Tools for the production of spatial audio within BINCI

André Kruh-Elendt[1], André Fiebig[1], Roland Sottek[1], and Julien De Muynke[2]

[1] *HEAD acoustics GmbH, 52134 Herzogenrath, Germany,*

Email: {andre.kruh-elendt, andre.fiebig, roland.sottek}@head-acoustics.de

[2] *Eurecat, Centre Tecnològic de Catalunya, Multimedia Technologies Group, 08005 Barcelona, Spain,*

Email: julien.demuynke@eurecat.org

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 732130 – BINCI project

Introduction

With the recent introduction of support for immersive audiovisual technologies by some major content sharing platforms, such as YouTube and Facebook, and the establishment of standards for the efficient encoding and transmission of spatial audio [1], interest in the composition and employment of these audio formats has extended beyond the academic and research fields to include artists, sound engineers and professional content creators. Amongst the different existing spatial audio formats, the binaural technology presents the advantage of being suited for the reproduction of immersive audio content across multiple platforms and devices, requiring, essentially, nothing more than a pair of headphones. Nonetheless, support for reproduction over loudspeaker arrays and interoperability between formats are still expected.

With this in mind, the need for an integrated and flexible solution for the production of spatial audio becomes clear. The development of such a set of tools and their incorporation into established audio production workflows are the main goals of the BINCI project. In this paper, an overview of the tools being developed is given. Furthermore, an HRTF selection process and the employment of room impulse responses recorded with spherical microphone arrays are highlighted as relevant techniques adopted in the project.

Spatial audio production tools

Within BINCI, two main software modules are being developed, each one targeting specific user-groups and tasks in the production and delivery of spatial audio content. A brief overview of these tools is given below, while a more detailed description can be found in [2].

The Binaural Home Studio (BHS) is a suite of production and post-production tools targeting sound engineers and creators involved in the process of designing, editing and mixing audio with Digital Audio Workstations (DAWs) in typical studio environments.

The BHS is composed of four main components (see Fig. 2): The *Audio Processing Server*, an ambisonics-based processing unit performing encoding, manipulation and rendering operations on the audio input; a set of *DAW Plug-ins* (see Fig. 1), which provide a simple graphical interface to control the processing performed in the server; the *Virtual Sound Card*, serving as a multichannel audio interface between the DAW and the audio processing server; and a *Visualizer*, which gives graphical feedback to the end-user about the current spatial configuration of the sound scene and can be used to synchronize the produced audio with 360° video for audiovisual applications.

The use of a modular structure with plugins controlling the background processing server allows a seamless integration of the spatial encoding and rendering engines into the workflows typically used for the production of stereo and multichannel content. Sound source panning and clustering, and the application of modulation and other audio effects are some of the sound manipulations known from typical audio production environments which have been extended for spatial audio in the set of BINCI tools.

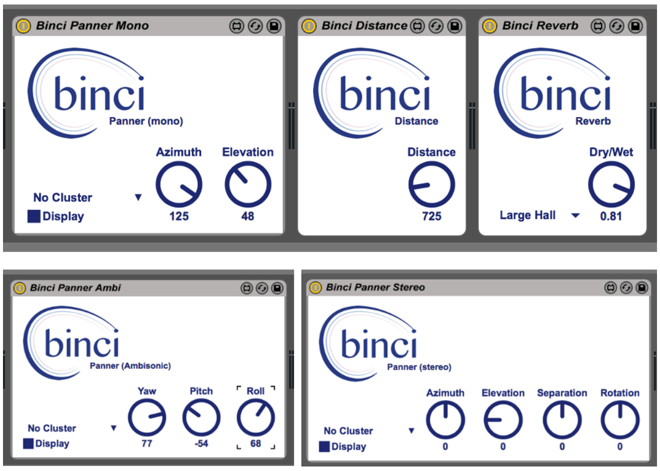

Fig. 1: Example of DAW plugins developed as part of BINCI

The Binaural Player (BP) is a cross-platform dynamic binaural rendering module designed for reproduction of ambisonics sound scenes. The BP is designed primarily for playback of content created with the BHS, nonetheless B-Format audio files and streams are also accepted as an input. This ensures compatibility with applications and platforms which support spatial audio formats.

By using ambisonics as an intermediate format, both in the BHS and the BP, sound scene manipulations, such

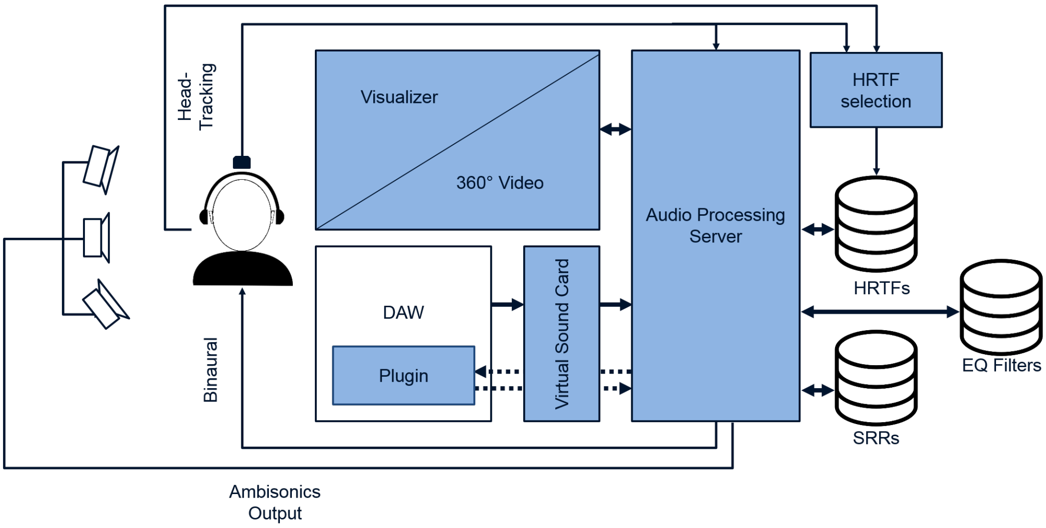

Fig. 2: Structure and main components of the Binaural Home Studio

as rotations, can be performed efficiently. Furthermore, rendering audio to a three dimensional loudspeaker array is possible, extending the flexibility for creators and end-users to monitor and listen to the created content.

All tools developed within BINCI use dynamic binaural rendering (i.e. support for head motion) by default and apply headphone equalization whenever a matching filter for the employed model is available. By combining these techniques with the use of high resolution HRTFs, an HRTF selection process, and the application of acoustic room information, localization errors are reduced and the realism of binaural scenes is improved.

HRTF selection and individualization

The BHS and BP modules perform binaural rendering using a database of HRTFs with a high spatial resolution (2,5° in both azimuth and elevation) measured for 25 adults. Head and torso dimensions extracted according to [3] for 79 individuals (see Fig. 3) were used in the selection of the 25 measured subjects. The selection process was based on an iterative clustering intending to create maximal variance of the extracted geometrical dimensions of the selected adults. The extracted anthropometric features for the chosen subjects accompany each of the corresponding HRTF sets. In addition to the measured HRTFs, the potential for modelling binaural transfer functions using analytical models, as described in [4], is currently being evaluated as a solution for obtaining HRTFs for children.

With the purpose of providing a more realistic or plausible binaural experience for the listener, a fast HRTF selection and individualization procedure is included as part of the tools developed in BINCI. The individualization procedure consists of a pre-selection step, were a reduced number of HRTF sets are presented to the end-user based on the input of easily measurable head dimensions, and an Interaural Time Difference (ITD) adjustment step, based on the method described in [5]. The ITDs of the pre-selected HRTFs are thereby adjusted until a stable sound source is perceived at a given location. Finally, the adjusted HRTF set is exported for use in the binaural rendering process.

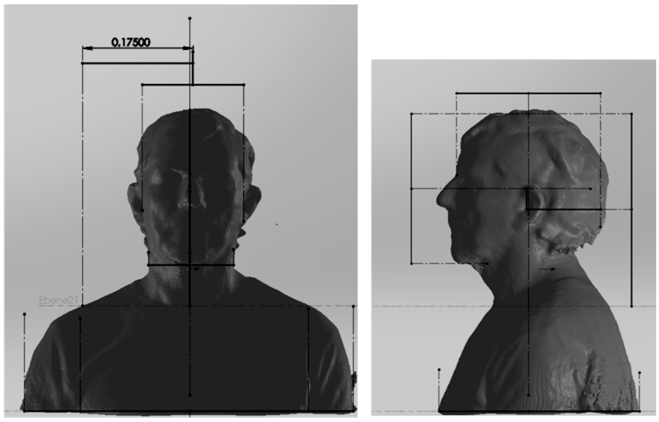

Fig. 3: Extraction of anthropometric parameters for one of 79 subjects

The described ITD scaling and adjustment procedure can be applied on values extracted from the HRIRs using, for example, the onset threshold or cross-correlation with the minimum-phase methods [6]. Alternatively, ITD values can be modelled for any azimuth and elevation angles using the provided head dimensions. Discontinuities and exaggerated values from the extracted ITDs (see Fig. 4) could be avoided. Further investigations in the project will provide information about the perceptual difference and possible advantages for the two ITD computation methods.

Spatial Room Impulse Responses

One important feature of the spatial audio production tools developed in BINCI is the capability to simulate different environments (i.e. rooms). This is achieved by convolving the audio input from the DAW with a filter describing the acoustic properties in the selected

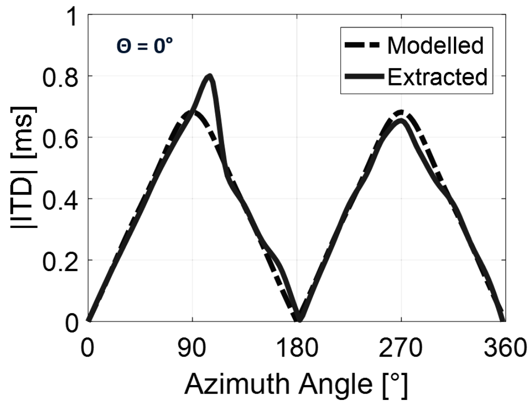

Fig. 4: Comparison of modelled and extracted ITD values for one subject and all measured azimuth positions in the horizontal plane

environment—a room impulse response (RIR). To increase compatibility with the employed spatial (ambisonics) format, sets of RIRs have been recorded using spherical (ambisonics) microphone arrays as so-called Spatial Room Impulse Responses (SRRs). Examples of ambisonics RIRs have already been published in [7].

The filters are thereby separated into directional components each containing the RIR corresponding to a specific direction in space. The number of components depends on the ambisonics order of the SRRs, which in turn depends on the model of ambisonics microphone used for the measurements: Soundfield DSF-2 MKII and Sennheiser Ambeo (both 1rst order), Zylia ZM-1 (3rd order), and MH Acoustics Eigenmike (4th order) are examples of commercially available ambisonics microphones of different orders.

The main advantage of using SRRs over omnidirectional RIRs is that the acoustic properties of the room are reproduced in accordance with the listener's head orientation with respect to the room. As an example, a very asymmetric room like a long and narrow corridor has very different echo patterns along the X and the Y axis. Depending on whether the user is looking along the X or Y axis, the series of echoes coming from the sides (90° and 270°) and coming from the front and the back (0° and 180°) are meant to be different. The dynamic binaural synthesis implemented in all BINCI tools allows for a continuous change of orientation of the RIR as the listener's head rotates.

It should be noted that a convenient way to store, exchange and use SRRs data is proposed in [8] as a new convention for the SOFA file format [9].

Summary

A current overview of the development status for the spatial audio production tools developed in BINCI has been given. Some of the most relevant technologies employed in the project and the advantages provided to members of the creative industry were highlighted. Further improvements and integration of the software solutions developed in BINCI are the next steps in the project. In addition, a current state of the production tools is being used by selected content creators with the purpose of evaluating their usability in creative environments and demonstrating their capabilities. To this end, the Fundació Juan Miró in Barcelona, the Alte Pinakothek in Munich and the St. Andrews Castle in St. Andrews have agreed to serve as demonstration sites for BINCI illustrating the potential benefit of binaural audio (guide) content for cultural sites and providing an option to tests the usability at large scale.

Acknowledgements

The authors would like to thank all partners involved in the BINCI project for the productive and cooperative work. Moreover, we thank Andreas Herweg, Frederic Allion and Matthias Reffgen for their support in the development of software prototypes, the optimization of computation models and the execution of measurements, respectively. The work presented in this document is part of BINCI, a Horizon 2020 innovation project funded by the European Union under the grant agreement No. 732130.

References

[1] J. Herre, J. Hilpert, A. Kuntz, and J. Plogsties, "MPEG-H 3D Audio — The New Standard for Coding of Immersive Spatial Audio," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 5, pp. 770–779, 2015.

[2] A. Garriga, M. E. Fuenmayor, and M. Caballero, "Binaural tools for 3D audio production at home," *NEM Summit*, 2017.

[3] V. R. Algazi, R. O. Duda, and D. M. Thompson, "The CIPIC HRTF Database," *Signal Processing*, no. October, pp. 99–102, 2001.

[4] R. Sottek and K. Genuit, "Physical modeling of individual HRTFs (Head Related Transfer Functions)," in *DAGA*, 1999.

[5] A. Lindau, J. Estrella, and S. Weinzierl, "Individualization of dynamic binaural synthesis by real time manipulation of the ITD," *Proc of the 128th AES Convention*, 2010.

[6] B. F. G. Katz and M. Noisternig, "A comparative study of interaural time delay estimation methods," *J. Acoust. Soc. Am.*, vol. 135, no. 6, 2014.

[7] "The openair database." Available at `http://www.openairlib.net`.

[8] A. Pérez-Lopez and J. De Muynke, "Ambisonics Directional Room Impulse Response as a new Convention of the Spatially Oriented Format for Acoustics," *Engineering Brief, AES Convention 2018*, to be published.

[9] "AES69-2015: AES standard for file exchange - Spatial acoustic data file format," 2015.