SPATIAL ENHANCEMENT FOR IMMERSIVE STEREO AUDIO APPLICATIONS

Andreas Floros[1] and Nicolas – Alexander Tatlas[2]

[1]Dept. of Audiovisual Arts, Ionian University, Corfu, Greece
[2]Dept. of Electronic Engineering, Technological Educational Institute of Piraeus, Greece
E-mail: [1] floros@ionio.gr, [2]ntatlas@teipir.gr

ABSTRACT

In this wok, we propose a stereo recording spatial enhancement technique which retains the original panning / source location, proportionally mapped into the perceived expanded sound stage. The technique uses a time-frequency domain metric for retrieving the panning coefficients applied during the initial stereo mixing. Panning information is also used for separating the original single channel audio streams and finally for synthesizing the expanded sound field using binaural processing. The technique is mainly intended for playback applications over short-distant loudspeaker setups or headphones and, in the context of this work, it is subjectively evaluated in terms of the perceived sound field expansion and the sound-source spatial distribution accuracy.

***Index Terms*—** Stereo widening, binaural mixing

1. INTRODUCTION

Recent advents on three dimensional (3D) visual representation technologies have emerged an increasing market requirement for audio-visual reproduction authenticity targeted to a wide range of applications, such as digital high-definition TV, cinema, video games and lately typical computer usage. Unlike the compact solutions provided for efficiently realizing 3D optical reproduction, spatial sound representation is still based on conventional multichannel surround techniques, or, alternatively, on ambisonics or wavefield synthesis. Nearly all these approaches require relatively complex loudspeaker setups, while additional limitations are imposed, mainly in terms of efficiently delimiting the sweet-spot area.

Nowadays, when considering 3D sound, two parameters have to be considered a) the existing, significant volume of the existing stereo recordings and broadcasts, as well as the wide-base of in-use stereo playback systems, and b) the fact that many users very frequently use portable devices (such as smart phones and ultra-portable computers) for carrying out audio reproduction, where the limited distance of the installed loudspeakers narrow the stereo stage and render the perceived playback monophonic. Hence, spatial enhancement of existing stereo audio material is now very demanding in order to provide a more natural listening environment in modern 3D audio/visual applications, and can be achieved in the context of broadening the perceived width and depth of the reproduced sound field, ideally to dimensions beyond the physical locations of the loudspeakers [1] or the headphones.

Towards this aim, binaural technology represents an attractive approach, since it offers efficient spatial sound reproduction over a pair of audio channels, based on transfer functions' models of the physical path between the sound source and each listener's human ear [2]. Hence, binaural technology is highly compatible with typical stereo playback setups, while, as it will be explained in detail in the next Section, it can also be incorporated for expanding the stereophonic stage by introducing spatially distributed virtual sound sources.

Currently, one may discriminate two different categories of applications for stereo spatial enhancement using binaural acoustics: the real-time ones that perform all the necessary audio processing tasks during the playback, and the non real-time. Although it seems that there is a clear advantage of the real-time approach, this work focuses on a technique, which pre-processes the original stereo material in order to produce the final enhanced stereo output. As it will be shown, this allows widening a given stereo sound field, taking into account and maintaining the original panning applied during the stereo master track production. Moreover, provided the increased usage of portable equipment for audio playback, this technique can be applied transparently to the user, during the transfer of the audio stereo data to the local storage device. Under this approach, the signal processing computational requirements of the portable platform are additionally relaxed, as the processing tasks are not performed concurrently with audio data (i.e. mp3) decoding.

The rest of the paper is organized as following: Section 2 provides a short description of binaural synthesis fundamentals, focusing mainly on its key aspects usually employed for stereo recording widening/enhancement. Next, in Section 3 the proposed stereo 3D enhancement algorithm is analytically described, followed by typical results obtained through specific demonstrative cases, as well as

978-1-4577-0274-7/11/$26.00 ©2011 IEEE

overall subjective outcomes, all presented in Section 4. Finally, Section 5 concludes this work and proposes a number of issues that can be further considered for technically and practically evolving the proposed technique.

2. BINAURAL SYNTHESIS: AN OVERVIEW

The ability of humans to localize sound in three dimensions (3D) represents a fundamental functional characteristic of our hearing mechanism. This ability is based on the well-known duplex-theory fundamental sonic cues, in particular a) the interaural intensity (or level) sound difference and b) the interaural time difference, all introduced by the difference in relative spatial paths between the ears and an active sound source. Both the above factors (usually abbreviated as IID and ITD respectively) satisfactory explain the sound localization properties of human, under an exception originating from specific positions geometrically organized as 3D "cones-of-confusion", where the IIDs and ITDs have the same values. However, even in these geometric areas, humans are able to localize sound, due to a filtering effect imposed by the presence of the head, torso and especially the outer ear (pinna).

This effect is modeled through the Head-Related Transfer Functions (HRTFs) [3]. HRTFs are transfer functions that strongly depend on the relative position of the human listener's ear canal and the specific sound source placement [4]. Hence, provided that they are appropriately selected, they can be employed for calculating the binaural signal that would be collected by both ears under the presence of one (or more) sound sources through signal convolution, a process frequently described under the term binaural synthesis [5].

Binaural synthesis has been extensively employed in many application areas, namely in virtual or augmented reality audio environments, for realizing authentic 3D sonic environments [6] – [7]. In these applications, the achievement of high-level user immersion is the fundamental requirement, raising several implementation issues related to sonic representation accuracy [8]. However, additionally to these obvious sound field synthesis applications, binaural technology was recently employed for parametric MPEG surround coding [9]. The spatial information is extracted from multichannel audio and a down-mix is produced and transmitted together with the low-rate spatial side information, allowing backward compatible representation of high quality audio at bitrates comparable to those currently used for representing stereo (or even mono) audio signals. The enhancement of the spatial characteristics of typical stereo recordings also represents an alternative binaural processing target application [10]. The aim is to obtain "spatially-enhanced" immersion under stereo playback through headphones or two loudspeakers, at the expense of possible negative effects at the overall sound quality, mainly in terms of the perceived spectral balance and induced distortion. Spatial enhancement is usually assessed in terms of the achieved soundstage width and the sound sources' image exact definition.

2.1. Overview of stereo spatial enhancement techniques

Various approaches have been employed towards the above aim. For example, cross-feed simulation and delay effects [11] or controlled reverberation addition corresponding to a listening space [12] were previously employed for improving the spatial stereo sound image produced through headphones. More recently, in [13], the authors proposed a real-time 3D audio enhancement algorithm that maps the stereo audio input to a number of virtual sound sources appropriately placed around the listener, assuming an open or closed space. The synthesis of the virtual sources is performed using binaural processing and the audio streams that are assigned to them are derived from the stereo data using simple two channel transformation / mixing. The same approach was recently applied in terms of a "Music Widening" algorithm targeted to hearing aid users who suffer from the "in-the-head locatedness" effect [14]. The algorithm achieves sound externalization by introducing simple mirror-image virtual sources within a room.

In all the above cases, the virtual sound source spatialization is controlled through the definition of the room geometry and reverberation properties, or by algorithmically distributing the sound sources, following the loudspeaker setup positions of common surround sound formats. As mentioned in Section 1, this work proposes a stereo spatial enhancement technique that maintains the stereo panning applied on the original audio recording, hence allowing a more authentic and natural sonic stage expansion in terms of sound field similarity presented in the initial recording.

Apart from the binaural synthesis process, the proposed technique also incorporates the cross-talk cancellation technique employed in [13]. In general, cross-talk cancellation (or compensation, CTC) is required when the binaural synthesis output playback is performed over a stereo loudspeaker setup, in order to cancel the undesired crosstalk acoustic paths inherently imposed between a specific loudspeaker and the acoustic receiver opposite ear [15]. The cancellation efficiency strongly depends on the relative positions between the loudspeakers and the listener's ears, which ideally should be known. Any practical deviation from the assumed geometry induces unpredictable signal amplitude and phase deviations and affects the overall cancellation efficiency.

3. PROPOSED STEREO SPATIAL EXPANSION

Fig. 1 illustrates the architecture of the proposed spatial expansion algorithm. The stereo input signal feeds the Panning Analysis sub-system, which aims to derive the panning gains applied during the initial stereo mix. This information is further used by the Source Stereo Unmix module, which separates the original stereo soundtrack sources. Finally, the binaural synthesis module derives the binaural 2-channel output (indicated as b-L(t) and b-R(t)), taking into account the stereo image position obtained by the panning gain information and by creating an appropriate number of virtual sound sources, located at the derived (or linearly mapped in the spatial domain) positions. The detailed analysis of each of the above modules is provided in the following three Sections.

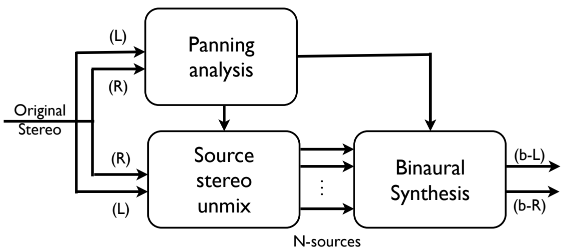

Fig. 1. Proposed algorithm architecture.

3.1. Stereo panning analysis

Stereo panning analysis is performed using the methodology proposed by Avendano in [16]. It aims to achieve source separation from stereo signals using a cross-channel metric applied in the frequency domain. This metric (namely the panning index) is directly related to the panning gains applied during the stereo mix and was recently used for music information retrieval tasks [17]. Here, we follow the same methodology, but the panning index estimation is performed in one third octave bands.

More specifically, assuming a stereo signal comprised of the left and right components $s_l(t)$ and $s_r(t)$ respectively, the panning index is obtained by comparing their time-frequency representation $S_l(m, k)$ and $S_r(m, k)$ obtained by consecutive short time fourier transforms (STFT). The comparison function is defined as

$$\psi(m,k) = 2\frac{\left|S_l(m,k)S_r^*(m,k)\right|}{\left|S_l(m,k)\right|^2 + \left|S_r(m,k)\right|^2} \tag{1}$$

where k is the frequency index, m the time index of the consecutive STFT blocks and '*' denotes complex conjugation. In [16] it is shown that this function results into values proportional to the applied panning coefficients a, in those frequency bands that correspond to the spectral content of the original sound sources' signals, that is

$$\psi(m,k) = 2a\sqrt{1-a^2} \tag{2}$$

provided that a sinusoidal energy-preserving panning law is applied. In order to resolve the ambiguity caused by the symmetrical values obtained by eq. (2) as a function of a, partial similarity expressions are defined as:

$$\psi_l(m,k) = \frac{\left|S_l(m,k)S_r^*(m,k)\right|}{\left|S_l(m,k)\right|^2},\ \psi_r(m,k) = \frac{\left|S_r(m,k)S_l^*(m,k)\right|}{\left|S_r(m,k)\right|^2}$$

and the panning index is finally obtained as

$$\Psi(m,k) = \left[1-\psi(m,k)\right]\tilde{\Delta}(m,k) \tag{3}$$

The $\tilde{\Delta}(m,k)$ is the ambiguity-resolving function calculated as

$$\tilde{\Delta}(m,k) = \begin{cases} 1, \text{ for } \Delta(m,k) > 0 \\ 0, \text{ if } \Delta(m,k) = 0 \\ -1, \text{ for } \Delta(m,k) < 0 \end{cases} \tag{4}$$

where:

$$\Delta(m,k) = \psi_l(m,k) - \psi_r(m,k) \tag{5}$$

The above panning index can be expressed in terms of 1/3rd octave bands as:

$$\Psi(m,\lambda) = \frac{(-1)^\rho}{L}\sqrt{\sum_{k=l(\lambda)}^{L(\lambda)} \Psi^2(m,k)} \tag{6}$$

with λ being the index of a specific frequency band defined by its starting frequency bin $l(\lambda)$ and having length equal to $L(\lambda)$ bins. Additionally, ρ defines the sign and equals to 2 provided that the summation of the $\Psi(m,k)$ values within the λ-th frequency band is positive, otherwise it equals to 1.

Based on the 1/3rd octave band representation of eq. (6), one may produce a representation of the distribution of energy as a function of the panning index. This is done by integrating the energy of these frequency bands that correspond to same (or similar within a specific range) panning index values. In the derived time-panning domain, the i-th panning coefficients applied during the original stereo mix are clearly identified (and denoted here as *panning_index(i)*). Hence, if the exact stereo loudspeaker layout is considered to be available, these panning coefficients can also be mapped to specific positions between them. For the purposes of the current work, the stereo loudspeaker angle is considered to be 90^0 (see Fig. 2), allowing a direct map of the panning index value to the i-th source position, expressed in terms of the angle between the left loudspeaker and the sound source $\theta_{left}(i)$ as:

$$\theta_{left}(i) = 45^0(panning\_index(i) + 1) \tag{7}$$

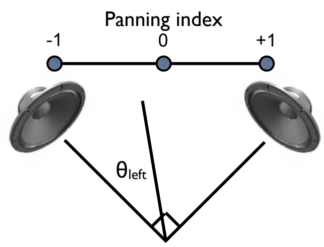

Fig. 2. Panning index to spatial position mapping for the considered stereo loudspeaker setup.

It should be also noted here that in order to obtain one panning coefficient per sound source, the latter should not significantly overlap on the frequency domain. If this is not the case, there will be an estimation error, which is however bounded as it is shown in [16].

3.2. Source unmixing

The original stereo-track sound source separation (or unmix) can be performed by spatially defining stereo source images. In particular, the original sound source signals can be separated by indicating time-frequency zones, where the time-frequency presented panning index $\Psi(m,k)$ has values equal to *panning_index(i)*. These specific time-frequency zones can then be employed for synthesizing the corresponding separated sources time-domain signals (through inverse STFT). As proposed in [18], in order to take into account the potential spectral overlapping of the sound source signals, windowing functions must be applied to each specific *panning_index(i)* value, expressed as:

$$W_i(\Psi(m,k)) = v-(1-v)e^{-\frac{1}{2\xi}(\Psi(m,k)-panning\_index(i))} \quad (8)$$

where ξ defines the window width, the selection of which represents a trade-off between sound source separation accuracy and distortion imposed by neighboring sources in the frequency domain. Moreover, the v parameter is used for always achieving non-zero STFT values. The separated sound source signals are finally obtained in the frequency domain, i.e.

$$S_i(m,k) = W_i\left[\Psi(m,k)\right]\left(S_l(m,k)+S_r(m,k)\right) \quad (9)$$

and the corresponding time domain signal $s_i(t)$ are derived by inverse STFT operations.

3.3. Binaural mixing

Aim of the binaural mixing subsystem is to assign the separated sound signals $s_i(t)$ derived by the source unmixing algorithm to virtual sound sources located at equally distant positions that follow the variations of $\theta_{left}(i)$ angles. In order to do so, the following angle transformation is used:

$$\theta_{virtual,i} = panning\_index(i)\theta_b \quad (10)$$

where $\theta_{virtual,i}$ denotes the angle of the virtual (binaural) source image relative to the central axis that corresponds to a panning index value equal to zero (see Fig. 2). θ_b represents the binaural image widening factor, a parameter that can be defined in the range of $[45^0, 90^0]$ and expands the allowable angular range of the virtual sound sources placement. Obviously, if θ_b equals to 45^0, the potential virtual sources positions are limited to those implied by the stereo loudspeaker setup.

Binaural synthesis is performed using the Amphiotik Technology library presented in [19], using the KEMAR set of HRTFs. Additionally, as mentioned in Section 2.1, cross talk cancelation is also provided, allowing the reproduction of the binaural output through a pair of stereo loudspeakers.

4. RESULTS

The performance of the proposed stereo enhancement scheme was initially evaluated using a stereo mix (see Fig. 3), produced by panning two different audio streams: a vocal track and a bass instrument. The initial panning positions of these sound sources where defined relatively to the left loudspeaker position as $\theta_{vocals}=30^0$ and $\theta_{bass}=50^0$.

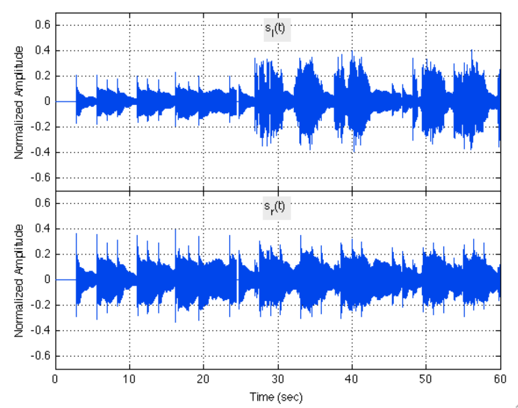

Fig. 3. Original stereo signal (left - $s_l(t)$ and right - $s_r(t)$).

Fig. 4 (upper diagram) illustrates the resulting variation of the stereo signal energy as a function of the panning index calculated using eq. (6) in 1/3[rd] octave bands ($\Psi(m,\lambda)$). Dark-colored points correspond to lower signal energy values, light gray correspond to high energy distribution and white color represents the absence of energy ditribution. This variation was calculated for possible panning index values ranging from -1 to 1 with an incremental step equal to 0.05. The existence of the two dominant panning coefficients that are clearly identified in this diagram can be also verified by calculating the average signal energy distribution as a function of $\Psi(m,\lambda)$. This is displayed in the lower graph of Fig. 4, where it can be observed that the maximum average signal energy is obtained for panning index equal to -0.35 and 0.1 respectively. Using eq. (7), these results correspond to panning angles θ_1=29.25^0 and θ_2=49.5^0, values that are very close to the original panning positions considered. Obviously, the achieved angle approximation accuracy strongly depends on the variation step of the possible panning index values. The selected step value (0.05) seems to provide acceptable accuracy (less than one degree).

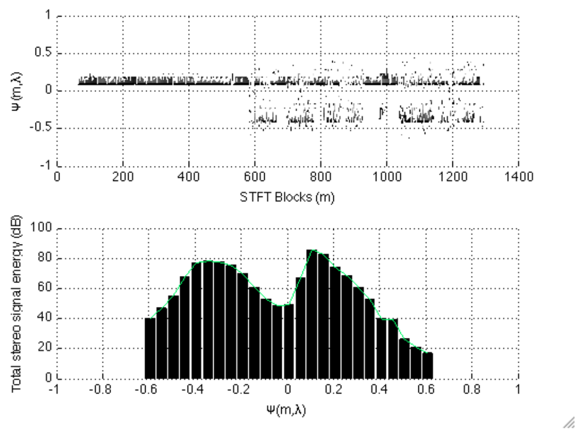

Fig. 4. Up: variation of the stereo signal energy as a function of the panning index $\Psi(m,\lambda)$ Down: average signal energy distribution as a function of $\Psi(m,\lambda)$.

Based on the extracted dominant panning index values, eq. (9) can be employed for calculating the unmixed $s_1(t)$ and $s_2(t)$ signals that were originally panned for producing the stereo input. Fig. 5 shows the corresponding waveforms. In eq. (9) ξ was selected equal to 0.006 and v = 0.

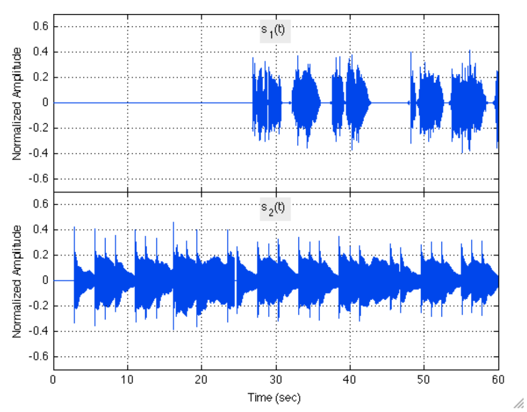

Fig. 5. Separated sound source signals $s_1(t)$ and $s_2(t)$.

Finally, the signals $s_1(t)$ and $s_2(t)$ were applied to the binaural synthesis subsystem. The virtual image angles were derived by eq. (10) for θ_b=45^0 and 90^0. Both binaural outputs were used for performing three different listening tests. During the first one, playback was performed through a pair of headphones. In the second test case, two high-quality active monitor speakers were employed, placed as explained in Section 3.1, while, in the third one, playback was performed through a mobile smartphone equipped with stereo loudspeakers. The tests were performed within a common office environment, since our purpose was to evaluate the performance of the proposed algorithm in everyday conditions. However, playback gains in all three tests were appropriately adjusted in order to obtain sound level differences less than 0.5dB, as larger deviations could bias the listening test results.

A total of 15 listeners were selected for participating in the tests. All of them were familiar with objective listening measurements; however, only 30% of them had participated in 3D sound tests before. The listeners had to firstly indicate the perceptually wider spatial soundstage. The processed results are shown in Table I showing a clear advantage of the proposed algorithm.

TABLE I
WIDER SPATIAL SOUND PERCEPTION

	Stereo	Binaural (θ_b=45^0)	Binaural (θ_b=90^0)
Headphones	4%	23%	73%
Loudspeakers	15%	21%	64%
Mobile	1%	20%	79%

Moreover, at a second test phase, the listeners were asked to evaluate the degree of the panning similarity between the stereo spatial soundstage and the two binaural outputs, under all playback types considered. It should be noted that this similarity was defined relatively, i.e. within the specific sound stage perceived in each stereo/binaural signal playbacks. The measurement was performed using a 5 scale integer score (1 to 5), where 1 denotes that no similarity exists in terms of the perceived panning positions and 5 that the two compared signals have identical panning positions within each perceived sound stage. Table II summarizes the obtained panning similarity scores for all playback cases under test. Clearly, the panning similarity is high in all test cases, with the best case being the reproduction through headphones, a fact that can be explained due to the employment of CTC and the impact of the playback room acoustics when employing the other two reproduction types.

TABLE II
PANNING SIMILARITY SCORES (REFERENCE: STEREO SIGNAL)

	1	2	3	4	5
Headphones (45^0)	-	-	6%	21%	73%
Headphones (90^0)	-	-	3%	12%	85%
Loudspeakers (45^0)	-	5%	14%	19%	62%
Loudspeakers(90^0)	-	3%	13%	17%	67%
Mobile (45^0)	-	-	11%	23%	66%
Mobile (90^0)	-	-	9%	17%	74%

The proposed algorithm efficiency was additionally measured in the case of a complete rock stereo recording. Two electric guitars, one bass, drums and vocals' channel tracks were panned in stereo. In this case, although a wider sound stage was perceived again, smaller panning similarity scores were measured (nearly 10% lower in high score categories), probably due to the high similarity of the two electric guitars in the frequency domain that resulted into a perceptual "fuzzy" placement of these sources. It should be noted however that these tests were not systematic and the number of participants evaluated the test signals was limited; hence no further detailed results are presented here.

5. CONCLUSIONS

Stereo signal spatial enhancement represents an approach that can be particularly significant in specific playback cases: short-distant loudspeaker pairs (such as those installed on portable devices and smartphones of limited physical dimensions) or even personalized audio reproduction through headphones may significantly benefit from the targeted sound stage widening, allowing an improvement of the overall user immersion in the reproduced stereo audio environment.

While a number of stereo expansion techniques have already been published in the literature, and various commercial products have been reported and evaluated [20], in this work we attempt to realize a technique that maintains the original panning information within the extended stereo sound stage. Panning information retrieval is a task that is performed based on an existing time-frequency cross-channel metric, calculated for the purposes of the current work in 1/3rd octave bands. This allows a more simplified and fast definition of the original stereo panning coefficients, without any accuracy losses. Sound source separation is also performed, taking into account the identified panning locations of the sound sources. The resulting audio streams are finally mapped into appropriately placed virtual sound sources. This spatial mapping considers both the original panning coefficients, as well as the targeted sound stage. The final stereo-expanded audio stream is synthesized using the well-known binaural synthesis technique.

The efficiency of the proposed stereo enhancement algorithm was evaluated through a sequence of listening tests that considered three playback scenarios defined by the potential employment of the technique in everyday life: through headphones, common stereo loudspeaker pairs and short-distant loudspeaker setups incorporated in mobile platforms (i.e. smartphones). In all test cases the subjective evaluation resulted into perceptually wider sound fields, while retaining the original panning nature of the initial stereo recording.

The proposed technique is not designed as a real-time processing algorithm. Instead, it can be applied in a way transparent to the potential user, i.e. during the transfer of the audio material to a specific (portable) sound device. Some performance assessment questions that should be investigated using a more thorough subjective evaluation process are related to the case of sound sources with similar frequency content but different panning positions. This is an open issue that should be considered in the future, probably by tracking the existence of these sources as a function of time.

12. REFERENCES

[1] R. Maher, E. Lindemann and J. Barish, "Old and new techniques for artificial stereophonic image enhancement", Audio Eng. Soc. 101st Convention, Los Angeles, California, preprint 4371, Nov. 8 – 11, 1996.

[2] H. Møller, "Fundamentals of binaural technology", *Applied Acoustics*, Vol. 36, No. 3 – 4, pp.171 – 218, 1992.

[3] J. Blauert, "Spatial Hearing: The psychophysics of human sound localization", Revised edition, Cambridge, Massachusetts, The MIT Press, 1997.

[4] V. Pulkki, "Virtual Sound Source Positioning Using Vector Base Amplitude Panning", *J. Audio Eng. Soc.*, Vol. 45, No. 6, pp. 456 – 466, 1997.

[5] J. M. Jot, S. Wardle, and V. Larcher, "Approaches to Binaural Synthesis", presented at the Audio Eng. Soc. 105th Convention, preprint 4861, September 1998.

[6] Ch. Tsakostas and A. Floros, "Optimized Binaural Modeling for Immersive Audio Applications", Audio Eng. Soc. 122nd Convention, Vienna, 2007.

[7] A. Härmä, M. Tikander, M. Karjalainen, T. Lokki, H. Nironen, S. Vesa, "Techniques and applications of wearable augmented reality audio", Audio Eng. Soc. 114th Convention, Amsterdam, The Netherlands, 2003.

[8] M. Tikander, and V. Riikonen, "Sound quality of an augmented reality audio headset", In. Proc. of the 8th Intl. Conference on Digital Audio Effects (DAFx'05), Madrid, Spain, 2005.

[9] J Breebaart, J. Herre, L. Villemoes, C. Jin, K. Kjorling, J. Plogsties and J. Koppens, "Multichannel Goes Mobile: MPEG Surround Binaural Rendering", Audio Eng. Soc. 29th International Conference, Seoul, Korea, 2006.

[10] S. Olive, "Evaluation of Five Commercial Stereo Enhancement 3D Audio Software Plug-ins", Audio Eng. Soc. 110th Convention, Amsterdam, The Netherlands, 2001.

[11] M. V. Thomas, "Improving the stereo headphone sound image", *J. of the Audio Eng. Soc.*, Vol. 25, No. 7/8, pp. 474 – 478, 1977.

[12] G. S. Kendall, W. L. Martens, and M. D. Wilde, "A spatial sound processor for loud- speaker and headphone reproduction", Audio Eng. Soc. 8th International Conference, pages 209–221, Washington, USA, May 3-6 1990.

[13] Ch. Tsakostas, A. Floros and Y. Deliyiannis "Binaural Rendering for Enhanced 3D Audio Perception", In Proc. of the AudioMostly 2nd Conference on Interaction with Sound, Fraunhofer Institute for Digital Media Technology IDMT, Ilmenau – Germany, Sept. 27 – 29, pp. 125 – 128, 2007.

[14] P. Minnaar, "Enhancing music with virtual sound sources", *The Hearing Journal*, Vol. 63, No. 9, pp. 38 – 43, Sept. 2010.

[15] M. Akeroyd, J. Chambers, D. Bullock, A. Palmer, A. Summerfield, P. Nelson, and S. Gatehouse, "The binaural performance of a cross-talk cancellation system with matched or mismatched setup and playback acoustics", *J. Acoustic Soc. Am.*, Vol. 121, No. 2, pp. 1056-1069, February 2007.

[16] C. Avendano, "Frequency-domain source identification and manipulation in stereo mixes for enhancement, suppression and re-panning applications", In Proc. of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp. 55 – 58, 2003.

[17] G. Tzanetakis, L. Martins, K. McNally and R. Jones, "Stereo Panning Information for Music Information Retrieval Tasks", *J. Audio Eng. Soc.*, Vol. 58, No. 2, pp. 409 – 417, May 2010.

[18] C. Avendano and J. Jot, "Frequency domain techniques for stereo to multichannel upmix", In Proc. of the Audio Eng. Soc. International Conference on Virtual, Synthetic and Entertainment Audio, Espoo, Finland, June 2002.

[19] Ch. Tsakostas and A. Floros, "Optimized Binaural Modeling for Immersive Audio Applications", Audio Eng. Soc. 122th Convention, Vienna, May 2007.

[20] S. Olive, "Evaluation of five commercial stereo enhancement 3D audio software plug-ins", Audio Eng. Soc. 110th Convention, Amsterdam, May 2001.