Real-time Spatial Mixing Using Binaural Processing

Christos Tsakostas[1], Andreas Floros[2] and Yannis Deliyiannis[3]
[1] Holistiks Engineering Systems, Athens, Greece, tsakostas@holistiks.com
[2] Dept. of Audiovisual Arts, Ionian University, Corfu, Greece, floros@ionio.gr
[3] Dept. of Audiovisual Arts, Ionian University, Corfu, Greece, yiannis@ionio.gr

***Abstract* — In this work, a professional audio mastering / mixing software platform is presented which employs state-of-the-art binaural technology algorithms for efficient and accurate sound source positioning. The proposed mixing platform supports high-quality audio (typically 96kHz/24bit) for an arbitrary number of sound sources, while room acoustic analysis and simulation models are also incorporated. All binaural calculations and audio signal processing/mixing are performed in real-time, due to the employment of an optimized binaural 3D Audio Engine developed by the authors. Moreover, all user operations are performed through a user-friendly graphical interface allowing the efficient control of a large number of binaural mixing parameters. It is shown that the proposed mixing platform achieves subjectively high spatial impression, rendering it suitable for high-quality professional audio applications.**

I. INTRODUCTION

In computer music synthesis and production, it is a very common approach that the accurate spatial positioning of the sound sources is performed by multiple loudspeaker systems. Amplitude panning represents the most frequently used technique for positioning the sound sources using such multiple speaker setups and nearly all audio mixing devices offer controls for manipulating the level of a specific sound source. In two-dimensional setups (where all loudspeakers are positioned in the same plane with the listener), panning is usually performed using pair-wise methods [1]. For the simple stereophonic playback case, a number of panning laws have been proposed [2] that result into the perception of a virtual sound source between the loudspeakers.

In all panning cases, the common problem is that the loudspeakers are usually placed in different positions inside the playback enclosure. However, an ideal panning system should be capable of creating identical spatial auditory scenes using any loudspeaker configuration. Towards this aim, enhanced panning and surround sound techniques have been proposed in the literature (such as the Vector-base amplitude panning – VBAP [3] and the Ambisonics matrixing sound reproduction technique [4]), which render the sound source positioning independent from the number of the loudspeakers employed for playback.

Lately, there has been a significant proliferation of three-dimensional (3D) audio technologies intended mainly for multimedia and portable consumer electronics applications [5]. Binaural source localization [6] represents a highly accurate technique for achieving 3D audio environment recreation by synthesizing a two-channel audio signal using the well-known Head Related Transfer Functions (HRTFs) [7] between the sound source and each listener's human ear. Hence, only two loudspeakers or headphones are required for binaural audio playback. The simple setup of a binaural reproduction system renders it convenient for a number of state-of-the-art applications, including mobile applications and communications, especially when headphones are employed.

In this work, a spatial mixing and audio mastering application (termed as Amphiotik Synthesis) is presented, which employs binaural technology for effectively producing 3D audio recordings. Based on a powerful binaural audio engine recently developed by the authors [8], the Amphiotik Synthesis application allows the real-time production of high-quality (24bit/96kHz) binaural signals, for a large number of virtual sources. Moreover, as it will be presented in the following paragraphs, transaural playback, room acoustics modeling and an enhanced HRTFs equalization algorithm is supported for efficiently positioning the active sound sources within typical enclosures.

The rest of the paper is organized as following: Section II presents a general overview of binaural technology while in Section III, the Amphiotik Synthesis application is described and in Section IV typical test cases are presented that demonstrate its usage and effectiveness. Finally, Section V concludes this work.

II. BINAURAL HEARING OVERVIEW

Binaural hearing is based on two basic 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. Both sound localization cues result into the reception of two different sound waves by the human ears that perceptually provide information on the direction of an active sound source [9].

In binaural modeling, the effect of the above basic cues is incorporated into directional-dependent HRTFs: convolving the mono sound source wave with the appropriate pair of HRTFs derives the sound waves that correspond to each of the listener's ears. The binaural left and right signals can be reproduced directly using headphones or a pair of conventional 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 cross-talk cancellation techniques [10]. The above binaural synthesis process can be also combined with existing sound field models producing binaural room simulations and modeling. In more detail, the above 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.

One major problem of binaural hearing is that for high spatial localization accuracy, the HRTFs must be measured for the targeted human head and for different azimuth and elevation angles. In order to overcome this problem, pre-measured HRTF sets are usually employed [11], [12]. However, this significantly reduces the spatial sound impression. Moreover, the recorded HRTFs measurements usually contain the frequency response of the measurement system. Hence, the HRTF measurements must be equalized. A number of HRTFs equalization techniques have been published [13], providing a reference response, which is then inverted and used to filter the entire data set.

III. Amphiotik Synthesis Application

Following the above description it is clear that binaural processing introduces high computational load, which proportionally increases with the number of the active sound sources and especially when binaural room simulations are additionally considered. The above high computational load usually represents the major reason for non real-time binaural processing implementations.

As mentioned previously, in this work, the Amphiotik Synthesis application real-time processing capabilities were achieved due the employment of a 3D binaural audio engine developed recently by the authors [8]. More specifically, the Amphiotik Technology is a software library, which has been designed in a manner that carefully balances authenticity for the listeners and real-time operations on typical PCs. It consists of two main modules: (a) The Amphiotik API, available for 3D-Audio applications development and (b) the Amphiotik 3D-Audio Engine, which is responsible for the signal routing, incorporating several public and custom binaural and signal processing algorithms in the processing chain. One of the most important advantages of the above engine is that almost every parameter of the binaural model is available through the API and can be accessed by the user or software developer in real-time.

Based on the Amphiotik Technology, the Amphiotik Synthesis application is a tool that supports a very convenient way for mixing audio streams, for the cases that binaural or even just standard stereo output is desired. It employs the Amphiotik Technology API mentioned previously, through a graphical user interface (GUI), in order to manipulate several parameters such as: audio streams, virtual sound sources, virtual binaural receiver, sound field models, HRTF, crosstalk cancellation algorithms, headphones and frequency equalization techniques, post equalization and automatic gain control. The GUI for accessing all the above parameters is shown in Fig. 1.

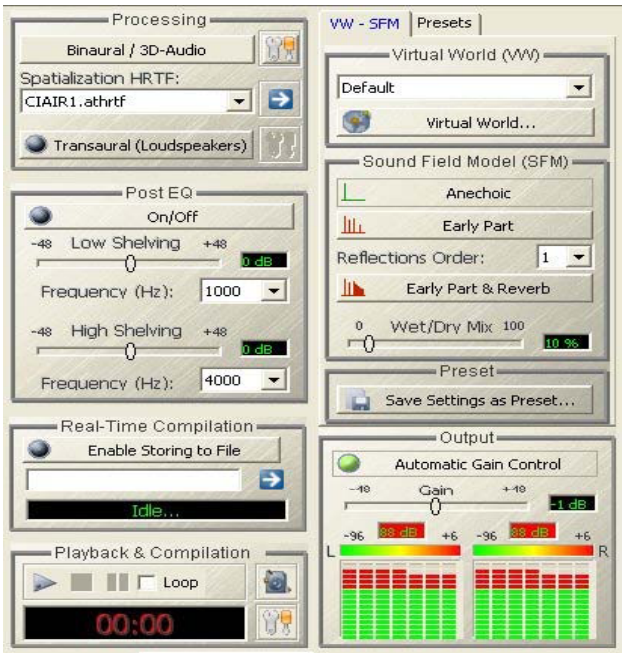

Fig. 1. Amphiotik Synthesis GUI for manipulating and controlling the basic mixing parameters

Using the Amphiotik Synthesis application it is assumed that both the sound sources and the receivers are positioned within a virtual world, which may be opened or closed. The virtual world acoustical 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, while the early part is simulated by means of the "Image Source Method" and "Image Receiver Method" [14]. Additionally, early part & pseudo-reverb uses a hybrid algorithm in which the early part is estimated as described earlier and the reverberation part of the room impulse response is estimated with digital audio signal processing algorithms.

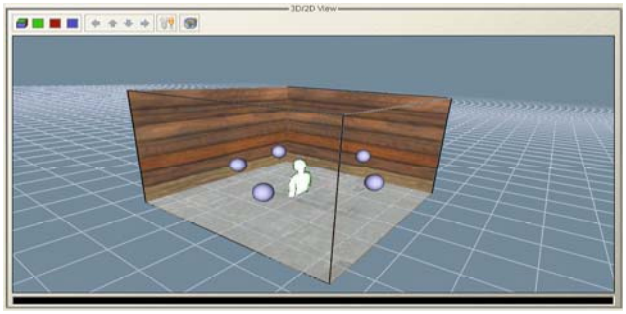

Fig. 2. 3-dimensional view of the virtual world

As it is shown in Fig. 2, the Amphiotik Synthesis GUI supports 3-dimensional (as well as 2D) view of this virtual world, for a better user-perception of the intended receiver and sound sources positioning. The shape of the virtual world can be arbitrary, but for the shake of real-time processing in moderate computers, "shoebox" like spaces are better supported. For this case, the GUI provides simple functions for defining the dimensions of the room (length, width and height) and the materials (absorption coefficients) of each surface (see Fig. 3). Additionally, an

internal materials database is utilized for the re-use of the user-defined materials parameters. The access to this database is performed through the room materials module illustrated in Fig. 4. The user can define the absorption coefficients of the materials for a wide range of frequencies, as well as their appearance during the visualization of the virtual world.

Fig. 3. Room parameterization module

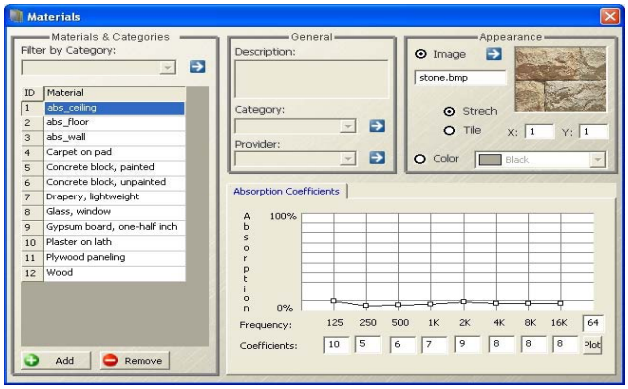

Fig. 4. The room materials module

Amphiotik Synthesis includes some extra features like A/B audio playback, and batch processing options. Moreover, for effective HRTF library manipulation the Amphiotik Technology HRTF module shown in Fig. 5 is also provided. This module represents a tool for efficiently monitoring the selected HRTF library.

In Fig. 6, the Amphiotik Synthesis cross-talk cancellation module is shown. As mentioned in Section II, cross-talk cancellation represents a required processing stage for listening to binaural 3D-audio over loudspeakers and is fully manipulated by defining both the physical location of the loudspeakers and the HRTF library to be used. For a better "musicality" of the processed audio, the Amphiotik Synthesis cross-talk cancellation module includes a custom algorithm recently developed by the authors that controls the "strength" of the required equalization processing. This equalization can be also performed in limited frequency bandwidth if desired. It should be also noted that non-symmetric loudspeaker positions are also supported by the Amphiotik Synthesis cross-talk cancellation algorithm.

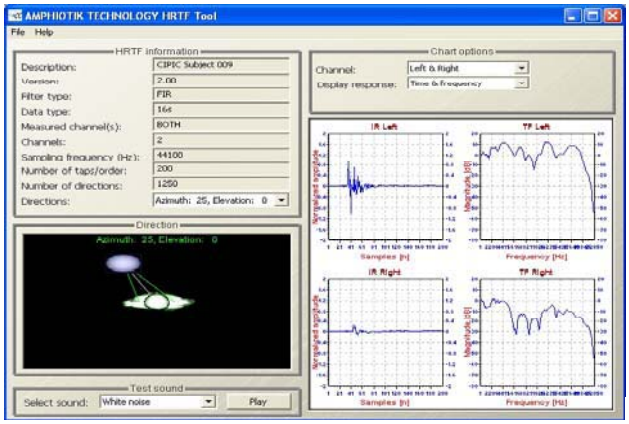

Fig. 5. Amphiotik Technology HRTF module

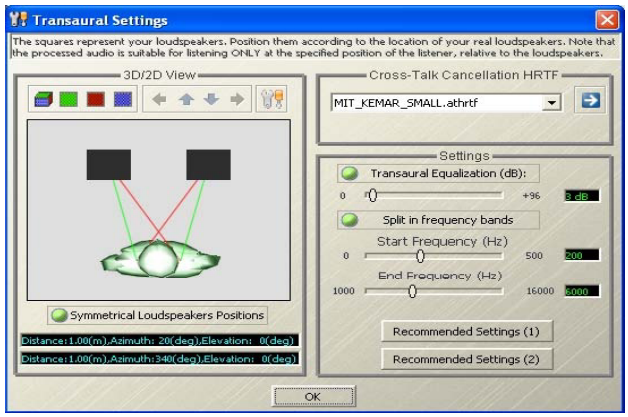

Fig. 6. The cross-talk cancellation module

IV. Typical Test Cases

Using the Amphiotik Synthesis application, all binaural calculations and processing/mixing processes can be performed in real-time, allowing the user / sound engineer to efficiently perform binaural mixing and monitoring concurrently. During this work and in order to demonstrate it's real-time capabilities, Amphiotik Synthesis was used for the production of binaural and standard stereo mixing for audio samples that were offered to the authors, including classical, jazz and rock music (see Table I). For the first three samples, binaural mixing for headphones, loudspeakers (for an angle of 20 degrees) and standard stereo were produced. For the last two, the effect of Amphiotik Synthesis as a mastering tool is indicated. Finally, after a sequence of thorough tests it was found that due to the employment of the Amphiotik 3D engine, up to 8 high-quality sound sources can be supported at 96kHz/24bit, covering most of the professional audio mastering and mixing requirements. Obviously, the above number of sound sources can be significantly increased for CD quality sound (44.1kHz/16bit).

TABLE I.
AUDIO SAMPLES AND NUMBER OF CHANNELS PER SAMPLE

Audio Sample - Gender	Number of Channels
Jazz	9
Rock	7
Rock Ballad	5
Classical	2
Milonga	1

Although no organized audience experiments took place for the subjective comparison of stereo versus binaural mixing, several listeners - musicians and audiophiles - were asked to describe their perception for each one of the samples and for each of the mixings. The obtained impressions can be briefly summarized as follows:

- For binaural mixing, the "out-of-head" effect is clearer when listening over headphones.
- When binaural reverberation is applied, the perceived auditory image is significantly "broader".
- In binaural mixing the position of the instruments is clearly perceived for the case of headphones, less clearly for the case of loudspeakers and even less for the case of standard stereo.
- In general, binaural mixing is more "pleasant" than the standard stereo.

Finally, it should be noted that in all the above test cases, it was found that the above perceptually assessed performance was further enhanced when playback was performed through headphones and a novel HRTF equalization algorithm described in [8] was employed.

V. CONCLUSIONS

In this work the Amphiotik Synthesis multichannel audio mixing application is presented which combines the well-known binaural technology and HRTF theory for creating and recording high-quality spatial audio signals in state-of-the-art virtual audio performing environments. The proposed Amphiotik Synthesis mixing application incorporates a number of novel and optimized algorithms and techniques for binaural signal processing, such as: a) a cross-talk cancellation method for non-symmetric placement and playback through stereo loudspeakers b) an HRTF equalization technique that significantly improves the spatial position perception of the active sound sources and c) synthesized binaural signal post-equalization algorithms which may optionally include headphones equalization, user-defined frequency equalization and Automatic Gain Control (AGC). In addition, pre-equalization is also supported for the stereo audio signals before they are spatialized.

All binaural calculations and processing/mixing are performed in real-time. Due to the efficient Amphiotik Engine implementation employed, up to 8 sound sources are currently supported at 96kHz/24bit, with concurrent room acoustics modeling and analysis. The above number of sound sources can be significantly increased for CD quality sound (44.1kHz/16bit), covering a wide range of professional audio mastering / mixing applications.

Some of the features that future versions of the Amphiotik Synthesis will include are the extension of the GUI for supporting moving sound sources and real-time processing of audio streams captured from analog or digital audio interfaces (such as a PCs sound card). Additionally, it is in the authors' near future intentions to integrate a motion tracking system with the Amphiotik Synthesis application, in order to allow automatic spatial mixing and high-quality audio mastering.

ACKNOWLEDGEMENTS

The authors wish to thank the following people who provided the audio samples for the typical test cases mentioned in Section IV:

- Spyridoula, rock group.
- Mr. Nikos Rokkos, sound engineer.
- Prof. Angelo Farina, Environmental Technical Physics, University of Parma, Italy.
- Mrs. Lito Christodoulopoulou, guitar performer.

REFERENCES

[1] J. Chowning, "The Simulation of Moving Sound Sources" *J. Audio Engineering Society*, Vol. 19, No. 1, 1971, pp. 2 – 6.

[2] J. Bennett, K. Barker anf F. Edeko, "A New Approach to the Assessment of Stereophonic Sound System Performance", *J. Audio Engineering Society*, Vol. 33, No. 5, May 1985, pp. 314 – 321.

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

[4] M. Gerzon, "Periphony: With-height Sound Reproduction", *J. Audio Engineering Society*, Vol. 21, No. 1, pp. 2 – 10, 1972.

[5] AES Staff Technical Writer, "Binaural Technology for Mobile Applications", *J. Audio Engineering Society*, Vol. 54, No. 10, Oct. 2006, pp. 990 – 995.

[6] H. Viste and G. Evangelista, "Binaural Source Localization", in *Proc. of the 7th Int. Conference on Digital Audio Effects (DAFx'04)*, Naples, Oct. 2004, pp. 145 – 150.

[7] H. Moller, M. Sorensen, D. Hammershoi and C. Jensen, "Head-Related Transfer Functions of Human Subjects", *J. Audio Engineering Society*, Vol. 43, No. 5, May 1995, pp. 300 – 321.

[8] Ch. Tsakostas and A. Floros, "Optimized Binaural Modeling for Immersive Audio Applications", to be presented at the *Audio Eng. Doc. 122nd Convention*, May 2007 (accepted).

[9] V. Pulkki, "Compensating Displacement of amplitude-panned Virtual sources", presented at the *AES 22nd International Conference on Virtual, Synthetic and Entertainment Audio*, 2002 Espoo, Finland, pp. 186-195.

[10] 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, May 1999, pp. 106-108.

[11] http://sound.media.mit.edu/KEMAR.html

[12] V. Algazi, R. Duda, D. Thompson and C. Avendano, "The CIPIC HRTF database", in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New York, Oct. 2001, pp. 99 – 102.

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

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