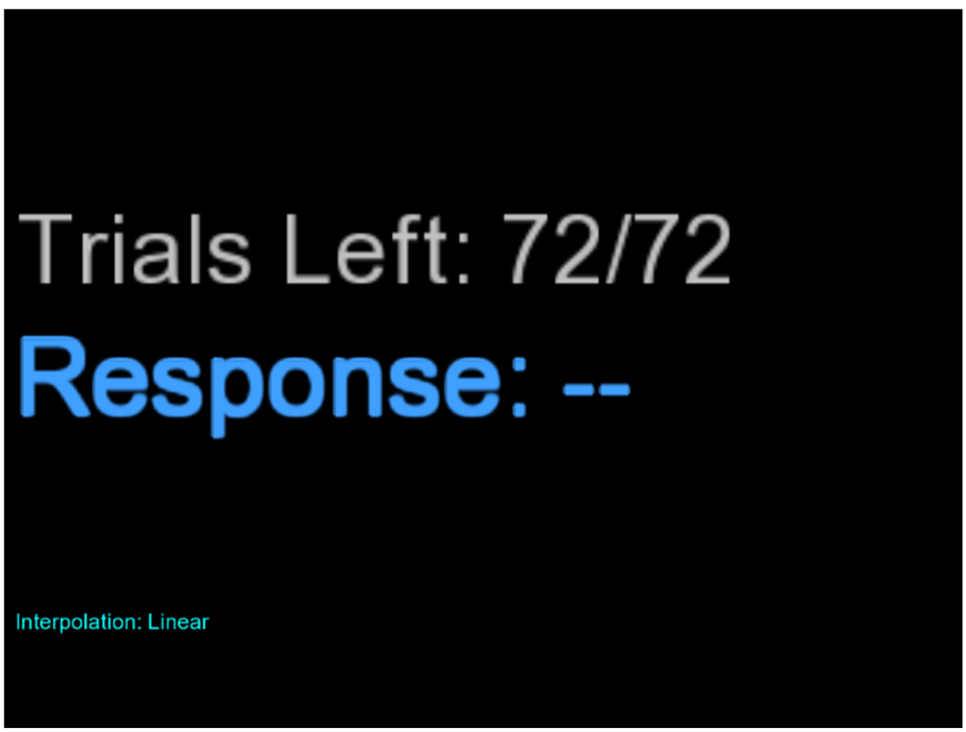
(a)

(b)

(c)

Figure 3. (a) Overview of the tabletop system. (b) Close-up of the touchscreen. (c) Inside the table.

(a)



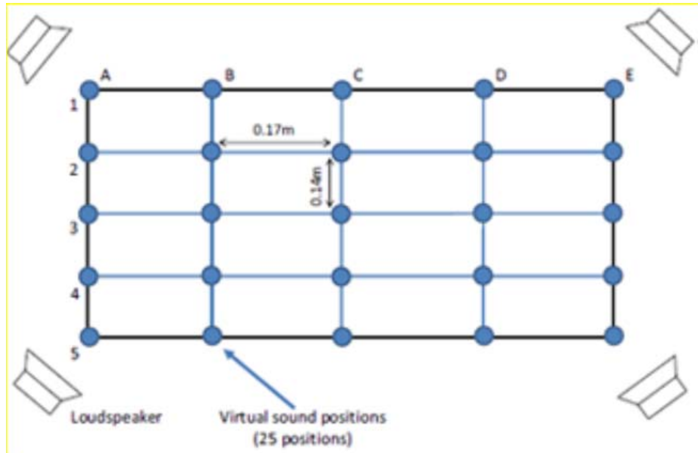
(b)

Figure 4. The user interface of the experiment software. (a) Full interface for testing loudspeakers and panning methods. (b) Minimal interface for running experiments.

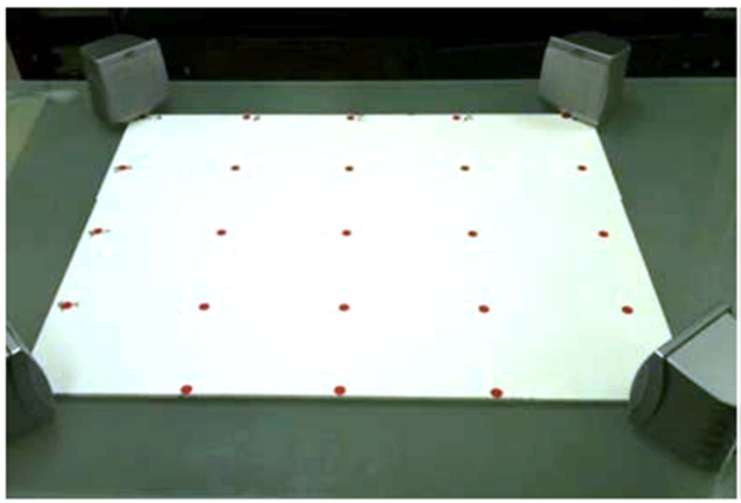
3.3 Experiment One: Bilinear Interpolation with the Rectangular Inward Loudspeaker Configuration

3.3.1 Auditory Stimulus

A total of eight unpaid volunteers participated in this experiment with an average participant age of 26. The auditory stimulus consisted of a broadband white-noise signal sampled at a rate of 44.1 kHz and band-pass filtered using a 256-point Hamming windowed FIR filter with low and high frequency cut-offs of 200 Hz and 10 kHz respectively. The auditory stimulus was output through JVC SX-XSW-31 loudspeakers (four loudspeakers in total). For the purpose of this experiment, the loudspeakers were placed on the four corners of the table (surface) as shown in Figure 5. The duration of the auditory stimuli was 2s and the average level (SPL) of the sound stimuli, measured with a Radio Shack sound level meter (model 33-2055) with an A-weighting, placed at the location where the participant's head would be was 68 dB. The experiment took place in a large laboratory at the University of Ontario Institute of Technology (room dimensions of 40.0m × 20.0m × 9.5 m). Although the room itself contained a variety of equipment including workstations, tables, chairs, etc. for the duration of the experiment effort was taken to limit the amount of external noise (e.g., equipment was turned off). The average background noise level, also measured at the location where the participant's head would be (and measured in the absence of the sound stimulus) was 59 dB (the maximum and minimum background noise level was 62 dB and 57 dB respectively). The loudspeakers for the surface computer setup are intended to be mounted on stands and positioned at each corner of the smart table at a height equivalent to the height of the seated participant's ears. However, for the purposes of this experiment, the loudspeakers were placed on the surface of the table itself. Doing so allowed each loudspeaker to be placed directly at the location of one of the virtual sound source positions (at each corner of the surface) and provided the participants with a simple and intuitive reference sound level for each corner position.



(a)



(b)

Figure 5. Experimental setup. (a) Grid of the 25 virtual sound source positions and loudspeaker setup. (b) Actual setup.

3.3.2 Experimental Method

The experiment consisted of 25 trials and each participant participated in the experiment individually. Participants were seated on a chair around the horizontal smart table setup as shown in Figure 5 (with four loudspeakers positioned at each corner of the surface of the smart table) for the duration of the experiment. Only auditory stimuli were present (i.e., no visual stimuli). In each trial, participants were presented with an auditory stimulus that was spatialized using the distance-based amplitude panning technique described above so that it appeared as if the sound source originates at one of 25 positions across on the surface of the table. The virtual sound sources were positioned on a grid where the horizontal and vertical separation was 0.17 m and 0.14 m respectively (see Figure 5(a)). Each of the 25 virtual sound source positions was indicated with a red dot. The participant's task for each trial was to indicate which of the 25 positions they believed the virtual sound source was emanating from. They indicated their choice by choosing one of the 25 positions and indicating this to the experimenter who was recording their choices. Once their choice was recorded, this indicated the end of the trial. The next trial began after the participant indicated to the experimenter that they were ready for the next trial. The ordering in which trials (virtual sound source positions) were presented to the participants was randomly chosen. Prior to the start of the experiment, participants were presented with the auditory stimulus at each of the four corner positions (individually, one after the other) to provide them with a reference. All participants were provided three test trials (where the virtual sound source position was randomly chosen) prior to beginning the experiment.

3.3.3 Results

The Euclidean distance between the *actual* virtual sound source position and the *perceived* virtual sound source position (i.e., the position that the participants perceived the sound source to be at) is used to measure the accuracy of the participants' ability to correctly determine the virtual sound source position. Ideally, the actual and perceived positions would be identical and the Euclidean distance (and hence error) is equal to zero. Graphical illustrations of the average error for each participant (averaged across each of the 25 virtual sound source positions) are provided in Figures 6 and 7. The average error (Euclidean distance) and standard deviation for each of the 25 virtual sound source positions (averaged across each of the eight participants) is summarized in Table 1. An examination of Table 1 indicates that for the majority of the virtual sound source positions, the perceived virtual sound source position was incorrect but close (within two positions) of the actual position. Further examination indicates that the largest errors occurred in the positions corresponding to rows 4, 5 and columns C, D (these positions are closest to the participant). The most accurate responses occurred for virtual sound source positions across the borders of the surface (the sides to the left, right, and top front) [Lam et al. 2010].

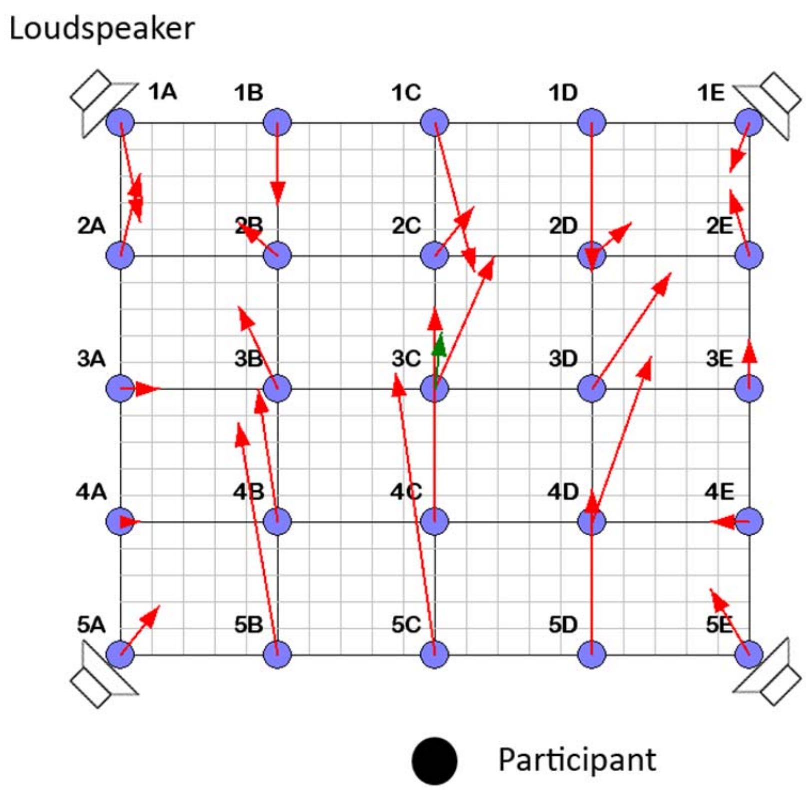


Figure 6. Results for inward-facing loudspeaker configuration. Error vector plot. The red arrows show the error for each of the virtual sound source positions while the green arrow in the middle represents the average of all the red arrows.

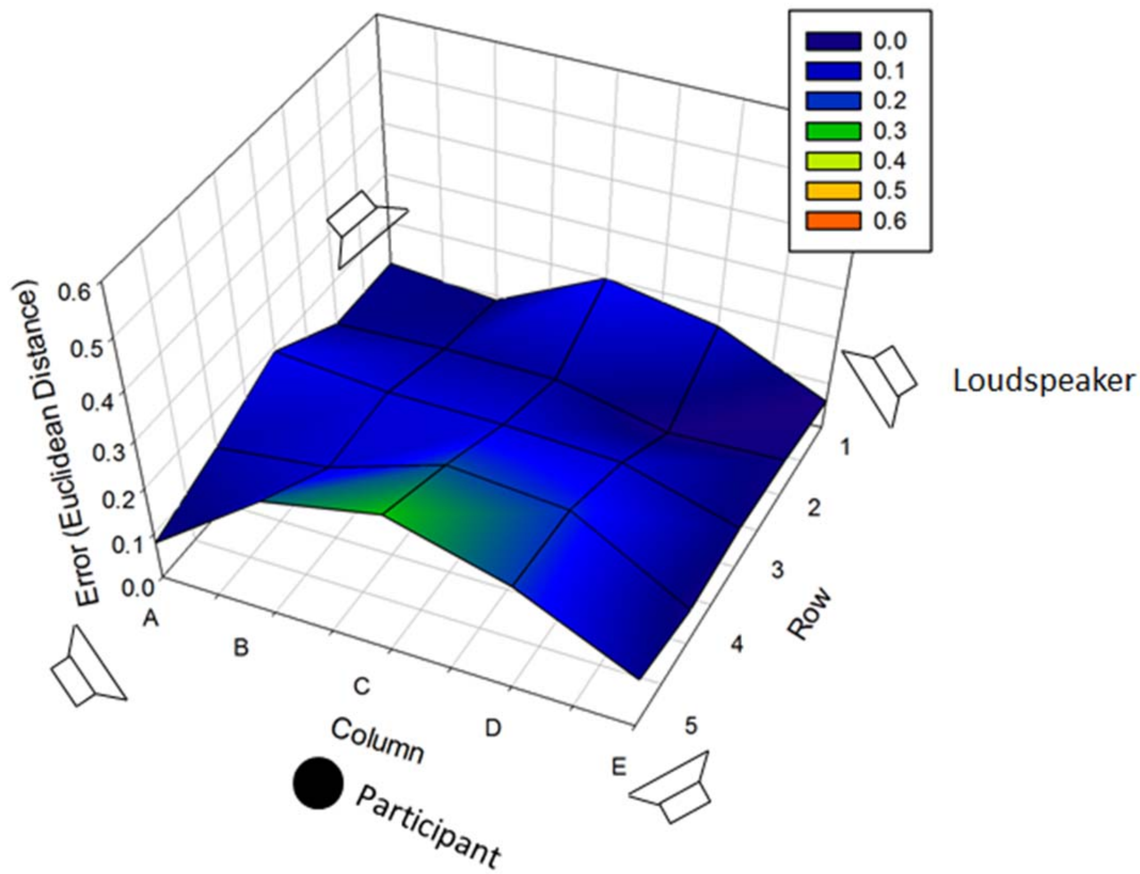


Figure 7. Results for inward-facing loudspeaker configuration. Average error (Euclidean distance or the difference between the actual and perceived virtual sound source positions, measured in meters) for virtual sound source position (averaged across each of the eight participants).

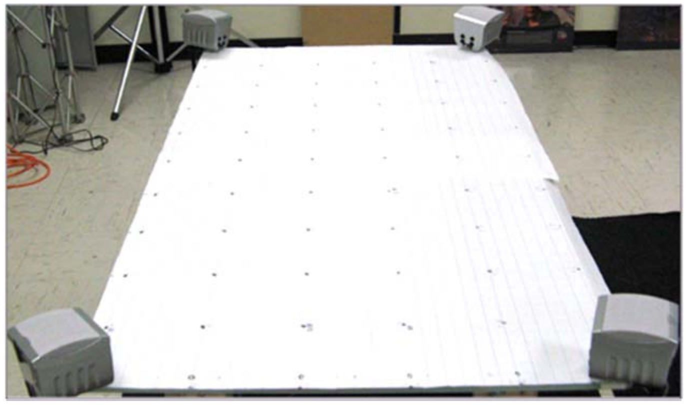
	A	B	C	D	E
1	0.11 ±0.15	0.09 ±0.10	0.20 ±0.19	0.16 ±0.19	0.06 ±0.09
2	0.11 ±0.07	0.12 ±0.08	0.12 ±0.08	0.07 ±0.09	0.08 ±0.09
3	0.19 ±0.08	0.17 ±0.15	0.17 ±0.16	0.16 ±0.15	0.09 ±0.10
4	0.13 ±0.13	0.16 ±0.12	0.24 ±0.15	0.22 ±0.14	0.08 ±0.08
5	0.08 ±0.14	0.25 ±0.18	0.30 ±0.21	0.23 ±0.15	0.11 ±0.13

Table 1. Average error of inward-facing loudspeaker configuration (Euclidean distance or the difference between the actual and perceived virtual sound source positions) and standard deviation for virtual sound source position (averaged across each of the eight participants).

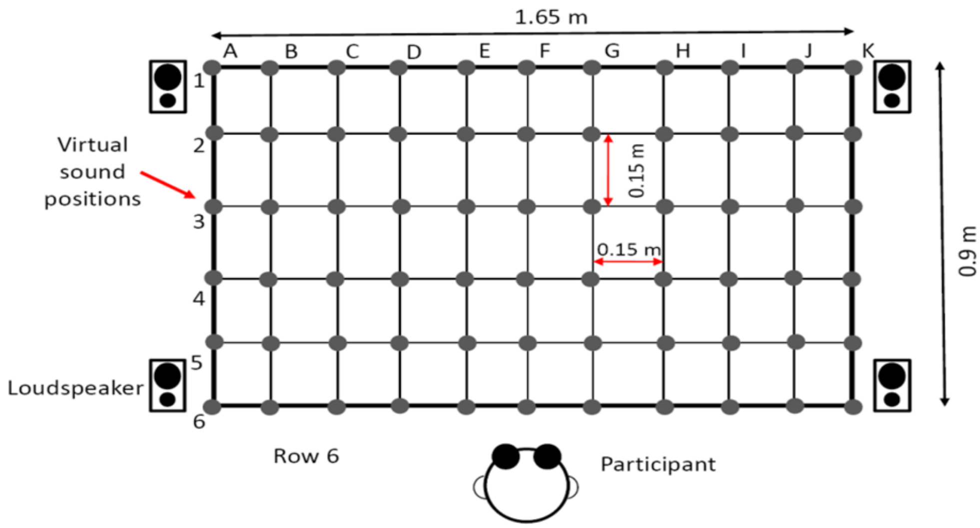
3.4 Experiment Two: Bilinear Interpolation with the Rectangular Upward Loudspeaker Configuration

3.4.1 Auditory Stimulus

Eight unpaid volunteers participated in this experiment with an average participant age of 25. The auditory stimulus consisted of a broadband white noise signal sampled at a rate of 44.1 kHz and band-pass filtered using a 256-point Hamming windowed FIR filter with low and high frequency cut-offs of 200 Hz and 10 kHz respectively. The auditory stimulus was output through JVC SX-XSW 31 loudspeakers (four loudspeakers in total). For the purpose of this experiment, a loudspeaker was placed at each of the four corners of the table and oriented such that they were facing upwards (see Figure 8). The loudspeakers were set to a height equal to 1.0 m (slightly higher than the 0.90 m of the table). The duration of the auditory stimuli was 2 s and the average level (SPL) of the sound stimuli, measured with a Radio Shack sound level meter (model 33-2055) with a weighting, placed at the location where the participant's head would be was 68 dB. The experiment took place in a large laboratory at the University of Ontario Institute of Technology (room dimensions of 40.0m x 20.0m x 9.5 m). Although the room itself contained a variety of equipment including workstations, tables, chairs, etc. for the duration of the experiment effort was taken to limit the amount of external noise (e.g., equipment was turned off). The average background noise level, also measured at the location where the participant's head would be (and measured in the absence of the sound stimulus) was 57 dB (the maximum and minimum background noise level was 63 dB and 55 dB respectively).



(a)



(b)

Figure 8. Upward loudspeaker setup. (a) Actual setup. (b) The 66 virtual sound source positions and loudspeaker setup.

3.4.2 Experimental Method

Participants were seated on a chair around the tabletop computer setup as shown in Figure 8(a) (with four loudspeakers positioned at each corner of the surface of the table) for the duration of the experiment. Only auditory stimuli were present (i.e., no visual stimuli) for the duration of the experiment. In each trial, participants were presented with an auditory stimulus that was spatialized using the bilinear interpolation amplitude panning technique described in Section 3 so that the apparent location of the sound source originates at one of 66 positions across the surface of the table. The virtual sound sources were positioned on a grid where the horizontal and vertical separation was 0.15m and 0.15m respectively. Figure 8(b) provides a graphical illustration of the experimental setup. The sound was spatialized to each of the 66 virtual sound source positions twice (i.e., 66 positions repeated twice for each position) yielding a total of 132 trials and the ordering of each trial was random. The experiment took approximately 25 minutes to complete and all participants completed it in a single session. The participant's task for each trial was to indicate which of the 66 positions they perceived the virtual sound source was emanating from. They indicated their choice by choosing one of the 66 positions and indicating this to the experimenter who was recording their choices. Once their choice was recorded, this indicated the end of the trial. The next trial began after the participant indicated to the experimenter that they were ready for the next trial. The ordering in which trials (virtual sound source positions) were presented to the participants was randomly chosen. Prior to the start of the experiment, participants were presented with the auditory stimulus at each of the four corner positions (individually, one after the other) to provide them with a reference. All participants completed three test trials (where the virtual sound source position was randomly chosen) prior to beginning the experiment.

3.4.3 Results

The Euclidean distance between the actual virtual sound source position and the perceived virtual sound source position (i.e., the position that the participants perceived the sound source to be at) is used to measure the accuracy of the participant's ability to correctly determine the virtual sound source position. Ideally, the actual and perceived positions would be identical and the Euclidean distance (and hence error) is equal to zero. Graphical illustrations of the average error for each participant (averaged across each of the 25 virtual sound source positions) are provided in Figures 9 and 10. The average error (Euclidean distance) and standard deviation for each of the 66 virtual sound source positions (averaged across each of the eight participants) is summarized in Table 2. Examination of Table 2 indicates that the average (mean) error ranged from 0.06m to 0.49 m. Furthermore, the largest errors occurred in Rows 5 and 6 towards the middle of the rows (the largest five errors being was 0.49 m, 0.44 m, 0.44 m, 0.43 m, and 0.38 m, at positions 6E, 5E, 6I, 6H, and 5I respectively). These positions are closest to the participants. The most accurate responses occurred for virtual sound source positions across the borders of the surface (the sides to the left, right, and top front).

Graphical illustrations of the average error for each participant (averaged across each of the 66 virtual sound source positions) are provided in Figures 9 and 10. For each participant, the average error ranged from 0.20m to 0.28 m. In other words, given the grid spacing of 0.15m x 0.15 m, participants were able to localize the sound source to within two positions of the actual virtual sound source.

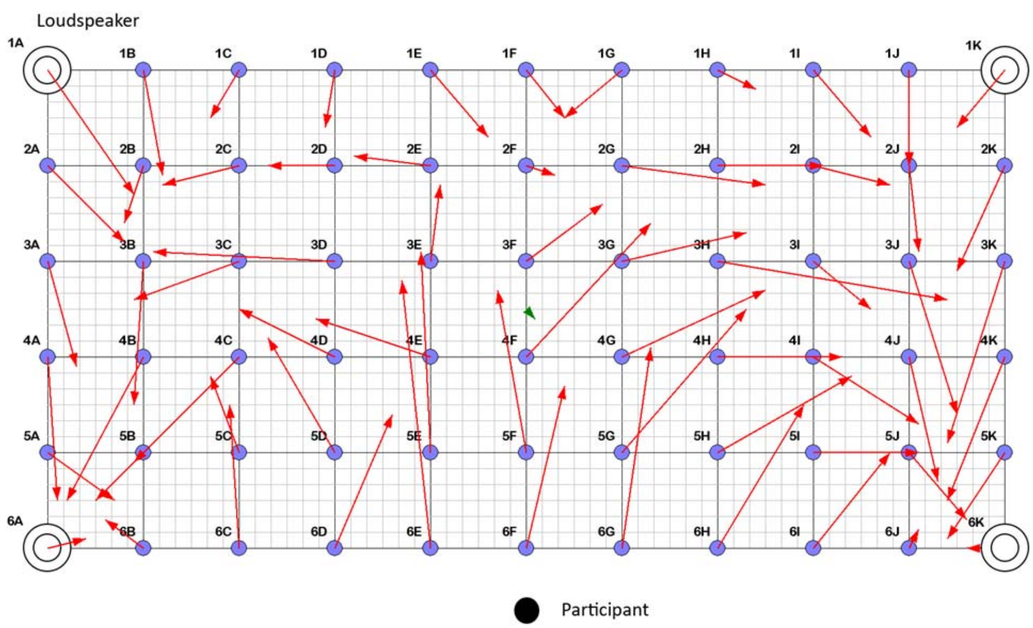


Figure 9. Results for the upward loudspeaker configuration. Error vector plot. The red arrows show the error for each of the virtual sound source positions while the green arrow in the middle represents the average of all the red arrows.

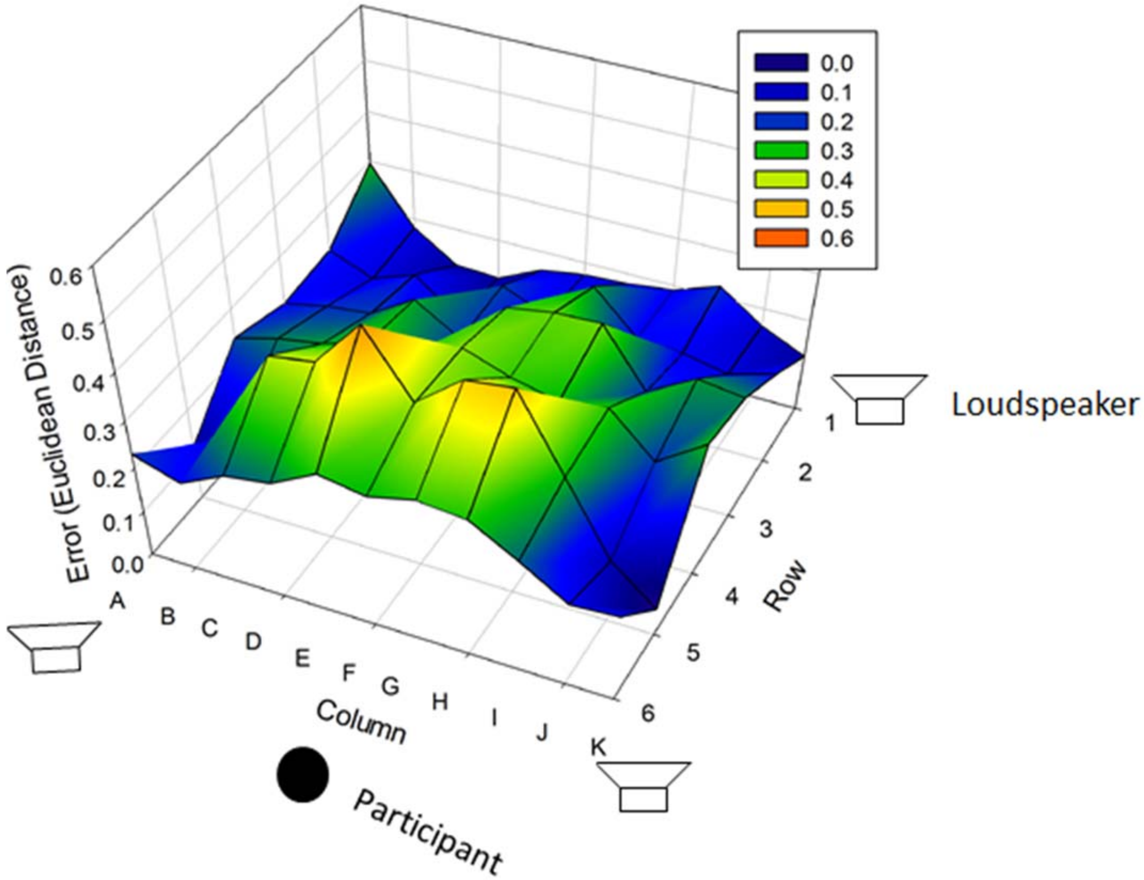


Figure 10. Results for the upward loudspeaker configuration. Average error (Euclidean distance or the difference between the actual and perceived virtual sound source positions, measured in meters) for virtual sound source position (averaged across each of the eight participants). Not shown in diagram is loudspeaker at position 1A.

	A	B	C	D	E	F	G	H	I	J	K
1	0.29 ±0.17	0.18 ±0.00	0.12 ±0.12	0.12 ±0.08	0.16 ±0.13	0.17 ±0.13	0.18 ±0.11	0.19 ±0.12	0.22 ±0.21	0.16 ±0.15	0.12 ±0.15
2	0.21 ±0.12	0.17 ±0.12	0.21 ±0.09	0.13 ±0.11	0.23 ±0.14	0.19 ±0.14	0.28 ±0.16	0.20 ±0.14	0.19 ±0.11	0.17 ±0.10	0.20 ±0.11
3	0.20 ±0.14	0.22 ±0.11	0.23 ±0.14	0.29 ±0.14	0.23 ±0.12	0.32 ±0.19	0.34 ±0.18	0.34 ±0.12	0.22 ±0.12	0.28 ±0.12	0.27 ±0.12
4	0.24 ±0.08	0.26 ±0.10	0.28 ±0.13	0.31 ±0.11	0.32 ±0.11	0.35 ±0.13	0.32 ±0.15	0.30 ±0.12	0.31 ±0.14	0.23 ±0.10	0.29 ±0.14
5	0.20 ±0.10	0.17 ±0.08	0.26 ±0.19	0.32 ±0.19	0.44 ±0.22	0.36 ±0.20	0.35 ±0.15	0.38 ±0.17	0.26 ±0.11	0.16 ±0.08	0.24 ±0.10
6	0.12 ±0.12	0.17 ±0.09	0.37 ±0.17	0.39 ±0.27	0.49 ±0.21	0.36 ±0.21	0.43 ±0.20	0.44 ±0.20	0.29 ±0.16	0.14 ±0.06	0.06 ±0.10

Table 2. Average error of the upward loudspeaker configuration (Euclidean distance or the difference between the actual and perceived virtual sound source positions) and standard deviation for virtual sound source position (averaged across each of the eight participants).

3.5 Experiment Three and Four: Bilinear and Inversed Distance-Based Interpolation with the Diamond Loudspeaker Configuration

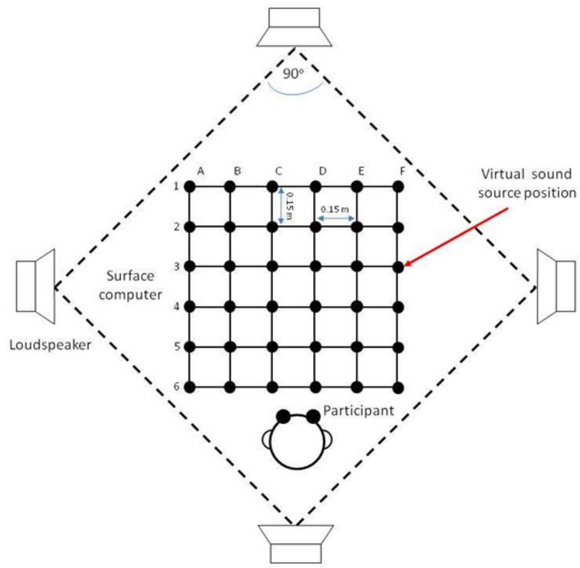
3.5.1 Auditory Stimulus

A total of 10 volunteers participated in this experiment. The average age of the participants was 24 years old. None of the participants reported any history of auditory disease or disorders. The auditory stimulus consisted of a broadband white noise signal sampled at a rate of 44.1 kHz and band-pass filtered using a 256-point Hamming windowed FIR filter with low and high frequency cut-offs of 200 Hz and 10 kHz respectively. The auditory stimulus was output through JVC SX-XSW 31 loudspeakers (four loudspeakers in total). The loudspeakers were placed one on each of the sides of the surface in a square diamond configuration (see Figure 11(a)). The loudspeakers were set to a height equal to 1.0 m (slightly higher than the 0.90 m of the table/surface). The duration of the auditory stimuli was 2 s and the average level (SPL) of the sound stimuli, measured with a Radio Shack sound level meter (model 33-2055) with an A-weighting, placed at the location where the participant's head would be was 68 dB. The experiments took place in a large laboratory at the University of Ontario Institute of Technology (room dimensions of 40.0 m × 20.0 m × 9.5 m). Although the room itself contained a variety of equipment including workstations, tables, chairs, etc. for the duration of the experiment effort was taken to limit the amount of external noise (e.g., equipment was turned off). The average background noise level, also measured at the location where the participant's head would be (and measured in the absence of the sound stimulus) was 57 dB (the maximum and minimum background noise level was 63 dB and 55 dB respectively).

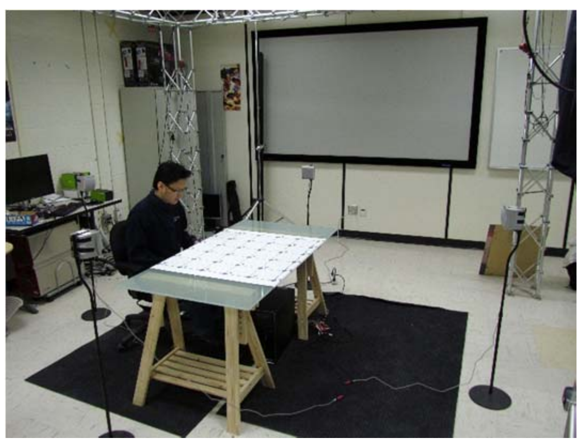
3.5.2 Experimental Method

Participants were seated on a chair on one of the sides of the horizontal surface. For each trial, participants were presented with an auditory stimulus that was spatialized using one of the positions on the surface using either the bilinear interpolation or the inverse distance amplitude panning method. Only auditory stimuli were provided but the subjects were not blindfolded and could view the room. The virtual sound sources were positioned on a grid where the horizontal and vertical separation was 0.15 m and 0.15 m respectively, resulting in a total of 36 virtual sound source positions (see Figure 11(a)). Each of the 36 grid positions was clearly labeled in large text (the grid itself was professionally printed and centered about a table; these experiments only considered sound localization on a horizontal surface hence, the actual surface computer itself was not needed and not used).

For each trial, a test sound was generated and participants were instructed to choose from a set of possible grid locations clearly marked on the surface (rows were marked with numbers beginning with "1" while columns were marked with letters beginning with "A", as shown in Figure 11(a)) and enter their choice of row and column using a standard computer keyboard. The next trial began after the participant entered their choice and pressed the "Enter" key on the keyboard to indicate that they were ready for the next trial. A total of 36 grid positions (spatial sound sources) were considered and audio simulation was repeated two times for each of the two amplitude panning methods considered leading to a total of 104 trials (i.e., 36 grid positions × 2 repetitions × 2 amplitude panning methods). Conditions were presented to the participants in random order. The experiment took approximately 20 minutes to complete and was completed in a single session. Prior to the start of the experiment, participants were presented with the auditory stimulus spatialized to each of the four corner positions (individually, one after the other) to provide them with a reference. All participants were also provided three test trials (where the virtual sound source position was randomly chosen) prior to beginning the experiment.



(a)



(b)

Figure 11. Diamond loudspeaker setup. (a) The 36 virtual sound source positions and loudspeaker setup. (b) Actual setup.

3.5.3 Results

The Euclidean distance between the actual virtual sound source position (i.e., the location that the sound was spatialized to using the bilinear interpolation/inverse distance amplitude panning method) and the perceived virtual sound source position (i.e., the position that the participants perceived the sound source to be emanating from) is used to measure the accuracy of the participant's ability to determine the virtual sound source position. Ideally, the actual and perceived positions would be identical and the Euclidean distance (and hence error) will equal zero indicating participants were able to correctly localize the virtual sound source each and every trial for both presentation techniques.

3.5.4 Bilinear Interpolation Amplitude Panning

The average error (Euclidean distance) and standard deviation for each of the 36 virtual sound source positions (averaged across each of the eight participants) is summarized in the plot of Figures 12 and 13 and Table 3. The average error across each of the 36 positions ranged from 0.11 m to 0.47 m with an average of 0.23 m $\pm$0.07 m. Given the grid spacing of 0.15 m × 0.15 m, participants were able to localize the sound source to within two positions of the actual virtual sound source. Examining the plot of Figure 12, it is evident that the error was largest for each of the four corners of the surface. The positions corresponding to the five largest errors are: (6A; 0.47 m $\pm$ 0.13 m), (1A; 0.37 m $\pm$ 0.13 m), (5A; 0.34 m $\pm$ 0.16 m), (6F; 0.31 m $\pm$ 0.10 m), and (1F; 0.31 m $\pm$ 0.18 m).

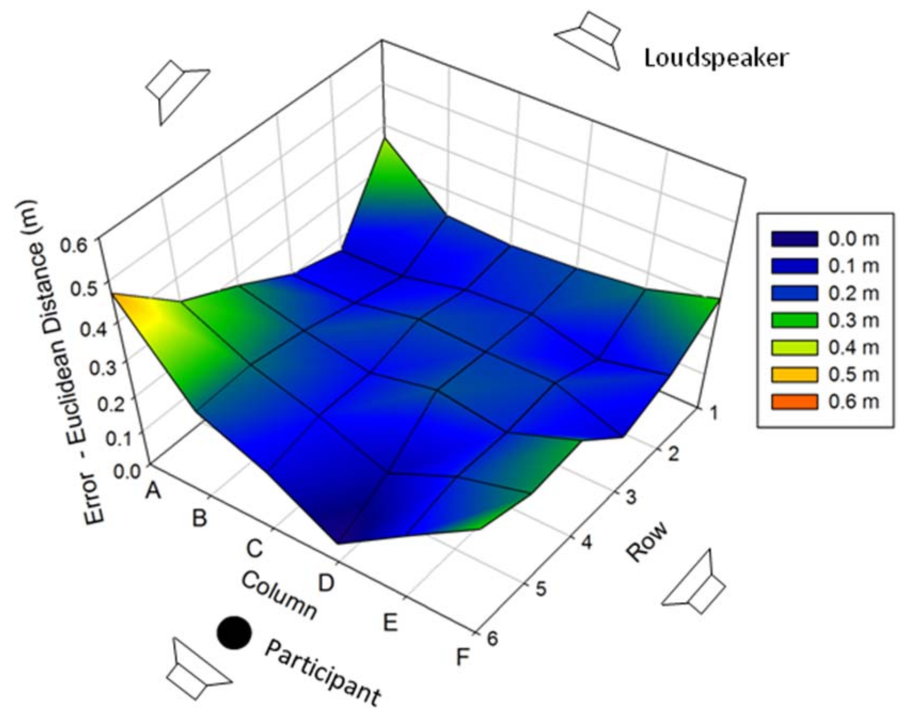


Figure 12. Results for the diamond configuration with bilinear interpolation amplitude panning. Average error (Euclidean distance or the difference between the actual and perceived virtual sound source positions, measured in meters) for virtual sound source position (averaged across each of the 10 participants).

	A	B	C	D	E	F
1	0.37 ± 0.13	0.23 ± 0.17	0.22 ± 0.14	0.24 ± 0.13	0.26 ± 0.16	0.31 ± 0.18
2	0.16 ± 0.11	0.17 ± 0.16	0.21 ± 0.12	0.22 ± 0.11	0.18 ± 0.15	0.22 ± 0.12
3	0.20 ± 0.10	0.20 ± 0.12	0.24 ± 0.18	0.23 ± 0.12	0.23 ± 0.15	0.16 ± 0.10
4	0.28± 0.17	0.21 ± 0.10	0.20 ± 0.12	0.24 ± 0.10	0.21 ± 0.11	0.28 ± 0.15
5	0.34 ± 0.16	0.26 ± 0.12	0.20 ± 0.08	0.11 ± 0.10	0.22 ± 0.13	0.27 ± 0.13
6	0.47 ± 0.13	0.25 ± 0.13	0.17 ± 0.12	0.06 ± 0.09	0.19 ± 0.13	0.31± 0.10

Table 3. Average error of the diamond loudspeaker configuration with bilinear interpolation amplitude panning (Euclidean distance or the difference between the actual and perceived virtual sound source positions) and standard deviation for virtual sound source position (averaged across each of the 10 participants).

Figure 13 provides a "vector plot" of the average error for each of the 36 positions whereby the magnitude and direction of the error associated with each of the 36 positions is shown. The red arrows show the error for each of the virtual sound source positions while the green arrow in the middle represents the average of all the red arrows.

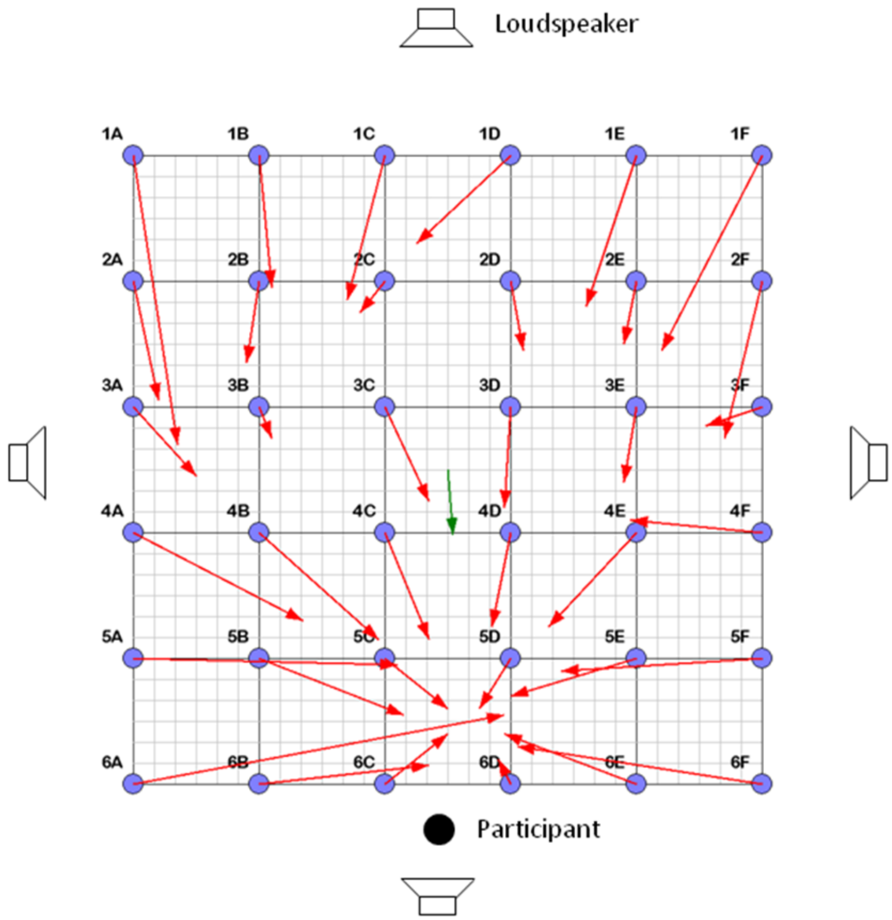


Figure 13. Results for the diamond configuration with bilinear interpolation amplitude panning. Error vector plot. The red arrows show the error for each of the virtual sound source positions while the green arrow in the middle represents the average of all the red arrows.

3.5.5 Inverse Distance Amplitude Panning

The average error (Euclidean distance) and standard deviation for each of the 36 virtual sound source positions (averaged across each of the 10 participants) is summarized in the plot of Figures 14 and 15 and Table 4. The average error across each of the 36 positions ranged from 0.13 m to 0.44 m with an average of 0.24 m ±0.07. Given the grid spacing of 0.15 m × 0.15 m, participants were able to localize the sound source to within two positions of the actual virtual sound source. Examination of the plot of Figure 14 reveals that the largest errors occurred at three of the surface corners (towards corners at position 6A, 1F, and 6F). The positions corresponding to the five largest errors are: (6A; 0.44 m ± 0.14 m), (5A; 0.36 m ± 0.17 m), (1E; 0.36 m ± 0.16 m), (6F; 0.34 m ± 0.09 m), and (1F; 0.34 m ± 0.17 m)

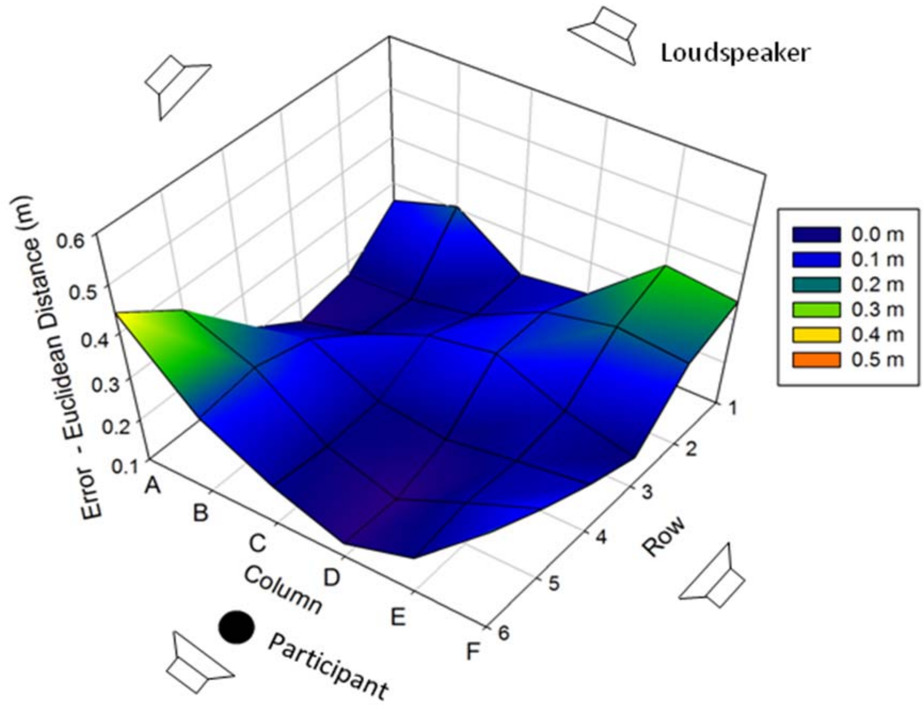


Figure 14. Results for the diamond configuration with inverse-distance amplitude panning. Average error (Euclidean distance or the difference between the actual and perceived virtual sound source positions, measured in meters) for virtual sound source position (averaged across each of the 10 participants).

	A	B	C	D	E	F
1	0.26 ± 0.20	0.31 ± 0.18	0.21 ± 0.16	0.23 ± 0.12	0.36 ± 0.16	0.34 ± 0.17
2	0.17 ± 0.11	0.17 ± 0.10	0.20 ± 0.10	0.28 ± 0.14	0.31 ± 0.17	0.30 ± 0.16
3	0.15 ± 0.09	0.18 ± 0.13	0.25 ± 0.15	0.28 ± 0.12	0.21 ± 0.11	0.17 ± 0.13
4	0.22 ± 0.13	0.27 ± 0.12	0.19 ± 0.12	0.17 ± 0.10	0.20 ± 0.12	0.22 ± 0.14
5	0.36 ± 0.17	0.30 ± 0.13	0.18 ± 0.09	0.13 ± 0.10	0.21 ± 0.07	0.28 ± 0.12
6	0.44 ± 0.14	0.28 ± 0.08	0.19 ± 0.12	0.13 ± 0.15	0.19 ± 0.12	0.34 ± 0.09

Table 4. Inverse distance amplitude panning results. Average error (Euclidean distance or the difference between the actual and perceived virtual sound source positions) and standard deviation for virtual sound source position (averaged across each of the 10 participants).

The vector of the average error for each of the 36 positions is provided in Figure 14. As previously described, the magnitude and direction of the error associated with each of the 36 positions is shown.

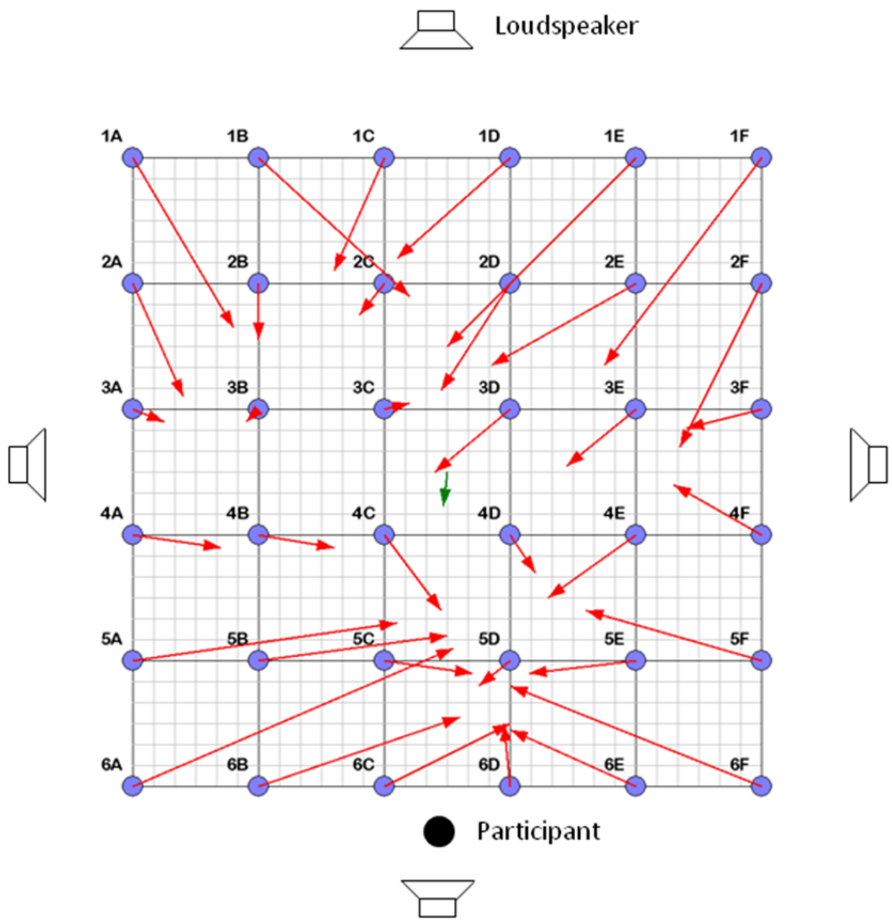


Figure 15. Results for the diamond configuration with inverse-distance amplitude panning. Error vector plot. The red arrows show the error for each of the virtual sound source positions while the green arrow in the middle represents the average of all the red arrows.

3.5.6 Comparison

Here, a comparison between the error (across each of the 36 positions considered) associated with the two amplitude panning methods is provided. Based on average error across all positions, both amplitude panning methods are very similar (0.23 m ±0.07 vs. 0.24 m ±0.07 respectively). A graphical comparison in the form of the absolute difference between the averages for each of the 36 positions considered is provided in Figure 16. A graphical inspection of the difference is fairly consistent and without any large spikes for any of the positions.

An independent-samples t-test was conducted to compare sound localization accuracy on a (horizontal) surface using bilinear interpolation and inverse distance amplitude panning methods, both with a diamond loudspeaker configuration. There was a non-significant difference in the scores between the bilinear interpolation amplitude ($M = 0.23$, $SD = 0.07$) and inverse distance ($M = 0.24$, $SD = 0.07$) amplitude panning methods, $t = 0.52$, $p = 0.60$. [Lam et al. 2012]

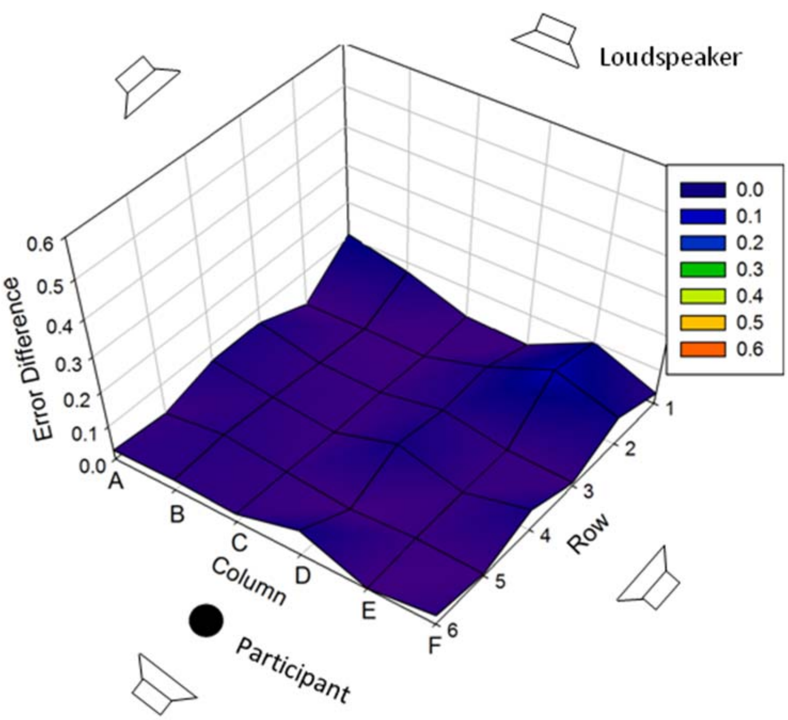


Figure 16. The difference in error between the results of bilinear interpolation and inverse distance amplitude panning methods for each of the 36 positions.

3.6 Experiment Five: "Ground Truth"

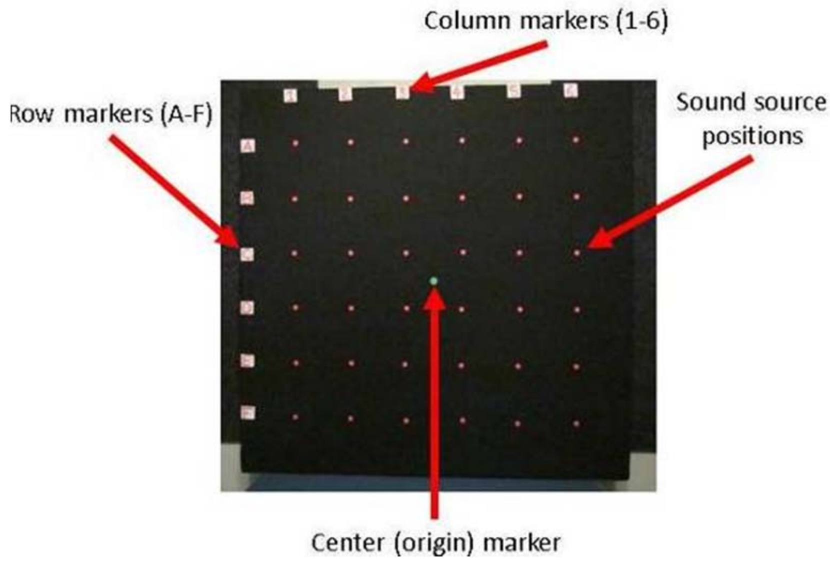
3.6.1 Auditory Stimulus

A total of five volunteers participated in the experiment. The average age of the participants was 29 years old. The auditory stimulus consisted of a broadband white noise signal sampled at a rate of 44.1 kHz and band-pass filtered using a 256-point Hamming windowed FIR filter with low and high frequency cut-offs of 200 Hz and 10 kHz respectively. The sound was output on an iHome iHM60 portable multimedia loudspeaker which was manually moved one of the 36 sound source positions by one of the experimenters (described below). The experiments took place in an Eckel audiometric room at the University of Ontario Institute of Technology (room dimensions of 2.3 m × 2.3 m × 2.0 m). The Eckel audiometric room provides (frequency dependent) noise reduction across a wide range of frequencies (e.g., at 19 dB at 125 Hz and 60 dB at 4 kHz). The average background noise level within the audiometric room measured with a Radio Shack sound level meter (model 33-2055) with an A-weighting, placed at the location where the participant's head would be in the absence of any sound stimuli was below 50 dB (the lowest level measurable with the sound level meter). The average sound level also measured at the location where the participant's head would be with the presence of the sound source was 68 dB.

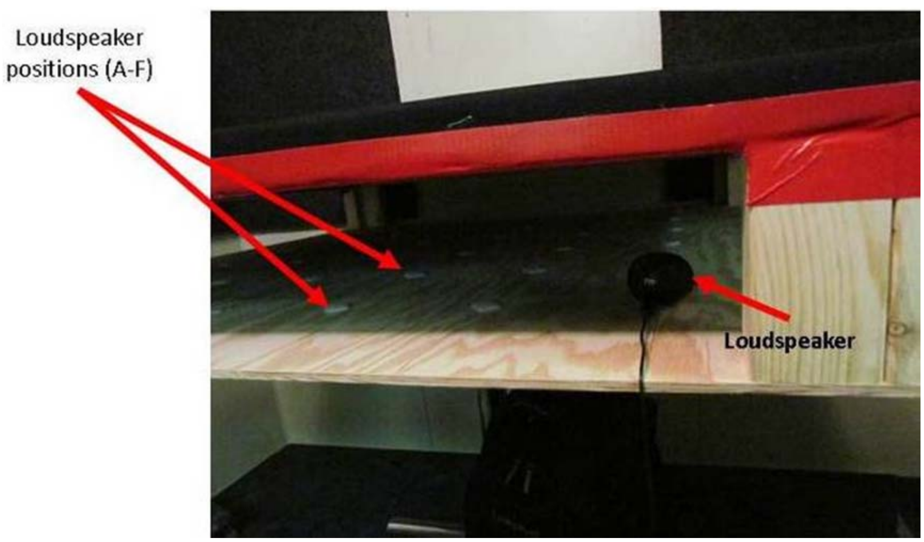
3.6.2 Experimental Method

Participants were seated on a chair 0.51 m from the surface at a height of 1.36 m and instructed to look forward at the green marker located at the center of the box. In an effort to limit and deviations from their intended positions, participants were asked to line up the tip of their nose with a thin piece of string (with a weight on its bottom) hanging from the ceiling of the audiometric room. For each trial, the loudspeaker was physically moved to one of the 36 sound source positions, the sound stimuli was presented and the participant’s task was to indicate which of the 36 positions they believed the sound was emanating from. Participants indicated their choice by stating the corresponding row and column to the experimenter who then recorded their choice. The sound was turned off and the next trial began. A total of 36 grid positions (spatial sound sources) were considered and each position was repeated two times leading to a total of 72 trials (i.e., 36 grid positions × 2 repetitions). Each of the 72 sound source positions will be considered in random order. Prior to the start of the experiment, participants were presented with the auditory stimulus at each of the four corner positions of the surface (individually, one after the other) to provide them with a reference.

In addition to the collection of ground truth data with respect to a horizontal surface, the experiment was repeated with the box positioned vertically (i.e., flipped 90 degrees) as shown in Figure 18(b). The vertical configuration was tested following testing with the horizontal configuration (and after a short break). The collection of ground truth data for sound sources positioned vertically allows us to compare our sound localization abilities on a horizontal screen.



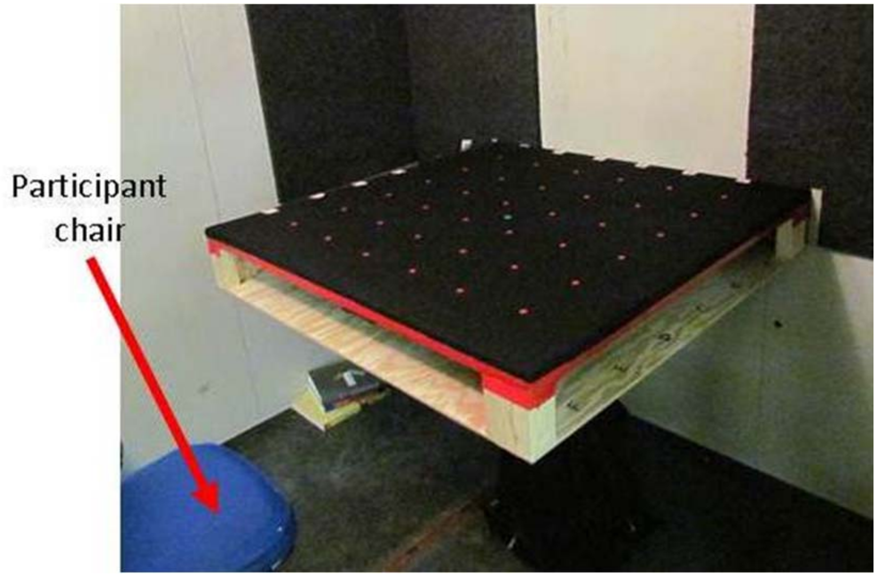
(a)



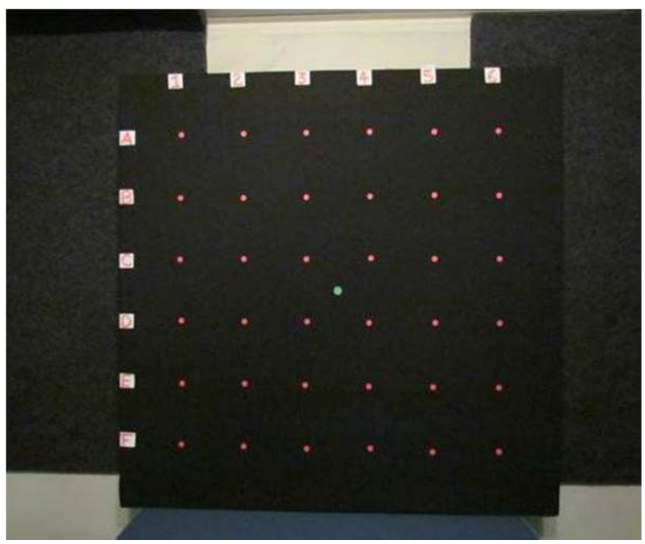
(b)

Figure 17. Hardware setup. Top view (surface and what the participants will see) of the surface of the "box" with the sound source positions, rows, and columns labelled (a). Side view of the box with the sound source positions and the sound source (b).

In our sound verification hardware setup, the surface and pre-defined sound source positions are modeled to imitate the configuration of our previous work and experiments with multiple physical sound sources. An illustration of the hardware setup is provided in Figure 17. As shown, the hardware setup essentially consists of a custom built box with openings on two of its sides. Inside the box there are 36 pre-defined loudspeaker locations; each location is labelled and allows for a loudspeaker to be easily attached (and later removed) to it in a simple manner. The top of the "box" is covered with loudspeaker grill cloth covering the inside of the box and therefore hiding the loudspeaker from the participants while allowing the sound to pass through. On the top of the box which is covered with loudspeaker grill and visible to the participants, the 36 sound source locations are clearly labelled (in red) as are the rows and columns (see the white labels on the side and top; the rows are labelled from A-F while the columns are labelled from 1-6; see Figure 17(a)). With this particular hardware configuration, a single (small) loudspeaker (see Figure 17(b)) can be moved to each of the 36 pre-defined loudspeaker locations thus allowing us to collect "ground truth" data for each of these locations by manually moving the loudspeaker within the enclosure. This, however, is a tedious and time consuming process that involves two operators at both sides of the box (since the box width does not allow a single operator to place the loudspeaker directly at places in the two far side columns). The sound verification setup is, thus, only suitable for collecting a limited volume of verification data and once this is done and the device from Experiment One is properly tuned, it will be used for more automated and accurate experiments.



(a)



(b)

Figure 18. Experimental setup within the audiometric room where the experiments are taking place. Horizontal experimental setup (a) and vertical experimental setup (b).

In addition to the collection of ground truth data with respect to a horizontal surface (see Figure 18(a)), the experiment will be repeated with the box positioned vertically (i.e., flipped 90 degrees) as shown in Figure 18(b). The collection of ground truth data for sound sources positioned horizontally and vertically will provide us with further insight into our sound localization abilities on a horizontal surface (e.g., whether better or worse when compared to sound localization on a vertical surface and if so, what implications does this have for virtual sound source generation). The experiments will take place in an Eckel audiometric room at the UOIT (room dimensions of 2.3 m × 2.3 m × 2.0 m) to reduce any potential effects of environmental noises (air condition "hums", etc.) and reverberation of the generated sounds within the environment. The Eckel audiometric room provides (frequency dependent) noise reduction across a wide range of frequencies (e.g., at 19 dB at 125 Hz and 60 dB at 4 kHz). [Dakano et al. 2012]

3.6.3 Results (Horizontal Configuration)

The average error (Euclidean distance) and standard deviation for each of the 36 virtual sound source positions (averaged across each of the five participants) for the horizontal configuration is summarized in the plot of Figure 19 and Table 5. The average error across each of the 36 positions ranged from 0.02 m to 0.32 m with an average of 0.18m ±0.07 m. Given the grid spacing of 0.15 m × 0.15 m, participants were able to localize the sound source to within approximately two positions of the actual virtual sound source. Inspection of Figure 19, indicates that the largest errors appear to be along row F (closest to the participants) and along row A (furthest from the participants). The positions corresponding to the five largest errors are: (3F; 0.32 m ± 0.17 m), (4F; 0.32 m ± 0.16 m), (3A; 0.29 m ± 0.18 m), (4A; 0.26 m ± 0.15 m), and (6B; 0.26 m ± 0.14 m).

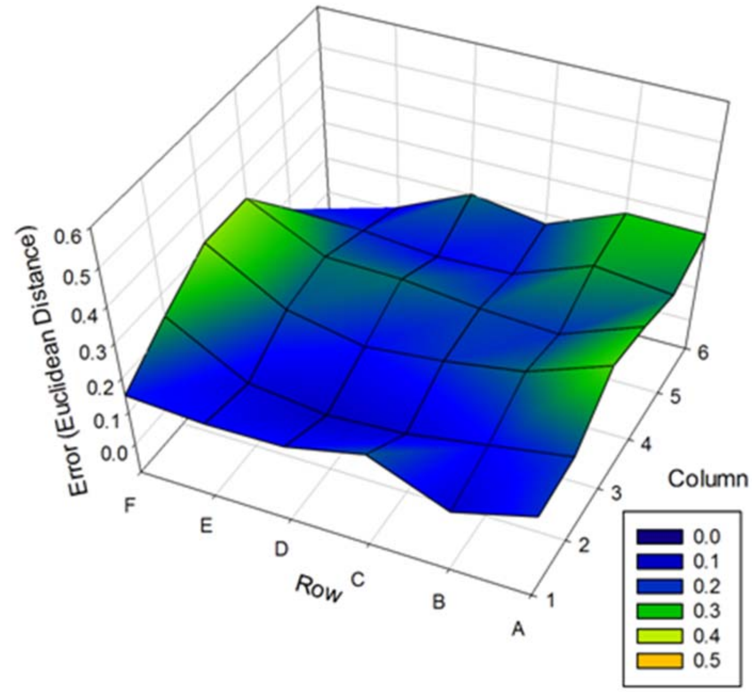


Figure 19. Results for the "ground truth" sound localization in horizontal position. Average error (Euclidean distance or the difference between the actual and perceived sound source positions, measured in meters) for the ground truth horizontal configuration experiment con (averaged across each of the five participants).

	1	2	3	4	5	6
A	0.16 ± 0.10	0.17 ± 0.13	0.29 ± 0.18	0.26 ± 0.15	0.21 ± 0.12	0.26 ± 0.12
B	0.10 ± 0.14	0.14 ± 0.11	0.20 ± 0.13	0.17 ± 0.15	0.23 ± 0.17	0.26 ± 0.14
C	0.19 ±0.19	0.10 ± 0.10	0.17 ± 0.12	0.18 ± 0.13	0.15 ± 0.09	0.17 ±0.16
D	0.14 ± 0.16	0.08 ± 0.08	0.14 ± 0.11	0.20 ± 0.14	0.14 ± 0.18	0.20 ± 0.14
E	0.14 ± 0.18	0.11 ± 0.10	0.19 ± 0.15	0.21 ± 0.14	0.16 ± 0.10	0.11 ± 0.10
F	0.15 ± 0.31	0.24 ± 0.19	0.32 ± 0.17	0.32 ± 0.16	0.17 ± 0.12	0.02 ± 0.05

Table 5. Average error for the "ground truth" sound localization results for the horizontal configuration. (Euclidean distance or the difference between the actual and perceived sound source positions) and standard deviation for sound source position (averaged across each of the five participants).

3.6.4 Results (Vertical Configuration)

The average error (Euclidean distance) and standard deviation for each of the 36 virtual sound source positions (averaged across each of the five participants) for the vertical configuration is summarized in the plot of Figure 20 and Table 6.

The average error across each of the 36 positions ranged from 0.02 m to 0.23 m with an average of 0.13m ± 0.05 m. Given the grid spacing of 0.15 m × 0.15 m, participants were able to localize the sound source to within approximately one position of the actual virtual sound source. Inspection of Figure 20 indicates that the largest errors appear to be along row F (the rows at the bottom end of the vertically placed board). The positions corresponding to the five largest errors are: (2E; 0.23 m ± 0.16 m), (3F; 0.20 m ± 0.12 m), (4F; 0.20 m ± 0.09 m), (2F; 0.19 m ± 0.13 m), and (2C; 0.18 m ± 0.11 m). In addition to smaller average error for the vertical configuration when compared to the horizontal configuration (0.13m ±0.05 vs. 0.18m ±0.07 m respectively), the smaller error associated with the vertical configuration is also evident through a visual inspection of the two resulting plots of Figures 19 and 20.

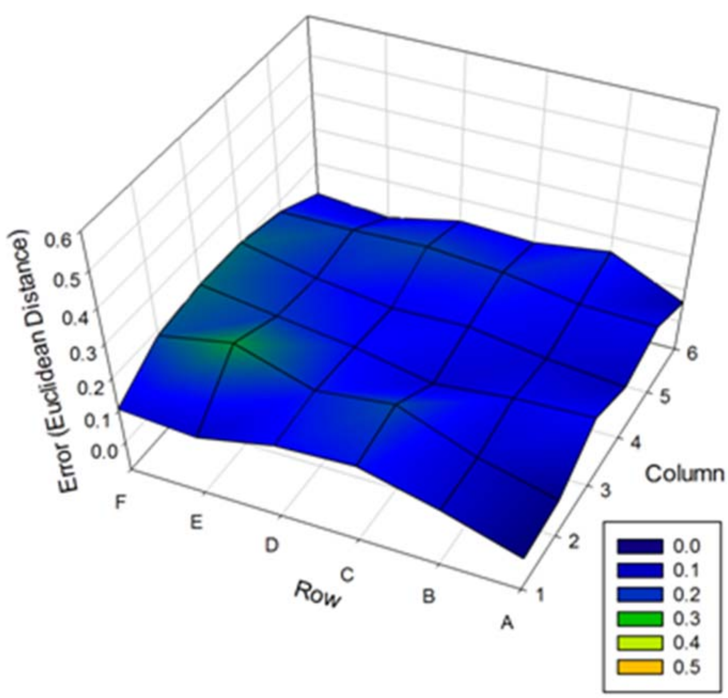


Figure 20. Results for the "ground truth" sound localization in vertical position. Average error (Euclidean distance or the difference between the actual and perceived sound source positions, measured in meters) for the ground truth vertical configuration experiment con (averaged across each of the five participants).

	1	2	3	4	5	6
A	0.02 ± 0.05	0.03 ± 0.06	0.13 ± 0.12	0.08 ± 0.08	0.12 ± 0.09	0.06 ± 0.07
B	0.09 ± 0.07	0.09 ± 0.07	0.12 ± 0.09	0.10 ± 0.10	0.13 ± 0.09	0.16 ± 0.10
C	0.15 ± 0.07	0.18 ± 0.11	0.10 ± 0.08	0.13 ± 0.10	0.16 ± 0.07	0.13 ± 0.09
D	0.14 ± 0.15	0.15 ± 0.14	0.14 ± 0.05	0.12 ± 0.10	0.18 ± 0.11	0.14 ± 0.08
E	0.09 ± 0.07	0.23 ± 0.16	0.17 ± 0.16	0.15 ± 0.18	0.17 ± 0.16	0.09 ± 0.07
F	0.11 ± 0.14	0.19 ± 0.13	0.20 ± 0.12	0.20 ± 0.09	0.17 ± 0.17	0.11 ± 0.16

Table 6. Average error for the "ground truth" sound localization results for the vertical configuration. Average error (Euclidean distance or the difference between the actual and perceived sound source positions) and standard deviation for sound source position (averaged across each of the five participants).

Figures 21 and 22 provides a "vector plot" of the average error for each of the 36 positions of both the horizontal and vertical configurations respectively, whereby the magnitude and direction of the error associated with each of the 36 positions is shown. The red arrows show the error for each of the virtual sound source positions while the green arrow in the middle represents the average of all the red arrows. A visual inspection of Figures 21 and 22 clearly illustrates that generally, there is a larger error associated for sound localization on the horizontal configuration. Furthermore, with respect to the vertical configuration (Figure 22), the error is clearly larger for locations associated with the lower half of the surface (i.e., rows 4, 5, and 6) whereas for the horizontal surface configuration, this pattern is not observed. For both configurations, the errors appear to be moved towards the center of the surface. In other words, participants incorrectly perceive the sound to be emanating towards the center of the surface.

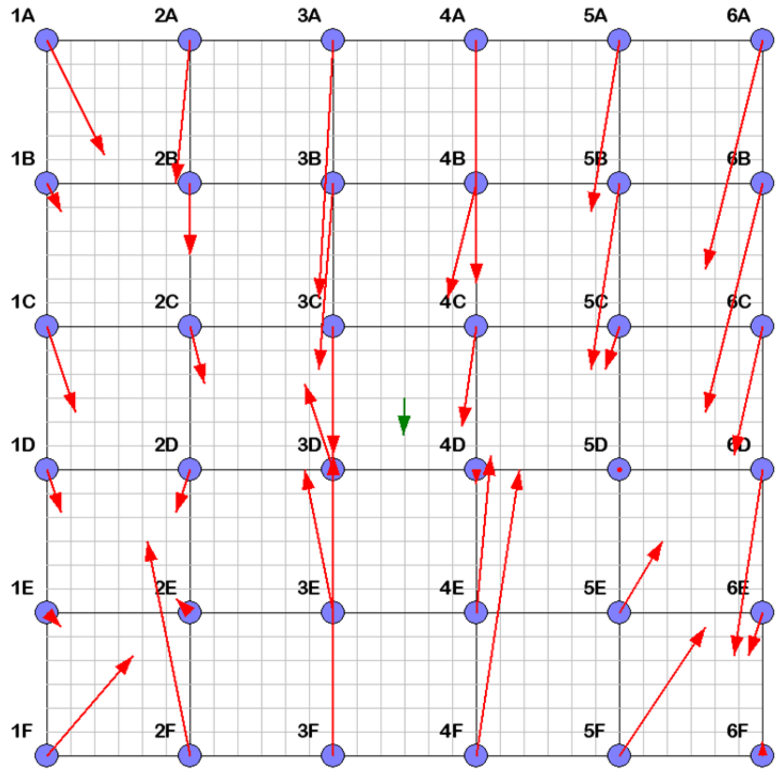


Figure 21. Results for the "ground truth" sound localization in horizontal position. Error vector plot. The red arrows show the error for each of the virtual sound source positions while the green arrow in the middle represents the average of all the red arrows.

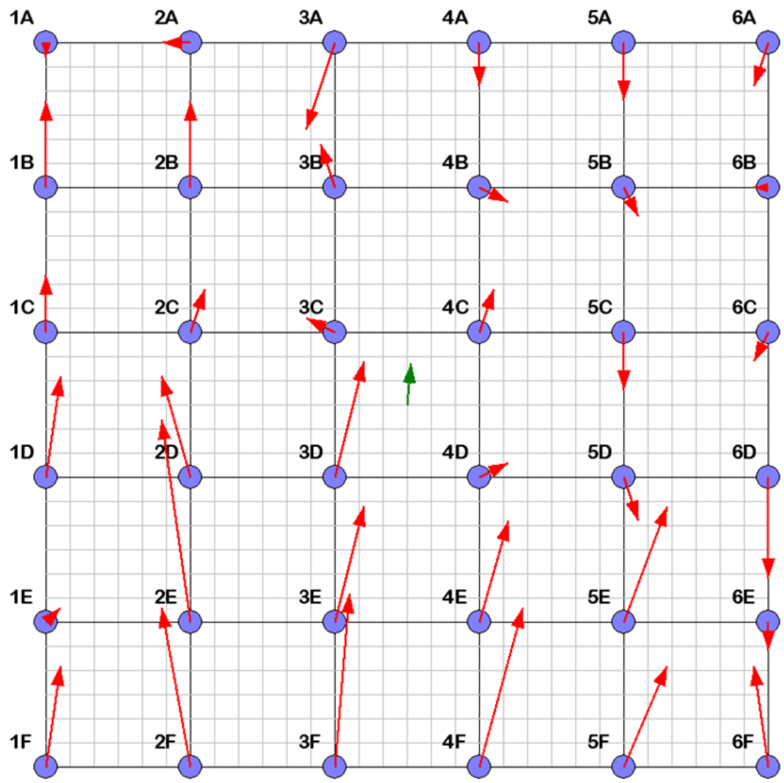


Figure 22. Results for the "ground truth" sound localization in vertical position. Error vector plot. The red arrows show the error for each of the virtual sound source positions while the green arrow in the middle represents the average of all the red arrows.

3.6.5 Comparison of the Horizontal Configuration Results

A visual comparison between the error (across each of the 36 positions considered) of the horizontal configuration and the bilinear amplitude panning methods is provided in Figure 23 in the form of a difference plot. The values in the resulting difference plot were obtained by subtracting the bilinear amplitude panning method values from the horizontal configuration values at each of the 36 positions considered; negative difference values indicate that the bilinear panning method values are greater than the corresponding horizontal configuration value. For the majority of the 36 positions considered, the bilinear amplitude panning method resulted in a greater error. The average values for the horizontal configuration and bilinear amplitude panning method are 0.18m ±0.07 m and 0.23 m ±0.07 m respectively.

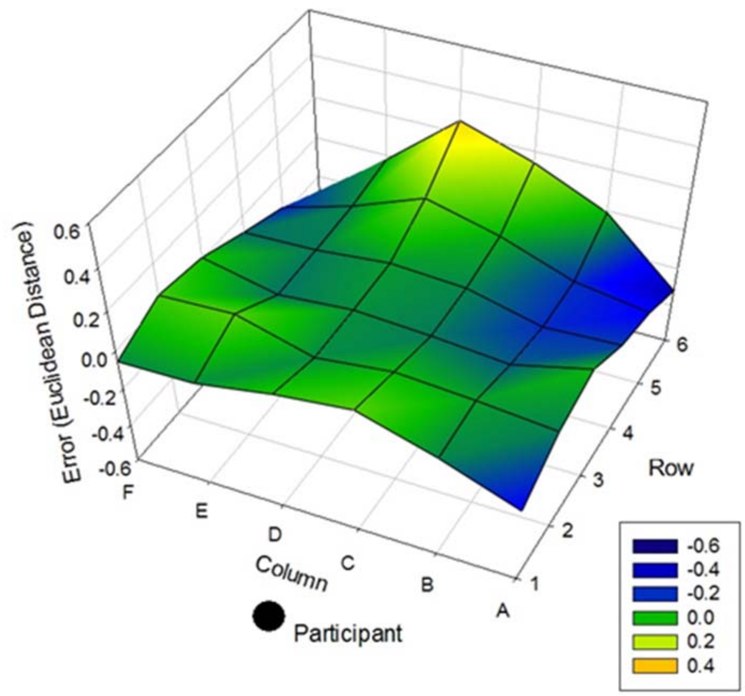


Figure 23. The difference in error between the results of the horizontal configuration and the bilinear amplitude panning method for each of the 36 positions considered.

A visual comparison between the error (across each of the 36 positions considered) of the horizontal configuration and the inverse distance amplitude panning methods is provided in Figure 24 in the form of a difference plot. The values in the resulting difference plot were obtained by subtracting the inverse distance amplitude panning method values from the horizontal configuration values at each of the 36 positions considered; negative difference values indicate that the inverse distance panning method values are greater than the corresponding horizontal configuration value. For the majority of the 36 positions considered, the inverse distance amplitude panning method resulted in a greater error. The average values for the horizontal configuration and bilinear amplitude panning method are 0.18m ±0.07 m and 0.24 m ±0.07respectively.

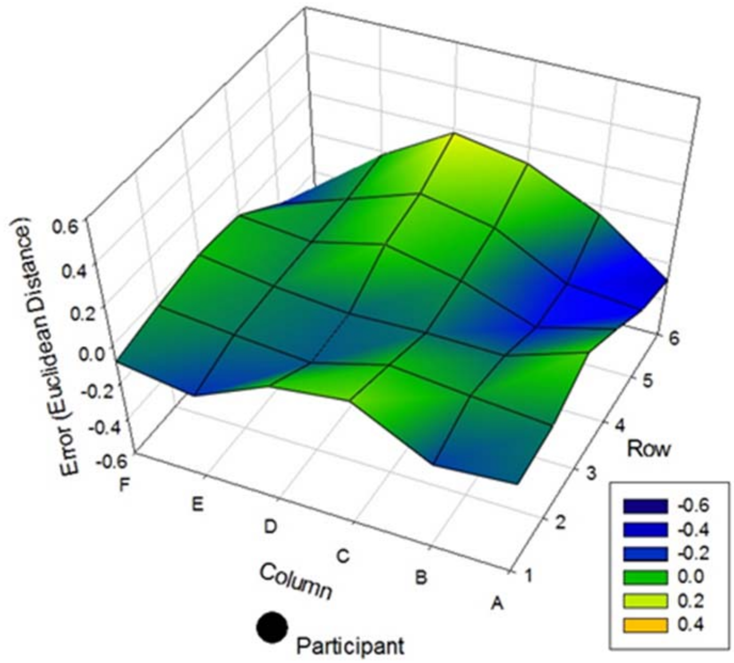


Figure 24. The difference in error between the results of the horizontal configuration and the inverse distance amplitude panning method for each of the 36 positions considered.

CHAPTER 4 – DISCUSSION

This chapter will summarize the results of the previously described experiments and discuss their impact on proper sound generation for tabletop displays.

4.1 Bilinear Interpolation with the Rectangular Inward Loudspeaker Configuration

In this experiment, sound localization on a horizontal surface was examined using a simple and computationally efficient amplitude panning method based on bilinear interpolation. Although the method is easy to implement and compute, results indicate that the method is prone to varying error across individuals particularly for the virtual sound source positions that are closest to the participant (user). However, the results do provide an indication of the potential issues game or interface designers face. More specifically, if users cannot localize sounds well when the sound sources are closest to them perhaps designers need to exaggerate placement when sounds are nearest to the user, use sounds that are more easily localized (sound source localization varies with frequency [Perrott and Saberi 1990] and changes in frequency [Ohta and Obata 2007]. Furthermore, sounds that have more formants/overtones are easier to localize than sine waves, and reverberation will also aid sound source localization (see [Roffler and Butler 1968]).

4.2 Bilinear Interpolation with the Rectangular Upward Loudspeaker Configuration

In the previous experiment, a preliminary user-based experiment tested the effectiveness of the bilinear interpolation amplitude panning method. In that experiment, using the bilinear interpolation method previously described in Chapter 2.1.1, the sound was spatialized to correspond to one of 25 pre-defined locations on the surface and the participants' task was to localize the sound. That experiment was similar to the experiment described here except that i) a smaller area of the table was used (25 virtual sound source positions considered in contrast to the 66 virtual sound source positions considered here), and ii) the loudspeakers were not facing upwards but rather, facing inwards towards the center of the table and therefore, the two loudspeakers closest to the participant were actually facing away from the participant. The results of this experiment indicated that the method is prone to varying error across individuals particularly for the virtual sound source positions that are closest to the participant (user). This error was attributed to the fact that for the positions resulting in the largest error (those closest to the participant), the two loudspeakers were facing away from the participants and the motivation for conducting the experiment described here (and flipping the loudspeakers such that they faced upwards) was to test whether the errors did in fact result from the fact that the two loudspeakers faced away from the participants. However, the results obtained here do not support this claim.

Despite the fact that the loudspeakers in the experiment described here did not face away from the participants, similar results were obtained and more specifically, the largest errors once again correspond to the locations closest to the participants while the most accurate responses corresponded to the locations around the edges and farthest away from the participants. This indicates that placing the loudspeakers at each of the four corners of the table may not necessarily be the optimal configuration regardless of the loudspeaker orientation. This was also demonstrated in the work of [Collins et al. 2010] that examined the use of audio-based games in providing subjective measures of player preference of two different loudspeaker configurations. More specifically, an audio-based, touch-table version of tabletop air hockey was developed (termed Audio Air Hockey) where the puck was not visible but rather emitted a continuous soft white noise sound as it moved. Participants played the game with two loudspeaker configurations: 1) the standard quadraphonic setup whereby a loudspeaker was placed at each corner of the table facing inwards towards the center on a 45°, and 2) the diamond configuration whereby the loudspeakers were placed midway on the sides of the table in a diamond shape. After playing the game with the two loudspeaker configurations, participants completed a questionnaire. Results showed that the majority of players preferred the diamond configuration to the standard quadraphonic configuration, and that the placement was perceived to impact their gameplay [Collins et al. 2010].

Although the results of this experiment in addition to the results of previous experiment and previous work (e.g., [Collins et al. 2010]) indicate that the standard quadraphonic loudspeaker configuration, regardless of loudspeaker orientation, may not be the optimal configuration for tabletop computers. We lack information regarding how accurately participants could localize "real" physical sound sources at the corresponding virtual sound source positions if present. This was addressed in the Ground Truth experiment (discussed below).

4.3 Bilinear and Inverse Distance-Based Interpolation with the Diamond Loudspeaker Configuration

This experiment investigated sound source localization on a horizontal surface using the bilinear interpolation, and inverse distance amplitude panning methods with a diamond loudspeaker configuration. Results of this experiment revealed that although both methods are prone to errors and they are not statistically different from each other ($p = 0.60$). The computational requirements are minimal for both methods hence; there is no difference in using one method over the other.

Unlike Experiment One where the error was largest for positions closest to the listener, here, for both panning methods, the errors were largest for the positions at the four corners of the surface (this is evident graphically when examining the three-dimensional plots of error vs. position of Figures 12 and 14 as well as the error vector plots of Figures 13 and 15). Participants faced forwards and were not allowed to move their heads hence, larger errors associated with the corner positions are of course expected given that human sound localization is most accurate for sounds directly in front (e.g., 0^{o} azimuth) and this accuracy decreases for azimuth angles off to the side [Durlach et al. 1993]. Examining the vector plots of Figures 13 and 15 for both panning methods, it is clear that both amplitude panning methods show a general and consistent bias towards the area directly in front of the listener (although the vector field for the inverse distance amplitude panning method is more chaotic and also shows a greater bias towards the bottom half of the grid). Graphically, the green arrow in the centre of each vector plot of Figure 13 and 15 represents the average error (magnitude and direction) across all positions indicating that participants consistently (and erroneously) perceived the sound sources to be located closer towards them and towards the centre of the surface. Each of the four loudspeakers was positioned such that its distance from the centre of the table was 1.2 m. However, given that the participants were seated in front of one of the loudspeakers (see Figure 11(b)), they were in fact closer to that particular loudspeaker than the other three and there was no correction made for this. In a "real-world" scenario, such corrections may not be possible particularly with multiple users and a static table set-up. The fact that the participants localized the sounds closer towards them may, in part, be due to the potentially greater influence this particular loudspeaker may have had on the participants' localization abilities, drawing the sound source position closer to them given that the sound emanating from this loudspeaker would be attenuated less before reaching the participants. Furthermore, here the "grid spacing" was set to 0.15 m × 0.15 m and participants were instructed to choose one of the 36 grid positions even if they actually perceived the sounds as emanating from a non-grid position; this grid spacing was chosen through informal listening tests yet, modifying the spacing between the virtual sound sources may also affect accuracy.

Although further experiments must be conducted to develop a better understanding of sound localization on a horizontal surface, the results presented here in addition to the results of Experiment One and Two and the results of Collins et al. [Collins et al. 2011]) indicate that it is very difficult to accurately determine the actual position of a virtual sound source on a horizontal surface using the tested listener-position independent spatialization techniques. In other words, sound localization on a horizontal surface is prone to localization error and developers/designers of applications on surface computers should be aware of such errors and aim to account for them or avoid them where possible.

4.4 "Ground Truth"

As shown with the results of the experiments previously described, the use of amplitude panning methods to spatialize a sound source on the surface of a horizontal surface is prone to large errors. However, there was no reference data to compare these references to. To address this issue, a sound verification hardware setup and methodology that allows for a single physical sound source to be moved to 36 pre-defined places (positioned on a grid with horizontal and vertical separations of 0.15 m) in a simple and efficient manner was devised (together with colleagues from York University in Toronto, Canada, and Shizuoka University in Hamamatsu, Japan). Using this novel hardware setup, the ground truth experiments were conducted to collect such reference measures and allow for meaningful conclusions/discussions to be drawn from the results of the previous amplitude panning methods. The surface and pre-defined sound source positions are modeled to imitate the configuration of the previous experiments with multiple physical sound sources. With this particular hardware configuration, a single (small) loudspeaker (see Figure 17(b)) was moved to each of the 36 pre-defined loudspeaker locations thus allowing for the collection "ground truth" data for each of these locations by manually moving the loudspeaker within the enclosure.

Although smaller than the errors arising from the amplitude panning methods, even in the presence of a physical sound source at the corresponding position, localizing the sound source is still prone to error hence, one should not expect to eliminate the errors when employing amplitude panning methods to spatialize a sound source to a position on a horizontal surface. That being said, it should also be stated that the process of conducting the ground truth experiments was a tedious and time consuming process that required three experimenters; two experimenters at both sides of the box to move the physical loudspeaker to one of the 36 positions (since the box width does not allow a single operator to place the loudspeaker directly at places in the two far side columns), and one experimenter to indicate which position to move the loudspeaker to. Given the complexity involved in conducting this experiment, only five participants completed the study and the results presented are preliminary but still indicate that sound localization on a horizontal surface is prone to error.

CHAPTER 5 – CONCLUSIONS

Tabletop displays represent a further step towards what is known as ubiquitous or pervasive computing. Given the collaborative nature of tabletop computers, gaming seems like a logical trajectory for tabletop computing technology and presents many opportunities for game designers. However, before tabletop computing becomes widely accepted, there are many questions, particularly with respect to audio interaction that require further investigation. Here, we examined sound localization on a horizontal surface using two amplitude panning methods (bilinear interpolation and inverse distance) with several loudspeaker configurations (setups) including i) the standard quadraphonic configuration whereby a loudspeaker is placed at each corner of the surface facing inwards towards the center of the surface, ii) a quadraphonic configuration where, instead of the loudspeakers facing inwards, they were "flipped upwards" such that the sound was emanating upwards, and iii) a diamond loudspeaker configuration whereby the loudspeakers were positioned on each of the four sides of the surface facing towards the center of the surface. Our results indicate that accurately localizing a virtual sound source on a horizontal surface is a difficult task and prone to error regardless the amplitude panning method used to spatialize the sound source or the loudspeaker configuration. This was confirmed in a "ground truth" experiment that was conducted to test people's ability to localize an actual sound source on a horizontal (and vertical) surface which demonstrated that sound source localization even with actual sound sources on a horizontal surface is prone to error. Given the presence of these errors, developers and designers of applications for tabletop displays must account for these errors and perhaps exaggerate placement when sound source positions correspond to positions with large error, or use sounds that are more easily localized. For example, sound source localization varies with frequency [Perrott and Saberi 1999] and changes in frequency [Ohta and Obata 2007], sounds that have more formants/overtones are easier to localize than sine waves, and reverberation will also aid sound source localization (see [Roffler and Butler 1968]). These are all potential areas that warrant further investigation.

Furthermore, images tend to "magnetize" sounds to a specific location [Chion 1994], and for video games in particular, it can be assumed that some errors can be corrected through the visualization of sounds.

The ground truth experiments conducted here included only five participants and were intended only to provide preliminary results. Future work will see the testing of a larger population. Future work will also involve multiple individuals seated around the table as opposed to a single participant considered here. Future work will also examine what, if any effect table size has on sound localization capabilities, and more specifically, is there an optimal size for 1, 2, 3, or 4 users? Conducting similar sound localization experiments with more than one participant seated around the table may present some difficulties. More specifically, how will each of the multiple participants indicate their choice of virtual sound source position without influencing each other? One potential solution to this problem is to provide each of the participants with a tablet-type computer (e.g., Apple iPad) where the pattern of virtual sound sources is replicated on the tablet-type computer and participants indicate their choice of sound source position by clicking/touching the corresponding position on the tablet-type computer. Tabletop computers are intended to be used with both visual and auditory stimuli. Therefore, future work will also examine the interaction of audio and visual cues and in particular, our ability to localize a sound source in the presence of visual stimuli (and potentially conflicting visual stimuli). Despite the inherent error observed here, in many gaming applications, "pin-point" sound localization accuracy may not necessarily be required. Rather, determining the direction to a sound source and whether the distance to the sound source is increasing or decreasing may be of more importance. Future work will thus investigate how well we can determine the angle/direction to a sound source using the bilinear interpolation amplitude panning method with the diamond loudspeaker configuration.

In this study only the auditory component of the tabletop display was considered. However, the tabletop computer represents a tangible device with a well-defined flat surface that is intended to be used for interactive media. In this context, the two components that must be considered are i) the surface touch functionality, and ii) the content presentation functionality. For surface touch functionality, standard digitizing techniques for absolute position sensing can be used. Unfortunately, existing tabletop computer display hardware presents serious limits to loudspeaker placement. Existing technology requires a hard display surface and many display and interaction monitoring techniques (e.g., rear-projection techniques and the Frustrated Total Internal Reflection (FTIR) [Han, 2005] interaction monitoring method) require access to the void below the table surface while front surface projection techniques require access to the void above the tabletop surface.

User confidence is an important issue and may be indicative for abilities of particular users as well as for some fundamental experiment environment deficiencies such as placement, external noise and disturbances, etc. In this regard, one issue may be the limitation of possible sound source selections to a predetermined grid (i.e., one of the 36 virtual sound source positions). This requires both appropriate display technology (e.g., front surface projection) as well as interaction monitoring technology that does not require access to the void beneath the interaction surface.

Future work will also examine the interaction of audio and visual cues and in particular, our ability to localize a sound source in the presence of visual stimuli (and potentially conflicting visual stimuli).

REFERENCES

[1] M. Chion. Audio-Vision. Columbia University Press, New York, NY, USA. 1994.

[2] K. Collins, K. Kanev and B. Kapralos. Using Games as a Method of Evaluation of Usability and User Experience in Human-Computer Interaction Design. In *Proceedings of the 13th International Conference on Humans and Computers (HC '10)*, Fukushima-ken, Japan, 2010.

[3] K. Collins, B. Kapralos, A. Hodge, and A. Hogue. Audio air hockey: A pilot study in using audio-based games for the measurement of loudspeaker placement preferences for smart tables. In Proceedings of the Audio Engineering Society 41st International Conference, London, UK, February 2011.

[4] N. I. Durlach, B. G. Shinn-Cunningham, R. M. Held. Supernormal auditory localization. *Presence*: *Teleoperators and Virtual Environments*, 2, 1993, pp. 89-103.

[5] J. Y. Han, 2005. Low cost multi-touch sensing through frustrated total internal reflection. In *Symposium on User Interface Software and Technology: Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology*, Seattle, WA, USA, October 1997, pp. 115–118.

[6] S. Izadi, S. Hodges, S. Taylor, D. Rosenfeld, N. Villar, A. Butler, J. Westhues. Going beyond the display: A surface technology with an electronically switchable diffuser. In Proceedings of the 21st annual ACM symposium on User interface software and technology, Monterey, CA, USA, October, 2008.

[7] B. Kapralos , M. R. M. Jenkin , and E. Milios. Auditory Perception and Spatial (3D) Auditory Systems. Technical Report CS-2003-07, Department of Computer Science and Engineering, York University, Toronto, Canada, 2003.

[8] B. Kapralos, M. Jenkin, E. Milios. Virtual Audio Systems. *Presence: Teleoperators and Virtual Environments*, 17:6, Dec. 6 2008, pp. 527-549.

[9] R. Kruger, S. Carpendale, S. D. Scott, S. Greenberg. Roles of orientation in tabletop collaboration: Comprehension, coordination and communication, Journal of Computer Supported Cooperative Work, 13, 501–537. 2004.

[10] J. Lam, B. Kapralos, K. Collins, A. Hogue, and K. Kanev. Amplitude panning-based sound system for a horizontal surface computer: A user-based study. In *Proceedings of the International Symposium on Haptic Audio-Visual Environments and Games*. Phoenix, AZ. USA, October 16-17, 2010, pp. 1-5.

[11] J. Lam, B. Kapralos, K. Collins, A. Hogue, K. Kanev, and M. Jenkin. Sound localization on tabletop computers: A comparison of two amplitude panning methods. *ACM Computers in Entertainment*. To appear in 2012.

[12] T. Lossius, P. Baltazar, T. de la Hogue. DBAP - Distance-Based Amplitude Panning. In *Proceedings of 2009 International Computer Music Conference*, Montreal, Canada, 2009.

[13] D. G. Malham, A. Myatt. 3-D Sound Spatialization using Ambisonic Techniques, *Computer Music Journal*, 19:4, Winter 1995, pp. 58-70.

[14] D. Nakano, J. Lam, K. Kanev, B. Kapralos, K. Collins, A. Hogue, and M. Jenkin. A framework for sound localization experiments and automation. In *Proceedings of The ACM Joint International Conference on Human-Centered Computer Environments (HCCE 2012)*. March 8-13, 2012, Hamamatsu, Japan.

[15] Y. Ohta and K. Obata. "A method for estimating the direction of sound image localization for designing a virtual sound image localization control system," in *Proceedings of the 123rd Convention of the Audio Engineering Society*, New York, NY. USA, October 5-8 2007, p. Preprint: 7230.

[16] D. R. Perrott and K. Saberi, "Minimum audible angle thresholds for sources varying in both elevation and azimuth," *Journal of the Acoustical Society of America*, 87:4, 1990, pp. 1728–1731.

[17] V. Pulkki. Localization of Amplitude-Panned Virtual Sources Ii: Two- and Three-Dimensional Panning, *Journal of the Audio Engineering Society,* Volume 49 Issue 9, September 2001, pp. 753-767.

[18] V. Pulkki. Uniform Spreading Of Amplitude Panned Virtual Sources, In *Proceedings of 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Pallz, New York**,** Oct. 17-20, 1999, pp. 187-190.

[19] V. Pulkki. Virtual Sound Source Positioning Using Vector Base Amplitude Panning, *Journal of the Audio Engineering Society,* 45:6, June 1997, pp. 456-466.

[20] V. Pulkki, M. Karjalainen. Localization of Amplitude-Panned Virtual Sources I: Stereophonic Panning, *Journal of the Audio Engineering Society,* 49:9, September 2001, pp. 739-752.

[21] S. K. Roffler and R. A. Butler. "Factors that influence the localization of sound in the vertical plane," *Journal of the Acoustical Society of America*, 43:6, 1968, pp. 1255–1259.

[22] S. D. Scott, S. Carpendale. Theory of tabletop territoriality. In Tabletops - Horizontal Interactive Displays, C. Muller-Tomfelde, Ed. Springer (HCI Series), London, UK, 375–406. 2010.

[23] R. Stricher, and F. A. Everest. The New Stereo Soundbook, Third Edition, Audio Engineering Associates, Pasadena, CA, USA.

[24] Un4seen Developments. August 21, 2012. "BASS Audio Library", *Un4seen Developments - 2MIDI / BASS / MID2XM / MO3 / XM-EXE / XMPlay*. 2003-2010. http://www.un4seen.com/

[25] A. D. Wilson, H. Benko. Combining Multiple Depth Cameras and Projectors for Interactions On, Above, and Between Surfaces. In Proceedings of the ACM Symposium on User Interface Software and Technology (UIST), New York, NY, USA, October 2010.