Binaural Mixing Using Gestural Control Interaction

Nikolas Grigoriou
Dept. of Audiovisual Arts
Ionian University
Tsirigoti 7, 49100 Corfu, Greece
+302661087725
nikolas.grigoriou@me.com

Andreas Floros
Dept. of Audiovisual Arts
Ionian University
Tsirigoti 7, 49100 Corfu, Greece
+302661087725
floros@ionio.gr

Konstantinos Drossos
Dept. of Audiovisual Arts
Ionian University
Tsirigoti 7, 49100 Corfu, Greece
+302661087725
kdrosos@ionio.gr

ABSTRACT

In this work a novel audio binaural mixing platform is presented which employs advanced gestural-based interaction techniques for controlling the mixing parameters. State-of-the-art binaural technology algorithms are used for producing the final two-channel binaural signal. These algorithms are optimized for real-time operation, able to manipulate high-quality audio (typically 24bit / 96kHz) for an arbitrary number of fixed-position or moving sound sources in closed acoustic enclosures. Simple gestural rules are employed, which aim to provide the complete functionality required for the mixing process, using low cost equipment. It is shown that the proposed platform can be efficiently used for general audio mixing / mastering purposes, providing an attractive alternative to legacy hardware control designs and software-based mixing user interfaces.

Categories and Subject Descriptors

J.5 [**Arts and Humanities**]: Performing arts.

General Terms

Algorithms

Keywords

Binaural mixing, gestural interaction

1. INTRODUCTION

Audio mixing procedures can be regarded as an essential and important part of music/audio production in many applications, including music production [9], video post processing [16] or cinematography [12]. Legacy audio mixing techniques typically employ panning and balance techniques [7] for common stereo applications. Moreover, provided the market requirements for spatial audio reproduction, surround sound coding techniques seem nowadays to be the most common used set up of sound reproduction systems, regarding the aforementioned applications.

In contrast to stereo sound systems' set up, surround sound systems [5] are likely to impose an increased complexity related to both the manipulation of multiple audio streams and the installation of the necessary hardware (i.e. speakers, amplifiers etc). However, in return, they yield to a more authentic spatial sound field compared to stereo sound reproduction systems. Typical enhancement examples include moving sound sources in the three dimensional space of a room/enclosure, as well as spatial audio effects.

The same audio reproduction efficiency can be achieved by binaural reproduction systems. It is well known that these systems take advantage of the human hearing binaural cues for constructing the sound waveforms that would excite the left and right ears, given the exact position of the sound source and receiver pair, as well as the closed room model (if required by the specific application) [7] [1]. Hence, in this case, spatial sound perception is achieved using only two audio channels, resulting into less complexity, regarding the installation components and the manipulation of audio streams [34]. The use of binaural techniques can be addressed in various commercial CD's, including for example the album "Binaural" of the well known music group "Pearl Jam" [34].

On the other hand, recently, a number of researches have focused on the growing need for employing gestural-based human – computer interaction [15] [22], since it is expected that it will allow the development of novel means for human-machine interfaces [18]. Gestural interaction can offer an easy – to – use and easy – to – learn tool for complex and rich data environments, allowing direct human – machine communication with the use of natural communication code [21] [23]. Moreover, and regarding sound as one of the main components of both multimedia/ communication applications, as well as for new media artistic expression, new gestural interaction applications have been already developed and applied, like Apple's iPhone [2] or GestureTek's interfaces [7] or Visual Conducting Interfaces (VCI) [26].

Within the context of developing new mixing spatial audio techniques, the combination of the advents of gestural interaction interfaces and binaural sound processing represents a novel approach that can extend the traditional mixing process in terms of multimodal interaction. Towards this aim, the current work proposes a novel binaural mixing platform based on gestural interaction for the control of mixing parameters in real-time. Hence, the requirements for specific mixing hardware compared to surround audio systems are diminished. This lead to reduced complexity of the complete system and, in addition, by the use of natural communication code, i.e. the employment of gestures, enhances the direct way of implementation for the mixing techniques.

AM'10, September 15–17, 2010, Piteå, Sweden.

The rest of the paper is organized as following: In Section 2, the state of the art regarding binaural sound mixing, along with gestural feedback methods and technologies, is presented, followed by the implementation analysis and details presented in Section 3. Typical results obtained during the experimental employment of the proposed platform are presented in Section 4. Finally, Section 5 concludes the work and provides the necessary approach that must be followed for enhancing and evaluating the efficiency of the interactive binaural mixer.

2. TECHNOLOGY OVERVIEW

It is well known that sound mixing is an essential part of audio production, which is based on a variety of techniques that depend on the target media. These techniques are realized on specialized hardware, such as multi-track consoles, appropriate multi-track editing/mixing software applications and require a skilled audio engineer to adjust the desired mixing parameters through a number of controls implemented as hardware (i.e. the sliders/faders on the mixing console) and/or as software.

As explained briefly in the previous Section, in this work we introduce an interactive audio mixing framework, based on binaural processing and gestural control. Under this approach, the audio mixing system can be controlled by the user without taking into account the requirements and limitations explained previously. In the next two Sections, a short presentation for binaural processing and gestural-based interaction is attempted, in order to allow the establishment of the topics that are necessary for introducing the proposed platform.

2.1 Binaural hearing, processing and mixing

Spatial hearing is a topic that has been extensively investigated in the past [4], resulting into a variety of spatial audio reproduction techniques. Aim of these techniques is to reproduce the exact - desired sound field in the three dimensional space. This can be performed for example using multiple loudspeaker setups, such as in the wavefield synthesis case [2]. In order to reduce the loudspeakers required number and the corresponding setup complexity, binaural technology can be alternatively employed, which aim to synthesize the stereo signal that will trigger the left and right human ears, under the specific sound reproduction conditions and parameters applied (i.e. the enclosure geometry, the relative position between the sound source and the receiver, etc).

Binaural reproduction is based on two cues that are responsible for human sound localization perception: a) the interaural time difference (ITD) imposed by the different propagation times of the sound wave to the two (left and right) human ears and b) the interaural level difference (ILD) introduced by the shadowing effect of the head. In binaural synthesis, the effect of the above cues is incorporated into Head Related Transfer Functions (HRTFs), which represent directional-dependent transfer functions between the listener's ear canal and the specific place of the sound source [20]. Combined with simulated closed room acoustic models, binaural synthesis can also produce binaural room models, suitable for a wide range of applications, including virtual and augmented reality applications [24] – [28]. The final binaural left and right signals are derived by convolving the mono sound source signals with the appropriate binaural impulse response and can be reproduced directly using headphones or a stereo pair of conventional loudspeakers. In the latter case, the undesired crosstalk paths that transit the head from each speaker to the opposite ear must be cancelled using cross-talk cancellation methods [33].

From the above summary it is clear that a significant drawback of binaural processing is the high computational load imposed by the increased number of active sound sources, especially when binaural room simulations are additionally considered, which represents the reason for non real-time binaural processing implementations. In this work, the binaural processing functionality was realized based on the Amphiotik Technology software library, a real-time 3D binaural audio engine introduced in [29]. More specifically, the Amphiotik Technology is designed in a manner that carefully balances authenticity for the listeners and real-time operations on typical PCs. Its core consists of the Amphiotik 3D-Audio Engine, which is responsible for the signal routing, as well as for incorporating several public and custom binaural and signal processing algorithms in the processing chain. The access to the Amphiotik Technology core is efficiently achieved through the Amphiotik API, allowing the rapid development of specialized 3D audio applications [31].

In real sound environments, it is very usual to listen to moving sound sources. Obviously, sound source (or receiver) movement introduces variable source – receiver spatial relative positions. A fact that imposes the employment of a set of HRTFs that corresponds to all possible spatial locations. However, HRTFs are usually measured for a specific finite set of horizontal and vertical angles. A proposed solution to overcome this restriction is HRTFs interpolation [10], at the expense of additional high computational load. Moreover, when moving sound sources (or receivers) are simulated, in order to attain authenticity, additional psychoacoustic cues must be considered. Typical cues are the Doppler illusion [17] and the precedence effect [13]. With respect to the above restrictions, the Amphiotik Technology framework allows the improved moving sound source representation using a time-varying binaural convolution / filtering algorithm introduced in [30].

2.2 Interaction through gestural feedback

In a typical mixing environment, the fundamental information that is taken into account is the actual (or the desired) position of the recorded sound sources. Focusing on binaural technology, this information has a significant impact on the final reproduction authenticity, since the spatial details are depicted in a more realistic way than in legacy panning / balancing techniques. Based on this fact, the fundamental requirement for the gestural interaction path is to be able to control and define the instantaneous sound sources' position. This can be performed by considering all the sound sources participating in the mixing process (i.e. in a live performance or a studio recording) placed into a virtual stage.

Nowadays the advent of ubiquitous computer and its increased calculation capabilities has made it possible to control a system via gestures. Gestural interaction is very frequently realized using specific parts, such as gloves, trackers, haptic devices attached to user's hands, or image processing / detection techniques. Typical cases of gestural interaction are employed in both digital arts and technological applications. For example, gestural interaction is widely employed in interactive installations, such as the SixthSense [15], a wearable gestural interface that augments the physical world with digital information and lets the user use natural hand gestures to interact with that information. Other representatives are a) the EyeTune [9], which represents an application of gesture recognition techniques aiming to manage large music collections, b) the GRASSP (Gesturally-Realized Audio, Speech and Song Performance) [20], an environment for gestural control of musical expression by creating a gesture-to-voice system that is currently being used by performers and c) the GeFighters [28] project.

3. IMPLEMENTATION DETAILS

This section analyzes the complete architecture (illustrated in Figure 1) and the implementation details of the interactive binaural mixer platform, focusing mainly on the gesture recognition and control subsystem functionality and on specific sub-modules integration issues.

As it is shown in Figure 1, the overall mixing consists of two main parts: one that is responsible for the visual input/output signal manipulation and one that is associated with the audio input/output signals collection and reproduction.

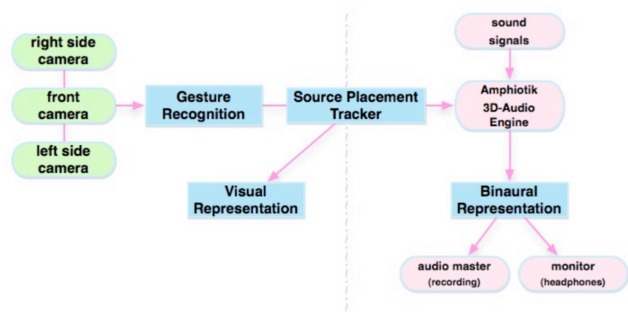

Figure 1 The complete system architecture

Being the core of the visual part, the gesture recognition module performs gesture identification using three digital video cameras: one placed at the front (relative to the user position) wall and two others on the left and right sides of the user. The gesture recognition algorithm was realized in software, based on the blob-detection library of the Processing[1] real-time platform. A detailed analysis of the gesture recognition module functionality will be provided in the next Section.

The audio part is responsible for three different procedures summarized as following: a) for mapping the input/recorded audio channels to virtual sound sources b) for the binaural signal formulation, based on the instantaneous positions of the virtual sound sources provided by the Source Placement Tracker sub-system and c) for providing the binaural output in both terms of acoustic feedback (monitoring) through headphones and 2-channel audio binaural master output.

3.1 Gesture recognition subsystem

As it was mentioned in Section 2.2, the gestural interaction should take into account the requirements for instantaneous sound source selection and spatial placement in a virtual stage environment. Hence, focusing on the sound source selection process, there is a strong dependency between the accuracy of the gesture recognition algorithm and the maximum number of the sound sources that can be manipulated concurrently (i.e. the interaction path "capacity"). In this work, a simple, color-based algorithm was selected for realizing the sound source selection: the user of the mixer is wearing a set of gloves that have different colors in the inner and outer sides. The two gloves have also different colors. For the purposes of the current work, these colors were selected to be solid, particularly red, green, blue and yellow for optimizing the color recognition process performance. It should be noted that this decision allowed the employment of low resolution and low cost digital cameras for providing visual input to the system.

More specifically, in order to select a specific sound source, the user must have his hand open, pointing towards the desired sound source visual image for at least one second (Figure 2, left). Then, the system "locks" onto the desired sound source. The user gets a selection acknowledgement in terms of a white stroke plotted around the virtual sound source visual image by the moment it is selected. Until deselected, any further movement of the specific user hand is directly mapped as a corresponding physical movement of the virtual sound source in the horizontal (X-Y) plane of the scene. These coordinates are fed in real-time to the Amphiotik 3D-Audio Engine for estimating the instantaneous binaural impulse response that corresponds to the specific sound source and the listener positions.

Additionally, the video signals provided by the side (left and right) cameras are used for determining the (left and right) hand position alternations in the vertical (X-Z) plane. These variations are directly mapped into sound source gain adjustment, a parameter that is related to the sound pressure level produced by the specific sound source at a distance equal to one meter. In this case, because of the fact that only the side part of the hand is detected, the gesture recognition subsystem employs a different processing algorithm based on two colors blob-detection.

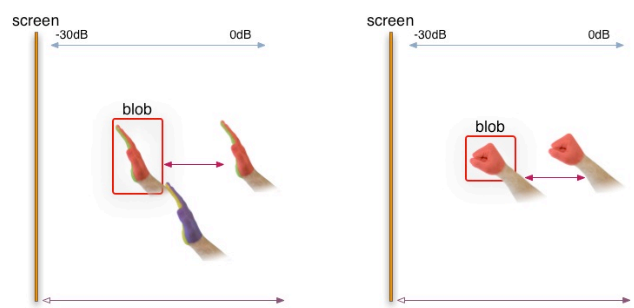

Figure 2 Sound source selection - deselection

The above positions' estimation procedure is supported by the visual projection of the virtual scene realized by the visual representation module appeared in Figure 1. Hence, the user is capable to optically monitor the current positions of the virtual sound sources, being represented as color-filled circle and rectangle shapes. It should be also noted that the sound source gain variations (i.e. the hands' movement in the vertical plane) are visually represented as corresponding variations of the corresponding color-filled shape area, combined with the display of the selected sound source gain value in dB.

In the case that the user needs to deselect a previously selected sound source, he has to close his hand (in a punching form) for at least one second, while pointing towards the particular pre-selected virtual sound source (see Figure 2, right). Closing the

---

[1] www.processing.org (last access: July 26th, 2010)

hand directly exposes the glove outer color to the front camera, providing the necessary feedback to the system for the virtual sound source deselection.

In order to finalize the gestural interaction set, an additional procedure for adding virtual sound sources into the virtual scene has to be defined. For the purpose of this work, this procedure was designed in the basis that the different sound sources were available through a list located at the upper side of the visual projection (Figure 4, right). This list automatically appears when any user's hand is detected to be pointing towards its exact position. All the sound sources that are attached onto the above list are muted. The user can then choose the desired sound source from the list by simply following the selection procedure described previously. The corresponding sound source is then activated (un-muted). It should be also noted that the number of the available to add virtual sources is determined by the number of the corresponding audio recording channels. In particular, each active audio recording channel is assigned an ascending integer index which is directly mapped to a virtual sound source that can be added to the virtual sound scene, following the previously described approach.

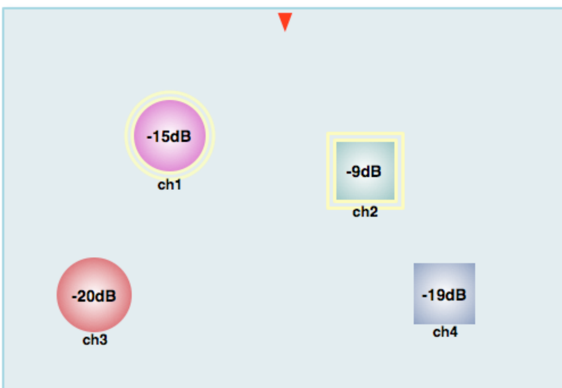

Figure 3 Visual representation of the virtual scene

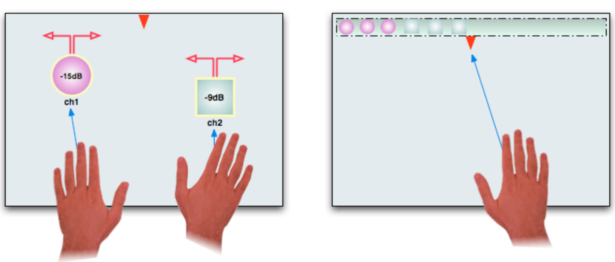

Figure 4 Concurrent virtual sound source control and addition

A final task that should be taken into consideration is system calibration, based on the physical dimensions of the enclosure in which the complete mixing system is installed. During this calibration process, specific human hand distances from the cameras are mapped to virtual scene distances, allowing a direct transformation between the actual and virtual space scale. The same calibration process is also applied for determining the range of the applicable audio channel gain values.

4. DEMONSTRATION AND RESULTS

In general, a critical parameter for the development of gestural-based human – machine interfaces and interaction schemes is the efficiency of the interaction design from the user point of view. Focusing exclusively on the proposed binaural mixing platform, it is necessary to investigate and determine the possible enhancements that can be achieved during the employment of the proposed system for the completion of a specific task, particularly compared with the legacy mixing means widespread used by audio professionals and engineers.

Towards this direction, in the present work a sequence of subjective tests was designed to answer fundamental questions related to specific implementation issues. Typical questions were in particular: a) what kind of hand gestures should be used, taking into account the trade-off between the user convenience and the restrictions/accuracy imposed by the gesture recognition algorithm? b) How can the selected gestures be optimally combined for maximizing the efficiency of the mixing process, mainly in terms of the number of the sound sources that can be active and simultaneously controlled? c) Finally, is there a clear advantage of the proposed audio mixing interface, compared to the employment of legacy hardware mixing devices? The tests were performed using a five-channel digital recording of rock material and the total duration of a mixing test was equal to 5 minutes.

The focus on the first two questions was continuous and resulted into the gestural interaction scheme described in Section 3.1 through a recursive investigation and optimization process. However, the major difficulty for successfully realizing the above sequence of tests was the limited number of the users who were selected to participate under the requirement that they should have the basic technical skills and knowledge in audio mixing practices. This fact resulted into the lack of consistent results, especially for those related to the proposed system efficiency assessment. What it became evident though from the above limited number of tests is the general impression of the direct ability to accurately place the sound sources in the virtual scene environment, even for complex sound source movement scenarios. This result was recorded by the majority of the users that are experienced with typical stereo panning techniques. Obviously, future work will have to focus on the subjective efficiency evaluation for verifying the above trend, following a well-organized assessment process.

Moreover, it was also concluded that the localization of a specific sound source in the virtual sound scene was significantly sub served by the virtual scene visual representation, rendering it a process that imposed no additional and unpredicted delays. This result is in close accordance with relative outcomes obtained by a number of previous studies that investigated the ability to construct gesture-controlled, spatially arranged auditory user interfaces [6], [25].

5. CONCLUSIONS

The continuous evolution in the area of multichannel audio reproduction and three dimensional sound field reconstruction represents the primitive point for the development of new forms of technological and artistic applications. Combined with novel interaction means that enhance the legacy user interfaces using feedback paths that are transparent to the user, a new generation of 3D interactive sonic applications can be emerged.

Under this perspective, in this work, a novel binaural mixing platform is presented which employs advanced gestural-based interaction techniques for controlling the mixing parameters. For binaural processing tasks, the Amphiotik 3D Audio Engine is employed which combines the well-known binaural technology and HRTF theory for creating state-of-the-art three dimensional audio environments in real-time, even for a large number of high-quality digital audio sources. Moreover, focusing on the path of interaction between the platform and the user, simple hand gesture-based patterns were selected for controlling the mixing parameters (i.e. for the instantaneous sound source selection and the corresponding positioning in the environment space). A limited number of assessment tests have indicated that the selected gestural set can sufficiently support the audio mixing process, mainly in terms of the more accurate and direct ability to place the sound sources in the virtual scene environment.

Future investigations and enhancements of the proposed platform will include analytical mixing accuracy and performance assessment under real-world usage conditions (i.e. during audio production phase), which will potentially lead to enhancements required for the optimization or redefinition of the interaction paths. Moreover, an additional task that can be considered is the adaptation of the proposed platform to new technological trends in gaming hardware such as the Project Natal for the Xbox 360 platform.

6. ACKNOWLEDGMENTS

The authors wish to thank Mr. Christos Tsakostas for his valuable contribution to this work.

7. REFERENCES

[1] AES Staff Technical Writer 2006. Binaural Technology for Mobile Applications. *Journal of Audio Engineering Society* 54, 10 (Oct. 2006), 990-995.

[2] Apple. Apple - iPhone 4 - Video calls, multitasking, HD video and more. http://www.apple.com/iphone/ (retrieved Mar. 06, 2010)

[3] Berkhout, A. J., Vogel, P., and de Vries, D., 1992. Use of wave field synthesis for natural reinforced sound. *AES 92nd Convention* (Vienna, Italy, Mar. 24 - 27, 1992).

[4] Blauert, J., 1997. Spatial Hearing (revised edition). The MIT Press (Cambridge, Massachusetts).

[5] Breebaart, J., Herre, J., Villemoes, L., Jin, C., Kjorling, K., Plogsties, J. and Koppens , J., 2006. Multichannel Goes Mobile: MPEG Surround Binaural Rendering. *AES 29th International Conference* (Seoul, Korea, 2006).

[6] Cohen, M., 1993. Throwing, pitching and catching sound: audio windowing models and modes. *International Journal of Man – Machine Studies* 39, 269-304.

[7] Gesturetek. Gesture Recognition & Computer Vision Control Technology & Motion Sensing Systems For Presentation and Entertainment. http://www.gesturetek.com/ (retrieved Mar. 06, 2010)

[8] Hammershøi, D., Møller, H., 2005. Binaural Technique — Basic Methods for Recording, Synthesis, and Reproduction. in *Communication Acoustics*, Ed. J. Blauert, Springer, Netherlands, 223-254.

[9] Harris, J.K., Delin, A.J.L., Naylor, J.R., Stewart, I.W. , Prince, J.M., 2002. The application of embedded transputers in a professional digital audio mixing system. *IEE Colloquium on Transputer Applications* (Aug. 06, 2002), 2/1-2/3.

[10] Hartung, K., Braasch, J., and Sterbing, S., 1999. Comparison of Different Methods for the Interpolation of Head-Related Transfer Functions. *AES 16th International Conference on Spatial Sound Reproduction*, 319-329.

[11] Hornof A., Rogers T., Halverson T., 2007, EyeMusic: Performing Live Music and Multimedia Compositions with Eye Movements, *NIME07* (New York, U.S.A., June 7 - 9, 2007).

[12] Izhaki, R., 2008. Mixing Audio: Concepts, Practices and Tools. Focal Press, Oxford.

[13] Litovsky, R., Colburn, S., Yost, W., and Guzman, S., 1999. The precedence effect. *Journal of the Acoustic Society of America* 106, 1633-1654.

[14] Mistry, P., Maes, P. and Chang, L. WUW - wear Ur world: a wearable gestural interface. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems* (Boston, MA, USA, April 4 - 9, 2009).

[15] Morrel-Samuels, P., 1990. Clarifying the distinction between lexical and gestural commands. *International Journal of Man-Machine Studies* 32, 5 (May 1990), 581-590.

[16] Mulhem, P., Kankanhalli, M.S., Ji Yi, Hassan, H., 2003. Pivot vector space approach for audio-video mixing. *IEEE Multimedia* 10, 2 (April 22, 2003), 28-40.

[17] Neuhoff, G., and McBeath, M. K., 1996. The Doppler Illusion: The Influence of Dynamic Intensity Change on Perceived Pitch. *Journal of Experimental Psychology, Human Perception and Performance* 2, 4, 970-985.

[18] Pavlovic, V.I., Sharma, R., Huang, T.S., 1997. Visual interpretation of hand gestures for human computer interaction: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19, 7 (July 1997), 667-695.

[19] Pritchard B., Fels S., 2006, GRASSP: Gesturally-Realized Audio, Speech and Song Performance, In *Proceedings of the 2006 International Conference on New Interfaces for Musical Expression* (Paris, France).

[20] Pulkki, V., 1997. Virtual Sound Source Positioning Using Vector Base Amplitude Panning. *Journal of Audio Engineering Society* 45, 6 (June 1997), 456-466.

[21] Quek, F.K.H., 2002. Unencumbered gestural interaction. *IEEE Multimedia* 3, 4 (Aug. 6, 2002), 36-47.

[22] Rehm, M., 2010. Nonsymbolic Gestural Interaction for Ambient Intelligence. In *Human-Centric Interfaces for Ambient Intelligence*, Ed. Aghajan, H., Lopez-Cozar Delgado, R., and Augusto, J. C., Academic Press, Oxford, 2010, 327-345.

[23] Sato, E., Yamaguchi T., Harashima F., 2007. Natural Interface Using Pointing Behavior for Human–Robot Gestural Interaction. *IEEE Transactions on Industrial Electronics* 54, 2 (April 2007), 1105-1112.

[24] Savioja, L., Huopaniemi, J., Lokki, T., and Vaanaen, R., 1999. Creating Interactive Virtual Acoustic Environments. *Journal of Audio Engineering Society* 47, 9 (Sept. 1999), 675-705.

[25] Sawhney, N. and Schmandt, C., 2007. Nomadic Radio: Speech and Audio Interaction for Contextual Messaging in Nomadic Environments. *ACM Transactions on Computer-Human Interaction* 7, 3, 353-383

[26] Segen, J.., Gluckman, J., Kumar, S., 2000. Visual interface for conducting virtual orchestra. *Proceedings of 15th International Conference on Pattern Recognition* 1, 276-297.

[27] Teixeira, J. M., Farias, T., Moura, G., Lima, J., Pessoa, S., and Teichrieb, V., 2006. Gefighters: an experiment for gesture-based interaction analysis in a fighting game. In *SBGames* (Brazil).

[28] Tikander, M., 2009. Usability Issues in Listening to Natural Sounds with an Augmented Reality Audio Headset. *Journal of Audio Engineering Society* 57, 6 (June 2009), 430-441.

[29] Tsakostas, Ch., and Floros, A., 2007. Optimized Binaural Modeling for Immersive Audio Applications. *Audio Engineering Society 122th Convention* (Vienna, Italy, May 2007).

[30] Tsakostas, Ch., and Floros, A., 2007. Real-time Spatial Representation of Moving Sound Sources. *Audio Engineering Society 123rd Convention* (New York, U.S.A., Oct. 2007).

[31] Tsakostas, Ch., Floros, A., and Deligiannis, Y., 2007. Real-time Spatial Mixing Using Binaural Processing. In *Proc. of the 4th Sound and Music Computing Conference* (Lefkada, Greece, July 11 - 13, 2007), 291-294.

[32] Tsakostas, Ch., Floros, A., and Deliyiannis, Y. 2007. Binaural Rendering for Enhanced 3D Audio Perception. In *Proceedings of the AudioMostly 2007 2nd Conference on Interaction with Sound* (Ilmenau, Germany, Sept. 27 - 29, 2007), 125-128

[33] Ward, A. B., Elko, G. W., 1999. Effect of loudspeaker position on the robustness of acoustic crosstalk cancellation. *IEEE Signal Processing Letters* 6, 5 (May 1999), 106-108.

[34] Wikipedia. Binaural (Album) - Wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Binaural_%28album%29 (retrieved Mar, 06, 2010)