

# Sound localization on a horizontal surface: virtual and real sound source localization

Jonathan Lam<sup>1</sup> · Bill Kapralos<sup>1</sup> · Kamen Kanev<sup>2</sup> · Karen Collins<sup>3</sup> · Andrew Hogue<sup>1</sup> · Michael Jenkin<sup>4</sup>

Received: 5 February 2015 / Accepted: 21 July 2015 / Published online: 30 July 2015  
© Springer-Verlag London 2015

**Abstract** As the technology improves and their cost decreases, tabletop computers and their inherent ability to promote collaboration amongst users are gaining in popularity. Their use in virtual reality-based applications including virtual training environments and gaming where multi-user interactions are common is poised to grow. However, before tabletop computers become widely accepted, there are many questions with respect to spatial sound production and reception for these devices that need to be addressed. Previous work (Lam et al. in *ACM Comput Entertain* 12(2):4:1–4:19, 2014) has seen the development of loudspeaker-based amplitude panning spatial sound techniques to spatialize a sound to a position on a plane just above a tabletop computer's (horizontal) surface. Although it has been established that the localization of these virtual sources

is prone to error, there is a lack of ground truth (reference) data with which to compare these earlier results. Here, we present the results of an experiment that measured sound localization of an actual sound source on a horizontal surface, thus providing such ground truth data. This ground truth data were then compared with the results of previous amplitude panning-based spatial sound techniques for tabletop computing displays. Preliminary results reveal that no substantial differences exist between previous amplitude panning results and the ground truth data reported here, indicating that amplitude panning is a viable spatial sound technique for tabletop computing and horizontal displays in general.

**Keywords** Tabletop computer · Surface computer · Spatial sound · Amplitude panning

---

✉ Bill Kapralos  
bill.kapralos@uoit.ca

Jonathan Lam  
jonathan.lam@uoit.ca

Kamen Kanev  
kanev@rie.shizuoka.ac.jp

Karen Collins  
collinsk@uwaterloo.ca

Andrew Hogue  
andrew.hogue@uoit.ca

Michael Jenkin  
jenkin@cse.yorku.ca

<sup>1</sup> University of Ontario Institute of Technology, Oshawa, ON, Canada

<sup>2</sup> Shizuoka University, Hamamatsu, Japan

<sup>3</sup> University of Waterloo, Waterloo, ON, Canada

<sup>4</sup> York University, Toronto, ON, Canada

## 1 Introduction

Tabletop computers (also known as surface computers, or smart tables) have become a popular interaction technology for group-based work and interactive simulations (e.g., games). The tabletop interaction surface provides a familiar metaphor for group-based work and entertainment allowing multiple users to position themselves around a horizontal computer display in a manner similar to sitting around a traditional table while interacting with the display itself. Tabletop computers naturally promote interaction amongst users, provide an engaging environment, and are an appealing option for applications beyond entertainment. They may provide an effective physical infrastructure for promoting collaborative interprofessional education of health professional or first responder teams (Dubrowski et al. 2015) and may have a multitude of other educational

uses as well (e.g., see Bortolaso et al. 2014; Wallace et al. 2009). For example, human anatomy training often occurs in a specialized laboratory with the instructor and the students standing (or seated) around the table where the cadaver is positioned. Such a scenario lends itself nicely to a tabletop computing platform whereby the cadaver table and the cadaver are replaced with a tabletop computer (display) and three-dimensional rendering of the cadaver, respectively. This allows the instructor and students to actively interact with the rendered model (e.g., remove anatomical layers, etc.) as a group as done in a traditional setting (see Anatomage Inc 2015). Such an approach eliminates the use of a cadaver (at least during the early stages of anatomy training) and the complications associated with cadavers (e.g., storage, acquisition and disposal, potential risk for pathogen transfer, and cost). This approach can be extended to virtual patients where various physiological functions can be simulated including respiratory and circulatory structures and the ability for the patient to interact (e.g., speak) with the users.

Although at first blush it may appear that a tabletop computing display is simply a traditional vertical display lying on a horizontal surface, there exist fundamental issues that must be addressed in order to fully operationalize a horizontal display. Of particular interest here is the ability of the user to integrate spatialized visual and audio cues into a coherent perception of a spatialized event on a horizontal surface. Other issues include questions regarding cooperation, orientation, viewing angle, and multi-touch sensing that will drive innovation in imagery, but such issues are being addressed elsewhere (see Han 2005; Kruger et al. 2004; Scott and Carpendale 2010).

Traditionally we experience our audiovisual media with screens that are oriented vertically with users/viewers sitting or standing in front of the screen looking directly at it. Audio mixing paradigms have been developed for this configuration. However, with tabletop computers the assumption that users stand in front of a vertical screen is no longer valid. The problem of conveying spatial sound for such computing platforms is further complicated by the fact that headphones, as commonly used in traditional computing platforms to deliver spatial sound, are not an attractive option for tabletop computing displays given that they restrict the users and interfere with the ability of users to easily communicate with each other and move around the table. Therefore, to maximize interaction amongst multiple users positioned around the tabletop computer, spatial sound should be delivered via a collection of loudspeakers positioned above, below, or around the tabletop computer. Prior work (Lam et al. 2014) investigated the generation of spatial sound using a loudspeaker array surrounding the tabletop computing display, with the perceived location of the sound created through a

constellation of loudspeakers coupled with an amplitude panning technique. These experimental results indicate that the localization of a virtual sound source on a horizontal surface (representing a computing display) is error prone. However, there is a paucity of studies on the localization of an actual (physical) sound source on a horizontal surface with which to compare these results and draw any meaningful conclusions. In other words, just how accurately are we able to localize a sound source on a horizontal surface? We refer to such reference data (i.e., sound source localization accuracy in the presence of a physical sound source on the horizontal surface) as “ground truth” data to distinguish these data from results obtained with fixed source virtual sound simulation approaches. Answering the question of accuracy with real loudspeaker locations across a horizontal surface is key to the development of effective virtual spatialized sound for such devices. Here, we provide details regarding an experiment that was conducted to collect such ground truth data. We also provide a comparison of this ground truth data with the prior work of Lam et al. (2014) that examined localization on a horizontal tabletop computing display of virtual sound sources generated using amplitude panning across four loudspeakers.

The remainder of this paper is organized as follows. In Sect. 2, background information is presented, beginning with a discussion of spatial sound followed by spatial sound generation and localization on tabletop computing display. Details regarding the experimental procedure conducted, including a description of the novel hardware setup constructed to examine the localization of a sound on a horizontal surface, are provided in Sect. 3. Experimental results are presented in Sect. 4, while a discussion and summary of the results in addition to plans for future work are provided in Sect. 5.

## 2 Background

Spatial sound technology refers to modeling the propagation of sound within an environment while accounting for the human listener. As Våljamäe (2005) describes it, the goal of spatial sound rendering is to “create an impression of a sound environment surrounding a listener in 3D space, thus simulating auditory reality.” Understanding sound spatialization in real versus virtual situations is particularly important now in the light of the recent rise of immersive virtual reality-based technologies such as the Oculus Rift and other three-dimensional virtual spaces such as CAVES that strive for a greater sensory experience. By understanding human strengths and weaknesses in sound localization, sound designers can make effective decisions with regards to virtual sound placement in the mix. Spatial

sound technology goes beyond traditional stereo and surround sound by allowing a virtual sound source to have such positional attributes as left–right, back–forth, and up–down (Cohen and Wenzel 1995). Spatial sound within interactive virtual and augmented reality environments allows users to perceive the position of a sound source at an arbitrary position in three-dimensional space, and when properly reproduced, it can deliver a very life-like sense of being remotely immersed in the presence of people, musical instruments, and environmental sounds (Algazi and Duda 2011). Spatial sound can add a new layer of realism (Antani et al. 2012) and contributes to a greater sense of presence (i.e., the sensation of “being there”) or immersion (Pulkki 2001b) (see Nordahl and Nilsson (2014) for a thorough discussion of presence and the influence of sound on presence). Spatial sound can also improve task performance (Zhou et al. 2007), convey information that would otherwise be difficult to convey using other modalities (e.g., vision) (Zhou et al. 2007), and improve navigation speed and accuracy (Makino et al. 1996). It has been suggested that auditory stimuli in general should be “regarded as a necessary rather than simply a valuable component of immersive virtual reality systems intended to make individuals respond-as-if real through illusions of place and plausibility” (Nordahl and Nilsson 2014).

With any auditory display, sound is output to the user using either headphones or loudspeakers. There are a number of advantages associated with headphone-based spatial sound delivery including the fact that headphones provide a high level of channel separation, thereby minimizing any crosstalk that arises when the signal intended for the left (or right) ear is also heard by the right (or left) ear. However, as previously described headphones are not an attractive option for tabletop computing displays and loudspeakers are typically employed instead. The traditional mechanism for generating spatialized sound using a loudspeaker array involves *amplitude panning* (Pulkki 2001a). Using the amplitude panning technique, two or more loudspeakers surround the listener and when the same sound signal is played through each of the loudspeakers with different amplitudes (via a gain factor applied to the signal of each loudspeaker), a new signal is formed that is perceived by the user to be emanating from a virtual sound source whose virtual position is dependent on the individual gain values. Various amplitude panning techniques exist which allow for a wide variety of loudspeaker setups including both two- and three-dimensional configurations (Pulkki 2001a). Regardless of the technique used, the general idea remains the same: For each real sound source compute appropriate gain factors to create the impression of a virtual sound source at a specific position relative to the listener. With the typical two-channel (stereo) configuration, the listener is placed symmetrically (in the

horizontal plane) equidistant between the left and right loudspeakers. By scaling the amplitude of the signal applied to the left and right loudspeakers by appropriate gain factors, 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 1997).

Amplitude panning can be extended to account for  $N > 2$  loudspeakers, as done in the popular *pair-wise amplitude panning* technique introduced by Chowning (1971). Pair-wise amplitude panning can produce sound sources in all azimuth directions given a sufficient number of loudspeakers. In this technique, despite the availability of  $N$  channels (loudspeakers), two loudspeakers are chosen and sound is applied to these two loudspeakers only in a manner similar to the conventional two-channel stereo panning technique. Three-dimensional panning is an extension of the two-channel, two-dimensional amplitude panning technique. However, rather than having all loudspeakers at the same height (e.g., on the same plane as the listener’s head), the height of some (or all) additional loudspeaker(s) differs. In this configuration, all loudspeakers are positioned equidistant from the listener. In a manner similar to pair-wise amplitude panning, sound is applied to a subset consisting of three loudspeakers only. A virtual sound source can be positioned anywhere on the triangle formed by the three loudspeakers. Currently, no general trigonometric method of three-dimensional amplitude panning for an arbitrary three-dimensional loudspeaker setup exists (Pulkki 2001b), and the calculation of the gains applied to the loudspeakers is configuration dependent. Another method of calculating the gain factors is the *vector base amplitude panning* (VBAP) technique, introduced by Pulkki et al. (1996). This technique can be used with an arbitrary number of loudspeakers and allows the loudspeakers to be placed in any position provided they are nearly equidistant around the listener and that the listening room is not very reverberant (Pulkki 1997). VBAP can be applied to both two- and three-dimensional loudspeaker configurations, including the traditional two-channel stereo setup and three-channel, three-dimensional setup. The VBAP method has been used to generate spatial sound for a multi-user tabletop computing interface (see Sasamoto et al. 2013). A complete discussion of spatial sound generation for virtual environments is beyond the scope of this paper but reviews covering various aspects of spatial sound are available. Blauert provides an overview of spatial hearing (Blauert 1996), Cohen and Wenzel (1995) discuss the design of multidimensional sound interfaces, while Kapralos et al. (2008) review virtual audio. Cohen (2010) provides an introduction to spatial sound “in the context of hypermedia, interactive multimedia, and virtual reality,” and a discussion of spatial

sound with an emphasis on loudspeaker-based amplitude panning including vector base amplitude panning is provided by Pulkki (2001a). Finally, Gardner (1998) discusses transaural audio (head-related transfer function-based loudspeaker displays) with two loudspeakers.

With respect to tabletop computing displays, the work of Sasamoto et al. (2013) saw the design of a prototype of tabletop interface that includes eight loudspeakers. Through the use of a tabletop interface, users are able to control the position of multiple sounds in a spatial sound environment. A novel feature of the system is that it allows multiple users to control the spatialization of independent sounds in real time. That being said, little work has focused specifically on spatial sound generation and localization. This may have to do with the fact that traditionally we experienced our audiovisual media with screens that have been aligned vertically with users/viewers sitting or standing in front of the screen looking directly at it. In such an approach, loudspeakers are typically mounted left, right, top, and bottom of the display, and with multiple loudspeakers, some form of amplitude panning is used for spatial sound generation. Although there are known issues with such an approach for vertical displays, how well does amplitude panning work for tabletop devices? Limited prior work has examined this issue, but with the growing use of tabletop computing displays and the need to deliver spatial sound, answering this question becomes important and may have implications for designers and developers of virtual reality-based applications that incorporate tabletop computing displays.

In order to address this issue, Lam et al. (2014) examined the localization of a virtual sound source spatialized to one of a set of 36 pre-defined locations on a horizontal surface (corresponding to the surface of a tabletop computer). The sounds were spatialized using the bilinear amplitude panning technique (i.e., the sounds were panned between loudspeaker pairs), under a diamond-shaped loudspeaker configuration, by means of loudspeakers that were placed at each of the four sides of the tabletop computing display facing inwards. In this experiment, participants were seated on a chair at the side of the horizontal surface, and for each trial, participants were presented with an auditory stimulus (white noise) that was spatialized to a position on the surface using the bilinear amplitude panning technique. The virtual sound source was synthesized on a grid where the horizontal and vertical separation was  $0.15\text{ m} \times 0.15\text{ m}$ , resulting in a total of 36 virtual sound source positions. Upon presentation of the auditory stimulus, participants indicated the location of the sound source verbally by choosing from the set of possible grid locations (their choice was recorded by the experimenters). 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

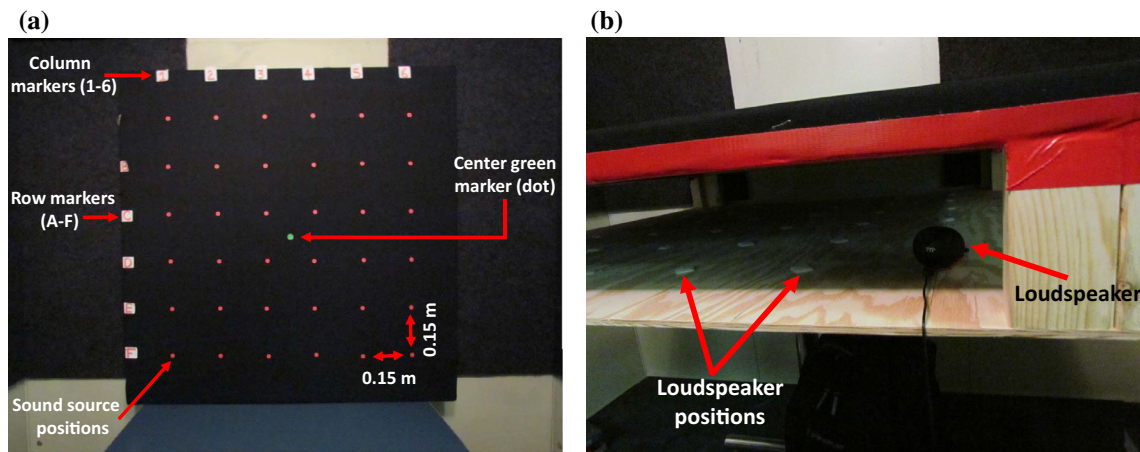
position (i.e., the position that the participants perceived the sound source to be emanating from) was used to estimate the accuracy of a participant's ability to localize the virtual sound source. Results indicated that the localization of a sound source spatialized to some position on a horizontal surface is quite error prone. More specifically, the average error across each of the 36 positions ranged from 0.11 to 0.47 m with an average of  $0.23 \pm 0.07\text{ m}$ . Given the grid spacing of  $0.15\text{ m} \times 0.15\text{ m}$ , participants were able to localize the sound source to within two positions of the actual virtual sound source position (Lam et al. 2014). Lam et al. (2014) repeated this protocol using the inverse-distance-based amplitude panning technique (i.e., the sound output at each loudspeaker was scaled by the distance between the corresponding loudspeaker and its distance to the virtual sound source), as an alternative to the bi-linear interpolation method. Results in this scenario were similar to the previously described experiment that employed the bilinear amplitude panning technique. More specifically, the average error across each of the 36 positions considered ranged from 0.13 to 0.44 m with an average of  $0.24 \pm 0.07\text{ m}$  or within two positions of the actual virtual sound source position (Lam et al. 2014). Both panning approaches resulted in substantive localization errors. However, are these errors a consequence of poor sound source localization in general, are they an artifact of the nature of tabletop computing displays, or are they a result of amplitude panning?

### 3 Experimental procedure

#### 3.1 Ground truth hardware

A tabletop “display” (loudspeaker surface) was constructed to obscure the true sound source position and to assist in recording the perceived position of the sound source (see Fig. 1). The surface and pre-defined sound source positions were modeled to imitate the configuration of previous work by Lam et al. (2014) that examined sound localization of virtual sound sources that were spatialized to one of 36 grid positions on a horizontal surface using amplitude panning techniques as described in the previous section. The hardware consists of a custom-built wooden box (loudspeaker surface) with openings on two of its sides (see Fig. 1). Inside the box, there are 36 pre-defined loudspeaker locations (the horizontal and vertical separation between each position is  $0.15\text{ m} \times 0.15\text{ m}$ ). Each location is labeled and allows for a loudspeaker to be easily attached to it (and later removed). The top of the loudspeaker surface (box) is covered with loudspeaker grill cloth, covering the inside of the box and thus hiding the loudspeaker from the participants, while allowing the





**Fig. 1** Hardware setup. **a** Top view of the loudspeaker surface with the sound source positions, rows, and columns labeled. **b** Side view of the box with the sound source positions and the sound source (color figure online)

sound to pass through. As shown in Fig. 1a on the top of the loudspeaker grill cloth and visible to the participants, the 36 sound source locations are clearly labeled (in red) as are the rows and columns (see the white labels on the side and top; the rows are labeled from A–F, while the columns are labeled from 1–6). With this hardware configuration, a single (small) loudspeaker could be moved outside the participant’s view to each of the 36 pre-defined loudspeaker locations, thus allowing for the collection of ground truth (reference) data for each of these locations.

### 3.2 Participants

Five male participants (with a mean age of 29 years) completed the experiment. The participants were students (undergraduate and graduate), or researchers at the University of Ontario Institute of Technology (UOIT). None of the participants reported visual or hearing defects. The experiment abided by the UOIT Research Ethics Review process.

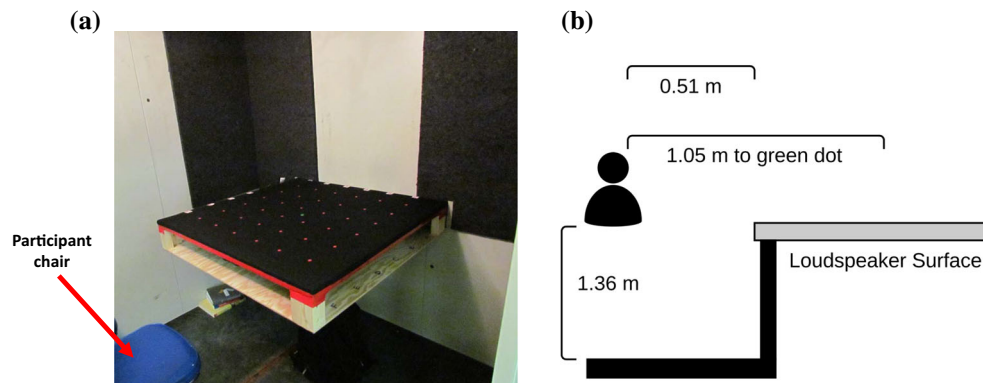
### 3.3 Auditory stimulus

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 finite infinite response (FIR) filter with low- and high-frequency cutoffs of 200 Hz and 10 kHz, respectively. The sound was output with an iHome iHM60 portable multimedia loudspeaker which was manually positioned at one of the 36 sound source positions [arranged in a  $6 \times 6$  grid], by the experimenter. The duration of the sound stimulus for each trial was two seconds (identical to the duration of the sound stimulus presented in the study by Lam et al. (2014)). The experiments took place in an Eckel audiometric room located at UOIT (room dimensions of

$2.3 \text{ m} \times 2.3 \text{ m} \times 2.0 \text{ m}$ ). According to the manufacturer, the Eckel audiometric room provides (frequency dependent) noise reduction across a wide range of frequencies (e.g., at 19 dB at 125 Hz and 60 dB at 4 kHz). The average background noise level within the audiometric room, measured using a Radio Shack sound level meter (model 33-2055) with an A-weighting, placed at the location where the participant’s head would be in the absence of any sound stimuli, was below 50 dB (the lowest level measurable with the sound level meter). The average sound level, also measured at the location where the participant’s head would be in the presence of the sound stimuli, was 68 dB.

### 3.4 Experimental method

Participants were seated on a chair and instructed to look forward at the green marker (dot) located at the center of the simulated display surface (see Fig. 2a). In order to limit deviations from their intended position, participants were asked to line up the tip of their nose with a thin piece of string (with a weight on its bottom), hanging from the ceiling of the audiometric room. The weight hung at a height of 1.36 m from the floor and 0.51 m from the edge of the simulated display surface and 1.05 m from the green marker at the center of the simulated display surface (see Fig. 2b). The placement of the chair was adjusted to allow the participant’s nose to be aligned with this weight. Participants were also instructed to limit movement and to maintain their alignment with the hanging weight, but participant movements were not tracked. Given this configuration along with the loudspeaker positions, the azimuthal angular deviation for each of the sound source positions was calculated and is summarized in Table 1. Angular deviation increases symmetrically, moving away from the midline (e.g.,  $0^\circ$ , directly in front of the participant), and decreases further away from the participant.



**Fig. 2** Experimental setup within the audiometric room where the experiments took place. **a** Configuration setup. **b** Configuration dimensions (color figure online)

**Table 1** Azimuth angle in degrees for each of the 36 sound source position

	1	2	3	4	5	6
A	14.7	8.9	3.0	3.0	8.9	14.7
B	16.3	10.0	3.4	3.4	10.0	16.3
C	18.4	11.3	3.6	3.6	11.3	18.4
D	20.9	12.9	4.4	4.4	12.9	20.9
E	24.3	15.2	5.2	5.2	15.2	24.3
F	28.9	18.3	6.3	6.3	18.3	28.9

For each trial, the loudspeaker was physically moved to one of the 36 sound source positions by the experimenters. To limit any potential cues, each participant was blindfolded and care was taken to limit any noise while the loudspeaker was moved to the next position. After the sound source was positioned, the blindfold was removed, the sound stimuli was presented, and the participant's task was to indicate which of the 36 positions they believed the sound was emanating from by verbally stating the corresponding row and column to the experimenter who recorded the information. A total of 36 grid positions (spatial sound sources) were considered, and each position was repeated twice leading to a total of 72 trials (36 grid positions  $\times$  2 repetitions). Each of the 72 trials were presented in random order. Prior to the start of the experiment, participants were presented with the auditory stimulus at each of the four corner positions of the surface (individually, one after the other) to provide them with a reference and to familiarize them with the experiment.

## 4 Results

For each participant, the error for each of the 36 sound source positions was averaged from the two responses considered for each position. The average error (Euclidean distance

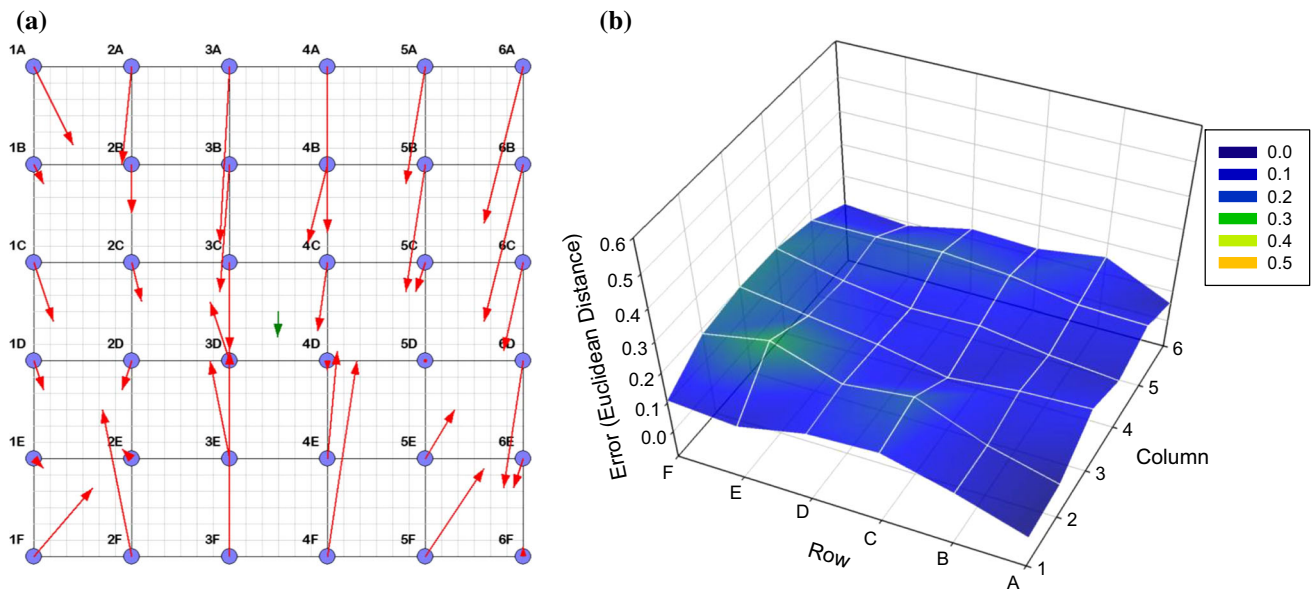
between the actual sound source position and the position indicated by the participant, measured in meters) for each of the 36 sound source positions (averaged across each of the five participants) is summarized in the plots of Fig. 3 and Table 2. In Fig. 3a, a vector plot of the average error for each of the 36 positions is shown. The red arrows denote 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. In Fig. 3b, the average error magnitude for each of the 36 sound source positions is presented in the form of a surface (3D) plot. The average error across each of the 36 positions ranges from 0.02 to 0.32 m with an average of  $0.18 \pm 0.07$  m. Given the grid spacing of  $0.15 \text{ m} \times 0.15 \text{ m}$ , participants were able to localize the sound source to within approximately two positions of the actual sound source (e.g., to within 0.32 m). The largest errors are along row F (closest to the participants) and along row A (furthest from the participants).

### 4.1 Summary of the results

A comparison between the resulting error plots associated with the localization of a sound source illustrates systematic differences in sound localization error for the ground truth data. The small sample sizes used in this preliminary study make it difficult to make sweeping recommendations or for any results to have much statistical power. That being said, it is possible to identify trends in the results obtained to date. Specifically, there appears to be a bias in responses where the sound is perceived closer to the center of the display.

## 5 Discussion and future work

Previous work has shown that the use of amplitude panning techniques to spatialize a sound source on to a horizontal surface is prone to large errors (Lam et al. 2014). However,



**Fig. 3** Sound source localization results. **a** Average error per sound source. **b** Height plot of the error magnitude. The participant was positioned midway between *columns* 3 and 4 with *row* F closest to them (color figure online)

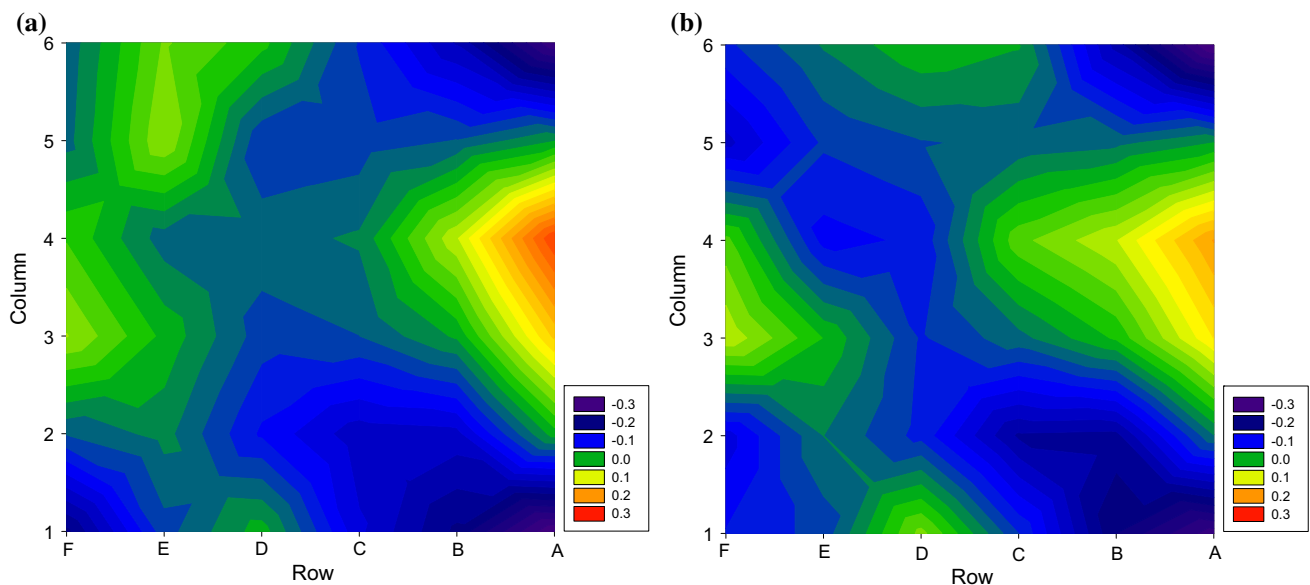
**Table 2** Results (average error in meters) averaged across all of the participants

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

prior to the study described here, there was a lack of any reference (ground truth) data with which to compare these results. In other words, how can one determine whether a large error was specific to the spatialization techniques employed or whether it was simply an inherently difficult task? Here, we have described an experiment that allowed us to collect such reference (ground truth) data that describe sound localization accuracy on a horizontal surface with actual sound sources. With the availability of these reference data, meaningful comparisons can now be made regarding the accuracy of virtual sound source localization on a horizontal surface with virtual sounds generated using a variety of methods including vector base amplitude panning. Although results must be regarded as preliminary given that the number of participants considered here (five) was small, we have shown that sound source localization on a horizontal surface with a physical (real) sound source is error prone. The observed error (the Euclidean distance between the actual sound source position and the perceived sound source position) ranged from 0.02 to 0.32 m. Although the direction of the positional

error was biased toward the centre of the display, the magnitude of the positional error was relatively constant over the surface.

Previous work has shown that virtual sound source localization on a horizontal surface is error prone. To better quantify the errors presented here, the ground truth data collected here were compared with earlier data on perceived sound source localization of virtual sound sources. Figure 4a, b shows contour plots of the signed difference between each of the 36 positions of the ground truth data obtained with the horizontal surface configuration and with the positional errors of each of the corresponding 36 positions previously found with the bilinear and inverse distance amplitude panning techniques, respectively (Lam et al. 2014). For each of the 36 positions, a negative difference corresponds to a larger error for the amplitude panning method for the corresponding position, while a positive difference corresponds to a larger error for the horizontal surface configuration ground truth error for that particular position. The average differences for the bilinear and inverse distance amplitude panning methods are



**Fig. 4** Contour plots illustrating the difference in average error (in meters) between the results for each of the 36 positions considered here with the corresponding virtual sound source localization

$-0.05 \pm 0.02$  and  $-0.06 \pm 0.02$  m, respectively. In both cases, the average difference is negative, indicating that on average, we are less accurate at localizing a virtual sound source on a horizontal surface. However, the magnitude of this difference is reasonably small, well below the grid spacing of  $0.15 \times 0.15$ , and similar in size to the (approximately) 0.04 m diameter of the loudspeaker used in this ground truth study.

Although results must be regarded as preliminary, we have shown that sound localization errors are not specific to amplitude panning. More specifically, the localization of a real sound source on a horizontal surface is prone to similar errors and similar error patterns. In other words, sound localization is a difficult and error-prone task for tabletop displays. Localization errors as large as 0.32 m can have serious consequences in application systems. For example, consider an anatomical human model visualized on a life-sized tabletop display. A positional error of 0.32 m in the localization of the sound associated with the beating heart (when considering a live virtual patient) may lead to the perception of the beating heart sound as emanating from the face or from the patient's midsection. That being said, amplitude panning, a simple technique that requires limited computational resources, is a viable spatial sound technique for tabletop computers and horizontal displays when accurate sound source localization is not required.

Developers and designers of applications for tabletop computers must recognize that spatialized sound accuracy is relatively poor on tabletop computing displays. To

obtained two amplitude panning techniques. **a** Bilinear amplitude panning and **b** inverse distance amplitude panning (color figure online)

overcome this limited localization accuracy, one option might be to exaggerate the simulated sound source placement when sound source positions correspond to positions associated with larger errors, or to use sounds that are more easily localized. Furthermore, sound source localization varies with frequency (Perrott and Saberi 1990), changes in frequency (Ohta and Obata 2007), and as a result of the filtering effects of the outer ear (pinna), sounds emanating from higher elevations (with respect to the listener) have more energy at high frequencies (Parise et al. 2014). Therefore, the frequency range of the emitted sound may also be a factor to consider [Parise et al. (2014) quantified the degradation on frequency-dependent sound source location judgements with listeners posed in a variety of configurations aside from being seated or standing]. These are all potential areas that warrant further investigation.

Here, we have considered a single user when in fact tabletop computers are generally intended to include multiple users. What, if any, effect does table size have on sound localization capabilities, and more specifically, is there an optimal table size for one, two, three, or four users? The results presented here are a step toward a model that can provide an estimate of sound localization accuracy based on spatial locations relative to the users' location. Future work will expand this study to incorporate multiple user orientations with respect to the surface to generalize sound localization on a horizontal surface with respect to a user's position and orientation. This will help us particularly with multi-user scenarios and allow us to determine the best parameters for sound generation and delivery to minimize



localization errors across all users. That being said, conducting similar sound localization experiments with more than one participant seated around a tabletop computing display will present some difficulties and will require addressing how multiple participants will indicate their choice of virtual sound source position without influencing each other. Generalizing this further and possibly providing greater insight into sound localization on a surface, future work will also examine sound localization on a surface that is diagonally slanted (e.g., oriented at an angle between  $0^\circ$  and  $90^\circ$ ). This can be facilitated using a drafting (drawing) table that allows the orientation of the surface to be easily adjusted.

Tabletop computers are intended to be used with both visual and auditory stimuli. Therefore, future work will also examine the interaction of audio and visual cues and, in particular, our ability to localize a sound source in the presence of visual stimuli (and potentially conflicting visual stimuli). Despite the inherent error observed here, in many applications (such as games), highly accurate sound localization may not actually be required. Rather, determining the direction of a sound source and whether the distance to the sound source is increasing or decreasing may be of greater importance. Moreover, the conjunction of visual image with sound has the tendency to pull the sound to the image, so making localization much easier (the ventriloquist effect, e.g., Alais and Burr 2004). Furthermore, auditory distance estimates are more accurate when visual cues are also present; thus, the addition of visual cues can potentially lead to more accurate sound localization and help alleviate the large errors observed here (Sodnik et al. 2006).

**Acknowledgments** Funding to support this work has been provided by the Research Institute of Electronics, Shizuoka University, in the form of a Cooperative Research Projects Grant, the Social Science and Humanities Research Council of Canada, and the Natural Sciences and Engineering Research Council of Canada.

## References

- Anatomege Inc. (2015) Anatomege table virtual dissection. [www.anatomege.com](http://www.anatomege.com). Accessed on 9 July 2015
- Alais D, Burr D (2004) The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol* 14(6):257–262
- Algazi VR, Duda RO (2011) Headphone-based spatial sound. *IEEE Signal Proc Mag* 28(1):33–42
- Antani L, Chandak A, Savioja L, Manocha D (2012) Interactive sound propagation using compact acoustic transfer operators. *ACM Trans Graph* 31(1):1–12
- Blauert J (1996) *The psychophysics of human sound localization* (revised ed). MIT Press, Cambridge
- Bortolaso C, Graham TCN, Scott SD, Oskamp M, Brown D, Porter L (2014) From personal computers to collaborative digital tabletops to support simulation based-training. In: *Proceedings of 19th international command and control research and technology symposium*, Alexandria, VA, USA, pp 37–45
- Chowning J (1971) The simulation of moving sound sources. *J Audio Eng Soc* 19(1):2–6
- Cohen M (2010) Under-explored dimensions in spatial sound. In: *Proceedings of 9th ACM SIGGRAPH conference on virtual-reality continuum and its applications in industry*, Seoul, Korea, pp 95–102
- Cohen M, Wenzel E (1995) The design of multidimensional sound interfaces. In: Barfield W, Furness T (eds) *Virtual environments and advanced interface design*. Oxford University Press Inc, New York, pp 291–346
- Dubrowski A, Kapralos B, Jenkin M, Kanev K (2015) Interprofessional critical care training: interactive virtual learning environments and simulations. In: *Proceedings of 6th international conference on information, intelligence, systems and applications*, Corfu, Greece
- Gardner W (1998) *3-D audio using loudspeakers*. Kluwer Academic Publishers, Norwell
- Han JY (2005) Low-cost multi-touch sensing through frustrated total internal reflection. In: *Proceedings of 18th ACM symposium on user interface software and technology*, Seattle, WA, USA, pp 235–238
- Kapralos B, Jenkin M, Milios E (2008) Virtual audio systems. *Presence Teleoper Virtual* 17(6):524–549
- Kruger R, Carpendale S, Scott SD, Greenberg S (2004) Roles of orientation in tabletop collaboration: comprehension, coordination and communication. *Comput Support Coop Work* 13(5–6):501–537
- Lam J, Kapralos B, Collins K, Hogue A, Kanev K, Jenkin M (2014) Sound localization on table-top computers: a comparison of two amplitude panning methods. *ACM Comput Entertain* 12(2):4:1–4:19
- Makino H, Ishii I, Nakashizuka M (1996) Development of navigation system for the blind using GPS and mobile phone communication. In: *Proceedings of 18th annual meeting of the IEEE engineering in medicine and biology society*, Amsterdam, The Netherlands, pp 506–507
- Nordahl R, Nilsson NC (2014) The sound of being there: presence and interactive audio in immersive virtual reality. In: Collins K, Kapralos B, Tessler H (eds) *The Oxford handbook of interactive audio*. Oxford University Press, New York, pp 213–233
- Ohta Y, 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 123rd convention of the audio engineering society*, New York, NY, USA, Preprint: 7230
- Parise CV, Knorre K, Ernst MO (2014) Natural auditory scene statistics shapes human spatial hearing. *Proc Nat Acad Sci* 111(16):6104–6108
- Perrott DR, Saberi K (1990) Minimum audible angle thresholds for sources varying in both elevation and azimuth. *J Acoust Soc Am* 87(4):1728–1731
- Pulkki V (1997) Virtual sound source positioning using vector base amplitude panning. *J Audio Eng Soc* 45(6):456–466
- Pulkki V (2001a) Localization of amplitude-panned virtual sources I: stereophonic panning. *J Audio Eng Soc* 49(9):739–751
- Pulkki V (2001b) Spatial sound generation and perception by amplitude panning techniques (Unpublished doctoral dissertation). Department of Electrical and Communications Engineering, Helsinki University of Technology, Helsinki, Finland
- Pulkki V, Huopaniemi J, Huutilainen T, Karjalainen M (1996) DSP approach to multichannel audio In: *Proceedings of international computer music conference*, Clear Water Bay, Hong-Kong, pp 93–96
- Sasamoto Y, Cohen M, Villegas J (2013) Controlling spatial sound with table-top interface. In: *Proceedings of international joint*

- conference on awareness science and technology and ubi-media computing, Aizuwakamatsu, Japan, pp 713–718
- Scott SD, Carpendale S (2010) Theory of tabletop territoriality. In: Muller-Tomfelde C (ed) *Tabletops—horizontal interactive displays*. Springer, London, pp 375–406
- Sodnik J, Tomazic S, Grasset R, Duenser A, Billinghurst M (2006) Spatial sound localization in an augmented reality environment. In: *Proceedings of 2000 international conference on auditory display*, Sydney, Australia, pp 1–8
- Väljamäe A (2005) Self-motion and presence in the perceptual optimization of a multisensory virtual reality environment (Tech. Rep. No. R037/2005). Department of Signals and Systems, Division of Communication Systems, Chalmers University of Technology, Sweden
- Wallace JR, Scott SD, Stutz T, Enns T, Inkpen KM (2009) Investigating teamwork and taskwork in single and multi-display groupware systems. *Personal and ubiquitous computing: special issue on interaction with coupled and public displays* 13(8):569–581
- Zhou ZY, Cheok A, Qiu Y, Yang X (2007) The role of 3-D sound in human reaction and performance in augmented reality environments. *IEEE Trans Syst Man Cybern* 37(2):262–272