Sound Localization on Tabletop Computers: A Comparison of Two Amplitude Panning Methods

BILL KAPRALOS and JONATHAN LAM, University of Ontario Institute of Technology, Oshawa, Canada
KAREN COLLINS, University of Waterloo, Waterloo, Canada
ANDREW HOGUE, University of Ontario Institute of Technology, Oshawa, Canada
KAMEN KANEV, Shizuoka University, Hamamatsu, Japan
MICHAEL JENKIN, York University, Toronto, Canada

Tabletop computers (also known as surface computers and smart tables) have been growing in popularity for the past decade and are poised to make inroads into the consumer market, opening up a new market for the games industry. But before tabletop computers become widely accepted, there are many questions with respect to sound production and reception for these devices that need to be explored, particularly when it comes to multimedia consumption on the devices. For example, which loudspeaker setups should be used to take into consideration the multi-user nature of tabletop computers, and which panning method(s) maximize the spatial localization abilities of the user(s)? Previous work suggests that a quadraphonic diamond-shaped loudspeaker configuration—whereby a loudspeaker is placed at each of the four sides of the tabletop computer—leads to more accurate localization results when compared with a traditional quadraphonic loudspeaker configuration—whereby a loudspeaker is placed at each of the four corners of the tabletop computer. Given this preference for a diamond loudspeaker configuration, we examine two amplitude-panning methods (bilinear interpolation and inverse distance) for spatializing a sound on the (horizontal) surface of the table-computer with a diamond loudspeaker configuration. Results from the study detailed in this paper indicate that there are no significant differences between the two methods and that both methods are prone to error.

Categories and Subject Descriptors: H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems—*Audio input/output*; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*Auditory (non-speech) feedback*

General Terms: Human Factors, Experimentation

Additional Key Words and Phrases: Tabletop computer, surface computer, audio interaction, amplitude panning, loudspeaker configuration, sound localization

ACM Reference Format:
Bill Kapralos, Jonathan Lam, Karen Collins, Andrew Hogue, Kamen Kanev, and Michael Jenkin. 2014. Sound localization on tabletop computers: A comparison of two amplitude panning methods. ACM Comput. Entertain. 12, 2, Article 4 (December 2014), 19 pages.
DOI: http://dx.doi.org/10.1145/2701657.2633415

1. INTRODUCTION

Tabletop computing has been growing in popularity for the past decade. Most major computer companies have now developed, or are working toward, a tabletop computer, and we can expect to see tabletop computers make inroads into the consumer market

---

Authors' email: J. Lam, B. Kapralos, and A. Hogue: {jonathan.lam, bill.kapralos, Andrew. hogue}@uoit.ca; K. Collins: collinsk@uwaterloo.ca; K. Kanev: kanev@rie.shizuoka.ac.jp; M. Jenkin: jenkin@cse.yorku.ca.

DOI: http://dx.doi.org/10.1145/2701657.2633415

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] detects where the surface of a table is being touched, allowing it to project an image 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 toward a consumer model. As with the majority of consumer computer technology, entertainment applications will likely drive the success of consumer-model tabletop computing. Audio is a key component of many entertainment applications, and the generation of effective audio for interactive multi-user tabletop displays is a novel application area with many open questions.

Video games are a logical application for tabletop-computing technology, given that games have been played on table-like surfaces for thousands of years (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 have disappeared, along with the arcades). The move to tabletop computers will likely introduce a new market for the games industry as the technology encourages multi-player social gaming—many users can crowd around a table quite naturally.

The move from vertical-screen video games to a horizontal tabletop introduces interesting questions with respect to the implementation of graphics and sound. Questions of cooperation, orientation, and viewing angle will drive innovation in imagery and have been explored elsewhere [Kruger et al. 2004; Scott and Carpendale 2010]. However, there are many questions that remain unanswered regarding the use of audio in an interactive tabletop setting. For decades we have experienced our digital audiovisual media on a vertical screen; our televisions, movie theaters, and computer screens typically present information vertically, in front of the user. When multiple users are present, all of the users sit in front of the vertical display surface. Content and rendering devices for such media have 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 toward them. With tabletop computing these assumptions are no longer valid. Users are now above the display surface, surrounding it, and viewing angles are typically oblique. Given that tabletop computers are intended for multiple users and emphasize collaboration, headphones are typically not an appropriate vehicle for audio since they may limit and even deter natural verbal interaction among 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 been able 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. There are several open problems with respect to audio interaction on a tabletop computer, including the following that our ongoing work is addressing:

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)?

Our previous work in audio displays for tabletop computers compared sound localization accuracy with two loudspeaker configurations [Collins et al. 2011; Lam et al. 2010]. Although preliminary, results of our previous work suggest that a diamond-shaped loudspeaker configuration allows for more accurate localization and is preferred by users, compared with the traditional quadraphonic configuration. In Lam et al. [2010] a user-based experiment tested the effectiveness of the bilinear interpolation amplitude panning method that involves panning of the sound between loudspeaker pairs, on a tabletop computer using a traditional quadraphonic configuration whereby a loudspeaker was placed at each corner of the surface of the display and faced inward toward the center of the surface on a 45-degree angle (this implied that two of the loudspeakers closest to the participants were actually facing away from the participants). Sounds were spatialized to one of 25 pre-defined locations on the horizontal surface; the participants' task was to localize the sound. Results indicate that bilinear interpolation amplitude panning with the traditional quadraphonic loudspeaker configuration is prone to varying error across individuals, particularly for virtual sound source positions that are closest to the participant [Lam et al. 2010]. Collins et al. [2011] examined listener preference of the traditional and diamond loudspeaker configurations. A touch-table video game version of tabletop air hockey similar to the two-player Pong game distributed by Atari in the 1970s 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 sounds 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 10 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 was not explicitly examined in this study, and this became the motivation for the current study.

The large spatialization errors associated with the standard quadraphonic loudspeaker configuration and the user preference for the diamond configuration observed in our previous work suggest that a diamond loudspeaker configuration could form the basis of an effective tabletop computer sound system. Here we present the results of an experiment that compared two amplitude panning methods, the bilinear interpolation method and the inverse distance method using a diamond loudspeaker configuration, with respect to sound localization on a horizontal surface. The purpose of the current study is to examine sound localization using the preferred diamond loudspeaker configuration with two amplitude panning methods to determine whether one or the other leads to more accurate sound localization and bring us closer to answering questions regarding the delivery of spatial audio on a tabletop computer.

1.1. Paper Organization

The remainder of the paper is organized as follows: Section 2 provides background information regarding spatial sound in video games and sound panning methods in general. A more detailed description of the bilinear interpolation and the inverse distance amplitude panning methods is provided in Section 3. Particulars of the experimental methods are provided in Section 4, and experimental results are presented in Section 5. A discussion of the experimental results is provided in Section 6, and finally, concluding remarks and plans for future research are provided in Section 7.

2. BACKGROUND

2.1. Spatial Sound in Games

Surround sound systems employ a collection of loudspeakers that surround a listener [Gardner 1998]. The sound output by each loudspeaker is recorded using distinct recording channels/microphones or created synthetically. By presenting these recorded sounds through their corresponding loudspeaker, the aim is to surround the listener with sound, providing them with the impression of sounds coming from all directions. Surround sound consumer setups have changed—and continue to change—as technology advances, but the core idea remains of having a series of loudspeakers placed on a horizontal plane at approximately ear height (or above) around the listener. For example, the quadraphonic surround sound system employs four loudspeakers: two in front of the listener (front-left and front-right) and two behind the listener (rear-left, and rear-right). The sound field (e.g., concert, performance, etc.) in this case is captured with four microphone elements/channels that are typically packed into a single housing, and each element is mapped to one of the four loudspeakers. The sound recorded in each of the four microphones is then output to the corresponding loudspeakers (see [Robjohns 2001a, b, c] for greater details regarding Quadraphonics and surround sound in general).

In contrast to traditional surround sound techniques, spatial sound reproduction provides a virtual sound source with such attributes as left-right, back-forth, and up-down, regardless of the number of loudspeakers used [Cohen and Wenzel 1995]. Spatial sound can provide effective non-visual cues in a variety of different applications, including the development of compelling virtual and game environments. Spatialized audio cues have been found to increase the sense of presence and immersion, compensate for poor visual cues, increase the subject's emotional involvement in the simulation, and provide higher realism and the ability to identify directionality and proximity within the simulated environment [Durlach and Mavor 1995; Shilling and Shinn-Cunningham 2002; Shilling et al. 2002]. Work from the field of virtual reality and simulation suggests the importance of audio to a realistic portrayal of three-dimensional space:

> [Spatial] sound can contribute to the sense of immersivity in a 3-D environment. One might be able to work more effectively within a VR environment if actions were accompanied by appropriate sounds that seemingly emit from their proper locations, in the same way that texture mapping is used to improve the quality of shaded images [Begault 1994].

In terms of video games and virtual-environment-based training systems, spatial sound is particularly important because the user must not only be aware of what is in front of their character, but also what may lie behind the character. Matthew Lee Johnston of Microsoft explains:

> Traditional stereo has been used to localize a sound in the player's forward visual field. What 3-D audio adds is the ability to localize the sound behind the player, which is arguably way more important, since the sound is usually

> the only way to provide the player with feedback about what's going on behind them. Some games use maps and have "rear view" options, or even let you pan your visual field around to look, but using 3-D audio to position an object behind the player is not only more immediate and instinctual, but it allows the player to focus simultaneously on the fore and aft perspective (cited in [Miller 1999]).

It is not surprising, therefore, that surround sound in games arose at about the same time as the first-person shooter genre (FPS). FPS games arrived with the release of Wolfenstein 3D (id software 1992). However, it was Doom (id software 1993) that was the first major FPS success [Collins 2008]. Sound effects were a key feature of Doom, alerting the player to not only the location of enemy demons, but also the type of demon. Prince describes the concept for sound in the game:

> There were several classes of sounds in Doom. One was general active sounds that were not attached to any one demon. These were more or less ambient sounds, but they didn't play until demons close to the player "woke up" (usually based upon the player making some noise in the area)... Another helpful thing about the sound driver was that the volume of sounds depended upon the distance from the player to the source of the sound. This helped keep the overall volume down during noncombat. It also stood to help scare the pants off the player when a demon in a dark niche woke up and immediately screamed his attack sound [Prince 2006].

A similar use of sound effects is common in stealth games where non-playing characters are alerted to the player's presence by sound (see Splinter Cell (Ubisoft 2002)). These types of games rely heavily on spatialized sound technology, helping the player to locate objects and people in a three-dimensional space. In other words, in many games it is incredibly difficult to play without spatialized sound.

2.2. Point-of-Audition and the Tradition of Vertical-Screen-Based Spatial Sound

Movie and game surround sound is mixed as discrete mixes, or it is matrixed. In discrete mixes, individual audio channels provide unique audio signals. The sound is encoded in a one-to-one manner with the intended set of audio output devices. This leads to issues when fewer (or more) audio output channels are present. A matrixed system encodes extra information into a two-channel stereo signal and recovers those extra channels through a decoding process (e.g., Dolby Pro Logic) if more than two output channels are available. Many video games now use matrixed surround. If the user has only a stereo system, the information can be played back, and amplitude adjustments and phasing is used to re-create the original mix effect. If the player has a decoder (e.g., Dolby Pro Logic or Dolby Pro Logic II), the sound will play back as a surround mix. Moreover, if the player is playing a game that was mixed only in stereo, the Pro Logic II system will guess where the sounds should be routed and transform the sound into a surround mix.

At issue, however, is that these systems have always been designed with a single viewpoint in mind and a vertical screen. Stephan Schütze describes the Xbox system and spatial sound:

> The main characteristic that sets a console game system apart from the real world is that the player is static, looking at a screen that is always positioned to the front, and experiencing audio output from speakers positioned around the listener [Schütze 2003].

This static listening position has been referred to in film studies as "point of audition," analogous to point of view in cinema, in which there is an implied distance from the listener to the sound source(s) [Langkjaer 1997]. In other words, the (presumed) viewing angle is closely tied to the loudspeaker configuration, and also to the ways in which the media is mixed. However, with tabletop computing, the viewer's position is not fixed in terms of the viewing angle to the screen. The viewer may position him or herself at any position around the screen (and with a variety of body positions and angles, from leaning over the tabletop to sitting normally and vertically). Given the importance of spatial sound, and that the accuracy of the perception of spatial sound is dependent on the presumed viewing angle, tabletop computers clearly introduce a variety of problems with respect to the loudspeaker positioning and sound mixing.

2.3. A Brief Overview of Amplitude Panning

Amplitude panning is a common approach to generating spatialized sound. By adjusting the amplitude (volume) of the signal applied to each of N loudspeakers through the use of a gain factor (that is, the factor by which the power is amplified), the listener perceives a virtual sound source emanating from a direction that is dependent on the gain factors [Pulkki 2001]. Mathematically, amplitude panning can be described by

$$b_i(t) = g_i(t) \times s(t) \quad i = 1 \ldots N,$$

where $b_i(t)$ is the signal output by loudspeaker i at time t; $s(t)$ is the unprocessed sound applied to each of the loudspeakers at time t; $g_i(t)$ is the gain factor applied to the signal delivered to loudspeaker i; and N is the total number of loudspeakers being used. $N = 2$ forms the basis of the popular two-channel (stereo) configuration whereby the listener is placed symmetrically (in the horizontal plane), equidistant between the left and right loudspeakers in the "sweet spot." By scaling the amplitude of the signal applied to the left and right loudspeakers (with the gains g_L and g_R respectively), the virtual sound source can be positioned anywhere on the active arc (a semicircle between the two loudspeakers with radius equal to the distance between the listener and each of the loudspeakers [Pulkki 2001]), and therefore can appear to be located in what is known as the "phantom centre," between the loudspeakers. Individual gains can be established in a variety of ways (e.g., in a stereo configuration, using the stereophonic law of sines [Bauer 1961a, b] or the tangent law [Bennett et al. 1985]) although commonly, the values of these gains are not listener-position dependent. Alternative listener-position-dependent techniques also exist (see vector-base amplitude panning (VBAP) [Pulkki 1997], as an example). A review of amplitude panning is available in [Pulkki and Karjalainen, 2001] and [Pulkki 2001].

3. TECHNICAL OVERVIEW

Figure 1 provides a conceptual overview of the tabletop hardware setup. The system is intended to accommodate multiple users (1 to 4) and consists of (i) a table (surface computer), (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 [Collins et al. 2009]). The multi-touch table is a custom-built display system developed internally. An ultra short throw projector is used for rear projection (Hitachi CP-A100 allows great control of 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. Although

Fig. 1. Conceptual overview of the tabletop system.

this is our tabletop computer setup, here we were concerned only with examining sound localization on a horizontal surface, and our experiments did not include any visuals or touch-based input.

3.1. Bilinear Interpolation Amplitude Panning

One computationally simple technique (and one of the methods used in the experiment presented here) is based on bilinear interpolation, where the sound is panned between loudspeaker pairs. The four loudspeakers are arranged in a rectangular formation (See Figure 2). Referring to Figure 2(a), first the "left-horizontal" scalar V_L is determined for the front-left and back-left loudspeakers (S_{FL} and S_{BL} respectively) by dividing the horizontal distance between the right-hand pair of loudspeakers (S_{FR} and S_{BR}) and the virtual sound source D_R (the virtual sound source is denoted by V_S) by the 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 back-right loudspeakers (S_{FR} and S_{BR} respectively), is determined by dividing the horizontal distance between the left-hand pair of loudspeakers (S_{FL} and S_{BL}) by the virtual sound source D_L and D_X (note that $\mathrm{D_R + D_L = D_X}$):

$$V_L = D_R/D_X$$
$$V_R = D_L/D_X$$

The loudspeakers are grouped into a front pair (S_{FL} and S_{FR}) and a back pair (S_{BL} and S_{BR}), and the scalars (V_F and V_B) are determined in a similar manner (see Figure 2(b)):

$$V_F = D_B/D_Y$$
$$V_B = D_P/D_Y$$

The amplitude (level) for each of the four loudspeakers is determined using the four panning coefficients $\mathrm{V_F}$, $\mathrm{V_R}$, $\mathrm{V_L}$, $\mathrm{V_R}$:

$$S_{FL} = V_F \times V_L$$
$$S_{FR} = V_F \times V_R$$
$$S_{BL} = V_B \times V_L$$
$$S_{BR} = V_B \times V_R$$

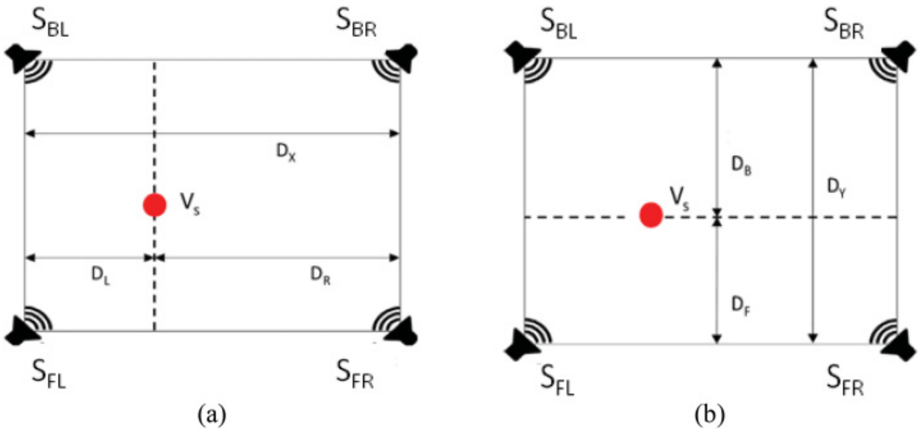

Fig. 2. Bilinear interpolation amplitude panning example. (a) Horizontal scalars. (b) Vertical scalars.

The diamond configuration is identical except that the coordinate system is rotated by 45 degrees.

3.2. Inverse Distance Amplitude Panning

With the distance amplitude method, the sound output at each loudspeaker is scaled by the distance between the corresponding loudspeaker and its distance to the virtual sound source (see [Lossius et al. 2009]). Here s_i is the signal (sound) applied to the i^{th} loudspeaker and is calculated as:

$$s_i = \frac{S_v}{(d_i^r + k)}$$

where s_v 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 the roll-off coefficient (derived empirically through informal testing), and k is a small constant value primarily for preventing errors from division by zero. Figure 3 provides a graphical example, and as shown, the greater the distance between the virtual sound source and a loudspeaker, the less the influence of that loudspeaker to the output sound. Each s_i is normalized, and this normalized signal is applied to the corresponding loudspeaker.

4. METHODS AND MATERIALS

4.1. Participants

Participants consisted of volunteer students and researchers from the University of Ontario Institute of Technology (UOIT). A total of 10 volunteers participated in the experiment. The average age of the participants was 24. None of the participants reported any 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.

4.2. 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 cutoffs of 200 Hz and 10 kHz, respectively. The auditory stimulus was output through JVC SX-XSW 31 loudspeakers (four loudspeakers in

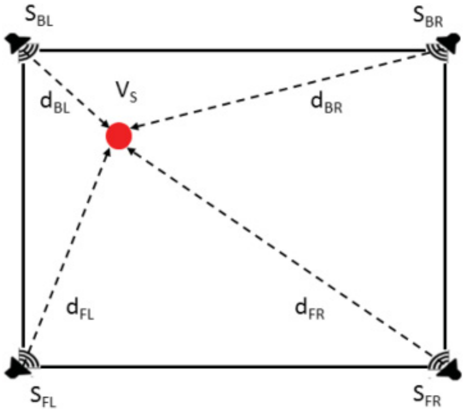

Fig. 3. Inverse distance amplitude panning example. The greater the distance between the virtual sound source (V_S) and a loudspeaker, the less the influence of that loudspeaker to the output sound. In this example, the sound from loudspeaker S_{BL} will be the strongest, while the sound from loudspeaker S_{FR} will be the weakest.

total). One loudspeaker was placed on the center of each of the sides of the surface in a diamond configuration (see Figure 4(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 40.0m × 20.0m × 9.5m). 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 not used in the experiment 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. 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 to a position on the surface using either the bilinear interpolation or the inverse distance amplitude panning methods. 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 4(b)). Each of the 36 grid positions was clearly labeled in large text (the grid itself was professionally printed and centered around a table; these experiments considered only sound localization on a horizontal surface hence, the actual surface computer itself was not needed and not used).

For each trial, a sound was generated and participants were instructed to choose from the set of possible grid locations, which were 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 4(a) and enter their choice of row and column using a standard computer keyboard). The next trial began after the participant entered their

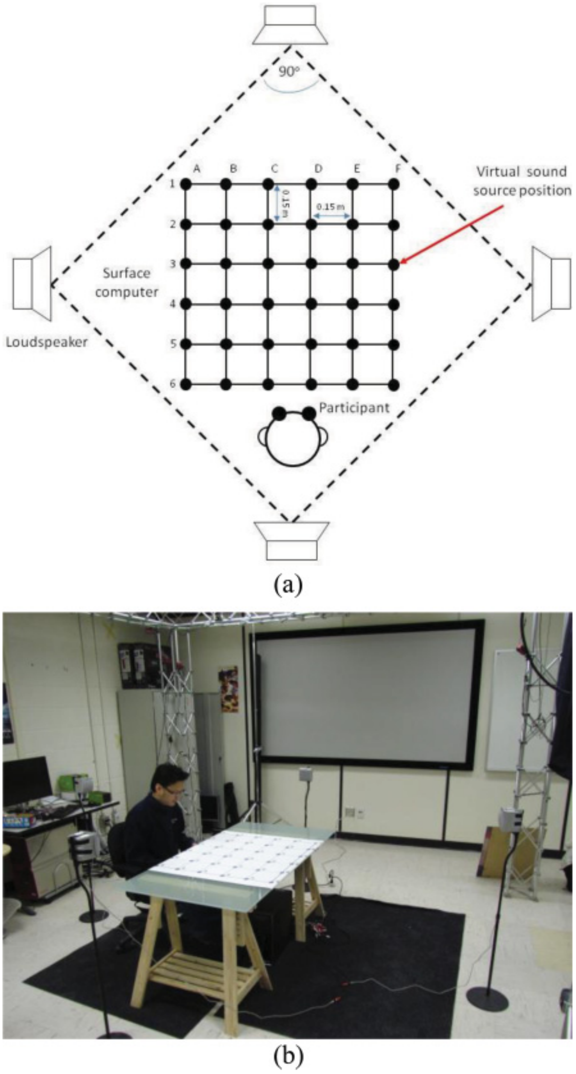

Fig. 4. "Diamond" loudspeaker setup. (a) The 36 virtual sound source positions and loudspeaker setup. (b) Actual setup.

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 each position was repeated two times for each of the two amplitude panning methods considered, leading to a total of 144 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 (test trial responses were ignored).

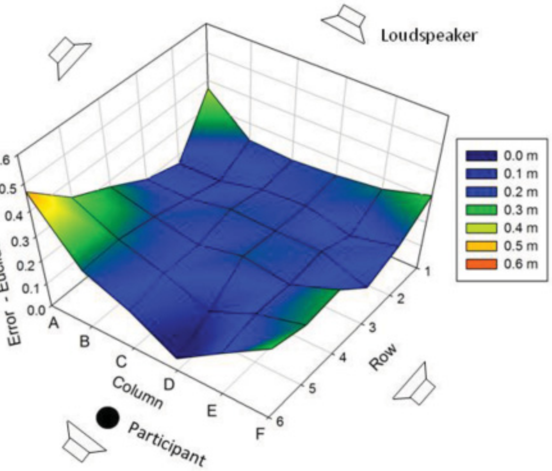

Fig. 5. Bilinear interpolation amplitude panning results. Average error (Euclidean distance or the difference between the actual and perceived virtual sound source positions, measured in meters) for virtual sound source position (averaged across each of the 10 participants).

5. RESULTS

The Euclidean distance between the *actual* virtual sound source position (i.e., the location that the sound was spatialized to) and the *perceived* virtual sound source position (i.e., the position that the participants perceived the sound source to be emanating from) was 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, indicating participants were able to correctly localize the virtual sound source in each and every trial for both presentation techniques.

5.1. 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 10 participants) is summarized in the plot of Figure 5 and Table I. 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 ± 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 5, 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 ± 0.13 m), (1A; 0.37 m ± 0.13 m), (5A; 0.34 m ± 0.16 m), (6F; 0.31 m ± 0.10 m), and (1F; 0.31 m ± 0.18 m).

Figure 6 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.

5.2. 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 Figure 7 and Table II. 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 m. Given the grid

Table I. Bilinear interpolation amplitude panning results. Average error (Euclidean distance or the difference between the actual and perceived virtual sound source positions) and standard deviation (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

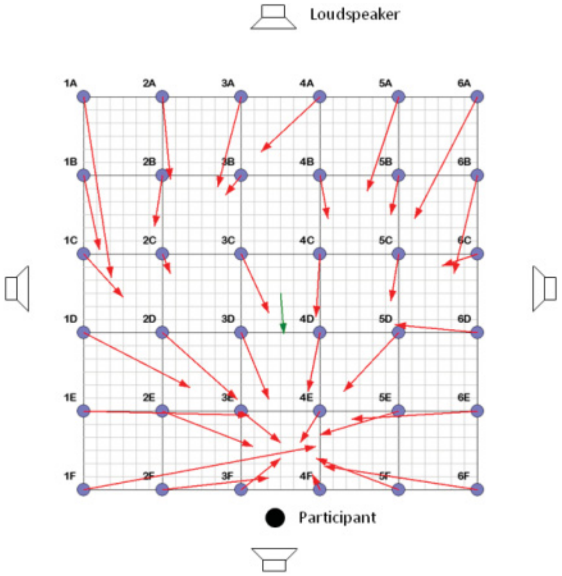

Fig. 6. Bilinear interpolation amplitude panning results. Error vector plot. The red arrows show the error for each of the virtual sound source positions, while the green arrow in the middle represents the average of all the red arrows.

spacing of 0.15m × 0.15m, participants were able to localize the sound source to within two positions of the actual virtual sound source. Examination of the plot of Figure 7 reveals that the largest errors occurred at three of the surface corners (toward 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).

The vector of the average error for each of the 36 positions is provided in Figure 8. As previously described, 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.

5.3. Comparison

An independent-samples t-test was conducted to compare sound localization error on a (horizontal) surface using bilinear interpolation and inverse distance amplitude panning methods, both with a diamond loudspeaker configuration. No significant difference

Table II. Inverse distance amplitude panning results. Average error (Euclidean distance or the difference between the actual and perceived virtual sound source positions) and standard deviation (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 III. The absolute difference in average error (measured in meters) between the results of bilinear interpolation and inverse distance amplitude panning methods for each of the 36 positions

	A	B	C	D	E	F
1	0.11	0.08	0.01	0.01	0.10	0.03
2	0.01	0.00	0.01	0.06	0.13	0.08
3	0.05	0.02	0.01	0.05	0.02	0.01
4	0.02	0.06	0.01	0.07	0.01	0.06
5	0.08	0.04	0.02	0.02	0.01	0.01
6	0.03	0.03	0.02	0.07	0.00	0.03

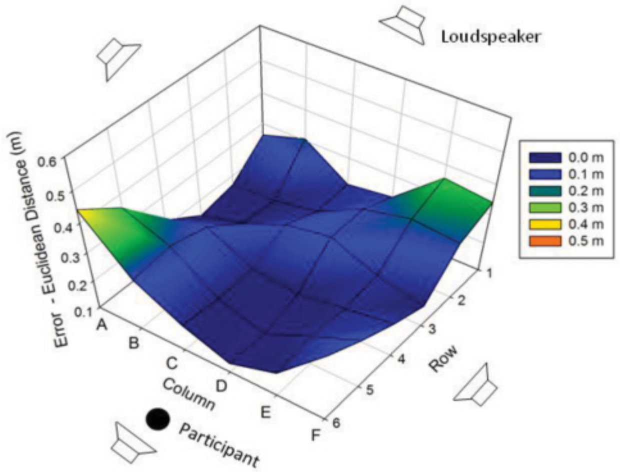

Fig. 7. Inverse distance amplitude panning results. Average error (Euclidean distance or the difference between the actual and perceived virtual sound source positions, measured in meters) for virtual sound source position (averaged across each of the 10 participants).

was found in the scores between the bilinear interpolation amplitude (M = 0.23, SD = 0.07) and the inverse distance (M = 0.24, SD = 0.07) amplitude panning methods; t = 0.52, p = 0.60.

6. GENERAL DISCUSSION

We investigated sound source localization on a horizontal surface using the bilinear interpolation, and inverse distance amplitude panning methods with a diamond loudspeaker configuration. Our results revealed that although both methods are prone to errors, they are not statistically different from each other (p = 0.60). The computational

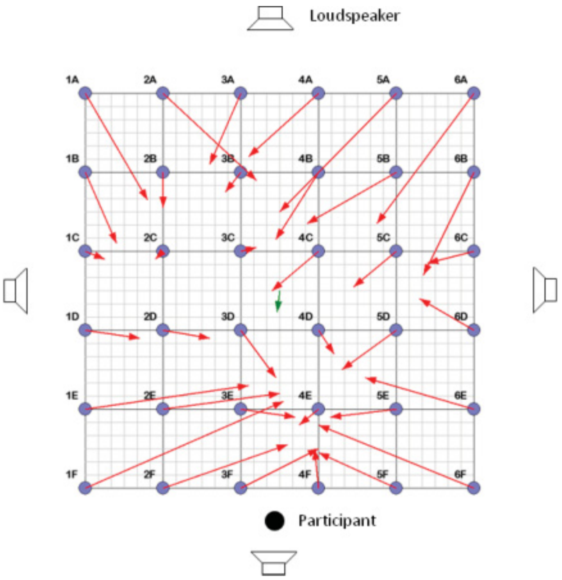

Fig. 8. Inverse distance amplitude panning results. Error vector plot. The red arrows show the error for each of the virtual sound source positions, while the green arrow in the middle represents the average of all the red arrows.

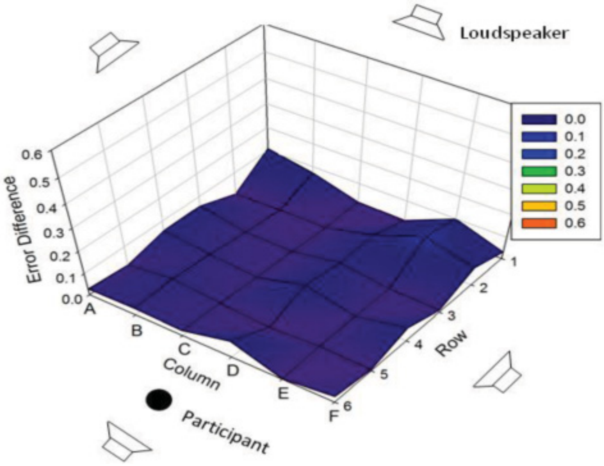

Fig. 9. The absolute difference in error between the results of bilinear interpolation and inverse distance amplitude panning methods for each of the 36 positions.

requirements are minimal for both methods, hence there is no difference in using one method over the other.

Unlike the work of Lam et al. [2010], 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 versus position of Figures 5 and 7, as well as the error vector plots of Figures 6 and 8). Participants faced forward 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° azimuth); this accuracy decreases for azimuth angles off to the side [Durlach et al. 1993]. Examining the vector plots of Figures 6 and 8 for both panning methods, it is clear that both amplitude panning methods show a general and consistent bias toward 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 toward the bottom half of the grid). Graphically, the green arrow in the center of each vector plot of Figures 6 and 8 represents the average error (magnitude and direction) across all positions, indicating that participants consistently (and erroneously) perceived the sound sources to be located closer to them and to the center of the surface. Each of the four loudspeakers was positioned such that its distance from the center of the table was 1.2 m. However, given that the participants were seated in front of one of the loudspeakers (see Figure 4(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 setup. The fact that the participants localized the sounds closer to 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.

In Section 1, the following three open problems with respect to audio interaction on a tabletop computer were described: "i) what loudspeaker constellations are appropriate for tabletop computers?", ii) how does our perception of spatial sound change with these different loudspeaker configurations?", and iii) "what panning methods should be used to maximize the spatial localization abilities of the user(s)?" 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 our previous results ([Lam et al. 2010; Collins et al. 2011]) provide us with greater insight and understanding regarding audio interaction on a tabletop computer. More specifically, we have shown that it is very difficult to accurately determine the actual position of a virtual sound source on a horizontal surface using listener-position-independent spatialization techniques, irrespective of the loudspeaker configuration or amplitude panning method employed (however, a diamond loudspeaker configuration is preferred over the traditional quadraphonic loudspeaker configuration). 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.

In this study we considered only the auditory component of the tabletop computer. However, the tabletop computer represents a tangible device with a well-defined flat surface that is intended for interactive-media use. 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.

7. CONCLUSIONS

Tabletop displays represent a further step toward 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 and spatial sound delivery in particular that require further investigation. Previous work regarding sound localization on a horizontal surface indicates large sound localization errors with bilinear interpolation using a standard quadraphonic loudspeaker configuration, and a preference for a diamond loudspeaker configuration. Given the preference for a diamond loudspeaker configuration and the lack of any studies investigating sound localization on a horizontal surface with such a loudspeaker configuration, here we examined sound localization on a horizontal surface using bilinear interpolation and inverse distance amplitude panning with a diamond loudspeaker configuration. Despite the preference for the diamond loudspeaker configuration, results indicate that accurately localizing a virtual sound source on a horizontal surface is a difficult task and prone to error, irrespective of the amplitude panning method.

7.1. Implications for Tabletop Content Designers

The results of previous work in conjunction with the results of this study indicate that sound localization on a horizontal surface is erroneous using either a bilinear interpolation or inverse distance amplitude panning method using either a traditional quadraphonic or the preferred diamond loudspeaker configuration. Although there are other panning methods to consider (e.g., vector base amplitude panning or VBAP [Pulkki 1997]), it is not anticipated that they will lead to much greater results, and therefore, developers and designers of applications for tabletop displays must account for this and not necessarily rely on accurate sound localization as a primary method conveying information. Rather, they might exaggerate placement when sound source positions correspond to positions with large error. It has been shown that sound source localization varies with frequency [Perrott and Saberi 1990] and changes in frequency [Ohta and Obata 2007], sounds that have more formants/overtones are easier to localize than sine waves, and reverberation also aids 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.

7.2. Future Work

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, we do not have “ground truth” data with which to compare our results. 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? Ongoing work is developing the hardware (a horizontal surface with actual sound sources at predefined positions) to collect this ground-truth data and allow a comparison to be made with the results described here. Future work will involve examining sound localization using different amplitude panning methods, such as the vector base amplitude panning (VBAP) method [Pulkki 1997], and drawing a comparison with the results presented here. Future work will examine multiple individuals seated around the table as opposed to a single participant considered here and will also examine what, if any, effect table size has on sound localization capabilities, and more specifically, is there an optimal size for one, two, three, or four 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 for use 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). Finally, despite the inherent error observed here, in many gaming applications, “pinpoint” 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.

ABOUT THE AUTHORS

Jonathan Lam is a graduate of the Game Development and Entrepreneurship program at the University of Ontario Institute of Technology (UOIT) in Oshawa, Canada. He is currently pursuing an MSc degree at the UOIT. Interests include developing unique and novel ideas for computer and video gaming, examining novel ways of utilizing digital audio in games and other areas, digital music composition, and appreciating vintage synthesizers and computer audio/music technology.

Bill Kapralos is an associate professor in the Game Development and Entrepreneurship Program at the University of Ontario Institute of Technology in Oshawa, Canada. His current research interests include multi-modal virtual environments/reality, serious games (and more specifically, examining the factors that lead to a maximum transfer of knowledge and retention), the perception of auditory events, and 3-D (spatial) sound generation for interactive virtual environments and video games. He chaired the ACM FuturePlay International Conference on the Future of Game Design and Technology from 2007 to 2010.

Karen Collins is a Canada Research Chair in Interactive Audio at the University of Waterloo, Canada, where she teaches sound for digital media and game design. Her research interests include interactive audio, mobile audio, game music, and gambling studies. She is the author of *Game Sound* (MIT Press, 2008) and has published extensively about sound in video games and slot machines.

Andrew Hogue is currently an assistant professor at the University of Ontario Institute of Technology in Oshawa, Canada, within the Game Development and Entrepreneurship program in the Faculty of Business and IT. His research interests include the development and evaluation of game design techniques for education. He collaborates with Bill Kapralos at UOIT and K. Collins at U. of Waterloo on serious games and audio interface research and is a co-investigator on a Google Faculty

Research Award looking at augmented reality and serious games in the mobile phone space. Dr. Hogue has also been involved in securing over $2M in research funds for a variety of robotics and simulation projects and has co-chaired the ACM Conference on Future Play since 2008.

Kamen Kanev received an MSc in mathematics and a PhD in computer science in 1984 and 1989, respectively. He is a professor with the Research Institute of Electronics, the Graduate School of Informatics, and the Graduate School of Science and Technology, Shizuoka University, Japan, where he teaches and supervises students majoring in computer and information science. His main research interests are in interactive computer graphics, vision information processing, and user interfaces and surface based interactions. On this and related topics he has authored and coauthored more than 100 scientific journal and conference papers and patents. Dr. Kanev is a member of the IEEE, the Association of Computing Machinery (ACM), and the Asia-Pacific Society for Computers in Education (APSCE).

Michael Jenkin is a professor of computer science and engineering, and a member of the Centre for Vision Research at York University, Canada. Working in the fields of visually guided autonomous robots and virtual reality, he has published more than 150 research papers, including co-authoring Computational Principles of Mobile Robotics with Gregory Dudek and a series of co-edited books on human and machine vision with Laurence Harris. Michael Jenkin's current research interests include work on sensing strategies for AQUA, an amphibious autonomous robot being developed as a collaboration between Dalhousie University, McGill University, and York University; the development of tools and techniques to support crime scene investigation; and the understanding of the perception of self-motion and orientation in unusual environments, including microgravity.

ACKNOWLEDGMENTS

The financial support of the Social Sciences and Humanities Research Council of Canada (SSHRC) in the form of a Research Development Initiatives (RDI) grant to Karen Collins and Bill Kapralos, and the Natural Sciences and Engineering Research Council of Canada (NSERC) in the form of individual Discovery Grants to Bill Kapralos, Andrew Hogue, and Michael Jenkin and the Research Institute of Electronics, Shizuoka University, Japan, is gratefully acknowledged.

REFERENCES

Bauer, B. 1961. Phasor analysis of some stereo phonic phenomena. *Journal of the Acoustical Society of America* 33, 1536–1539.

Bauer, B. 1961. Stereophonic earphones and binaural loudspeakers.. *Journal of the Audio Engineering Society* 9, 148–151.

Begault, R. *3-D Sound for Virtual Reality and Multimedia*. 1994. Academic Press, Burlington, MA, USA.

Bennett, J. C., Barker, K., and Edeko, F. O. 1985. A new approach to the assessment of stereophonic sound system performance. *Journal of the Audio Engineering Society* 33, 314–321.

Chion, M. 1994. *Audio-Vision*. Columbia University Press, New York, NY, USA.

Cohen, M., and Wenzel, E. 1995. The design of multidimensional sound interfaces. In *Virtual Environments and Advanced Interface Design*, Barfield, W., and Furness, T. Eds. Oxford University Press Inc., New York, NY, USA, 1995, 291–346.

Collins, K., Kapralos, B., Hodge, A., and Hogue, A. 2011. 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.

Collins, K., Kapralos, B., and Kanev, K. 2009. Sound interface design for smart table computer interaction. In *Proceedings of the 12th International Conference on Humans and Computers*, Hamamatsu, Japan, December 2009.

Collins, K. 2008. *Game Sound: An Introduction to the History, Theory and Practice of Video Game Music and Sound Design*. MIT Press, Cambridge, MA, USA.

Durlach, N. I., and Mavor, A. S. 1995. *Virtual Reality: Scientific and Technological Challenges*. National Academies Press, Washington, DC, USA.

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

Gardner, W. 1998. *3-D Audio Using Loudspeakers*. Kluwer Academic Publishers, Norwell, MA, USA.

Han, J. Y. 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, 115–118.

Izadi, S., Hodges, S., Taylor, S., Rosenfeld, D., Villar, N., Butler, A., and Westhues, J. 2008. 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.

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, 501–537.

Lam, J., Kapralos, B., Collins, K., Hogue, A., and Kanev, K. 2010. Amplitude panning-based sound system for a horizontal surface computer: A user-based study. In *Proceedings of the IEEE International Symposium on Haptic Audio Visual Environments and Games 2010*, Phoenix, AZ, USA, October 2010, 174–178.

Langkjaer, B. 1997. Spatial perception and technologies of cinema sound, *Convergence*, 3, 92–107.

Lossius, T, P. Baltazar and T. de la Hogue. 2009. DBAP - Distance-based amplitude panning. In *Proceeding of the International Computer Music Conference (ICMC)*, Montreal, Quebec, Canada, August 2009.

Miller, M. 3D Audio. 1999. *Gamasutra: The Art and Business of Making Games*, November 1999.

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 2000, (preprint 7230).

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

Pulkki, V. 1997. Virtual sound source positioning using vector base amplitude panning. *Journal of the Audio Engineering Society* 45, 456–466.

Pulkki, V., and Karjalainen, M. 2001. Localization of amplitude-panned virtual sources I: Stereophonic panning, *Journal of the Audio Engineering Society*, 49, 739–751.

Pulkki, V., 2001. Localization of amplitude-panned virtual sources II: Two- and three-dimensional panning, *Journal of the Audio Engineering Society*, 49, 753–767.

Prince, B. 2006. Tricks and techniques for sound design. In *Proceedings of the Game Developer's Conference*, San Jose, CA, USA, March 2006.

Robjohns, H. 2001. You are surrounded: Surround sound explained - part 3, *Sound on Sound*, October 2001.

Robjohns, H. 2001. You are surrounded: Surround sound explained - part 2, *Sound on Sound*, September 2001.

Robjohns, H. 2001. You are surrounded: Surround sound explained - part 1, *Sound on Sound*, August 2001.

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, 1255–1259.

Schütze, S. 2003. The creation of an audio environment as part of a computer game world: the design for Jurassic park operation genesis on the XBOX as a broad concept for surround installation creation, *Organized Sound*, 8, 171–180.

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, 375–406.

Shilling, R. D., and Shinn-Cunningham, B. 2002. Virtual auditory displays. In *Handbook of Virtual Environment Technology*, K. Stanney, Ed. Lawrence Erlbaum Associates, Mahwah, NJ, USA, 65–92.

Shilling, R., Zyda, M., and Wardynski, E. C. 2002. Introducing emotion into military simulation and video game design America's army: Operations and VIRTE. In *Proceedings of the 3rd Annual European Conference GAME-ON'2002: Simulation and AI in Computer Games*, London, UK, November 2002.

Wilson, A. D., and Benko, H. 2010. 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.