Amplitude Panning-Based Sound System for a Horizontal Surface Computer: A User-Based Study

Jonathan Lam*, Bill Kapralos* Karen Collins†, Andrew Hogue*, and Kamen Kanev‡
*Faculty of Business and Information Technology, and Health Education Technology Research Unit,
University of Ontario Institute of Technology. Oshawa, Ontario, Canada.
Email: {jonathan.lam, bill.kapralos, andrew.hogue}@uoit.ca
†Canadian Centre of Arts and Technology, University of Waterloo. Waterloo, Canada.
Email: collinsk@uwaterloo.ca
‡Research Institute of Electronics, Shizuoka University. Hamamatsu, Japan.
Email: kanev@rie.shizuoka.ac.jp

***Abstract*—Given the growing popularity of multi-touch mobility devices (e.g., iPods, smartphones), the move to multi-user touch screens and horizontal surfaces is a likely trajectory of the technology. Before smart tables become widely accepted, there are many questions particularly with respect to sound production and reception and multi model interaction for these devices that need to be explored (i.e., the interaction of sound and video cues). With respect to the sound interface, this introduces several design issues that must be addressed. More specifically, where do we position the loudspeakers and where should we position sounds in the mix, (in which speaker) for best reception? Here we describe a simple, and computationally efficient bilinear interpolation-based amplitude panning method designed specifically for horizontal surface computers with four loudspeakers. Preliminary user-based experiments were conducted to test the effectiveness of the method. Preliminary results indicate 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.**

I. Introduction

Sound in entertainment applications such as video games plays an important role: it can help to communicate important information to the player; it can serve as a sound symbol or leitmotif; it can help to situate the player in a specific location; and it reduces learning curves and creates a 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 such, sound (music, dialogue, and sound effects) for television, film, software, and games has been designed accordingly, with the placement of the speakers and the sound mixing all developed based on this format.

Recently, smart table-top touchscreen computers (also known as surface computers, smart table computers, or smart tables), where users position themselves around a horizontal computer screen in a manner similar to sitting around a "traditional" table, have been introduced (e.g., the Microsoft surface is a multi-touch computer that responds to natural hand gestures and real-world objects[1]. Although smart tables have yet to be primarily designed as consumer models, with the growing popularity of multi-touch mobility devices (e.g., iPods, smartphones), the move to multi-user touch screens and a horizontal surface is a likely trajectory of the technology. Moreover, these devices may well become a part of social entertainment, where families and friends can interact with each other around a table-like surface.

Before smart tables become widely accepted, there are many questions particularly with respect to sound production and reception and multi model interaction for these devices that need to be explored (i.e., the interaction of sound and video cues). No longer is there just one person in front of a screen, these surface and multi-touch computers are designed as multi-user devices. However, this introduces several design issues, particularly with respect to the sound interface (i.e., input/output of sound), that must be addressed in order to ensure consistent and effective sound interface. More specifically, where do we position the loudspeakers when there are two users opposite each other playing a game (i.e., where is the "front")? How does our perception of sound change when we are leaning over our computer screen versus facing it? Where should we place the loudspeakers? Where should we position sounds in the mix, (in which speaker) for best reception? In designing user interfaces for multi-user applications on touch screens that rely on hand gestures for more natural interaction, how should the applications respond sonically to these interactions?

Recently, we have begun investigating sound interaction techniques for horizontal surface computing platforms. In order to develop novel and innovative sound interaction techniques for smart table computers we will be drawing on the very new field of sonic interaction design, which studies the use of sound to convey information, meaning, aesthetic, and emotional qualities in an interactive context [1]. Sonic interaction design encompasses a range of technical and theoretical concepts, including sound design, interaction design, the interactive arts, auditory display, psychology, usability,

[1] http://www.microsoft.com/surface/en/us/Pages/Product/WhatIs.aspx

978-1-4244-6508-8/10/$26.00 ©2010 IEEE

user experience, design, and perception. We have begun experimenting with a number of sound generation techniques and have proposed a simple amplitude panning method (which we term *distance-based amplitude panning* 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 its distance to the virtual sound source (see [2] for greater details). Here, we present the results of a preliminary user-based experiment that was conducted to examine the effectiveness of distance-based amplitude panning method.

A. Paper Organization

The remainder of the paper is organized as follows. In Section 2, an overview of the importance of sound in interactive displays/applications and an introduction to sound interface/interaction design is provided. A description of the amplitude panning method is provided in Section 3. A description of the user-based experiment conducted to test the effectiveness of the method in addition to the results of the experiment is provided in Section 4. Finally, concluding remarks and plans for future research are provided in Section 5.

II. Background

The smart table is a collaborative computer with a horizontal screen, raising interesting problems for surround sound, which has been developed for vertical screens where users all face the same direction. The most common configuration of home theater surround systems (Dolby 5.1) is characterized by additional speakers to a stereo pair. In a theatre, the central speaker (which handles the majority of sound) between the front stereo pair "locks" the dialogue to the screen and improves performance for listeners who are seated off-centre. Two channels are used in the rear on side walls or back wall. The ".1" is the *low frequency effects* (LFE) channel-separate subwoofer (see [3]). In home systems, loudspeakers can be carefully configured for optimal sound (usually with stereo pairs forming an equilateral triangle with the listener and speakers at ear height)

These specifications for placement have been created over several decades based on trial and error and experience, and an agreed upon specification for ideal placement is now promoted by the Grammy awards, the Audio Engineering Society, and other organizations. New ways of testing optimal configurations are still in development, as the industry moves to create a more three-dimensional space [4], and moves to 7.1 and 8.1 surround sound. The placement of speakers is important in terms of audio reception, since poor speaker placement can result in undesirable "colourations" and "lost" frequencies. The subwoofer's placement in particular can increase bass by reflecting off hard surfaces (walls). A change in placement, such as putting a speaker in a corner or against a wall, can mean certain frequencies (i.e., the bass) are emphasized.

The difficulty when it comes to smart tables, however, is that the position of the listeners has changed, so that two listeners may be faced opposite to each other in front of the table. 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 speakers (and therefore sound) on a horizontal plane around the users, how can we give the perception of depth? Is it a matter of altering the pitch of sounds, or is it necessary to physically change the speaker location to achieve this affect (essentially, tilting the whole 5.1 set-up on an axis)? There are two possible solutions to this problem: i) move the speakers, 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 speakers around based on the number of people using the smart table, and so the second option offers the most viable alternative, although we will test both solutions in a variety of ways.

A. Bilinear Interpolation-Based Amplitude Panning Method

Fig. 4 provides an overview of the smart table hardware setup. The system is intended to accommodate multiple users (1-4) and consists of i) a smart table, ii) a video camera, iii) four loudspeakers (currently, JVC SX-XSW31 are being used), and iv) four microphones. The (external) video camera is used specifically for object recognition and will not be described here (see [2]). The multitouch table is a custom built display system developed internally. It uses a standard dining room table from IKEA with frosted glass as the projection surface. An ultra short throw projector is used for rear projection (we use the Hitachi CP-A100 as it enables great control of the projection size. Optitrack cameras are used as they provide direct illumination with their built-in IR LEDs, operate at 100 fps and provide decent resolution/performance tradeoffs. The cameras have 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.

Fig. 1. Conceptual overview of the proposed system.

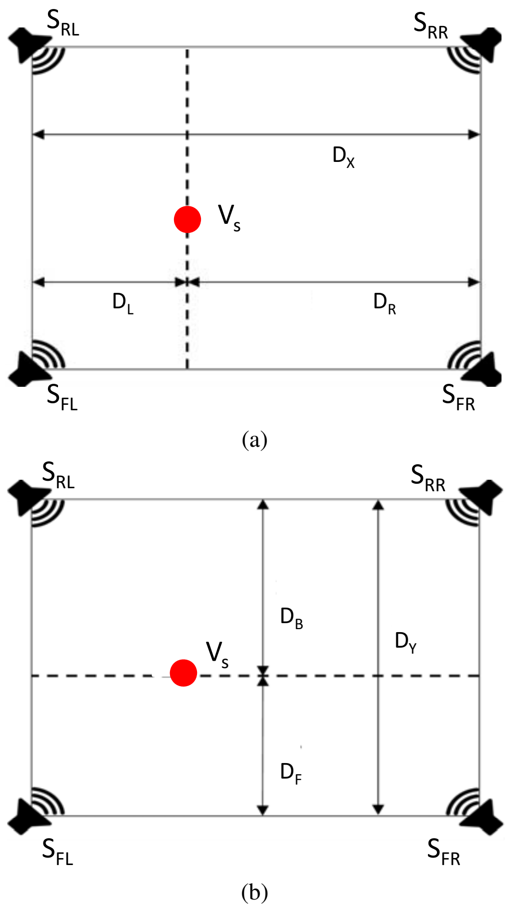


Fig. 2. Bilinear interpolation-based amplitude panning example. (a) Horizontal scalars. (b) Vertical scalars.

As described, the system consists of four loudspeakers (each at one of the four smart table corners) facing the table computer 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. We have begun experimenting with a number of amplitude panning 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 its distance to the virtual sound source [2].

Another simple and easy to calculate technique is based on bilinear interpolation and the sound is panned between loudspeaker pairs. Referring to Fig. 2(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 horizontal distance between them and the virtual sound source D_R and D_X:

$$V_L = D_L/D_X \tag{1}$$
$$V_R = D_R/D_X \tag{2}$$

Next, in a similar manner and referring to Fig. 2(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:

$$V_F = D_F/D_Y \tag{3}$$
$$V_B = D_B/D_Y \tag{4}$$

Finally, the amplitude (level) for each of the four loudspeakers is determined as follows:

$$S_{FL} = V_F \times V_L \tag{5}$$
$$S_{FR} = V_F \times V_R \tag{6}$$
$$S_{RL} = V_B \times V_L \tag{7}$$
$$S_{RR} = V_B \times V_R \tag{8}$$

III. METHODS

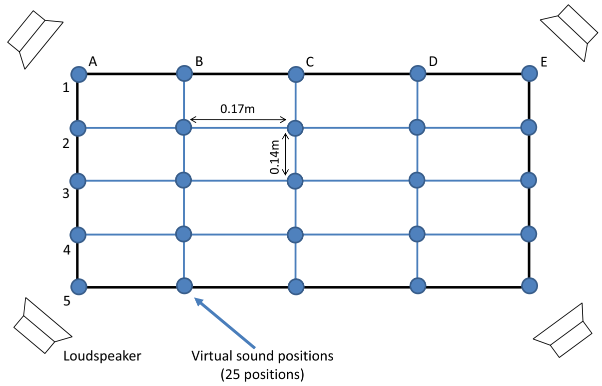
A. Participants

A total of eight unpaid volunteers participated in the experiment. Participants were either researchers, or students from the University of Ontario Institute of Technology and the average participant age was 26 ($\sigma = 4.7$). Participants reported no history of auditory disease or disorders. The experiment abided by the University of Ontario Institute of Technology Ethics Review process for experiments involving human participants.

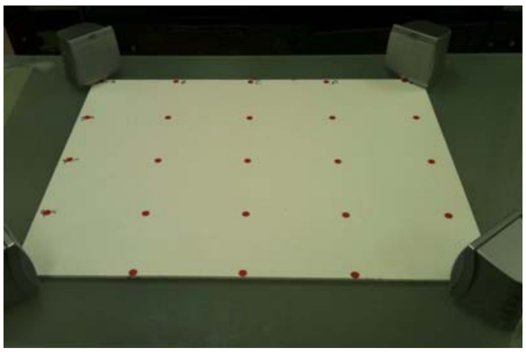
B. Auditory Stimulus

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 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). 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 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



(a)



(b)

Fig. 3. Experimental setup. (a) Grid of the 25 virtual sound source positions and loudspeaker setup. (b) Actual setup.

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.

C. 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 Fig. 1(a) (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 stimuli 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 Fig. 3). 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.

IV. 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. 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 I. An examination of Table I 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).

A graphical illustration of the average error for each participant (averaged across each of the 25 virtual sound source positions) is provided in Fig. 4.

V. Conclusion

Smart tables represent a further step towards what is known as ubiquitous or pervasive computing - that is, in the very near future we will not 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 smart-table computing becomes widely accepted, there are many questions, particularly with respect to sound production and reception, and multi model interaction for these devices that needs to be explored. In this paper we described a simple and

	A (m)	B (m)	C (m)	D (m)	E (m)
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 I

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 EIGHT PARTICIPANTS.

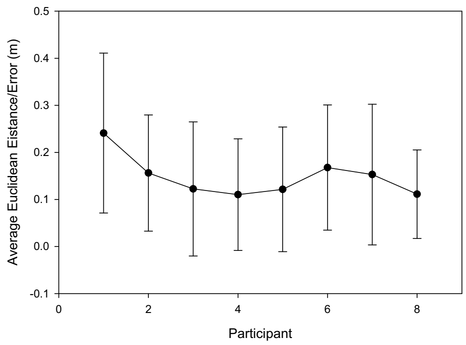


Fig. 4. Average error (Euclidean distance or the difference between the actual and perceived virtual sound source positions) and standard deviation for each participant (averaged across each of the 25 virtual sound source positions).

computationally efficient amplitude panning method based on bilinear interpolation. We also conducted a preliminary user study with eight participants to provide an indication of the method's effectiveness and to also provide avenues for further investigation. Although the method is easy to implement and compute, preliminary 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). Although preliminary, the results do provide an indication of the potential issues presented 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 [5] and changes in frequency [6]. Furthermore, sounds that have more formants/overtones are easier to localize than sine waves, and reverberation will also aid sound source localization (see [7]).

Future work will examine variations/modifications to the bilinear interpolation method followed by further, more thorough and extensive user-based experiments that will also include multiple individuals seated around the table as opposed to a single participant considered here. Surface 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 conflicting visual stimuli.

ACKNOWLEDGMENT

The financial support of the Social Sciences and Humanities Research Council (SSHRC) of Canada in the form of a Research Development Initiatives (RDI) grant to B. Kapralos and K. Collins, and the Natural Sciences and Engineering Research Council of Canada (NSERC) in the form of a Discovery Grant to B. Kapralos is gratefully acknowledged.

REFERENCES

[1] D. Rocchesso, S. Serafin, F. Behrendt, N. Bernardini, R. Bresin, G. Eckel, K. Franinovic, T. Hermann, S. Pauletto, P. Susini, and Y. Visell, "Sonic interaction design: sound, information and experience," in *Proceedings of the ACM CHI 2008: Extended Abstracts*, Florence, Italy, April 5 -10 2008, pp. 3969–3972.

[2] K. Collins, B. Kapralos, and K. Kanev, "Sound interface design for smart table computer interaction," in *Proceedings of the 12th International Conference on Humans and Computers*, Hamamatsu, Japan, December 9-10 2009.

[3] F. Rumsey, *Spatial Audio*. Woburn, MA. USA: Focal Press, 2001.

[4] S. N. Goodwin, "How players listen," in *Proceedings of the of the Audio Engineering Society 35th International Conference*, London, UK, February 11-13 2009.

[5] 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*, vol. 87, no. 4, pp. 1728–1731, 1990.

[6] Y. Ohta and K. Obata, "A method for estimating the direction of ound 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 200y, p. Preprint: 7230.

[7] 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*, vol. 43, no. 6, pp. 1255–1259, 1968.