Binaural Rendering for Enhanced 3D Audio Perception

Christos Tsakostas[1], Andreas Floros[2] and Yiannis Deliyiannis[2]

[1]HOLISTIKS Engineering Systems, Digeni Akrita 29, 122 43 Athens, Greece
tsakostas@holistiks.com

[2]Dept. of Audiovisual Arts, Ionian University, 49 100 Corfu, Greece
{floros, yiannis}@ionio.gr

Abstract. Despite the recent advantages on multichannel audio coding and playback technologies, 3D audio positioning is frequently required to be performed over legacy stereo loudspeaker setups, due to certain application limitations and restrictions mainly presented in portable and low cost systems. In this work, a 3D audio platform is introduced which allows the real-time conversion of any stereo audio signal to a high-quality immersive audio stream using binaural technology. Due to a number of advanced binaural processing algorithms and features, the proposed conversion platform demonstrates perceptually improved 3D audio spatialization, rendering it suitable for implementing high-quality immersive audio applications.

1. Introduction

Two-channel stereophony has been the basic representative of audio reproduction systems for more than half a century. Today, the advent of high-resolution storage formats (such as Super Audio CD [1], DVD-Video/Audio BluRay and HD-DVD) has significantly boosted the proliferation of many surround sound coding schemes used for audio-only, movies, home theatre, virtual reality and gaming applications. However, despite the advantages in multichannel sound, a number of parameters have to be considered for developing high-quality spatial audio representation: a) the volume of the existing stereo recordings and stereo reproduction systems, including TV sets, CD players, laptop computers and mobile phones b) the large number of loudspeakers with specific placement specifications that are usually required for multichannel reproduction (e.g. for 5.1, ambisonics or wavefield synthesis setups) c) the frequently complicated loudspeakers' cable interconnections and d) the decreased application portability, due to the receiver position limitations imposed by the multichannel coding schemes.

Today, the advent of many portable devices (including small audio/video players, Personal Digital Assistants – PDAs and mobile phones and computers) usually equipped with ordinary stereo loudspeakers necessitates the development of low-cost, low-power algorithms for 3D audio immersion using legacy stereo setups. Towards this perspective, binaural technology [2] represents a very attractive format, as it allows accurate 3D audio spatialization by synthesizing a two-channel audio signal using the well-known Head Related Transfer Functions (HRTFs) between the sound source and each listener's human ear [3], [4]. Hence, only two loudspeakers or headphones are required for binaural audio playback.

Binaural technology was recently employed for parametric MPEG surround coding [5], where the spatial information is extracted from multichannel audio and a downmix 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. In this work, a novel real-time 3D audio enhancement platform is presented which converts any stereo audio material to a high-quality immersive audio stream using binaural rendering. The proposed platform incorporates novel HRTF equalization methods that significantly improve the spatial position perception of the active sound sources compared to previously reported methods [6]. Efficient crosstalk cancellation algorithms are also incorporated for stereo loudspeaker support, with a large number of FIR coefficients (2048 at 44.1kHz sampling frequency). Moreover, Sound Field Models are supported, allowing the definition of fully customizable virtual auditory environments or the selection of predefined virtual world templates. As it will be described later, an unlimited number of virtual sound sources can dynamically be linked to the channels of the stereo input, allowing accurate and fully parameterized 3D audio positioning.

The rest of this paper is organized as following: In Section 2, an overview of the binaural technology is provided, followed by the brief description of the proposed algorithm for 3D enhancement of stereo audio signals presented in Section 3. In Section 4, a typical implementation of the 3D audio enhancement technique is provided. Finally, Section 5 concludes this work.

2. Binaural technology overview

It is well known that using binaural technology the accurate spatial placement of any virtual sound source is achieved by filtering monaural recorded (or synthesized) sound with appropriately selected Head Related Transfer Function (HRTFs) [7]. In general, the latter functions describe the paths between a sound source and each ear of a human listener in terms of 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 different propagation path lengths, as well as the shadowing effect of the human head. Both ITD and ILD sound localization cues result into two different sound waveforms arriving to the human ears, allowing the perception of the direction of any active sound source.

When using binaural technology, the above basic localization cues are incorporated into HRTFs, which represent directional-dependent transfer functions between the human listener's ear canal and the specific sound source placement [8]. Hence, convolving the mono sound source wave with the appropriately selected pair of HRTFs produces the sound waves that correspond to each of the listener's ears. This process is called

binaural synthesis. Binaural synthesis can be also combined with existing sound field models producing binaural room simulations and modelling. This method facilitates listening into spaces that only exist in the form of computer models. In more detail, the sound field models can output the exact spatial-temporal characteristics of the reflections in a space. In this case, the summation of binaural synthesis applied to each reflection produces the Binaural Room Impulse Response. Finally, the binaural left and right signals are reproduced using headphones or a pair of conventional stereo loudspeakers. In the latter case, the additional undesired crosstalk paths that transit the head from each speaker to the opposite ear must be cancelled using crosstalk cancellation techniques [9].

3. 3D Audio spatialization using binaural rendering

By employing the above binaural synthesis approach, one can create virtual sound sources placed around a listener within an open or closed space. This is the well-known concept of binaural rendering, which allows the creation of 3D virtual sound environments. More specifically, as explained previously, the convolution of the original audio signal with a pair of HRTFs for each sound source and the final mixing of the resulting binaural signals produces the desired 3D audio perception.

In order to achieve 3D audio enhancement of typical stereo material, the stereo audio input must be mapped to a number of virtual sound sources appropriately placed into the desired virtual world. Figure 1 illustrates this mapping procedure. Prior to mapping, the input stereo signal is pre-processed in order to produce the necessary audio streams that will be mapped to the selected virtual sound sources. Apart from the profound audio streams of the left and right channels, more audio streams can be acquired such as the summation or the difference of them.

Each audio stream is then mapped to an arbitrary number of virtual sound sources. The benefit of this approach is that different aspects of the audio signal can be freely manipulated, while any spatial effect can be theoretically achieved. For example a low passed version of the (L+R)/2 audio stream can be mapped to a virtual sound source mimicking a low frequency playback unit (subwoofer).

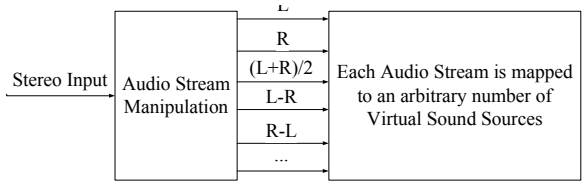

Figure 1: Stereo audio signal to virtual sound sources mapping

After audio stream to virtual sound source mapping, the binaural impulse responses are calculated for both left and right ears, and the final binaural signal is derived by convolving them with the derived audio streams as described in the previous Section. Obviously, in the case of a closed virtual world with specific geometry and material properties, the binaural room impulse response are calculated and employed for deriving the final binaural signal.

4. Implementation and performance evaluation

Using the proposed 3D audio enhancement platform, all the required binaural processing of the original stereo material is performed using the Amphiotik Technology framework developed by the authors and presented in detail in [10]. The benefit of this approach is that the Amphiotik Technology core (namely the Amphiotik 3D Audio Engine) offers the capability of rapid 3D-Audio applications development, yet preserving a carefully designed balance between authenticity and real-time operations / calculations. More specifically, the Amphiotik 3D Audio Engine incorporates state-of-the-art binaural processing algorithms, such as a novel algorithm for HRTFs equalization, cross-talk cancellation techniques and room acoustics modeling for accurate acoustic representation of virtual environments. The Amphiotik 3D Audio Engine state-machine is also responsible for the signal routing that must be performed in order to perform all calculations required for producing the binaural signal within the defined virtual world.

The communication of the 3D audio enhancement platform with the Amphiotik 3D audio engine is performed using the Amphiotik API, which provides easy to use software methods for defining the binaural model and the virtual world parameters in real-time. Figure 2 illustrates the Amphiotik Technology architecture employed. The Amphiotik API provides the necessary functions for the definition of the overall virtual auditory environment, that is: (a) the geometry and materials of the virtual world, (b) the sound field model and (c) the virtual sound sources and receivers characteristics and instantaneous position. In addition, it provides functions that interact directly with the 3D-Audio engine to parameterize various aspects of the engine such as the HRTF data set to be used, the cross-talk cancellation algorithm activation, the headphones equalization as well as user-defined parametric frequency equalization.

The shape of the virtual world can be arbitrary, but for the shake of real-time processing in moderate power computers, “shoebox” like spaces are better supported. For this case, the API provides simple functions for defining the dimensions of the room (length, width and height) and the materials (absorption coefficients) of each surface. An internal materials database is utilized for the re-use of the above materials parameters.

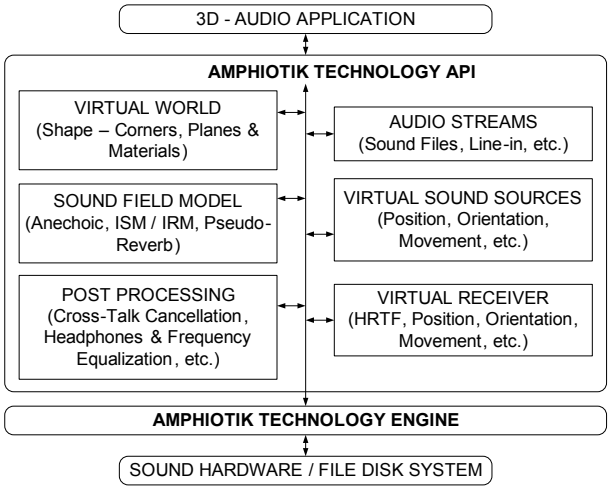

Figure 2: Architecture of the Amphiotik Technology framework

The API also allows the definition of one virtual binaural receiver, while there are no limitations on the number of the defined sound sources. Both the virtual receiver and sound sources can be placed arbitrarily and in real-time in the 3D virtual auditory environment, concerning both their position and orientation. An arbitrary number of audio streams can be defined, which are linked to the virtual sound sources. An audio stream can be associated with more than one virtual sound sources. Audio streams are usually sound files on the local disk system but they can also originate from the soundcard’s line-in input or even an Internet media file link.

Using the Amphiotik 3D audio engine, the acoustical environment modeling can be performed using one of the following sound field models: (a) anechoic, (b) early part, and (c) early part & pseudo-reverb. For the anechoic case reflections are not considered, consequently the geometry of the room is ignored. On the other hand, the early part is simulated by means of the "Image Source Method" and "Image Receiver Method" [11]. The order of the reflections can be altered in real-time and its maximum value is limited to five. Early part & pseudo-reverb uses a hybrid algorithm in which the early part is estimated as described earlier and the reverberation part (i.e. the late part) of the room impulse response is estimated with digital audio signal processing algorithms. Specifically, two reverberation algorithms are currently supported: (a) Schroeder [12] and (b) Moorer [13]. According to the Schroeder algorithm the late part is calculated by the means of comb-filters and all-pass filters, whilst for the case of the Moorer technique the late part is approximated with an exponentially decaying white noise. The reverberation time is calculated using the Sabine equation [14]. A proprietary algorithm has been employed in order to combine the binaural early part with the monaural late part.

Figure 3 depicts a general overview of the Amphiotik 3D-Audio Engine. For each pair of virtual receiver and sound source, a binaural IR is calculated, taking under consideration their instantaneous positions, orientation and room geometry and materials as well, if the early part option is enabled. Real-time convolution is accomplished by the means of un-partitioned and partitioned overlap-and-add algorithms. Each convolution produces two channels: Left (L) and Right (R). All the L and R channels, produced for each virtual sound source, are summed up producing finally only two channels.

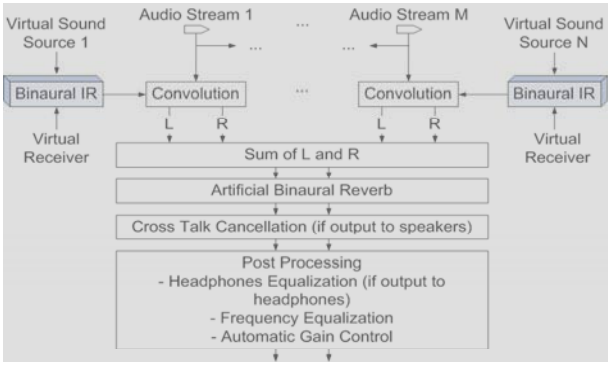

Figure 3: Amphiotik 3D audio engine general structure

Crosstalk cancellation (CTC) is applied if the audio playback is performed over stereo loudspeakers. CTC filters are built, in real-time, with HRTFs. The Amphiotik API gives the capability to select a different set of HRTFs for the crosstalk cancellation processing than the one used for spatialization. In general, CTC filters built with HRTFs impose the problem that the low-power low frequencies are excessively amplified and vice-versa. Two strategies have been used in order to overcome this problem: (a) the employment of band-limited CTC and (b) a special equalization method called "Transaural Equalization". Band-limited CTC simply partially overcome the above problem by not using the very low and the very high frequencies. The outcome is that musicality becomes significantly better, and at the same time the loss of the very low and the very high frequencies is not particularly perceptible. Transaural equalization on the other hand, is based on post-filtering of the transaural audio channels, in order to approximate the magnitude that they would have if listening over headphones was selected. In addition, the software gives the capability to apply the crosstalk cancellation filters directly to the stereo input without prior processing through the 3D-Audio engine. It is also important to note that the Amphiotik API gives the capability to select non-symmetric loudspeakers positions (e.g. for loudspeakers in cars).

Additionally, as it shown in Figure 3, post-equalization is applied to the synthesized binaural signal, which may optionally include headphones equalization, user-defined frequency equalization and Automatic Gain Control (AGC). Pre-equalization is also supported for the stereo audio signals before they are spatialized.

Concerning the sound motion simulation, a time-varying filtering method is employed that minimizes the need of additional computations for accurate moving sound-sources representation. This mechanism additionally takes into account psychoacoustic criteria and cues for perceptually optimizing the 3D audio representation performance [15].

Finally, for effective real-time operation and interaction with the user, the Amphiotik engine checks for any possible change of the parameters in time frames, which are defined by the block length used (typically 512 - 8192 samples at a sampling rate equal to 44.1 KHz) and re-initializes all the appropriate modules.

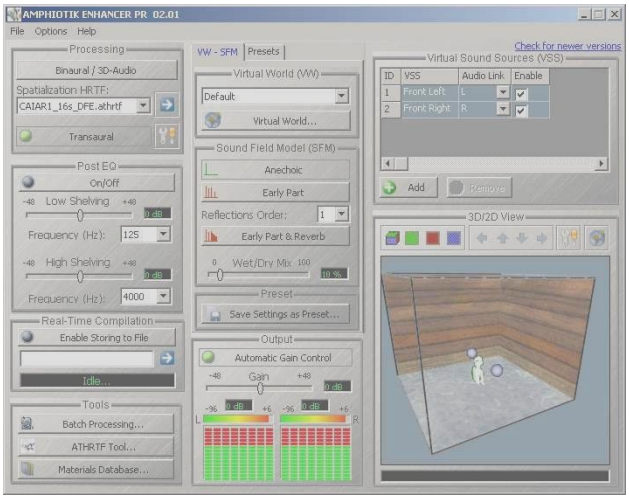

Figure 4: The Amphiotik Enhancer module

In order to demonstrate the capabilities and the performance of the 3D audio enhancement platform, the Amphiotik Enhancer plug-in was developed during this work, which uses the Amphiotik API and engine (Figure 3). As mentioned previously, its main purpose is to enhance the standard stereo audio signals, offering a much more pleasant 3D acoustical experience. The Amphiotik Enhancer module is compatible and can be plugged in several popular audio hosts (like Winamp, Windows Media Player, VST Hosts, and DirectX).

The Amphiotik Enhancer incorporates all the API features that were described previously, through a graphical user interface (GUI). Typical user interaction procedures include manipulation and definition of the audio streams, definition and placement of the virtual sound sources and the virtual binaural receiver, the analytic description of the desired sound field models, as well as the selection of the HRTF library (as shown in Figure 5). Cross-talk cancellation or headphones playback options are also available while frequency equalization can be additionally selected. It is important to note that the GUI supports 3-Dimensional view of the virtual world, for better user-perception of the intended simulations and the final binaural playback.

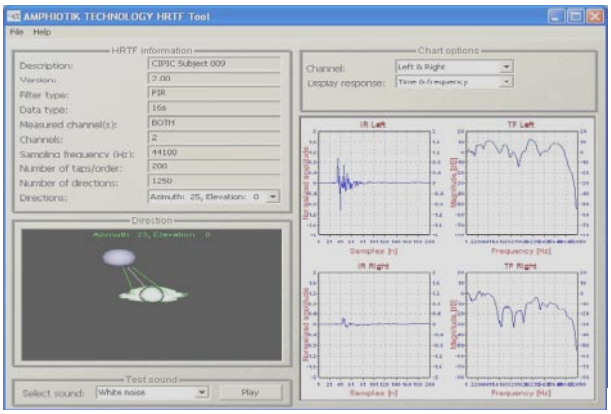

Figure 5: HRTF library selection GUI

The perceptual performance of the proposed 3D audio enhancement platform was evaluated through a number of subjective audio tests. During these tests, the Amphiotik Enhancer module was used for producing enhanced 3D audio material from typical stereo (CD-quality) recordings. The responses of the experienced audience that participated into these tests have shown a significant improvement of the perceived audio immersive impression. Moreover, from these tests it was found that for real-time operation and for sound field models activated, up to 8 virtual sources in average can be defined, allowing the high-quality and flexible stereo to virtual sound source mapping.

5. Conclusions

The advent of many multichannel audio coding and playback formats and schemes is nowadays leading towards the surround sound prospect. However, due to the large volume of existing legacy stereo audio material as well as the usage of electronic / consumer playback devices equipped with stereo loudspeakers, the employment of appropriate signal processing algorithms for 3D enhancement of stereo audio signals is now more demanding.

In this work, a novel 3D audio enhancement platform is presented which converts any stereo audio material to a high-quality immersive audio stream. The conversion is performed using binaural rendering, which allows the creation and accurate representation of various forms of virtual auditory worlds. The proposed platform may be employed for the development of both software plug-ins as well as for hardware-based applications, using an Application Protocol Interface (API) available. Here, it is shown that, in terms of 3D audio perception, the proposed platform achieves high degree of authenticity, due to a number of optimized algorithms, such as a novel HRTF equalization scheme. Moreover, the 3D enhancement platform is optimized for real-time operation, allowing its employment to consumer devices, game engines etc.

Future research work will consider in more detail moving virtual sound sources, taking into account additional psychoacoustic parameters and cues that are required for authentic sound reproduction, such as the Doppler effect.

References

[1] E. Janssen and D. Reefman, *"Super-Audio CD: an introduction",* IEEE Signal Processing Mag., Vol. 20, No. 4, pp. 83–90 (2003)

[2] H. Møller, "Fundamentals of Binaural Technology", Appl. Acoustics, Vol. 36, pp. 171 – 218, (1992).

[3] H. Møller, M. F. Sørensen, D. Hammershøi, C. B. Jensen, *"Head-related transfer functions of human subjects",* J. Audio Eng. Soc., Vol. 43, No. 5, pp. 300-321 (1995)

[4] E. Wenzel, M. Arruda, D. Kistler and F. Wightman, *"Localization using nonindividualized head-related transfer-functions",* J. Acoust. Soc. Am., Vol. 94, pp. 111-123 (1993)

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

[6] S. Olive, *"Evaluation of five commercial stereo enhancement 3D audio software plug-ins"*, Presented at the 110th Convention of the Audio Engineering Society, Amsterdam, The Netherlands, Preprint 5386 (2001)

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

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

[9] A. B. Ward, G. W. Elko, *"Effect of loudspeaker position on the robustness of acoustic crosstalk cancellation"*, IEEE Signal Processing Letters, Vol. 6, No. 5, pp. 106-108 (1999)

[10] Ch. Tsakostas and A. Floros, "Optimized Binaural Modeling for Immersive Audio Applications", Presented at the 122nd Convention of the Audio Engineering Society, Vienna, Preprint 7100 (2007)

[11] C. Tsakostas, *"Image Receiver Model: An efficient variation of the Image Source Model for the case of multiple sound sources and a single receiver"* presented at the HELINA Conference, Thessaloniki Greece (2004)

[12] M. R. Schroeder, *"Natural Sounding Artificial Reverberation"*, J. Audio Engineering Society, Vol. 10, No 3, (1962)

[13] J. A. Moorer, *"About This Reverberation Business"*, Computer Music Journal, Vol. 3, No 2. (1979)

[14] W. C. Sabine, *"Reverberation"* originally published in 1900. Reprinted in Acoustics: Historical and Philosophical Development, edited by R. B. Lindsay. Dowden, Hutchinson, and Ross, Stroudsburg, PA, (1972)

[15] Ch. Tsakostas and A. Floros, "Real-time Spatial Representation of Moving Sound Sources", to be presented at the Audio Engineering Society 123rd Convention, New York, (2007)