Audio Engineering Society

Convention e-Brief 431

Presented at the 144th Convention
2018 May 23 – 26, Milan, Italy

The Anaglyph binaural audio engine

David Poirier-Quinot and Brian F.G. Katz

Sorbonne Université, CNRS, Institut Jean Le Rond d'Alembert, UMR 7190, F-75005 Paris, France

Correspondence should be addressed to David Poirier-Quinot (`david.poirier-quinot@upmc.fr`)

ABSTRACT

Anaglyph is part of an ongoing research effort into the perceptual and technical capabilities of binaural rendering. The Anaglyph binaural audio engine is a VST audio plugin for binaural spatialisation integrating the results of over a decade of spatial hearing research. Anaglyph has been designed as an audio plugin to both support ongoing research efforts as well as to make accessible the fruits of this research to audio engineers through traditional existing DAW environments. Among its features, Anaglyph includes a personalizable morphological ITD model, near-field ILD and HRTF parallax corrections, a Localisation Enhancer, an Externalisation Booster, and SOFA HRIR file support. The basic architecture and implementation of each audio-related component is presented here.

1 Introduction

"Binaural hearing" refers to the capability of integrating information from the two ears to perceive a sound in three-dimensional space. Psychophysical studies have shown that various mechanisms are involved in the human auditory system for sound localisation [1]. To infer the angular direction of a sound source, these mechanisms rely on direction-dependent audio cues, resulting from the propagation of an acoustic wave from the source to both ears. Using digital signal processing, these cues can be applied to any audio input to simulate a sound source at a virtual position in a listener's 3D auditory space (experienced over headphones). This process is referred to as "binaural synthesis", and is used to create 3D soundscapes in virtual and augmented reality [2] or applied to audio tracks to create spatial binaural compositions [3].

Based on more than a decade of research in spatial hearing and binaural rendering, the Anaglyph project encapsulates various efforts spanning previous and ongoing studies in the field. The Anaglyph engine, described below, represents the latest evolution of an academic binaural rendering engine [4] incorporating results of recent studies while also being packaged in a more distributable and user-friendly format.

The Anaglyph binaural audio engine is being distributed as a VST (Virtual Studio Technology) plugin, compatible with most Digital Audio Workstations (DAW). Its primary function is the real-time rendering of a dry mono audio input buffer to a 2-channel bin-

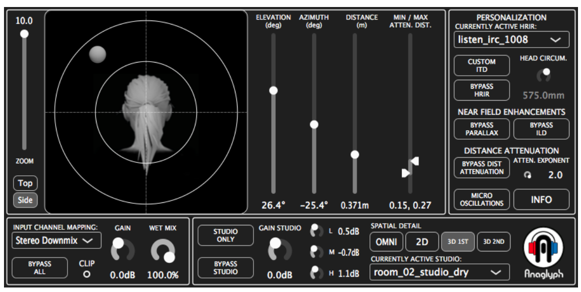

Fig. 1: GUI of the Anaglyph plugin.

aural stereo audio output. A cross between a research platform and a professional solution, Anaglyph is conceived as a proof of concept test for state-of-the-art spatialisation techniques to be compared against production environment expectations. The Anaglyph VST plugin is made available as a freeware for non-commercial use, and can be downloaded from the project webpage: `anaglyph.dalembert.upmc.fr`.

Anaglyph is implemented in C++, using the JUCE framework for cross-platform compilation, Graphical User Interface (GUI) design, audio classes (buffers, filters, etc.), and standard VST features (automations, presets, I/O management, etc.). The plugin GUI is illustrated in Figure 1. The architecture of the plugin is described in the next section while a detailed description of its main components is given in Section 3.

2 General Architecture

The overall architecture of the Anaglyph plugin is shown in Figure 2. The input audio buffer is encoded by the binaural module into a stereo buffer. The binaural module is composed of a Head Related Impulse Response (HRIR) convolution engine, an Interaural Time Difference (ITD) modifier, and an Interaural Level Difference (ILD) modifier. ITD, ILD, and HRIR respectively represent the time, amplitude, and spectral cues resulting from the propagation of a sound wave from a point source, travelling around the listener's head, arriving at the ear canal entrance.

Parallel to the binaural encoding, the input audio is fed to an Externalization Booster, based on a modified

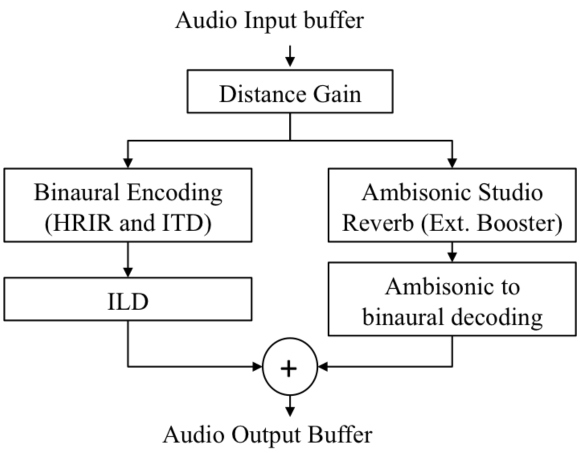

Fig. 2: Simplified architecture of the Anaglyph plugin.

3D room reverberation simulator [5]. The Finite Impulse Response (FIR) filter is loaded in the form of a 2nd order Ambisonic Room Impulse Response (RIR). The resulting Ambisonic buffer is decoded using a virtual speaker approach [6] and rendered binaurally to a stereo buffer, which is then combined with the the output of the direct sound binaural encoder.

3 Detailed Components

This section details the main components of the Anaglyph plugin.

3.1 Binaural Encoding

Whenever a new position is defined, through DAW automation lines or via the plugin's GUI, the convolution engine converts the pair of HRIR filters returned by the MySofa library into FIR filters, applying both right and left Impulse Responses (IR) in the frequency domain to reduce CPU consumption. The resulting stereo buffer is passed to the ITD modifier. This module adjusts the delay of the right or left input buffer to correspond to the estimated personalized ITD of the user relative to the ITD of the selected HRIR dataset [7, 8] for the current position. The user's head circumference is used to provide a reliable ITD model to personalize the ITD values [9]. If disabled, the original HRIR's ITD is used.

3.2 Distance Attenuation

The default distance attenuation model is based on the inverse square law. The law exponent can be modified, as can be both start and stop attenuation distances.

3.2.1 HRIR Near-Field Parallax

Due to technical constraints, most HRIR sets are measured for a unique source to head distance, providing an HRTF for the surface of given sphere. Anaglyph implements a "parallax HRIR correction" option to compensate for variable distances without the need for multiple HRIR sets at different distances by simulating the impact of source distance on the HRIR [10]. The parallax correction is illustrated in Figure 3. The correction affects source positions close to the subject compared to the HRIR measurement ring. The ring radius value is set based on the measurement distance indicated in the loaded SOFA HRIR.

3.3 ILD Near-Field Model

After binaural encoding, the stereo buffer is passed to the ILD modifier. The ILD modifier applies a frequency-dependant adjustment to both left and right buffers, using on a pair of cascaded biquad filters. The ILD modifier is applied for the current audio source position (azimuth, elevation, distance), reproducing ILD variations (via a spherical head model [11]).

3.4 Externalization Booster

In essence, the externalization booster is an Ambisonic room reverberation module with the direct sound removed, as this is processed by the binaural encoding module. As the name suggests, this module is used to improve source externalization, separate from any artificial reverberation added for artistic goals. Based on FIR convolution, this 'reverb' is applied on the mono input buffer, parallel to the binaural encoding. The FIR is interpolated from a set of 2^{nd} order Ambisonic RIRs. A spherical array of source positions surrounding the listener allows the convolution reverberator to provide coherent early reflection patterns (in time and space) with respect to the actual source direction used in the Binaural Encoding module. Such reflection patterns help support the sense that the source is located in a space away from the listener while also reinforcing the spatial cues for source localization.

The resulting Ambisonic buffer is decoded based on a virtual speaker approach [6] prior to summation with the binaural buffer. The virtual speakers used for Ambisonic to binaural decoding represent a 12 channel array, uniformly distributed on a sphere.

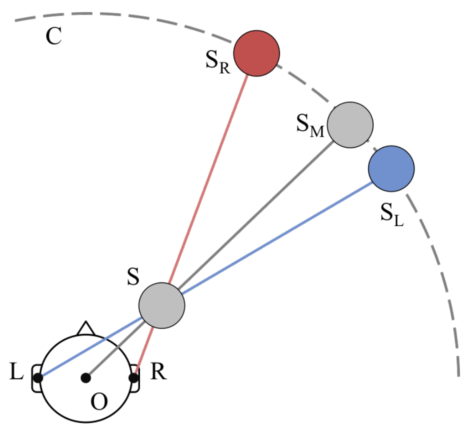

Fig. 3: Illustration of the parallax HRIR correction principle. **O** is the center of the subject's head, **L** and **R** are the positions of the subject's left and right ears respectively. **S** is the position of the audio source as defined in the plugin GUI, **C** represents the HRIR measurement ring, and $\mathbf{S_M}$ the projection of **S** on **C**. $\mathbf{S_M}$ is the HRIR position that would be selected without the parallax correction option enabled. $\mathbf{S_L}$ and $\mathbf{S_R}$ are the positions of the HRIRs actually selected for the left ear and right ear respectively when the parallax correction is enabled.

3.5 SOFA Loader

Upon startup, the plugin loads HRIR, ITD, ILD, and room IR SOFA files [12] using the MySofa C++ library [13]. Prior to loading the SOFA HRIR, a Matlab utility script is used for time-alignment of the HRIR while the ITD value is stored in the "Delay" section of the SOFA file. The MySofa library internal routines are used for linear interpolation between neighbouring IRs, (HRIR, room IR), time (ITD), and biquad (ILD) values.

3.6 Localisation Enhancer

The localisation enhancer module applies a subtle figure-of-8 movement to the audio source in the azimuth-elevation plane (fixed distance). The radius of these micro-oscillation can be modified, set to a default 1 degree amplitude in both dimensions. This oscillation is designed to compensate for the absence of listener's head micro-movements in non-tracked conditions, proved to improve localisation accuracy [14, 15]. The optimization of the control parameters and the quantification of the effect through formal studies of this module is currently the subject of study.

4 Conclusion

Anaglyph, and its previous incarnations, have been used in numerous research projects [16, 17, 18, 19, 20] as well as several production tests involving spatialised music[1], radio drama fiction[2], and immersive video game audio[3]. Anaglyph is currently being use by Bloober Team in several upcoming suspense/horror game projects. Anaglyph will be updated based on results of the authors and peers in binaural spatialisation.

[1] www.youtube.com/watch?v=R9Vx4PaE1ZA
[2] www.youtube.com/watch?v=V6N-6DxAVQw
[3] www.youtube.com/watch?v=q6muds1qW-w

5 Acknowledgements

Some upstream development support was provided through funding from the European Union's Horizon 2020 program under grant agreement No 644051.

References

[1] Blauert, J., *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT press, 1997.

[2] Begault, D. R. and Trejo, L. J., *3-D Sound for Virtual Reality and Multimedia*, Academic Press, 2000.

[3] Rumsey, F., *Spatial Audio*, CRC Press, 2012.

[4] Katz, B., Rio, E., and Picinali, L., "LIMSI Spatialisation Engine," Intl Deposit Digital Number: IDDN.FR.001.340014.000.S.P.2010.000.31235., 2010.

[5] Guastavino, C. and Katz, B., "Perceptual evaluation of multi-dimensional spatial audio reproduction," *J Acous Soc of Am*, 116(2), pp. 1105–1115, 2004.

[6] Noisternig, M., Musil, T., Sontacchi, A., and Holdrich, R., "3D binaural sound reproduction using a virtual Ambisonic approach," in *Virtual Environments, Human-Computer Interfaces and Measurement Systems*, pp. 174–178, IEEE, 2003.

[7] Andreopoulou, A. and Katz, B., "Identifying a perceptually relevant estimation method of the inter-aural time delay," *J Acous Soc of Am*, 141(5), pp. 3635–3635, 2017.

[8] Katz, B. and Noisternig, M., "A comparative study of interaural time delay estimation methods," *J Acous Soc of Am*, 135(6), pp. 3530–3540, 2014.

[9] Aussal, M., Alouges, F., and Katz, B., "HRTF interpolation and ITD personalization for binaural synthesis using spherical harmonics," in *Aud Eng Soc Conf 25*, 2012.

[10] Otani, M., Hirahara, T., and Ise, S., "Numerical study on source-distance dependency of head-related transfer functions," *J Acous Soc of Am*, 125(5), pp. 3253–3261, 2009.

[11] Duda, R. O. and Martens, W. L., "Range dependence of the response of a spherical head model," *J Acous Soc of Am*, 104(5), pp. 3048–3058, 1998.

[12] Majdak, P., Iwaya, Y., Carpentier, T., Nicol, R., Parmentier, M., Roginska, A., Suzuki, Y., Watanabe, K., Wierstorf, H., Ziegelwanger, H., et al., "Spatially oriented format for acoustics: A data exchange format representing head-related transfer functions," in *Aud Eng Soc Conv 134*, 2013.

[13] Hoene, C., Patino Mejia, I. C., and Cacerovschi, A., "MySofa – Design Your Personal HRTF," in *Aud Eng Soc Conv 142*, 2017.

[14] Nykänen, A., Zedigh, A., and Mohlin, P., "Effects on localization performance from moving the sources in binaural reproductions," in *Intl Cong and Exposition on Noise Control Eng*, volume 4, pp. 3193–3201, 2013.

[15] Hendrickx, E., Stitt, P., Messonnier, J.-C., Lyzwa, J.-M., Katz, B., and De Boishéraud, C., "Influence of head tracking on the externalization of speech stimuli for non-individualized binaural synthesis," *J Acous Soc of Am*, 141(3), pp. 2011–2023, 2017.

[16] Katz, B., Rio, E., Picinali, L., and Warusfel, O., "The effect of spatialization in a data sonification exploration task." in *Intl Conf on Auditory Display*, pp. 1–7, Paris, 2008.

[17] Katz, B. F., Kammoun, S., Parseihian, G., Gutierrez, O., Brilhault, A., Auvray, M., Truillet, P., Denis, M., Thorpe, S., and Jouffrais, C., "NAVIG: augmented reality guidance system for the visually impaired," *Virtual Reality*, 16(4), pp. 253–269, 2012, doi:10.1007/s10055-012-0213-6.

[18] Parseihian, G. and Katz, B., "Rapid head-related transfer function adaptation using a virtual auditory environment," *J Acous Soc of Am*, 131(4), pp. 2948–2957, 2012.

[19] Parseihian, G., Jouffrais, C., and Katz, B. F., "Reaching Nearby Sources: Comparison Between Real and Virtual Sound and Visual Targets," *Frontiers in Neuroscience*, 8, pp. 269:1–13, 2014, doi: 10.3389/fnins.2014.00269.

[20] Menelas, B.-A. J., Bourdot, P., Picinali, L., and Katz, B. F. G., "Non-visual identification, localization, and selection of entities of interest in a 3D environment," *J. Multimodal User Interfaces*, pp. 1–14, 2014, doi:10.1007/s12193-014-0148-1.