Advanced Sound Integration for Toy-Based Computing

Bill Kapralos, Kamen Kanev and Michael Jenkin

Abstract Despite the growing awareness regarding the importance of sound in the human-computer interface and the potential interaction opportunities it can afford, sound, and spatial sound in particular, is typically ignored or neglected in interactive applications including video games and toys. Although spatialized sound can provide an added dimension for such devices, one of the reasons that it is often overlooked is the complexity involved in its generation. Spatialized sound generation is not trivial, particularly when considering mobile devices and toys with their limited computational capabilities, single miniature loudspeaker and limited battery power. This chapter provides an overview of sound and spatial sound for use in human-computer interfaces with a particular emphasis on its use in mobile devices and toys. A brief review outlining several sound-based mobile applications, toys, and spatial sound generation is provided. The problems and limitations associated with sound capture and output on mobile devices is discussed along with an overview of potential solutions to these problems. The chapter concludes with an overview of several novel applications for sound on mobile devices and toys.

Keywords Audio interaction · Spatial sound · 3D sound · Mobile devices · Virtual environments · Augmented reality · Toy

B. Kapralos (✉) · K. Kanev
Faculty of Business and Information Technology, University of Ontario Institute of Technology, Oshawa, Canada
e-mail: bill.kapralos@uoit.ca

B. Kapralos · K. Kanev · M. Jenkin
Graduate School of Informatics, Shizuoka University, Hamamatsu, Japan

M. Jenkin
Electrical Engineering and Computer Science, York University, Toronto, Canada

P. C. K. Hung (ed.), *Mobile Services for Toy Computing,* International Series on Computer Entertainment and Media Technology, DOI 10.1007/978-3-319-21323-1_6

Introduction

Our world is filled with sounds. The sounds we hear provide us with detailed information about our surroundings and can assist us in determining both the distance and direction of objects (Warren 1983). This ability is beneficial to us (and many other species) and at times, is crucial for survival. In contrast to the visual sense, we can hear a sound in the dark, fog, and snow, and our auditory system is omni-directional, allowing us to hear sounds reaching us from any position in three-dimensional space in contrast to the limited field-of-view associated with vision. Given this omni-directional aspect, hearing serves to guide the more "finely tuned" visual attention system (Shilling and Shinn-Cunningham 2002) or as Cohen and Wenzel describe (1995), "the function of the ears is to point the eyes". It has been shown that sounds can be superior to visual stimuli for gaining attention (Posner et al. 1976), and certain sounds (e.g., the sounds of a baby crying) immediately activates mental images and schemata providing an effective means of attention focus (Bernstein and Edelstein 1971). 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 (Bishop and Cates 2001). Sounds also serve to elaborate the perception of visual information by providing us with information on invisible structure, dynamic changes, and abstract concepts that may not be expressed visually (Bishop and Cates 2001).With respect to mobile services and human-computer interactions with application software, and video games, sound can play a vital role in the communication of information. Sound can also play a vital role in toys, that is, products intended for use in learning or play (although here we are particularly interested in such devices that have been augmented with computational and acoustic input/display capabilities). In such applications, sound can be used to convey alarms, warnings, messages, and status information (such as an incoming email, or an error) (Buxton 1990). Sound effects, known as Foley sounds, associated with a particular visual imagery (such as footsteps, a door opening, glass breaking, a ball bouncing, etc.), are often added in the post-production of live action film and animation (see Doel et al. 2001) to enhance the effect of the moving imagery. It is commonly accepted within the audio/entertainment industry that "sound is emotion" and a visual interface without an appropriately designed audio component will be "emotionally flat" (Doel et al. 2001). Studies regarding the role of sound in media have shown that there is an increased physiological response in players when playing video games with sound versus those playing without sound (Shilling et al. 2002).

Despite the important role of sound within the human-computer interface, visual-based interactions comprise the majority of human-computer interactions. Recently, however, there has been a large push to exploit other modes of interaction, particularly the use of sound for human-machine interaction. This push has been motivated by a number of factors including the following (Frauenberger et al. 2004): (i) in the content of human-machine interfaces, the information being communicated is becoming more complex, making it difficult to express essential information using visual cues alone, (ii) in many applications, visual cues are restricted by the user's mobility, form factors, or by the user's visual attention being employed for other tasks, and (iii) given society's reliance on computers, such devices should be available to all members of society, including the visually impaired who cannot make use of visual-based displays. Fortunately, there is a growing effort to include those with disabilities of various types including visual impairment, in all areas of the "technological revolution" (e.g., video games). For example, audio-only video games that are played and perceived using sound, music, and acoustics only, provide access to video games by the visually impaired. One driver to this is the introduction of government regulations in many countries requiring software systems to address the specialized needs of the disabled and require systems to be accessible.

With the latest technological advances and ubiquitous use of mobile devices, sound can be used to provide unique, engaging, and interactive user experiences. Consider the Google Glass wearable computer that includes an optical head-mounted display. Google Glass provides information to the user in a hands-free format and allows users to communicate with the Internet using speech-based commands. After issuing an Internet search, for example, the speech-based search response is output to the user using bone conduction through a small loudspeaker located beside the ear, thus ensuring that the sound is (almost) inaudible to others (Arthur 2013). Sound is often included in toys, and in many video games sound is an integral part of the toy or game (e.g., the Guitar Hero series of music rhythm video games). Traditionally, toy sound was confined to the output of pre-recorded sound (sound effects and/or speech) through a single, poor quality loudspeaker (a discussion regarding loudspeaker-based sound is provided in Section "Headphones versus Loudspeaker Output") and offered little, if any, interaction possibilities. However, more recent toys and games often incorporate mobile devices and augmented reality technologies to provide far greater interactive and engaging user experiences, particularly with respect to sound. For example, the Tek Recon blaster toy (2014) features "real triggers and recoil action" and fires specially designed reusable "soft rounds" that can reach up to 23 m without endangering the public. Tek Recon has developed a freely available iOS and Android apps that make use of global positioning system (GPS) and mobile technology to provide an interface complete with interactive heads up display, live chat, radar tracking, and sound effects to accompany each firing of the blaster. The smartphone is attached to the blaster allowing for a "battle" amongst multiple players in the real world (see Tek Recon 2014).The software serves as a heads-up display allowing players to see how much ammunition (ammo) is left in their blaster, provide access to different vision modes such as night vision and a heat sensing view, and to "see" the location of enemies and teammates.

Sound-Centered Games and Toys

Sound effects and speech have played a central role in various influential toys, games, and video games over the years including the very popular Speak & Spell, Simon, Merlin, and Operation. The Speak & Spell educational toy was introduced in 1978 by Texas Instruments as a tool for assisting children to learn to spell and pronounce commonly misspelled words. It was the first toy to incorporate electronically synthesized speech (Ostrander 2000). The user is presented with a spoken word (generated using a speech synthesizer) and their task is to correctly spell the word using the device's keyboard. When the user spells the word correctly, they are verbally praised but if they spell the word incorrectly they receive encouragement to try again. Speak & Spell supported the use of cartridges allowing for additional content to be added (words to be spelled/pronounced). Another early example of the use of sound in computer games is Simon. Simon was an assembly language-based electronic game developed by invented by Ralph H. Baer and Howard J. Morrison (1980) and released in 1978. The game device itself includes four colored buttons (green, red, blue, and yellow) each of which generates a particular tone when pressed or activated by the device (all tones are harmonic). The device automatically lights up a random sequence of buttons and the tone associated with each button. The player must then press the buttons in the same order as they were presented. After each round, the difficulty of the audio-visual pattern is increased by increasing the number random buttons presses required by the player (Baer and Morrison 1980). Some early hand-held devices also supported audio generation. Developed by Parker Brothers in 1978, Merlin is one of the earliest and most popular hand-held (touch phone-like) gaming devices. Through a series of buttons and sound output, Merlin supported six games: (i) Tic-Tac-Toe, (ii) Music Machine, (iii) Echo (a game similar to the game Simon), (iv) Blackjack 13, (v) Magic Square (a pattern-based game), and (vi) Mindbender (a game similar to the game Mastermind). Prior to the development of electronic games supporting audio generation, a number of mechanical and electro-mechanical games included sound as a key component of the game. For example, in the "Operation" game, the player takes on the role of the doctor and must make Cavity Sam (the patient) better or "get the buzzer". Cavity Sam is cured by picking funny ailment pieces out of the game tray using a pair of tweezers. However, if while attempting to remove a piece the player touches the sides of the openings with the tweezers, they will get the buzzer and light up Cavity Sam's nose; the player who removes the most ailments wins (Hasbro 2014).

Although many video games are graphic-centered, video games can also be sound-centred. A sound-centered game can be an audio-only video game (a video game that is played and perceived using sound, music, and acoustics only), or it can be an audio-based video game where visuals are included as part of the game but are not the focus (rather, sound is the focus). As described below, a number of audio-only and audio-based video games have been developed. Many 3D audio video games allow the user to explore an imaginary three-dimensional world of some form using sound. For example, Roden and Parberry (2005) present an engine for mobile game development that employs spatial sound and speech recognition. Motivation for the development of the engine hinges on the belief that generating realistic spatial sound is easier and less computationally demanding in contrast to the generation of visuals (graphics). The engine is designed for narrative-based adventure video games in which speech is used to provide information about a story or scenario. In addition to speech, the engine provides support for background music and sounds that correspond to objects or processes. The concept of a "sound stage" is used to locate the various characters in the video game: the narrator is always directly in front of the player while the non-player characters (NPCs) and other players are placed 60° to the right and left of the player. The engine provides support for the creation of "worlds" which set the context for the storyline. In each world, the developer can establish attributes (e.g., volume and position) for the sounds and apply obstruction, occlusion, or exclusion filters to them. The user can interact with the system using speech, the keyboard, or both. The authors suggest that audio-based video games be explicitly designed such that the player is immersed in an invisible world rather than be a "blind explorer in a virtual world". Another example is Papa Sangre, is an audio-based video game developed by Somethin' Else specifically for the Apple iOS platform. This game places the player—who is dead and exists in the afterlife—in Papa Sangre's palace, in complete darkness. The player's task is to save his/her love and get out of the palace while avoiding dangerous monsters. The player must navigate through the palace using only sound cues; they must determine which direction or how far/close objects are using binaural audio conveyed over headphones (Collins and Kapralos 2012).

Finally, Ranaweera et al. (2012) have developed a mobile-based (smartphone) virtual concert application whereby instruments are arranged around a virtual conductor (the user) located at the center of the arranged instruments. Using the smartphone as a simplified baton, the user is able to control the instruments in the concert. For example, selecting a specific instrument by pointing at it and tapping to select or start playing it. Upon selecting an instrument, the conductor is provided with visual cues regarding the ensemble (e.g., the instrument is "jiggled" or its components dilated and contracted, and a spotlight appears until the instrument is muted). Although not necessarily a video game, this particular application highlights the novel and engaging sound-based interactions on mobile platforms.

Augmented Reality Audio-Based Games

In addition to 3D video games, there also exist audio-based games that are designed as augmented reality games, that is, games that combine real-world elements with game worlds. In Guided By Voices, for instance, players navigate a 3D world using a wearable computer interface (Lyons et al. 2000). Guided By Voices uses a wearable computer and radio frequency-based location system to play sounds corresponding to the user's location and state. Players move around in the real world and trigger actions in the game world. These actions may be triggered on friend and foe characters and on objects that can be collected. If the player collects the necessary objects when they reach a specific location, changes in the narrative occur. The sound design was created to answer specific questions that the user may have in each space: Where am I? What is going on? How does it feel? And what happens next? The authors note that "When creating the sounds for this environment 'real' sounds were not appropriate. It is not enough to simply record a sword being drawn from a sheath. Instead, a sound effect must match the listener's mental model of what a sword should sound like. This is especially important in this and similar games that lack visual cues. It is known that if a sound effect does not match the player's mental model, no matter how 'real', he/she will not be able to recognize it" (Lyons et al. 2000). One key addition in Guided by Voices is the use of a narrator who explains what actions have taken place when the player's character dies, since it's not always clear from sound effects alone.

Ekman et al. (2005) present the design of the mobile game, The Songs of the North. Although the game includes a visual component (graphics), sound remains the primary output mode. Similar to the work described by Friberg and Gardenfors (2004), the developers argue that sound is especially suitable for mobile devices because of the limited graphical rendering capabilities of such devices. Furthermore, they argue that using sound can free the user from having to attend to the visual display of the device and enables them to engage in games in which their movement can be an important part of the gameplay. In The Songs of the North, players explore a spirit world and interact with virtual objects and characters. Players take on the role of spirit wolves looking for magical artifacts. These "artifacts" are placed within locations in the game-world, which correspond to geographical locations in the real world. In this massively multiplayer game, players interact—collaborate or fight—with non-player characters and other human players in the game. Sound is used primarily to provide information regarding the current state of the spirit world and to provide specific information about objects, characters, and actions. Based on the mobile phone's GSM positioning, when a player approaches a location containing an artifact, his/her phone plays a different sound depending on the location and the artifact in the vicinity. At this point, the player can interact with the artifact in a number of ways (e.g., collect it, or return it to the world). The selected interaction is accomplished by generating a particular tune from the phone's drum interface. The game takes the player's position as input, and outputs notification sounds through the phone, based on a player's proximity to something he/she can interact with.

Collins et al. (2010) describe a preliminary mobile audio positioning game that employs marker-based interactions. Players are assigned a particular sound to their mobile device and provided with a corresponding positional marker to wear on a hat on their head. In a room configured with cameras, the position of the players is tracked using the ARToolKit and a large screen placed on one of the walls displays the results of the positioning in a 3D representation. Players create a soundscape to match the landscape by positioning themselves in the room according to a series of directions. They may be required to continuously move throughout the room to match the positioning of a moving item in the landscape. Points are awarded for quickly locating the correct position of their sound-making object within the landscape.

Spatialized Sound and the Human-Computer Interface

Although the awareness of sound within the human-computer interface is growing, sound, and spatial sound in particular, is often ignored and/or neglected in interactive virtual environments (such as virtual simulations, video games, and serious games), and toys despite the great interaction opportunities it can provide. Spatial sound technology refers to modeling the propagation of sound while within an environment while accounting for the human listener, or as Väljamäe (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". Spatial sound technology goes beyond traditional stereo and surround sound by allowing a virtual sound source to have such 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 2001) (see Nordahl and Nilsson (2014) for a thorough discussion on 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). When properly reproduced over headphones, spatial sound can deliver a very life-like sense of being remotely immersed in the presence of people, musical instruments, and environmental sounds whose originals are either far distant, virtual, or a mixture of local, distant, and virtual (Algazi and Duda 2011). Nordahl and Nilsson (2014) suggest that auditory stimuli 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".

Despite the importance of sound in the human-computer interface and the novel interaction opportunities sound can provide, there are a number of problems and issues that must be addressed particularly with respect to spatial sound generation on the currently available computing platforms such as interactive surface computers (table-top computers, touchscreen tablets such as the Apple iPad). Not only are such computing platforms typically limited computationally (although this is improving), but they represent challenging technical issues given the assumption of such devices that the user will be positioned in front of the display and the information will be presented to them vertically, may not necessarily hold. Furthermore, mobile platforms have generally relied on a single (and often poor quality) loudspeaker to convey sound to the user greatly limiting what can be done with respect to spatial sound. However, incorporating sound within the user interface of mobile platforms/devices is important particularly given the reduced capacity inherent with the smaller visual displays associated with such devices limits the presentation of ordinary text and graphics and makes visual immersion with such devices difficult. In other words, as described by Fernando et al. (2007), the size of the visual interface is intimately related to the amount of information that can be conveyed. With its independence of the visual display, spatial sound can be used to provide a high level of auditory immersion making it a "natural choice for mobile applications" (Algazi and Duda 2005).

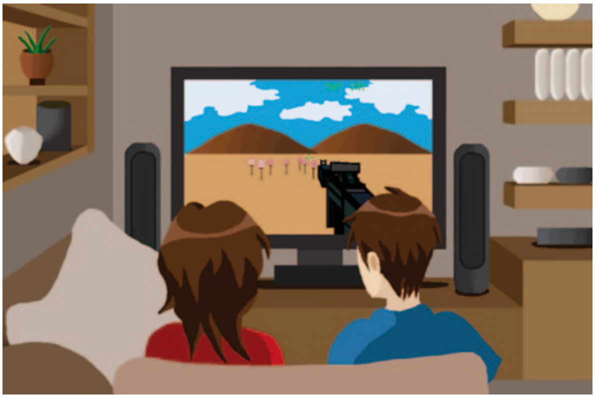

Fig. 1 An example of a vertically aligned television screen with users viewing it directly ahead of them while seated

Finally, for many decades now we have experienced our audio-visual media on screens that have been aligned vertically with users/viewers sitting or standing in front of the screen looking directly at it. As illustrated in Fig. 1, televisions, movie theaters screens, and computer screens have all presented information vertically in front of us and as a result, sound (music, dialogue, and sound effects) for television, film, software, and video games has been designed accordingly, with the placement of the loudspeakers and the sound mixing all developed based on this "vertical" format. However, with the recent surge in the use of mobile devices (smartphones, and tablets), and (to a lesser degree) tabletop 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; see Fig. 2), while look down at the screen, the assumption that users stand in front of the screen vertically cannot always be made. Furthermore, tabletop computers are designed as multi-user devices with viewers at several different angles and positions. In other words, no longer is there just a single user in front of a display looking at it vertically (see Fig. 2). The move from vertical screen-based

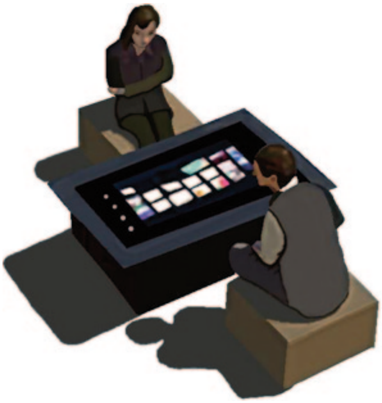

Fig. 2 Two users seated around a tabletop computer with a horizontally-aligned display

digital games to a tabletop introduces interesting open questions with respect to the delivery of visuals and sound. More specifically, as outlined by Lam et al. (2015), when we move to a horizontal screen, where do we position the loudspeakers when there are two users opposite each other playing a game (i.e., where is the "front")? How does our perception of spatial sound change when we are leaning over our computer screen versus facing it? Where should we place the loudspeakers? Where should we position sounds in the mix, (in which speaker) for best reception by the participants?

Chapter Overview

In this chapter we provide an overview of sound and spatial sound for use in the human-computer interface with an emphasis on mobile devices. In particular we concentrate on audio input/display. Such audio capabilities may be part of a suite of interaction modalities, or they may be the primary mode of interaction with the device. To take just one example of the latter, Dolphin (2014) refers to a "sound toy" as a device specifically intended as a playful medium for composition that provides access to music composition and sound creation. Sound toys can be considered as "compositional systems that allow players access to parameters of composition, various types of musical experiences, and sound worlds". They describe playful, accessible, and exploratory sonic-centric audiovisual interactive composition systems and software applications where the term "toy" suggests playful interactions (Fernando 2007). We provide an overview of the problems/limitations associated with sound capture and output on mobile devices along with potential solutions to these problems. The remainder of the chapter is organized as follows. An overview of spatial sound generation (auralization) and its associated limitations/issues along with a discussion of these limitation can be potentially overcome is provided in Section "Spatial Sound Generation". Section "Sound Capture and Output" provides a discussion regarding the capture (recording) and output of sound. The focus of the discussion is on mobile devices and more specifically, the issues inherent with the miniature inexpensive loudspeaker common on mobile devices.

Spatial Sound Generation

Collectively, "the process of rendering audible, by physical or mathematical modeling, the sound field of a sound source in space, in such a way as to simulate the binaural listening experience at a given position in the modeled space" is known as auralization (Kleiner et al. 1993). The goal of auralization is to recreate a particular listening environment, taking into account the acoustics of the environment and the characteristics of the listener (a thorough review of virtual audio and auralization is provided by (Kapralos et al. (2008)). Auralization is typically accomplished by

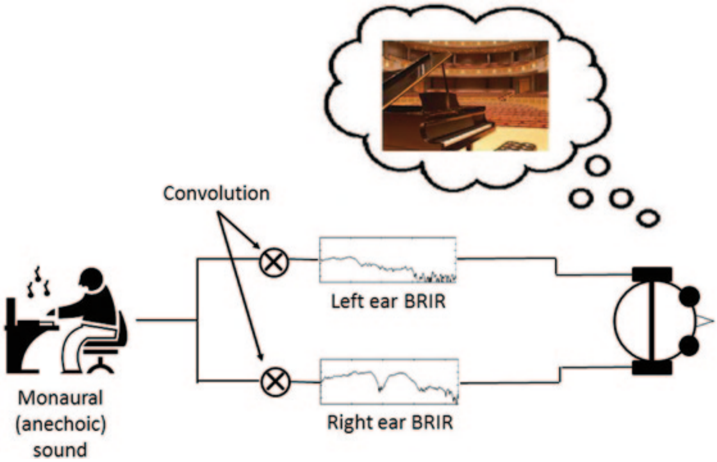

Fig. 3 Auralization overview

determining the binaural room impulse response (BRIR). The BRIR represents the response of a particular acoustical environment to sound energy and captures the room acoustics for a particular sound source and listener configuration. Once obtained, the BRIR can be used to filter, typically through a convolution process, the desired anechoic sound. When this filtered sound is presented to a listener the original sound environment is recreated (see Fig. 3). The BRIR can be considered as the signature of the room response for a particular sound source and human receiver. Although interlinked, for simplicity and reasons of practicality, the room response and the response of the human receiver are commonly determined separately and combined via a post-processing operation to provide an approximation to the actual BRIR (Kleiner et al. 1993). The response of the room is known as the room impulse response (RIR) and captures the reflection properties (reverberation), diffraction, refraction, sound attenuation and absorption properties of a particular room configuration (e.g., the environmental context of a listening room or the "room acoustics"). The response of the human receiver captures the direction dependent effects introduced by the listener due to the listener's physical make-up (e.g., pinna, head, shoulders, neck, and torso) and is known as the head related transfer function (HRTF). HRTFs encompass various sound localization cues including interaural time differences (ITDs), interaural level differences (ILDs), and the changes in the spectral shape of the sound reaching a listener. The HRTF modifies the spectrum and timing of sound signals reaching each ear in a location-dependent manner.

Various techniques are available for determining (measuring, calculating) both the HRTF and the RIR however, a detailed discussion of these techniques is beyond the scope of this chapter (see Kapralos et al. (2008) for greater details). The output of the techniques used to determine the HRTF and the RIR is typically a transfer function which forms the basis of a filter that can be used to modulate source sound material (e.g., anechoic or synthesized sound) via a convolution operation. When the filtered sounds are presented to the listener, in the case of HRTFs, they create the impression of a sound source located at the corresponding HRTF measurement position while when considering the RIR, recreate a particular acoustic environment. However, as with BRIR processing, convolution is a computationally expensive operation especially when considering the long filters associated with HRTFs and RIRs (filters with 512 coefficients are not uncommon) thus limiting their use in real-time applications and to "high" performance computational platforms.

Fundamental to the generation of virtual audio is the convolution operation that is typically performed in software in the time domain, a computationally intensive process. In an attempt to reduce computational requirements, a number of initiatives have investigated simplifying the HRTFs and RIRs. With respect to the HRTF, dimensionality reduction techniques such as principal components analysis, locally linear embedding, and Isomap, have been used to map high-dimensional HRTF data to a lower dimensionality and thus ease the computational requirements (e.g., see Kistler and Wightman (1992); Kapralos et al. (2008). Despite the improvements with respect to computational requirements, even dimensionality reduced HRTFs are still not applicable for real-time applications and although the amount of reduction can be increased thus improving performance, reducing the dimensionality of HRTFs too much may lead to perceptual consequences that render them impractical. Further investigations must be conducted in order to gain greater insight. With respect to the RIR, it is usually ignored altogether and approximated by simply including reverberation generated with artificial reverberation techniques instead. These techniques are not necessarily concerned with recreating the exact reflections of any sound waves in the environment. Rather, they artificially recreate reverberation by simply presenting the listener with delayed and attenuated versions of a sound source. Although these delays and attenuation factors do not necessarily reflect the physical properties of the environment being simulated, they are adjusted until a desirable effect is achieved. Given the interactive nature of video games and their need for real-time processing, when accounted for, reverberation effects in video games are typically handled using such techniques.

More recent work has seen the application of the graphics processing unit (GPU) to spatial sound rendering. The GPU is a dedicated graphics rendering device that provides a high performance, interactive 3D (visual) experience by exploiting the inherent parallelism in the feed-forward graphics pipeline (Luebeke and Humphreys 2007). For example, GPU-based methods that operate at interactive rates for acoustical occlusion (Cowan and Kapralos 2013), reverberation modeling (Rober et al. 2007), and one-dimensional convolution have been developed (Cowan and Kapralos 2013). Recent work has also taken advantage of perceptual-based rendering whereby the rendering parameters are adjusted based on the perceptual system (often vision), to limit computational processing. Both of these approaches have shown definite promise in their ability to provide computationally efficient spatial sound generation for interactive, immersive virtual environments and although greater work remains, they present a viable option for spatial sound generation on mobile devices. Greater details regarding the influence of sound over visual rendering and task performance is provided by Hulusic et al. (2012) while an overview of "crossmodal influences on visual perception" is provided by Shams and Kim (2010).

Sound Capture and Output

Headphones Versus Loudspeaker Output

With any auditory display, sound is output to the user using either headphones or loudspeakers. There are advantages and disadvantages with the use of either headphones or loudspeakers for sound display and one or the other may produce more favorable results depending on the application. The majority of mobile devices (notebook/tablet computers, cell phones, toys, MP3 players, gaming devices, toys, etc.) will generally have a single (miniature) loudspeaker. The proliferation of these devices has also led to a growing demand of higher quality sound output/reproduction, and this has presented manufacturers of these devices with increasing challenges while they attempt to provide louder, higher quality sound from typically small, lightweight, and inexpensive loudspeakers (Llewellyn 2011). Being low-cost, these loudspeakers rarely allow for the ideal flat frequency response across the entire audible frequency spectrum but rather exhibit significant variations in their output level across the audible spectrum, particularly at lower frequencies (Llewellyn 2011). In addition, smaller loudspeakers are generally restrictive with respect to their overall acoustical output and this too is particularly evident at lower (bass) frequencies (Larsen and Aarts 2000). Various methods and techniques have been devised to address some of these limitations inherent in small loudspeakers. For example, Larsen and Aarts (2000) describe a method that takes advantage of several human psychoacoustic phenomena that evokes the illusion of a greater lower frequency response while the power output by the loudspeaker at lower frequencies remains the same or is even lowered. A detailed discussion regarding this method in addition to other methods to improve the audio quality of mobile devices loudspeakers is provided by Llewellyn (2011). Aside from the technical issues associated with miniature loudspeakers, in the presence of others (e.g., on a plane, train, public space), loudspeaker-based sound may be bothersome to others. Furthermore, a single loudspeaker, common on the majority of mobile devices, cannot convey spatial sound (a minimum of two loudspeakers is required); often, headphones must be used instead.

Sound output via headphones offers several potential advantages over loudspeaker based systems. In particular, headphones provide a high level of channel separation and this is 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; this is particularly problematic when employing HRTF-based spatial sound via multiple loudspeakers. Headphones also isolate the listener from external sounds and reverberation which may be present in the environment (Gardner 1998), ensuring that the acoustics of the listening room or the listener’s position in the room, do not affect the listener's perception (Huopaniemi 1999). Headphones are typically used to output spatial sound where the goal is to control the auditory signals arriving at the listener’s ears such that these signals are perceptually equivalent to the signals the listener would receive in the environment being simulated (Ward and Elko 2000).

There are various drawbacks associated with using headphones. More specifically, after wearing headphones for an extended period of time, they can become very uncomfortable (Kyriakakis et al. 1999). Headphones may interfere with potentially important environmental sounds. Furthermore, the listener is constantly reminded that they are wearing the headphones and with respect to immersive virtual environments, this may negatively influence immersion. In addition, headphones may limit natural interactions amongst multiple users of an application.

There are a number of issues with headphones specific to the delivery of spatial sound. More specifically, sound delivered via headphones may lead to (see (Kapralos et al. 2008)) (i) ambiguous cues arising when the sound source is positioned on the median plane or directly above or below the listener, (ii) inside-the-head localization resulting in the false impression that the sound is originating from inside the listener's head (Kendall 1995), (iii) small movements of the headphones themselves, while being worn by the listener can change the position of the sound source relative to the listener, and (iv) the use of non-individualized HRTFs may be problematic.

Sound Capture with Mobile Devices

In addition to the output of sound via headphones or loudspeakers, sound can form part of the input interface, allowing users to interact with their application using sound. With recent technological advances, particularly with respect to speech synthesis, speech-to-text/text-to-speech conversion, and natural language user interfaces, sound, and speech in particular, is now a common method of interaction and widely available, included with a variety of applications/platforms. For example, Google Chrome (version 11) includes a speech-to-text feature that allows users to "talk" (issue instructions) to their Chrome browser. Microsoft's Speak is a text-to-speech converter included with Word, PowerPoint, Outlook, and OneNote, and allows for the user's typed words to be played as spoken words while Apple's Siri intelligent personal assistant (available on all iPhones, 4S or later) allows users to issue speech-based commends (e.g., send messages, schedule meetings, place phone calls, etc.). Unlike traditional speech recognition software that requires users to remember keywords and issue specific speech commands, Siri understands the user's natural speech, asking the user questions if it requires further information to complete a task.

The use of speech-based interaction techniques in addition to the growing use mobile-based conference calling and video telephony (which is growing in popularity given recent advancements in wireless technologies and mobile computing power), does require the ability of capturing the user's speech from a relatively close distance (e.g., one at about arm's length (Tashev et al. 2008)). Despite great improvements with respect to computing power, wireless technologies, and speech-based interaction methods and techniques particularly with respect to mobile platforms (phones, tablets, etc.), sound capture on mobile devices is typically limited. More specifically, such devices typically employ a single poor quality omni-directional microphone (which ideally responds equally to sounds incident from all directions) that picks up excessive ambient noise and reverberation, and has a limited range (typically under one meter). However, with respect to video-based streaming and capture, microphones should be able to capture sound at distances up to three meters (Tashev et al. 2008). As Tashev et al. (2008) describe, sound capture quality can be improved (an increase in the signal-to-noise ratio) by replacing the omni-directional microphone with a directional microphone (which will respond to sounds incoming from a specific direction(s)) and by employing multiple microphones (i.e., a microphone array) with beamforming techniques. Beamforming techniques allow a microphone array to localize sound-based events and/or be steered (focused) to a specific sound source location to capture any emanating sound from a sound source while attenuating any environmental noise that may be present (Zhang et al. 2008). Essentially, the goal of beamforming is to individually adjust the phase and/or the amplitude of the signal received at each microphone such that the combined output signal maximizes the signal-to-noise ratio (Boyce 2012a, b). Although various methods of beamforming have been developed, the simplest and perhaps most common, is *delay and sum beamforming* (Johnson and Dudegeon 1993). Consider a sound source located at some position x_s in three dimensional space. Furthermore, consider an array of M microphones (each microphone is denoted by m_i for $i=1...M$) where each microphone is at a unique position x_i and each is in the path of the propagating waves emitted by the sound source. In general, the time taken for the propagating sound wave to reach each microphone will differ and the signals received by each of the microphones will not have the same phase. The differences in the time of arrival of the propagating wave at the sensors depend on the direction from which the wave arrives, the positions of the sensors relative to one another, and the speed of sound v_{sound}. Beamforming takes advantage of these time differences between the time of arrival of a sound at each sensor (Rabinkin 1994), and allows the array signals to be aligned after applying suitable delays to steer the array to a particular sound source direction of arrival. Beamforming consists of applying a delay and amplitude weighting to the signal received by each sensor $s_i(t)$ and then summing the resulting signals:

$$z(t) = \sum_{i=0}^{M-1} w_i s_i (t - \Delta_i),$$

where $z(t)$ is the beamformed signal at time t, w_i is the amplitude weighting, Δ_i is the delay, and M is the number of sensors (Johnson and Dudgeon 1993).

The number of microphones employed in a microphone array on a portable device is typically small (two, or three) and the spacing between the microphones is typically also limited thus limiting the utility of a microphone array. Nonetheless, improvements over a single microphone generally result. Tashev et al. (2008) introduced a microphone array geometry for portable devices. This microphone array consisted of two unidirectional microphones placed back to back facing away from each other and provided well balanced noise suppression and speech enhancement that leads to an increase in the overall perceptual sound quality.

Finally, in addition to sound capture, the microphone on a mobile device can be used to pick up sounds arising from a variety of physical phenomena. This can, for example be used to support mobile music performance whereby performers can whistle, blow, and tap their devices as a method of musical expression (Zhang et al. 2008).

Beyond Virtual and Augmented Reality and Toys

The use of sound in virtual reality (VR) and augmented reality (AR) environments and toys that we have considered so far assume direct involvement of humans and their engagement in producing and/or listening to sounds. However, there are other uses of sound that may play an important role in VR and AR environments without direct engagement of humans. In our natural environment, sound properties can be instrumental for redefining the notion of presence and engagement/involvement and this can be extended to VR and AR environments and corresponding entities. The determination of physical presence based on the Euclidean distance between two entities is often not readily applicable, given the potential complexity of the involved environments. For example, consider a typical ("traditional") lecture. Here student attendance or "presence" can be verified by the instructor simply calling out a roll call of student names and receiving oral confirmations even if students are sitting at the far end of the room. Yet, a student who is just a meter or two away from the instructor in an adjacent room may not hear the instructor and thus not be "present" even if they are physically close. In the "normal/traditional" world we deem people that can hear us, and can be heard by us, as "present" and thus available for engagement in various interactions. That is, two people are present if they are within the sound of each other's voices. It is instructive to observe that this notion of presence is different from geographical nearness. Two physically close agents may not be able to hear each other. Validating this type of presence using radio signals (e.g., by Bluetooth or Wi-Fi signals) is problematic, as these signals can pass through walls and as a result, under the right conditions, agents can communicate over considerable distances. However, presence can be validated in a straightforward manner using ultrasound signals. More specifically, lower-band ultrasound signals that are beyond the range of normal human interaction can be readily transmitted and received with standard audio equipment thus providing viable mechanism for verifying presence without interfering with traditional human-agent interactions. An appropriately equipped cell phone, for example, could both emit and receive lower-band ultrasound signals to identify the agents that are 'within the sound of my voice'. In the text that follows we discuss how a similar approach can be applied to VR and AR environments and how VR and AR toy entities could be enabled with a sense of "presence" and augmented interaction capabilities through sound enhancements.

While playing, children often treat their toys (e.g., a doll) as living entities, speaking to their toys. Similarly to the voice-based communications in the classroom example discussed above, in a perfect playground the "presence" and availability of a toy should be easy to establish just by calling it. Furthermore, we can imagine an extension to this paradigm in which suitably equipped toys that want to "play" might call out to children that are present. Indeed this concept can be taken even further by observing that the toys themselves might communicate with other toys that are present.

What are the technical and physiological limits to augmenting devices with an ultrasound channel for the determination of whom and what is present? The human audible frequency spectrum falls approximately within the range of 20 –20 kHz (Moore 1989) and as a result, sound-based interaction within the human-computer interface is carried out within this range (it should be noted there is substantive evidence that humans can perceive sounds well above this frequency range (Lenhardt et al. 1991)). Consequently, sound input and output hardware is designed to cover this range within the limitations of current technologies. As discussed earlier in this chapter, the typical miniature loudspeakers embedded in mobile devices, gadgets, and toys are not always capable of delivering high-fidelity sound, especially in the low-frequency range. Furthermore, human hearing capabilities change with age (e.g., we lose sensitivity to higher frequency sounds as we age (Brant and Fozard 1990)). Obviously, such considerations do not apply to sound communications between toys and VR/AR entities where we can potentially employ a wider sound spectrum, beyond the human frequency perceptual range (e.g., ultrasound; frequencies greater than 20 kHz). In such a case, the frequency spectrum beyond the audible human frequency spectrum (e.g., *ultrasound*) could be reserved for VR/AR and toy communications while the more limited human audible frequency spectrum can be reserved for human-computer interaction-based applications. This enables VR/AR applications and toys to communicate amongst themselves without interfering any human-computer interactions, or the environment.

What are the limits and capabilities of utilizing ultrasound in VR/AR and toy environments? Our own current research and experimental work involving the use of ultrasound for VR/AR and toy communication aims to answer this question. First it is important to recognize that the deployment of ultrasound-based communication systems requires enhancements to current toy technologies. More specifically, ultrasound extends well beyond the 20 kHz human hearing limit. However, the higher the frequency, the less likely it becomes to reliably communicate the signal through the standard human-oriented hardware (e.g., loudspeakers intended for output in the human audible frequency range). Results of our own series of informal experiments indicate that input amplitudes drop 2–3 times when the signal frequency increases beyond 17–20 kHz. A feasible compromise may be to employ the upper frequencies of the human audible frequency spectrum (e.g., above 10 kHz); this range is supported by standard human-oriented sound equipment while not directly heard by most of the people. Secondly, within a VR/AR environment, a number of entities could be simultaneously engaged in different interactions. This obviously requires a mechanism for parallel communications between such entities and broadcasting. This can be addressed by adopting a simple frequency-based sound channel management system. Although limited, it allows for at least several VR/AR entities to announce their own presence and to detect the presence of others.

Our ultrasound-based work for VR/AR and toy communication has focused on serious gaming and in particular, class/lecture and/or group oriented educational gaming activities. A typical scenario involves a group of students working together in a classroom or another designated area that is supervised by an instructor. In the learning process students employ AR technologies, for example, by installing and using designated AR applications on their mobile device (smartphone). Let's consider for example a test or a quiz where traditionally students are handed a printed copy of the test, given some time to complete the test, and then the test is collected and graded. Now, what if we want students to complete the test electronically on their mobile device (e.g., smartphone, tablet, or notebook)? The test should obviously be made accessible online but only to those students that are physically present in class. Therefore, an application is required that allows downloading and opening a given test only if the device running the application is physically present in the designated room. Functionally, this can be achieved through an ultrasound beacon (the instructor's smartphone) broadcasting the access code for the test. Note that Bluetooth and Wi-Fi are not suitable for this purpose since they easily penetrate through walls unlike ultrasound.

The unique features and advantages of the proposed ultrasound communications method are clearly seen in other applications and in particular in AR games. The Tek Recon AR game as discussed earlier in this chapter employs the built-in location capabilities of the host smartphone for proximity warnings and other interaction enhancements when players are close to each other. However, there are problems with the determination of "closeness", especially indoors where GPS signals are often too weak or even non-existent. The proposed ultrasound-based method in contrast ensures reliable proximity determination both indoors and outdoors and can significantly enhance the gaming experience. By using ultrasound for two-way communications and broadcasting, for example, AR groups of partners and enemies can be established so that the behavior of humans in combat situations could be simulated more realistically in AR.

Conclusions

Although sound is a critical cue to perceiving our environment, and despite vital role sound can play in the communication of information in the human-computer interface, it is often overlooked in interactive applications such as video games, virtual environments, and toys where emphasis is typically placed on visuals. Sound within the human-computer interface is also vital when considering the visually impaired; the majority of which rely on, and are well practiced with, the use of sound (hearing) to gather information about their surroundings. Generally, human-computer interfaces have ignored the visually impaired. For example, the majority of video games are visual based and require the use of a graphical interface to allow for interaction. This poses a problem for the visually impaired who cannot make use of visual interfaces and therefore, cannot access or have limited access to videogames and all they can offer (Braud et al. 2002). However, substantive effort is currently underway to include those with disabilities of all types including visual impairment, in all areas of the "technological revolution" (including video games). This is in part to new legislation in many countries addressing the laws of the disabled and accessibility. As a result, audio-based interfaces are the natural substitution to vision and in fact, this is often exploited in the visually impaired's interaction with user interfaces (Copper and Petri 2004).

As we have described in this chapter, aside from conveying basic information to the user (e.g., an alarm sound to indicate a device's low power level), there is a great potential for novel uses of sound on AR/VR toys, beyond the traditional use of conveying information to the user. However, there are still a number of issues and limitations that must be overcome particularly when considering sound processing (and spatial sound generation) on mobile devices. Although constantly improving, the computational power in mobile devices is still a bottleneck for many interactive spatial sound applications particularly when considering that sounds are also accompanied by detailed graphics as well. Furthermore, mobile devices are typically restricted to a single miniature (and low quality) loudspeaker limiting the sound quality (particularly when considering lower frequencies), and the generation of spatial sound which requires a minimum of two loudspeakers. Further complicating matters is the limited battery life inherent in mobile devices; sound processing in combination with graphics processing can quickly deplete battery power and to avoid dropouts, sound processing must be kept at a high priority (Collins et al. 2010; Bossart 2006). However, as technology improves these problems will become less of an issue opening the door to a wide array of immersive and engaging audio-based interfaces.

Acknowledgment This work was supported in part by the *Natural Sciences and Engineering Research Council of Canada* (NSERC), the *Social Sciences and Humanities Research Council of Canada* (SSHRC), Interactive & Multi-Modal Experience Research Syndicate (IMMERSe) initiative, and the *Canadian Network of Centres of Excellence* (NCE), Graphics, Animation, and New Media (GRAND) initiative, and a KAKENHI Grant Number 25560109 (JSPS). The support of the *Research Institute of Electronics, Shizuoka University* in the form of a Cooperative Research Project Grant is also acknowledged.

References

Warren, R. M.: Auditory perception: A new analysis and synthesis, Cambridge University Press, New York, NY. USA (1983)

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

Cohen, M., and Wenzel, E.: The design of multidimensional sound inter- faces, Virtual Environments and Advanced Interface Design. In: W. Barfield and T. Furness (Eds.), Oxford University Press Inc., pp. 291–346, New York, NY. USA, (1995)

Posner, M. I., Nissen, M. J., Klein, R. M.: Visual dominance: An information-processing account of its origins and significance. Psychological Review, 83(2): 157–171, (1976) Bernstein, I. H., Edelstein, B. A.: Effects of some variations in auditory input upon visual choice reaction time. Journal of Experimental Psychology, 87(2): 241–247, (1971)

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

Buxton, W.: Using our ears: An introduction to the use of nonspeech audio cues. In: Proc. SPIE: Extracting Meaning from Complex Data: Processing, Display, Interaction, pp. 124–127, Santa Clara, CA. USA, (1990)

Doel, K., Kry, P. G., Pai, D. K.: Foleyautomatic: Physically-based sound effects for interactive simulation and animation. In: Proc. 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 2001), pp. 537–544, Los Angeles, CA. USA, (2001)

Shilling, R., Zyda, M., Wardynski, C. Introducing emotion into military simulation and videogame design: America's Army: Operations and VIRTE, In: Proc. GameOn Conference, London, (2002)

Frauenberger, C., Putz, V., Höldrich, R.: Spatial auditory displays: A study on the use of virtual audio environments as interfaces for users with visual disabilities In: Proc. 7th International Conference on Digital Audio Effects (DAFx'04), Naples, Italy, (2004)

Arthur, C.: Google Glass—hands-on review. *The Guardian* (London), July 2 (2013)

Tek Recon. http://www.tekrecon.com. Retrieved May 25, 2014

Ostrander, F.: The serious business of sound for toys. Los Angeles Audio Engineering Society Chapter Meeting, April 25, (2000)

Baer, R. H., and Morrison, H. J.: Microcomputer controlled game. U.S. Patent US4207087 (A), (1980)

Hasbro.: Operation game. http://www.hasbrogames.com/en-us/product/operation-game:86309B1C-5056-9047-F544-77FCCCF4C38F. Accessed on: Friday, July 18, 2014

Roden, T., Parberry, T.: Designing a narrative-based audio only 3D game engine. In: Proc. 2005 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology, pp. 274–274, Valencia, Spain, (2005)

Collins, K., Kapralos, B.: Beyond the screen: What we can learn about game design from audio-based games. In: Proc. Computer Games Multimedia and Allied Technology (CGAT 2012) Conference, Bali, Indonesia, (2012)

Ranaweera, R., Cohen, M., Endo, S.: iBaton: Conducting virtual concerts using smartphones. In: Proc. 2012 Joint International Conference on Human-Centered Computer Environments (HCCE 2012), Aizu, Japan, pp. 178–183, (2012)

Lyons, K., Gandy, M., and Starner, T. Guided by Voices: An Audio Augmented Reality System. In: Proc. International Conference on Auditory Display (ICAD 2000), Sydney, Australia, (2000)

Ekman, I., Ermi, L., Lahti, J., Nummela, J., Lankoski, P., Mayra, F.: Designing sound for a pervasive mobile game. In: Proc. 2005 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology, pp. 110–116, Valencia, Spain (2005)

Friberg, J., Gärdenfors, D.: Audio games: new perspectives on game audio. In: Proc. of the ACM International Conference on Advances in Computer Entertainment Technology, pp. 148–154, Jumanji, Singapore (2004)

Collins, K., Kapralos, B., Hogue, A., Kanev, K.: An exploration of distributed mobile audio and games. In: Proc. FuturePlay 2010 International Academic Conference on the Future of Game Design and Technology, pp. 253–254,Vancouver, BC, Canada, (2010)

Väljamäe, A.: Self-motion and presence in the perceptual optimization of a multisensory virtual reality environment. Technical Report No. R037/2005, Department of Signals and Systems, Division of Communication Systems, Chalmers University of Technology, Göteborg, Sweden,(2005)

Algazi, V. R., Duda, R. O.: Headphone-based spatial sound. IEEE Signal Processing Magazine, 28(1): 33–42, (2011)

Antani, L., Chandak, A., Savioja, L., Manocha, D.: Interactive sound propagation using compact acoustic transfer operators. ACM Transactions on Graphics, 31(1): Article 7, (2012)

Pulkki, V.: Spatial sound generation and perception by amplitude panning techniques. PhD Thesis, Electrical and Communications Engineering, Helsinki University of Technology, Finland, (2001) Nordahl, R., Nilsson, N. C.: The sound of being there: Presence and interactive audio in immersive virtual reality. In: K. Collins, B. Kapralos, and H. Tessler (Eds.), The Oxford Handbook of Interactive Audio, Oxford University Press, pp. 213–233, New York, NY, USA, (2014)

Zhou, Z. Y., Cheok, A. D., Qiu, Y., Yang, X.: The role of 3-D sound in human reaction and performance in augmented reality environments. IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans, 37(2): 262–272, (2007)

Makino, H., Ishii, I., Nakashizuka, M.: Development of navigation system for the blind using GPS and mobile phone communication. In: 18th Annual Meeting of the IEEE Engineering in Medicine and Biology Society, pp. 506–507, Amsterdam, the Netherlands, (1996)

Fernando, O. O. N., Cohen, N., Cheok, A. D.: Mobile spatial audio interfaces. In Proc. Mobile HCI '07, pp. 345–37, Singapore, (2007)

Algazi, V. R., Duda, R. O.: Immersive spatial sound for mobile multimedia. In: Proc. 7th IEEE International Symposium on Multimedia (ISM 2005), pp. 12–14, Irvine, CA, USA, (2005)

Lam, J., Kapralos, B., Collins, K., Hogue, A., Kanev, K., Jenkin, M.: Sound localization on tabletop computers: A comparison of two amplitude panning methods. ACM Computers in Entertainment (2015)

Dolphin, A.: Defining sound toys. Play as composition. In: K. Collins, B. Kapralos, and H. Tessler (Eds.), The Oxford Handbook of Interactive Audio, Oxford University Press, pp. 45–61, New York, NY , USA, (2014)

Kleiner, M., Dalenbäck, B., Svensson, P.: Auralization—an overview. Journal of the Audio Engineering Society, 41(11):861–875, (1993)

Kapralos, B., Jenkin, M., Milios, E.: Virtual audio systems. Presence: Teleoperators and Virtual Environments, 17(6): 527–549, (2008)

Kistler, D. J., Wightman, F. L.: A model of head-related transfer functions based on principle components analysis and minimum phase reconstruction. Journal of the Acoustical Society of America, 91(3): 1637–1647, (1992)

Kapralos, B., Mekuz, N., Kopinska, A., Khattak, S.: Dimensionality reduced HRTFs: a comparative study. In: Proc.2008 International Conference on Advances in Computer Entertainment Technology (ACE '08), pp. 59–62, Yokohama, Japan, (2008)

Luebke, D., Humphreys, G.: How GPUs work. IEEE Computer, 40(2): 96–100, (2007)

Cowan, B., Kapralos, B.: GPU-based real-time acoustical occlusion modeling. Virtual Reality, 14(3): 183–196, (2010)

Rober, N., Kaminski, U., Masuch, M.: Ray acoustics using computer graphics technology. In: Proc. 10th International Conference on Digital Audio Effects, Bordeaux, France, (2007)

Cowan, B., Kapralos, B.: Interactive rate virtual sound rendering engine. In: Proc. 18th IEEE International Conference on Digital Signal Processing (DSP 2013). Santorini, Greece, pp. 1–6, (2013)

Hulusic, V., Harvey, C., Debattista, K., Tsingos, N., Walker, S., Howard, D., Chalmers, A.: Acoustic rendering and auditory-visual cross-modal perception and interaction. Computer Graphics Forum, 31(1): 102–131, (2012)

Shams, L., Kim, R.: Crossmodal influences on visual perception. Physics of Life Reviews, 7(3): 295–298, (2010)

Llewellyn, W.: Audio quality improvement for mobile-device loudspeakers. Planet Analog, June 26, 2011

Larsen, E., and Aarts, R. M.: Perceiving Low Pitch through Small Loudspeakers, In: Proc. Audio Engineering Society Convention 108, paper 5151, Paris, France, (2000)

Gardner, W.: 3-D audio using loudspeakers. Norwell, MA, Kluwer Academic, (1998)

Huopaniemi, J.: Virtual acoustics and 3-D sound in multimedia signal processing. Doctoral Thesis, Faculty of Electrical and Communications Engineering, Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, Helsinki, Finland, (1999)

Ward, D. B., Elko, G. W.: A new robust system for 3D audio using loudspeakers. In: Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2000), pp. II781–II784, Istanbul, Turkey, (2000) Kyriakakis, C., Tsakalides, P., Holman, T.: Surrounded by sound. IEEE Signal Processing Magazine, 16(1), 55–66, (1999)

Kendall, G.: A 3D sound primer: Directional hearing and stereo reproduction. Computer Music Journal, 19(4), pp. 23–46, (1995)

Tashev, I., Mihov, S., Gleghorn, T., Acero, A.: Sound capture system and spatial filter for small devices. In: Proc. Interspeech 2008 International Conference, Queensland, Australia, (2008)

Zhang, C., Florêncio, D., Ba, D. E., Zhang, A.: Maximum Likelihood Sound Source Localization and Beamforming for Directional Microphone Arrays in Distributed Meetings. IEEE Transactionson Multimedia, 10(3):5 38–548, (2008)

Boyce, K.: Generating spatial audio from portable products—Part 1: Spatial audio basics. Electronic Engineering Times, March 1, (2012)

Boyce. K.: Generating spatial audio from portable products—Part 2: Acoustic beamforming using the LM48901. Electronic Engineering Times, March 16, (2012)

Johnson, D. H., Dudgeon, D. E.: Array Signal Processing: Concepts and Techniques, Prentice Hall (1993)

Rabinkin, D. V.: Digital hardware and control for a beamforming microphone array. Master's thesis, Electrical Engineering, The State University of New Jersey, New Brunswick, NJ, USA, (1994)

Misra, A., Essl, G., Rohs, M.: Microphone as sensor in mobile phone performance. In: Proc. 8th International Conference on New Interfaces for Musical Expression, Genova, Italy, (2008)

Moore, B. C. J.: An introduction to the psychology of hearing. San Diego, CA, USA, Academic Press (1989)

Lenhardt, M. L., Skellett, R., Wang, P., Clarke, A. M.: Human ultrasonic speech perception. Science, 253(5015), 82–5, (1991)

Brant, L. J., Fozard, J. L.: Age changes in pure-tone hearing thresholds in a longitudinal study of normal human aging. Journal of the Acoustical Society of America, 88(2), pp. 813–20, (1990)

Buaud, A., Svensson, H., Archambault, D., Burger, D.: Multimedia games for visually impaired children. In: K. Miesenberger, J. Klaus, W. Zagler (Eds.): Springer lecture Notes in Computer Science, 2398, pp. 173–180 (2002)

Cooper, M., Petri, H.: Three dimensional auditory display: Issues in applications for visually impaired students. In: Proc. International Community for Auditory Display, Sydney, Australia, (2004)

Bossart, P. S.: A survey of mobile audio architecture issues. In: Proc. Audio Engineering Society (AES) 29th International Conference., Seoul, South Korea, (2006)