A Framework for Sound Localization Experiments and Automation

Daichi Nakano[1], Jonathan Lam[2], Kamen Kanev[1], Bill Kapralos[2],

Karen Collins[3], Andrew Hogue[2], and Michael Jenkin[4]

[1]Shizuoka University, Hamamatsu, Japan
{dnakano|kanev}@rie.shizuoka.ac.jp

[2]University of Ontario Institute of Technology, Oshawa, Canada
{Jonathan.Lam|Bill.Kapralos|Andrew.Hogue}@uoit.ca

[3]University of Waterloo, Waterloo, Canada
collinsk@uwaterloo.ca

[4]York University, Toronto, Canada
jenkin@cse.yorku.ca

ABSTRACT

Table-top computing has been growing in popularity slowly for the last decade and is poised to make in-roads into the consumer market soon, opening up another new market for the games industry. However, before surface computers become widely accepted, there are many questions with respect to sound production and reception for these devices that need to be explored. Here, we describe two experiments that examine sound localization on a horizontal (table-top computer) surface. In the first experiment we collect "ground truth" data regarding physical sound source localization by employing a computer controlled grid of 25 equally spaced loudspeakers. In the second experiment we investigate virtual sound source localization using bilinear interpolation amplitude panning method and a modified quadraphonic loudspeaker configuration whereby four loudspeakers are positioned at each corner of the surface in a manner such that they emanate sound in an "upwards" direction. Obtained results indicate that sound localization of virtual sound sources on a horizontal surface is prone to errors and this is confirmed with our physical sound source "ground truth" data.

Categories and Subject Descriptors

H.5.5 [Information Interfaces and Presentation]: Sound and Music Computing- *Methodologies and Techniques*

General Terms

Measurement, Experimentation, Human Factors, Verification.

Keywords

Surface computer, audio interaction, loudspeaker configuration, amplitude panning, sound localization, horizontal surface.

1. INTRODUCTION

Table-top touch-screen computers (also known as surface computers, smart table computers, or smart tables), provide the opportunity for the development of innovative, interactive, and highly collaborative applications in a tabletop setting. However, before these opportunities can be realized, there are many questions related to display hardware and human factors that must be addressed. Hardware-related questions include issues related to display generation and interaction monitoring within the confines of a table-top display: What is the appropriate display technology for a surface that will be viewed from oblique angles? How can multi-touch gestures be properly captured given the large interaction surface and the desirability of a thin display device? Human-factors-related questions are critical as well. How should information be displayed given the multiple orientations from which the display will be viewed? How can directional audio be properly simulated given the potential distribution of users? How can displays be designed so that they foster interaction and collaboration from the participants? Many of these and other questions related to table-top displays are now beginning to be answered. For example, questions related to co-operation, orientation and viewing angle to the screen, driving innovation in terms of the visual display have been explored elsewhere [1,2], but with respect to sound, there are many questions that remain unanswered. It is interesting to observe that unlike traditional screen displays where the direction 'towards the front' is the same for all users, table-top displays have a different front direction for different users. This raises interesting problems for acoustic generation. For example, where do we physically position the loudspeakers when there are two people opposite each other surrounding a table-top display? How does our perception of the direction to a sound source change when users lean over a computer screen from different directions versus facing it? Where should we place the loudspeakers and how should sounds be positioned in the mix, (in which loudspeaker(s)) to best synthesize directional audio cues?

In order to address these and related questions we have established a collaborative research network with participants from Canada and Japan. The research network spans two continents and several research institutions with various

experimental facilities and table-top displays. Each site brings its own unique experimental facilities and expertise. For example, research conducted in Canada at the University of Ontario Institute of Technology (UOIT), the University of Waterloo, and York University deals with sound mixing and virtual sound perception, while researchers at Shizuoka University in Japan are developing and conducting experiments with physical loudspeaker grids. These experimental environments complement each other and allow the combined research team to match and verify gathered virtual and physical sound localization data.

Work to date has concentrated on the development and perception of directional audio cues within a table-top display setting. One critical question in the actual generation of directional audio cues in a table-top setting is where to actually place the loudspeakers. Only placing the loudspeakers 'within' the table limits the ability to generate simulated sources that are behind individual users, yet little theoretical or practical work has considered the issues related to placing loudspeakers so that they surround the table-top and the users. Our previous work in audio displays for table-top computers examined sound localization accuracy on a horizontal surface using a standard quadraphonic configuration where a loudspeaker was placed at each corner of the surface and faced inwards towards the center of the surface [3]. Sounds were localized to one of 25 pre-defined virtual positions on the surface using bi-linear amplitude panning whereby the sound was panned between loudspeaker pairs. Loudspeakers were mounted so that they faced towards the table at table height. Results indicated that the method is prone to varying error across individuals, particularly for virtual sound source positions that are closest to the participant. We hypothesize that the observed errors, particularly those closest to the participant, are due to the fact that the participant was 'behind' two of the loudspeakers, and thus the sound emanated away from them.

Here, we build upon our previous study with the goal of synthesizing directional audio for all participants in a table-top setting. Although we are measuring sound localization of virtual sound sources on a horizontal surface, and can make comparisons between different panning methods or different loudspeaker configurations, little "ground-truth" data exists in terms of human audio-direction perception for table-top displays to compare with. In other words, just how accurately can humans localize a sound on a horizontal surface when the sound is emanating from an actual sound source at the corresponding location, rather than a virtual location?

In Section 3, we describe a specialized software and hardware for computer control of a sound-transparent 25-speaker grid surface with multiple physical sound sources at pre-defined fixed positions. This system allows us to accelerate and increase the reliability of the collected data by significantly automating and simplifying the data gathering process.

In Section 4, we evaluate the bilinear interpolation panning method for table-top displays using a modified loudspeaker configuration. Rather than orienting the loudspeakers so that they face towards the participants as in [3], here the loudspeakers are positioned at each corner of the table and "flipped-up" so that they emanate sound in an "upwards" direction. This ensures that there are no loudspeakers facing away from the participants. This allows us to confirm if the errors present in our initial study [3] were due to the fact that the loudspeakers faced away from the listener.

Finally, we describe methods from preliminary experiments with a movable physical loudspeaker behind sound transparent horizontal and vertical surfaces. Such data allows us to account for existing differences in physical parameters of employed loudspeakers and to determine if that might have any significant bearing on the obtained experimental results.

2. PHYSICAL VS. VIRTUAL SOUND

The previous experiments described in [3,4] with virtual sounds on a flat surface created by different panning methods and physical loudspeaker configurations indicate that human perception within a desk-top setting might be somehow biased in respect to perceived origins of simulated sounds. Such biases might stem from the inherently different nature of the sound fields created with a single sound source and with multiple panned sound sources. Indeed, mixing sound from multiple loudspeakers to perfectly recreate both the amplitude, and the phase of the sound field of a single loudspeaker is far from trivial. Previous work by Bronkhorst has shown that localization of virtual sources is not as accurate as for real sources [5]. It might be difficult, in other words, to bring significant improvements in virtual sound localization perception just by higher fidelity recreations.

Ideally one would like to have an experimental environment where both, fixed physical sound sources and corresponding virtual sound sources can be associated with points on a given surface. For physical sound sources we employ small loudspeakers attached to a metal frame construction. The front of the frame is attached to a perforated board covered with sound-transparent white fabric that can also be used as a projection surface for presenting visual stimuli. The constructed device allows us to generate physical sounds precisely positioned at any of the 25 predefined grid points on the frame surface. Each of the loudspeakers is connected to a computer-controlled sound source via a relay board and a logical control circuit. Specialized software has been developed for controlling and directing output sound to appropriate loudspeakers. More detailed specifications of this experimental environment along with descriptions of conducted experiments and gathered experimental data are given in Section 3.

For the virtual sound sources we use four physical loudspeakers around the table-top surface either in a standard quadraphonic configuration where a loudspeaker is placed at each corner of the surface or in diamond configuration whereby a loudspeaker is placed at each side of the surface, and apply a standard amplitude panning method. Greater details of the virtual sound environment, the experiments conducted with it, and the gathered experimental data are given in Section 4.

To further investigate the nature and true origins of observed biases we have also set up an experimental environment for gathering reference localization data with a movable physical sound source.

3. EXPERIMENT ONE: LOCALIZATION OVER A GRID OF MULTIPLE PHYSICAL SOUND SOURCES

Given the novel nature of table-top displays little ground truth data exists as to the ability of subjects to localize sound sources in such a display. The experiment reported here helps to develop our background understanding of sound source localization. It relies

on a collection of different loudspeakers positioned at different locations

3.1 Participants

A total of four unpaid volunteers participated in the experiment. Participants were students from the Faculty of Informatics, Shizuoka University. The average participant age was 23 (none of the authors participated in the experiment). Participants reported no history of auditory disease or disorders.

3.2 Auditory Stimulus

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 stimuli were output through F77C123-1 3-inch loudspeakers (25 loudspeakers in total) with nominal power of 2 W and maximum power of 5 W. For the purpose of this experiment, loudspeakers were mounted at each of the 25 grid positions on the metal frame (Figure 1(a)) so that they faced up when the frame was put in horizontal position (Figure 1(b)).
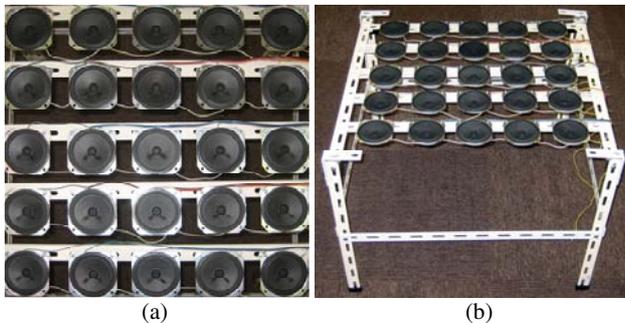

(a)   (b)

Figure 1. The metal frame in horizontal position with the mounted 25 loudspeakers facing up.

The loudspeakers were set to a height equal to 0.45 m (the height of the table-top when frame is placed horizontally as shown in Figure 1(b)). The auditory stimuli was presented and continuously repeated until the subject made his choice. A comfortable level of the sound stimuli was initially chosen and then maintained for all experiments by ensuring repeatability of the initial sound level setup.

Figure 2. Horizontal surface setup with a subject seated at the long edge of the table-top.

The experiment took place in a laboratory room at the Research Institute of Electronics, Shizuoka University (room dimensions of 6.6 m × 3.3 m × 2.7 m). Although the room itself contained a variety of equipment including desktop and laptop computers, printers, 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).

3.3 Methods

Each participant was seated on a chair at about 60cm from the long edge of the table-top as shown in Figure 2. Only auditory stimuli were present (i.e., no visual stimuli) for the duration of the experiment. In each trial, participants were presented with auditory stimuli from a sound source at one of the 25 positions across the surface of the table. The physical sound sources were positioned on a grid where the horizontal and vertical separation was 0.11 m and 0.11 m respectively.

Figure 3 provides a graphical illustration of the experimental setup with dimensions and loudspeaker numbering information included.
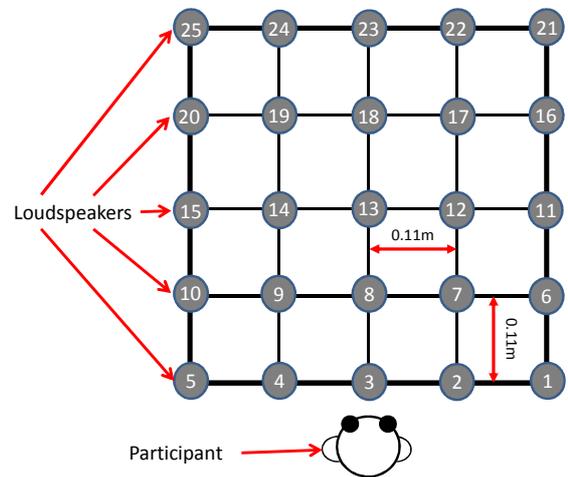

Figure 3. Grid arrangement dimensions and numbering of employed physical loudspeakers

The sound was presented to each of the 25 virtual sound source positions three times (i.e., 25 positions × 3 times per position) yielding a total of 75 trials and the ordering of each trial was random. The experiment took approximately 15 minutes to complete and all participants completed it in a single session.

(a)   (b)

Figure 4. Portable touch-screen interface for automated registration of participant choices (a) and its operation (b).

The participant's task for each trial was to indicate which of the 25 positions they perceived the physical sound source was

emanating from. They indicated their choice by selecting one of the 25 positions through a portable touch screen interface (Figure 4) that communicated and recorded their choices on a central server. Once their choice was recorded, this indicated the end of the trial and the next trial began automatically.

The ordering in which trials (physical 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.

4. EXPERIMENT TWO: LOCALIZATION OF VIRTUAL SOUND SOURCES CREATED BY PANNING

The goal of this experiment was to characterize sound source localization errors in a table-top display using amplitude planning from omnidirectional sound sources. Rather than utilizing the directional sound sources used in [3], here we approximated omnidirectional sound sources by orienting the loudspeakers so that they pointed out of the plane of the table-top.

4.1 Participants

A total of eight unpaid volunteers participated in the virtual sound experiment. Participants were either researchers, or students from the University of Ontario Institute of Technology (UOIT) and the average participant age was 25 (none of the authors participated in the experiment). Participants reported no history of auditory disease or disorders. The experiment abided by the UOIT Ethics Review process for experiments involving human participants.

4.2 Auditory Stimulus

The auditory stimulus consisted of the same 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 that was used in Experiment One. The auditory stimuli were 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 it was facing out of the plane of the table top ("upwards") (see Figure 5). 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 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 UOIT (room dimensions of 13. 0 m × 13.0 m × 3.0 m). As with the experiments at Shizuoka University, the room contained a variety of equipment including workstations, tables, chairs, etc. and 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).

4.3 Methods

Participants were seated on a chair at the long edge of the table-top setup shown in Figure 5(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 were used) for the duration of the experiment. In each trial, participants were presented with auditory stimuli that were spatialized using the the bilinear interpolation amplitude panning technique [3] so that it appeared as if 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 0.15 m and 0.15 m respectively. Figure 5(b) illustrates the experimental setup. The sound was spatialized to each of the 66 virtual sound source positions twice (i.e., 66 positions × 2 times per 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.

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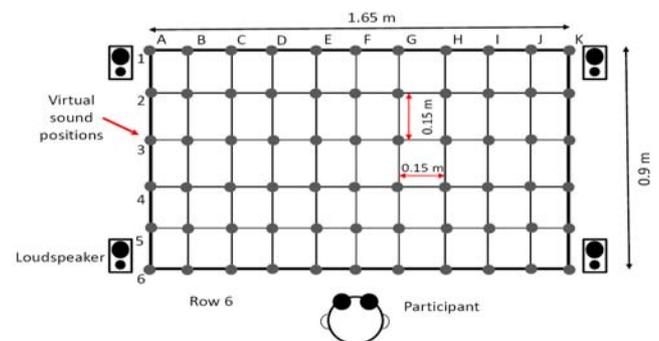
(b) The 66 virtual sound source positions and loudspeaker setup.

Figure 5. "Upward" loudspeaker setup.

The participant's task was to indicate which of the 66 positions they perceived the virtual sound source was emanating from. They indicated their choice by selecting one of the 66 positions and communicating it 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 of trials was chosen randomly. 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 chosen randomly) prior to beginning the experiment.

5. RESULTS AND DISCUSSION
5.1 Physical Sound Source Localization

The Euclidean distance between the *actual* sound source position (i.e., the location at which the physical loudspeaker the sound emanated from was placed) and the *perceived* 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 correctly 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.
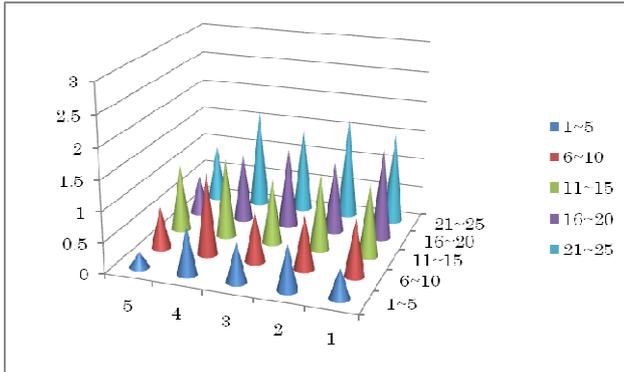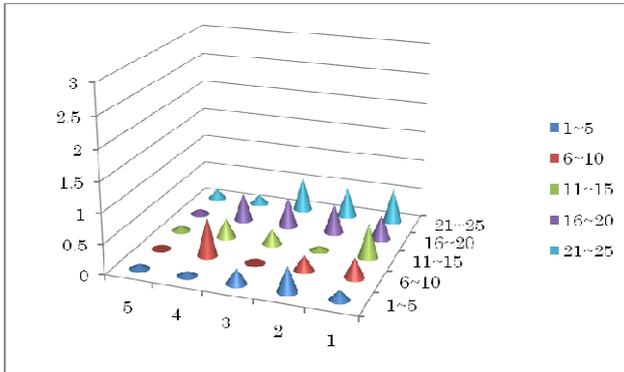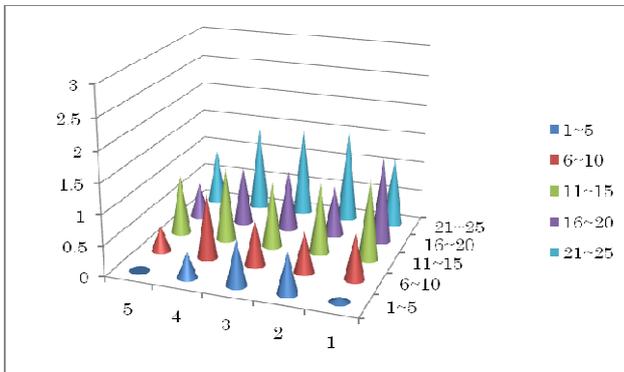
Figure 6. Average (Euclidean distance in grid units) error for each of the 25 physical sound positions combined for all four participants.
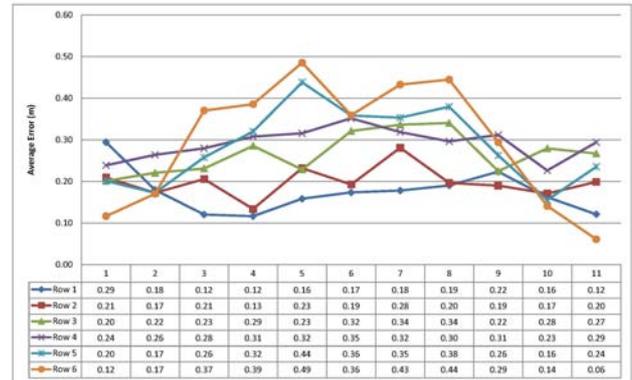
(a)

(b)

Figure 7. Average (Euclidean distance in grid units) error for each of the 25 physical sound positions in X (a) and Y (b) directions combined for all four participants.

The average error (Euclidean distance) for each of the 25 physical sound source positions (averaged across each of the four participants) is summarized in the plot of Figure 6. Neither the cumulative data nor the data associated with each of the subjects seem to expose well pronounced error distribution trends although errors tend to increase with the distance from the subject to the row of loudspeakers.
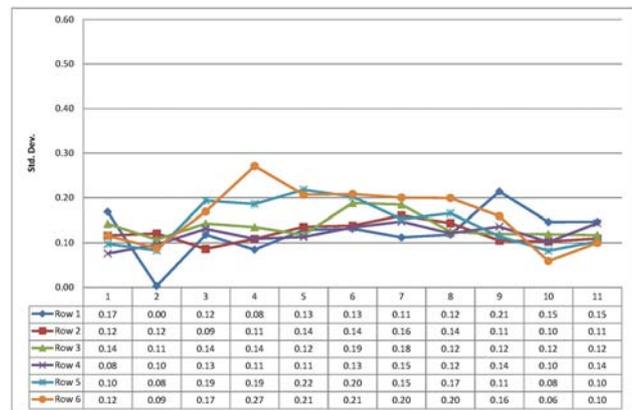
When average errors in the X and Y directions are presented separately, however, it becomes clear that they are quite different (Figure 7). The maximum average error in the X direction is 0.07 m while the error in the Y direction is over 0.16 m (more than twice). The total average error in the X direction is 0.03 m with a standard deviation of 0.02 m; three times smaller than the total average error in the Y direction (0.10 m with a standard deviation of 0.05 m). Calculated averages and their standard deviations for sound sources aligned in the X and Y directions, however, indicate that error distribution differences in respect to rows and columns might be negligible.

5.2 Virtual Sound Source Localization

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.

	1	2	3	4	5	6	7	8	9	10	11
Row 1	0.29	0.18	0.12	0.12	0.16	0.17	0.18	0.19	0.22	0.16	0.12
Row 2	0.21	0.17	0.21	0.13	0.23	0.19	0.28	0.20	0.19	0.17	0.20
Row 3	0.20	0.22	0.23	0.29	0.23	0.32	0.34	0.34	0.22	0.28	0.27
Row 4	0.24	0.26	0.28	0.31	0.32	0.35	0.32	0.30	0.31	0.23	0.29
Row 5	0.20	0.17	0.26	0.32	0.44	0.36	0.35	0.38	0.26	0.16	0.24
Row 6	0.12	0.17	0.37	0.39	0.49	0.36	0.43	0.44	0.29	0.14	0.06

(a)

	1	2	3	4	5	6	7	8	9	10	11
Row 1	0.17	0.00	0.12	0.08	0.13	0.13	0.11	0.12	0.21	0.15	0.15
Row 2	0.12	0.12	0.09	0.11	0.14	0.16	0.14	0.11	0.11	0.10	0.11
Row 3	0.14	0.11	0.14	0.14	0.12	0.19	0.18	0.12	0.12	0.12	0.12
Row 4	0.08	0.10	0.13	0.11	0.11	0.13	0.15	0.12	0.14	0.10	0.14
Row 5	0.10	0.08	0.19	0.19	0.22	0.20	0.15	0.17	0.11	0.08	0.10
Row 6	0.12	0.09	0.17	0.27	0.21	0.21	0.20	0.20	0.16	0.06	0.10

(b)

Figure 8. Average error (Euclidean distance of the difference between the actual and perceived virtual sound source positions) for virtual sound source position (averaged across each of the eight participants (a). Standard deviation (b).

Ideally, the actual and perceived positions would be identical and the Euclidean distance (and hence error) is equal to zero. 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 Figure 8(a) and Figure 8(b) respectively where each of the lines in each plot represents one row of virtual sound source positions (Rows 1-6; Row 1 represents the virtual sound sources farthest from the participant and Row 6 the virtual sound sources closest to the participant). Examination of Figure 8(a) indicates that the average (mean) error ranged from 0.06 m 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 0.49 m, 0.44 m, 0.44 m, 0.43 m, and 0.38 m, at positions (6, E), (5, E), (6, I), (6, H), and (5, I) 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).
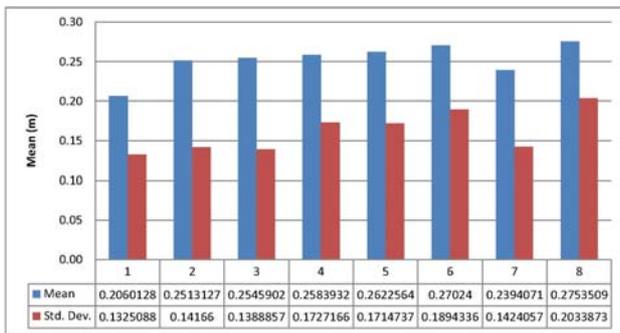

Figure 9. Average error (Euclidean distance of the difference between the actual and perceived virtual sound source positions) and standard deviation for each participant (averaged across each of the 66 virtual sound source positions).

A graphical illustration of the average error for each participant (averaged across each of the 66 virtual sound source positions) is provided in Figure 9. For each participant, the average error ranged from 0.20 m to 0.28 m. In other words, 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.

In previous work [3], a preliminary user-based experiment tested the effectiveness of the bilinear interpolation amplitude panning method. In that experiment, using the bi-linear interpolation method 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 indicated that the method is prone to varying error across individuals, particularly for the virtual sound source positions that are closest to the participant [3]. 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 actually 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. [5] that examined the use of audio-based games in providing subjective measures of player preference of two different loudspeaker configurations.

6. GENERAL DISCUSSION

In this paper we presented the results of two experiments that investigated sound localization on a horizontal surface that replicates the surface of a table-top computer.

In the first experiment, preliminary reference ("ground truth") data in respect to localization of physical sound sources was collected and observed errors were analyzed. The observed errors were generally smaller and the largest errors did not correspond to locations closest to the participants as was observed in our previous experiments [3] (rather, the error appears to increase with increasing distance from the participant).

In the second experiment, sound localization of virtual sound sources was generated with four loudspeakers at each of the four corners of the surface facing upwards. Similar to the results of our previous work which considered a similar setup but with the loudspeakers in a normal ("non-flipped") configuration facing inwards towards the center of the table [3], we observed large errors particularly for sound sources closest to the participants (errors ranged from 0.06 m to 0.49 m depending on the virtual sound source position).

The results from our preliminary studies described above suggest that we need to conduct further work to investigate human ability to localize sound sources more thoroughly. Results from Experiment Two are so far limited to only four participants but we have discovered that disparities in the physical loudspeakers might influence the experiment outcome. Indeed, arranging a large number of physical sound sources and ensuring consistency amongst the individual loudspeakers (e.g., frequency range, etc.), appears to be quite complex. Even for loudspeakers of the same make and model human ear seems to be able to tune to subtle differences in produced sound that are unconsciously used as cues for tagging the physical sound sources. Such tagging, in certain cases, may take precedence and even override the spatial cues leading to identification of a physical sound source based merely on its signature and not on its spatial localisation.

To address this issue we have developed a sound verification hardware setup and methodology which 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. Data collected with this setup will help identifying any outstanding discrepancies in the physical characteristics of the loudspeakers used in Experiment One that might have bearing the sound localization. Once

identified, physical loudspeakers exposing discrepancies will be detached from the grid and replaced with new ones from the same make and model. In this way we will ensure consistency with respect to the sound emanating at each physical sound source location on the grid of the device employed in Experiment One.
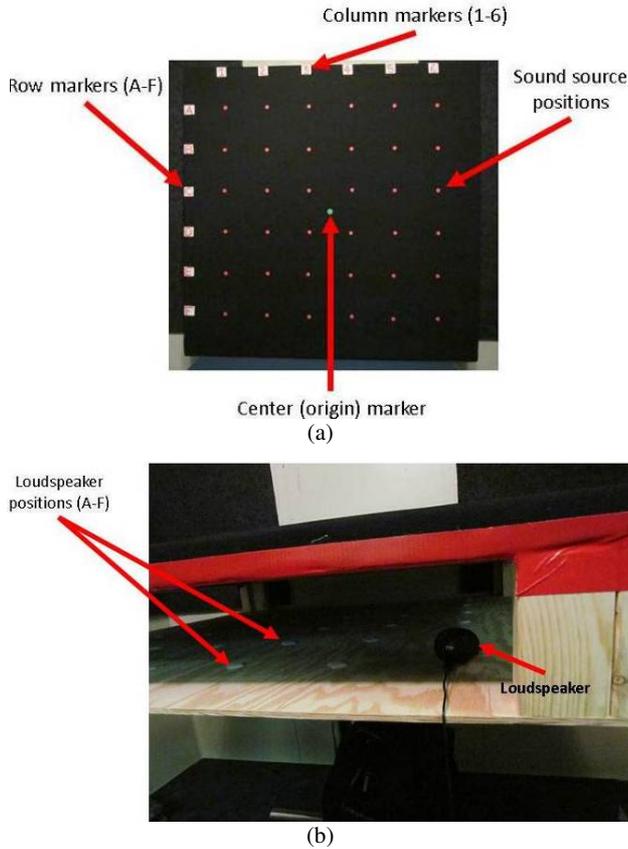

(a)

(b)

Figure 10. 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 10. 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 10(a)). With this particular hardware configuration, a single (small) loudspeaker (see Figure 10(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.
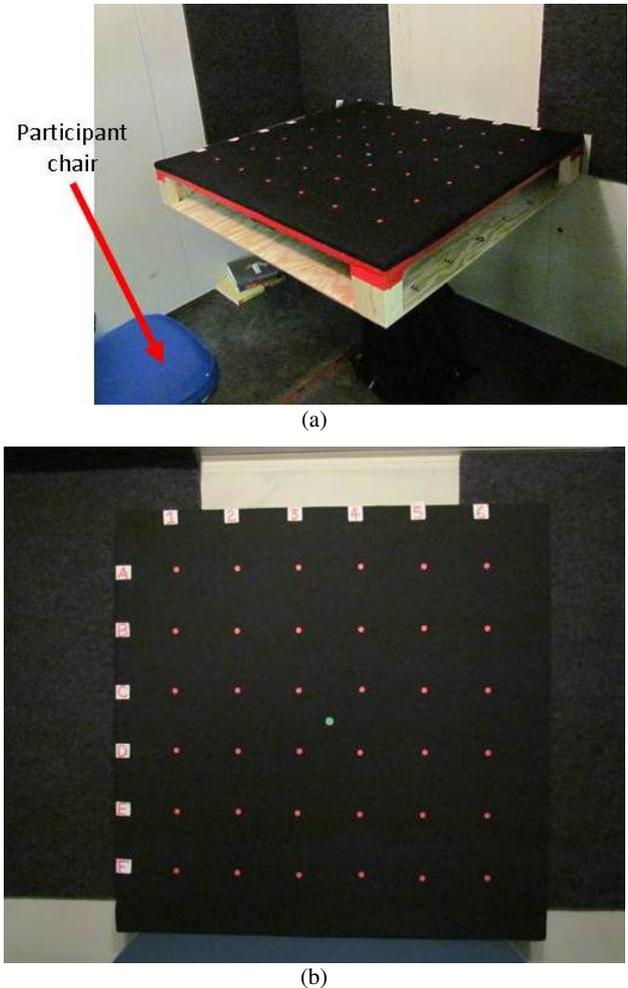

(a)

(b)

Figure 11. 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 11(a)), the experiment will be repeated with the box positioned vertically (i.e., flipped 90 degrees) as shown in Figure 11(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).

7. CONCLUSIONS

Table-top computers represent a further step towards what is known as ubiquitous or pervasive computing. In the very near future we will no more rely on a desktop model in which a single user employs a single desktop computer, but, rather, computers will be integrated into most aspects of our lives. Before table-top computing becomes widely accepted, there are many questions, particularly with respect to sound production and reception, and multi-modal interaction for these devices that need to be explored. Here we have described the results of a series of experiments conducted to test human capabilities for localization of both virtual and physical sound sources on a horizontal surface and our results indicate that the process is prone to errors with both virtual and physical sound sources (albeit, the average error is larger when considering virtual sound sources). Given that the localization of sound sources on a horizontal surface is prone to errors, particularly when considering sound source locations closest to the listener, perhaps designers of media for table-top computers should exaggerate placement when sounds are nearest to the user, use sounds that are more easily localized (sound source localization varies with frequency [6] and changes in frequency [7]. Furthermore, sounds that have more formants/overtones are easier to localize than sine waves, and reverberation will also aid sound source localization [8]. The results presented here are preliminary and further work remains.

Future work will examine what, if any effect table size has on sound localization capabilities, and more specifically, if there is any optimal size for 1, 2, 3, or 4 users. Conducting similar sound localization experiments with more than one participant seated around the table is already on our agenda. We are currently completing the development of the automated system for concurrent registration of participants' choices of virtual sound source positions without influencing each other. This is an extension of our previous version where a single participant registered his choice via a portable touch screen device.

In the extended, multiuser version each of the participants is provided with a tablet-type computer where the pattern of virtual sound sources is replicated on the device touch screen and participants indicate their choice of sound source position by clicking/touching the corresponding positions on the touch screens of their individual tablet-type computers. Note that since arrangement and labelling of sound sources on the table surface look different for each user (dependent on the seating placement around the table) the system needs to adjust presented content accordingly.

Finally, table-top 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).

8. ACKNOWLEDGMENTS

9. REFERENCES

[1] Kruger, R., Carpendale, S., Scott, S. D., and Greenberg, S. 2004. Roles of orientation in tabletop collaboration: Comprehension, coordination and communication, *Journal of Computer Supported Cooperative Work*, 13(5-6):501-537.

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

[3] Lam, J., Kapralos, B., Collins, K., Hogue, A., Kanev, K. 2010. Amplitude Panning-Based Sound System for a Horizontal Surface Computer: A User-Based Study, In *Proceedings of the 9th IEEE International Symposium on Haptic Audio-Visual Environments and Games HAVE 2010*, Phoenix, AZ, USA, October 16-17, 2010, pp.174-178.

[4] K. Collins, B. Kapralos, A. Hodge, and A. Hogue. 2010. 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 2-4, 2010.

[5] Bronkhorst, A. W. 1995. Localization of real and virtual sound sources. *Journal of the Acoustical Society of America,* 98(5, Pt 1), November, 1995, 2542-2553.

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

[7] Ohta, Y., and Obata, K. 2007. 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, preprint:7230.

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