Sound Interface Design for Smart Table Computer Interaction

Karen Collins[1], **Bill Kapralos**[2,3], **and Kamen Kanev**[4]

[1]Canadian Centre of Arts and Technology, University of Waterloo. Waterloo, Canada.
[2]Faculty of Business and Information Technology, [3]Health Education Technology Research Unit, University of Ontario Institute of Technology. Oshawa, Ontario, Canada.
[4]Research Institute of Electronics, Shizuoka University. Hamamatsu, Japan.

e·mail: Karen Collins: `collinsk@uwaterloo.ca`, Bill Kapralos: `bill.kapralos@uoit.ca`
Kamen Kanev: `kanev@rie.shizuoka.ac.jp`

Abstract

Recently, "smart" table-top touchscreen computers, in which users position themselves around a horizontal computer screen, have been introduced. Although the use of touchscreen computers is still not widespread, given the growing popularity of multi-touch mobility devices (e.g., iPods, smartphones), the move to multi-user touchscreens and a horizontal surface is a likely trajectory of the technology. However, before table-top touchscreen computing becomes widely accepted, there are many questions, particularly with respect to sound production and reception, and multi-modal interaction for these devices that need to be explored. In this paper we provide an overview of a table-top touchscreen computer setup and describe a simple amplitude panning method for the output of sound amongst four loudspeakers. The paper begins with a background on sound interaction design.

Keywords: Smart table computer, table-top computer, audio-video interaction, sound interaction design, human-computer interaction.

1 Introduction

For many decades now, we have experienced our audio-visual media on a vertical screen; our televisions, movie theaters, and computer screens have all presented information vertically in front of us. As such, sound (music, dialogue, and sound effects) for television, film, software, and games has been designed accordingly, with the placement of the speakers and the sound mixing all developed based on this format. Recently, smart table-top touchscreen computers (also known as surface computers, smart table computers, or smart tables), where users position themselves around a horizontal computer screen in a manner similar to sitting around a "traditional" table, have been introduced. Although smart tables have yet to be primarily designed as consumer models, with the growing popularity of multi-touch mobility devices (e.g., iPods, smartphones), the move to multi-user touch screens and a horizontal surface is a likely trajectory of the technology. Moreover, these devices may well become a part of social entertainment, where families and friends can interact with each other around a table-like surface. Before smart tables become widely accepted, there are many questions that need to be explored, particularly with respect to sound production and reception, and multi modal interaction, (e.g., the interaction of sound and video cues).

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

These questions are partly technical and partly perceptual, crossing the disciplines of psychology, interaction design, and areas of physics and computer science. While physics helps in positioning of loudspeakers in a room (incorrect placement can result in the reduction or altogether loss of various frequency components), the use of those loudspeakers (where we position the sounds and why, as well as the types of sound we use), relies on human auditory perception and usability. The development of audio positioning for a vertical screen has largely come from trial and error and conventionalization over decades of use. While there are ample studies of room acoustics and loudspeaker positioning, there are few formal methods of research that focus specifically on how the placement of sound in the mix impacts the listener, and there are no known studies of the perception of audio as we move to horizontal-based viewing screens.

We have recently begun to investigate the problems/issues described above. The goal of this work is to develop novel and innovative sound interaction techniques for smart table computers. In order to accomplish this, we will be drawing on the very new field of sonic interaction design, which studies the use of sound to convey information, meaning, and aesthetic and emotional qualities in an interactive context [RSB$^+$08]. Sonic interaction design encompasses a range of technical and theoretical topics and concepts, including sound design, interaction design, the interactive arts, auditory display, psychology, usability, user experience, design, and perception. With few formal methodologies in the field, we will draw from a range of methods to explore our research questions about sound on smart table computers.

The remainder of the paper is organized as follows. In Section 2, an overview of the importance of sound in interactive displays/applications and an introduction to sound interface/interaction design is provided. An overview of the proposed table-top touchscreen computer setup (with an emphasis on the sound interface) is provided in Section 3. Finally, concluding remarks are provided in Section 4.

2 Background

The sounds we hear provide us with detailed information about our surroundings and can assist us in determining both the distance to, and direction of objects [War83]. This ability is extremely beneficial for humans as well as other species and in many situations, can be crucial for survival. It has been shown that sounds can be superior to visual stimuli for gaining attention [PNK76], and certain sounds (e.g., a baby crying) immediately activate mental images and schemata providing an effective means of attention focus [BE71]. In fact, sounds not only help to focus our attention, but once the attention system is focused, sounds can help maintain our attention on appropriate information while avoiding distractions, thus engaging our interest over time [BC01, TJ95]. Sounds can also elaborate on visual information by providing us with information on invisible structure, dynamic changes, and abstract concepts that may not be expressed visually [BC01]. As in the real world, sound plays a vital role in the communication of information in computer applications and entertainment, such as office application software, and video games. Surround sound in particular helps to achieve the immersive, emotional impact of new media. One recent study found that 40% of users rated surround sound systems as "very important" to video game play [Goo06].

In human-computer interfaces, sound can convey alarms, warnings, messages, and status information (such as an incoming email, an error, or a critical battery alarm) [Bux90]. In interactive applications such as virtual environments and simulations, auditory cues can help a user to orient themselves, add a better sense of presence or immersion, compensate for poor visual cues (graphics), and at the very least, add a pleasing quality to the simulation [AC97, SSC02]. Sound in entertainment applications such as video games has many important functions. More specifically, it can help to communicate important information to the player ("quick!–take out your weapon, there's an enemy around the corner!"); it can serve as a sound symbol or leitmotif; it can help to situate the player in a specific location; it reduces learning curves and creates a sense of realism. Foley sound effects associated with particular visual imagery (such as footsteps, a door opening, glass breaking, a ball bouncing, etc.) can be added in the post-production of animations [DKP01, OCE01], film and video games, leading to a greater sense of presence, realism, and quality [DKP01]. Sound also plays an emotional role that helps increase immersion, to emotionally invest them in a narrative. It is often said in the entertainment industry that "sound is emotion" and a visual display without a properly designed audio component will be "emotionally flat". Studies regarding the role of sound in media in the last few decades have shown that the presence of sound increases the emotional and physiological response [SZ02].

The study of sound's role and how it interacts with the listener has been divided amongst many different disciplines. Information retrieval computer scientists have engaged in trying to quantify perceptual aspects of sound in order to be able to classify, search, and retrieve sounds from databases and to automatically find dynamic scenes in films [MDV01]. Music therapists have searched for ways to alter mood using sound and music (see [MKT09, PS74]), psychologists have studied the roles that sound plays in influencing perception and cognition (see [BEJ06]), and psychoacousticians have sought to understand the physical aspects of sound perception (see [AR97]). Marketing experts have studied the role sound plays in altering our perception of products (see [Bru90, YM04]) and semioticians have sought to understand sound as a symbolic language (see [TC03]). In just the past few years, an emerging field of study has been combining these disciplines with interaction design, to study sound perception in interactive media.

Interaction design studies how people use objects and systems, usually in the context of human-computer interaction. Interaction designers attempt to understand the usability of an object/application (does it function the way it is supposed to? How is it being used?) as well as user experience (whether it is enjoyable, satisfying and motivating) [WM74]. This relatively new field has evolved from computer science (and most notably, human-computer interaction) but has recently recognized the need for reception-based research with a humanities-based influence. Peter Wright and John McCarthy, for instance, note that "our account of what it might mean to design for experience leans heavily on ways of thinking about design that emerge from arts and humanities rather than from sciences and engineering... If future interaction design is to take user experience seriously, it must look farther afield to the fields of humanities and the arts to find a pragmatically useful conception of experience" [WM74]. Indeed, the study of human perception, communication and interaction has a long history in the social sciences and humanities, and new methods emerging in interaction design are drawn largely from social science and humanities-based disciplines.

Sound interaction design encompasses a range of technical and theoretical topics and concepts, including interaction design, affect theory, sound design, acoustics, and semiotics. It is our goal to combine this very recent area of inquiry with more traditional methods drawn from musicology and music psychology to the use of sound on smart table computers. Questions that we will explore include: since we are used to interacting with media that is vertical and designed to be consumed from a single angle (frontally), does this change in angle alter the way that we perceive sounds, and, if so, what are the ways that we might want to alter the sound, or alter the sound mix to best use sound in a smart table computer? Would applications built for smart tables benefit from more realistic interface sound effects? Or, instead, should they remain in the stylized fashion currently used in vertical-screen computers?

3 Proposed Method

Figure 1 provides an overview of the smart table hardware setup. The system is intended to accommodate multiple users (1-4) and consists of i) a PQLabs smart table, ii) a video camera, iii) four loudspeakers, and iv) four microphones. The (external) video camera is used specifically for object recognition (see Section 3.2).

3.1 Audio System

As shown in Figure 1 and as previously described, the system consists of four loudspeakers (each at one of the four smart table corners) facing the table computer surface (the additional option of a centre channel and LFE will be explored at a later time). This setup is similar to a traditional

Figure 1: Conceptual overview of the proposed system.

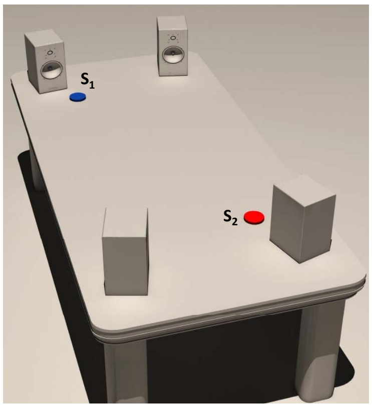

Figure 2: Virtual sound example.

quadraphonic surround sound system. However, traditional quadraphonic stereo techniques are intended for one listener and therefore, not applicable in this work. We have begun experimenting with a number of sound generation techniques and have implemented a simple amplitude panning method whereby the sound is output at each of the four loudspeakers but the level of the sound output at each loudspeaker is scaled by the distance between the corresponding loudspeaker and its distance to the virtual sound source given mathematically below

$$s_n = \frac{s_v}{r_{n_v}}, \tag{1}$$

where s_n is the signal (sound) applied to the n^{th} loudspeaker, s_v is the sound output from the virtual sound source, and r_{n_v} is the distance between the n^{th} loudspeaker and the virtual sound source. Figure 2 illustrates the four speakers surrounding the simulated environments playing area. Included in the diagram are two markers denoting the position of two virtual sound sources, s_1 and s_2. In this particular example, the level of the sound corresponding to virtual source s_1 will be strongest from the upper left loudspeaker; the loudspeaker closest to it. Similarly, the level of the sound corresponding to virtual source s_2 should be greatest at the loudspeaker closest to it (the lower right loudspeaker). This method is independent of each of the user's (listener's) physical position; the sounds are being simulated as coming from their position on the table, and if the user were to move farther away from the desk, the volume would naturally get quieter in relation to their distance away from the table.

3.2 Vision System: Object Recognition

In the set-up described so far, sound sources are associated with various virtual characters and events, presented on the smart table. Virtual sound sources are then software-controlled to match the activities of the characters in the virtual world as projected on the flat tabletop. Users interact with the virtual world and characters through the multi-touch smart table surface. In this model, a flat tabletop surface is employed to interact with an essentially three-dimensional virtual world. For proper accounting of the third dimension, additional interface components are obviously needed. This is a two-fold requirement. Firstly, tracking of user movements around the smart table is essential for the analysis of user-reported sound experience. Real-time positions and orientations of user heads can also provide feedback when adjusting the sound field parameters during the experiment. Secondly, physical objects that have impact on the sound environment should be registered with their three-dimensional shapes, positions, and orientations. Besides the loudspeakers, such objects, for example, may be small partitions, dividing the space and providing more privacy for the users. Interior objects that may not be present in the experimental environment but need to be simulated for fidelity reasons should also be accounted for. In summary, the key role of the additional interface components is to establish certain mapping of physical and virtual objects which will be employed in the sound simulations and experimentation.

Interface components with desired characteristics and functionality often emerge in Tangible User Interface (TUI) and Augmented Reality (AR) research where appropriate tools and development environments are commonly available. The ARToolKit, for example, is a computer vision-based augmented reality library that can be used to recognize specific real-world objects [KB99]. Such objects are then substituted with appropriate virtual counterparts and a live, augmented reality video stream is generated in result. In this work, we are primarily interested in the real-world object recognition capabilities of the ARToolKit which is based on special markers. To recognize positions and orientations, and to track real-world objects, one or more ARToolKit compatible markers need to be rigidly attached to each physical object. This requirement poses certain problems when non-rigid objects are considered. For human bodies, in particular, a large number of markers would be needed to track the body motions, along with several cameras to compensate for the natural occlusion of the markers. However, here we are only interested in tracking positions and orientations of user heads, which can be achieved with a single marker per person. In practice, markers are attached to special hats that users wear during experiments.

Markers are also attached to the smart table, the loudspeakers, and to other environmental objects. The smart table markers are used for determining the position and orientation of the global scene view camera in respect to the tabletop surface. This essentially establishes a global coordinate system where all other objects are registered. Positions and orientations of the four loudspeakers in respect to the global view camera are also determined, based on the markers attached to them. Loudspeaker coordinates are then converted to the global coordinate system and used for sound field calculations during the experiment. In this way, moving the loudspeakers in respect to the smart table can be immediately reflected into the corresponding sound calculations and compensated for, as much as possible. This allows us to check different loudspeaker arrangements around the smart table in real time and thus to compare user experiences in a reliable manner.

Objects from the physical world and their corresponding scaled models can be used as tangible interface components during the experiment. The virtual world and characters beneath the tabletop surface can, for example, be linked to physical objects placed onto the smart table. Technically, any object touching the smart table surface leaves its footprint so its coordinates can be determined. The smart table functionality, however, does not allow identification of such objects, which creates certain difficulties when multiple physical interface components are placed on the tabletop. We address this problem by attaching ARToolKit compatible markers to all physical objects on the smart table and deriving their positions and orientations through the global view camera. The smart table touch and position sensing functionality is still employed for invoking actions and for increased precision calculations and error compensation. There are other technologies and SDK environments for TUI and AR implementations that suit the needs of our experimental environment. The Cluster Pattern Interface (CLUSPI) [KK05, KK09b] co-invented by one of the authors, for example, supports direct point-and-click functionality [KK06a], based on digital carpet encoding of physical surfaces. Employed CLUSPI codes are practically invisible for naked human eyes and thus allow preserving the natural appearance of the physical objects. This feature is quite important in applications in education [KK09a], in multimedia presentations [Kan08], and others [KCM07, KK06b]. At this stage we conduct our initial experiments by employing well visible and sometimes obtrusive ARToolKit compatible markers but we expect to replace them with the invisible surface encoding supported by CLUSPI or other methods.

4 Summary

Smart tables represent a further step towards what is known as ubiquitous or pervasive computing - that is, in the very near future we will not rely on a desktop model in which a single user employs a single desktop computer, but, rather, computers will be integrated into most aspects of our lives. In addition, smart tables may also become the family dining table. Given the potential significance and practicality, understanding the method whereby we might best use these devices is critical. Sound in particular is often overlooked in terms of research, and yet is an important component in the usability and enjoyment of new computer devices. We will employ existing research methods and techniques for testing auditory perception of sounds generated with vertical and mobile devices, while developing and exploring new methods to understand how the change from a vertical to horizontal screen impacts the listener's perception of auditory events.

Here we proposed the setup of a table-top touchscreen computer and described a simple amplitude panning method to convey sound amongst the four loudspeakers. Although simple, the method is independent of user position and doesn't require extensive computational resources. Future work will explore further, more complex sound output/interaction techniques and new methods for testing sound perception by users, combining elements from a broad range of disciplines. In examining the specific problems, this work will lead to a significant broadening of our understanding of sound interaction design, of multi-user interaction, of how users perceive sound on horizontal surface devices, of how to implement sound for games on multi-touch computers, and about methodologies we can use to test user preferences of sound in computer applications. Since sonic interaction design is a very new and emerging field, this research will make a significant contribution towards methods and practice.

Acknowledgments

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

References

[AC97] D. B. Anderson and M. A. Casey. The sound dimension. *IEEE Spectrum*, pages 46–50, March 1997.

[AR97] C. Arias and O. A. Ramos. Psychoacoustic tests for the study of human echolocation ability. *Applied Acoustics*, 51(4):399–419, 1997.

[BC01] M. J. Bishop and Ward M. Cates. Theoretical foundations for sound's use in multimedia instruction to enhance learning. *Educational Technology Research and Development*, 49(3):5–22, 2001.

[BE71] I. H. Bernstein and B. A. Edelstein. Effects of some variations in auditory input upon visual choice reaction time. *Journal of Experimental Psychology: Human Perception and Performance*, 87(2):241–247, 1971.

[BEJ06] T. Baumgartner, M. Esslen, and L. Jancke. From emotion perception to emotion experience: Emotions evoked by pictures and classical music. *International Journal of Psychophysiology*, 60(1):34–43, 2006.

[Bru90] G. C. Bruner. Music, mood, and marketing. *Journal of Marketing*, 54(4):94–104, 1990.

[Bux90] W. Buxton. Using our ears: An introduction to the use of nonspeech audio cues. In E. Farrell, editor, *Proceedings of the SPIE: Extracting Meaning from Complex Data: Processing, Display, Interaction*, volume 1259, pages 124–127, Santa Clara, CA. USA, February 14-16 1990.

[DKP01] K. Doel, P. G. Kry, and D. K. Pai. Foleyautomatic: Physically-based sound effects for interactive simulation and animation. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 2001)*, pages 537–544, Los Angeles, CA. USA, August 12-17 2001.

[Goo06] S. N. Goodwin. How players listen. In *Proceedings of the 35th International Conference of the Audio Engineering Society: Audio for Games*, London, UK, February 11-13 2006.

[Kan08] K. Kanev. Tangible interfaces for interactive multimedia presentations. *International Journal of Mobile Information Systems: Special Issue on Information Assurance and Advanced Human-Computer Interfaces*, 4(3):183–193, 2008.

[KB99] H. Kato and M. Billinghurst. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In *Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR 99)*, pages 85–94, San Francisco, CA, USA, October 20-21 1999.

[KCM07] K. Kanev, Z. Cheng, and N. Mirenkov. Methods and technologies for enhancing and optimizing power plant inspection procedures. In *Proceedings of the 7th IEEE International Conference on Computer and Information Technology (CIT 2007)*, pages 605–609, Aizu Wakamatsu, Japan, October 16-19 2007.

[KK05] K. Kanev and S. Kimura. Digital information carrier, 2005. Japan, Patent No. 3635374.

[KK06a] K. Kanev and S. Kimura. Direct point-and-click functionality for printed materials. *Journal of Three Dimensional Images*, 20(2):51–59, 2006.

[KK06b] K. Kanev and S. Kimura. Surface based spatial awareness for mobile robots. In *Proceedings of the Eleventh International Symposium on Artificial Life and Robotics AROB '06*, Beppu, Japan, January 23-25 2006.

[KK09a] K. Kanev and S. Kimura. Combining ubiquitous direction-sensitive digitizing with a multimedia electronic dictionary for enhanced understanding. *International Journal of Imaging Systems and Technology*, 19(2):39–49, 2009.

[KK09b] K. Kanev and S. Kimura. Digital information carrier, 2009. Japan, Patent No. 4368373.

[MDV01] S. Moncrieff, C. Dorai, and S. Venkatesh. Affect computing in film through sound energy dynamics. In *Proceedings of the 9th ACM International Conference on Multimedia*, pages 525 – 527, London, UK, September 30 - October 5 2001.

[MKT09] N. Mirenkov, K. Kanev, and H. Takezawa. Mobile music therapy with multimedia quality of life supporters for elderly and disabled. *Journal of Mobile Multimedia: Special Issue on Multimedia Networking and Applications*, 5(1):29–44, 2009.

[OCE01] J. F. O'Brien, P. R. Cook, and G. Essl. Synthesizing sounds from physically based motions. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 2001)*, pages 529–536., Los Angeles, CA. USA, August 12-17 2001.

[PNK76] M. I. Posner, M. J. Nissen, and R. M. Klein. Visual dominance: An information-processing account of its origins and significance. *Psychological Review*, 83(2):157–171, 1976.

[PS74] P. O. Peretti and K. Swenson. Effects of music on anxiety as determined by physiological skin responses. *Journal of Research in Music Education*, 22(4):278–283, 1974.

[RSB+08] D. Rocchesso, S. Serafin, F. Behrendt, N. Bernardini, R. Bresin, G. Eckel, K. Franinovic, T. Hermann, S. Pauletto, P. Susini, and Y. Visell. Sonic interaction design: sound, information and experience. In *Proceedings of the ACM CHI 2008: Extended Abstracts*, pages 3969–3972, Florence, Italy, April 5 -10 2008.

[SSC02] R. D. Shilling and B. Shinn-Cunningham. Virtual auditory displays. In K. Stanney, editor, *Handbook of Virtual Environment Technology*, pages 65–92. Lawrence Erlbaum Associates, Mahwah, NJ. USA, 2002.

[SZ02] R. Shilling and M. Zyda. 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, 2002.

[TC03] P. Tagg and B. Clarida. *Ten Little Title Tunes, Mass Media Music*. Scholars' Press, New York, NY. USA, 2003.

[TJ95] F. Thomas and R. Johnston. The Disney sounds. In W. Rawls, editor, *Disney Animation: The Illusion of Life*, pages 145–161. Abbeville Press, New York, NY. USA, 1995.

[War83] R. M. Warren. *Auditory Perception: A New Analysis and Synthesis*. Cambridge University Press, New York, NY. USA, 1983.

[WM74] P. Wright and J. McCarthy. The value of the novel in designing for experience. In A. Pirhonen, H. Isomaki, C. Roast, and P. Saariluoma, editors, *Future Interaction Design*, pages 9–30. Springer-Verlag, 1974.

[YM04] E. Yorkston and G. Menon. A sound idea: Phonetic effects of brand. *Journal of Consumer Research*, 31(11):43–51, 2004.