Spatial audio productions using different sets of HRTFs

André Kruh-Elendt, André Fiebig, Roland Sottek
HEAD acoustics GmbH, Herzogenrath, Germany

Summary
The past decade has seen an increase in the demand of 3D audio-visual material for numerous applications. Channel-based, object-based and scene-based approaches have been developed and improved for the reproduction of 3D audio using loudspeaker arrays, while the technology of binaural synthesis provides the best immersive audio experience for headphone reproduction. Although consumer-grade devices and applications now support many of these technologies, the scarcity of high-quality 3D audio content and the lack of adequate production tools are issues to be addressed. Binaural Tools for the Creative Industries (BINCI) is an EC-funded research project, aiming to create user-friendly tools for the production of spatial audio, which are integrated in typical workflows with digital audio workstations used by sound engineers and other members of the creative industry. The progress of the project and an overview of the underlying technologies are presented in this paper. Furthermore, focus is given to the measurement and individualization of HRTFs as well as on how different sets of HRTFs relate to the perceived level of immersion and presence.

PACS no. 43.60.Dh, 43.66.Pn, 43.75.Zz

1. Introduction

Over the course of the past decade, technologies for the representation and efficient delivery of spatial audio content have been greatly improved and are currently being introduced into electronic products and content distribution platforms targeting the consumer market. Examples of this can be seen in the introduction of A/V receivers with support for multichannel 3D audio, the addition of spatial audio formats into international broadcasting standards and the adoption of spatial audio formats by internet platforms providing audiovisual content. Through this, spatial audio is becoming more accessible for a large group of users which demand high quality content and consistency across devices and platforms. This expectations most certainly include portable devices, where audio content is usually presented via headphones.

While each of the many available 3D audio formats provide unique features and advantages to the end user, sound engineers and creators are faced with the task of designing and mixing audio for different applications, often requiring detailed knowledge about the employed format or specialized tools for editing and mixing audio content. The Binaural Tools for the Creative Industry (BINCI) project aims to target these shortcomings by providing a set of tools which integrate seamlessly in typical studio environments centered around the employment of Digital Audio Workstations (DAWs) for the manipulation of recorded and synthesized audio signals. Although the main objective is to produce spatial (binaural) audio presented over headphones—hence the project's name—, compatibility with formats for loudspeaker reproduction is a relevant feature enforced by the use of ambisonics technology.

In the following, the basic components and functionality of the developed production tools are shortly presented, furthermore the issue of producing individualized binaural audio is addressed and a brief examination of approaches to this topic are discussed. Finally, the selected individualization procedure is described and an overview of ongoing investigations is given.

2. Audio production tools in BINCI

The tools developed by the project, consist of two main modules: The *Binaural Home Studio* (BHS), targeting the content creators, and the *Binaural Player* (BP), intended for consumers [1, 2]. Both modules use ambisonics as a core technology for efficient audio scene manipulation and flexible rendering.

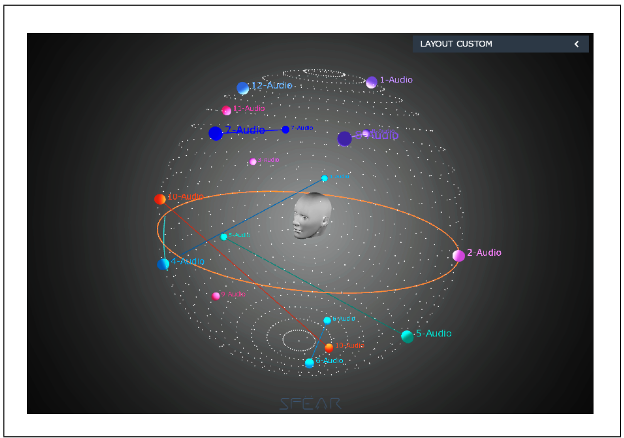

Figure 1. The BHS visualizer showing the spatial location of loudspeakers and sound sources with respect to the user's position.

2.1. The Binaural Home Studio and Binaural Player modules

The BHS is a tool developed for audio production environments composed of four main elements: The ambisonics based *Audio Processing Server* (APS), performing all encoding, signal processing and rendering on the audio channels (i.e. tracks) provided as inputs from the DAW; a collection of *DAW Plugins* which integrate seamlessly into the most common DAWs and allow the producer to control the processing carried out in the APS; a *Virtual Sound Card*, acting as an audio interface between DAW and APS; and the *Visualizer* (see Figure 1), providing visual feedback about the spatial sound source composition and options to manipulate the audio scene or to synchronize audio with 360° video content.

By outsourcing the signal processing to the APS, all implemented functionalities remain independent from the capabilities of the chosen DAW. This includes the ambisonics decoding and encoding, the dynamic binaural rendering, as well as some typical audio effects (e.g. panning, modulation, reverb) commonly employed by sound engineers, which have been adapted for use with spatial audio. An example for one of these effects is the (spatial) modulation plugin which, as a BINCI plugin, does not change the signal's waveform but rather modulates the pattern of motion for single or clustered sound sources around their spatial position.

One of the major advantages of the BHS is it's capability to render the ambisonics scene to a binaural stream, for dynamic headphone reproduction, or to loudspeaker signals, for playback over a 3D loudspeaker array. This provides a flexible and portable monitoring solution with the opportunity to compare spatial sound scene compositions in different reproduction scenarios.

On the playback side, the BP offers a solution for the reproduction of spatial audio content created with the BHS. Nonetheless, standard ambisonics formats, such as the B-Format (employed by major content sharing platforms) are also supported. The BP's main task is to perform dynamic binaural rendering of the ambisonics input scene and apply equalization for headphone reproduction.

2.2. Technological aspects

Both of BINCI's main modules employ state-of-the-art techniques aiming to improve the spatial audio experience. Apart from using ambisonics for efficient rotations and sound scene manipulations, a simulation of different acoustic environments is made possible by applying ambisonics based room impulse responses, so called Spatial Room Impulse Responses (SRR) [2, 3]. Furthermore, an HRTF individualization procedure (described in more detail later on) can be performed for the case where no individual filters are available. Together with head-tracked binaural rendering and the equalization of supported headphone models, better externalization and improved localization performance are expected.

3. Individualized binaural rendering

3.1. HRTF individualization methods

When dealing with binaural synthesis, the use of non-individual HRTFs will often lead to localization errors (e.g. front-back confusions) and a perceived instability of the sound sources. Different approaches aiming to improve localization performance and the overall quality of the binaural experience through HRTF individualization are therefore proposed in literature. Some approaches intend to compute a new set of HRTFs parting from some representation of an individual's head geometry, e.g. by capturing 3D representations of the head and pinnae [4, 5], by using measured anthropometric features [6, 7] or by approximating a head's geometry through geometrical models [8, 9]. Other approaches include the acquisition of individual HRTFs using direct (in-situ) measurements [10, 11], the adjustment of non-individual HRTFs [12, 13, 14] and the selection of an HRTF set based on subjective perception [15, 16].

While the computation of HRTFs based on an individual's head geometry can be performed faster as computer technology advances and the acquisition of accurate 3D models and head geometries may now be performed using consumer-grade devices and software tools, this approach is usually time consuming and often requires multiple devices, software solutions and manual intervention by the user. Similar constraints apply for in-situ HRTF measurements, which also require a specific measurement setup (i.e. in-ear microphones and a loudspeaker) and an adequate acoustical environment. In comparison, the selection and adjustment of a non-individual HRTF from a database can be performed quite fast and can be designed to require only minimal input from the user. Nonetheless, the mismatch of spectral cues might still lead to a wrong localization along the median plane or cone of confusion errors.

3.2. Individualization in BINCI

Since BINCI addresses primary members of the creative community working on different devices (including portable setups) and the final consumer of audiovisual media, the selection and adjustment of non-individual HRTFs from a database has been deemed as a suitable individualization procedure for the project. To this purpose, 25 different sets of HRTFs were measured with a spatial resolution of 2,5° in both azimuth an elevation. These binaural filters accompany BINCI's production tools and are complemented with additional anthropometric information extracted from respective 3D scans of head and torso.

The individualization procedure consists of a preselection step, where a few HRTF sets are selected based on easily measurable head dimensions, an individual selection step, where the most suitable HRTF set can be picked from the pre-selected set based on individual perceptual preferences, and finally an adjustment step, where the stability of virtual sound sources is improved through scaling of modeled ITD values (see Figure 2) following the method described in [14]. This procedure has the advantage of providing a better binaural experience while requiring only a minimal effort from the user.

By delivering anthropometric information together with the measured HRTFs and supporting the standardized SOFA format [17], the employment of other HRTF databases or binaural filters obtained from more complex individualization approaches is still possible. This is specially relevant for professional sound engineers working in studio environments with defined acoustic conditions, who will usually expect individually measured HRTFs or BRIRs to accurately simulate the loudspeaker-room configuration they are accustomed to.

4. Evaluation

An evaluation of the production tools' usability and of the performance of the technologies employed is a relevant aspect considered in the project's scope. For the former, a user group of professional content creators has been selected while the latter is achieved by using corresponding analysis and listening tests.

4.1. Assessment of production tools

Selected content creators and demonstration sites have been asked to participate in BINCI with the purpose of evaluating the set of audio production tools.

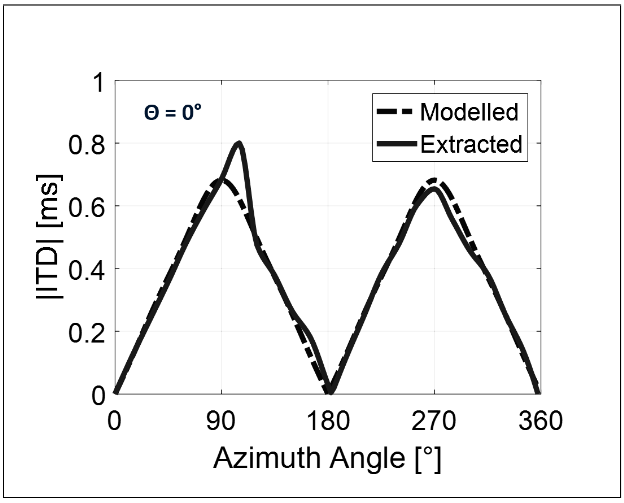

Figure 2. Comparison between modelled and extracted ITD values across all azimuth angles at 0° elevation for one subject. Modeled values computed from anthropometrical parameters are prefered as they match extracted values without producing discontinuities.

Development versions of the BHS have therefore been provided to sound engineers throughout the project's duration with the objective of creating sound compositions for four cultural institutions including the *Alte Pinakothek* in Munich, the *Fundacó Juan Miró* in Barcelona and the *St. Andrews Castle* in St. Andrews. Continuous feedback regarding the functionality and usability of BINCI's tools have led to improvements and additions to the project's scope. An example of this is the inclusion of video synchronization in the aforementioned BHS visualizer.

The binaural compositions created within BINCI will be demonstrated to the general public in the form of audio guides at the locations mentioned above. This should also provide valuable information about the deployment and performance of the BP.

4.2. Assessment of implemented technologies

Relevant technologies such as the SRRs and the HRTF individualization procedure are currently being evaluated. With regards to the HRTF individualization, an experiment has been initiated where participants were asked to perform an ITD adjustment on three pre-selected HRTF sets chosen from BINCI's HRTF database after finding the closest head geometries. Following this step, each participant was requested to rate the perceived sound source stability, plausibility and colouration of synthesized content using each of the HRTFs before and after individualization. The current results show a trend towards an improved stability for individualized HRTFs (i.e. where the ITD has been adjusted) as seen in Figure 3. Final results for this experiment and an additional evaluation on other relevant aspects such as externalization and localization performance will be presented in the final stages of the project.

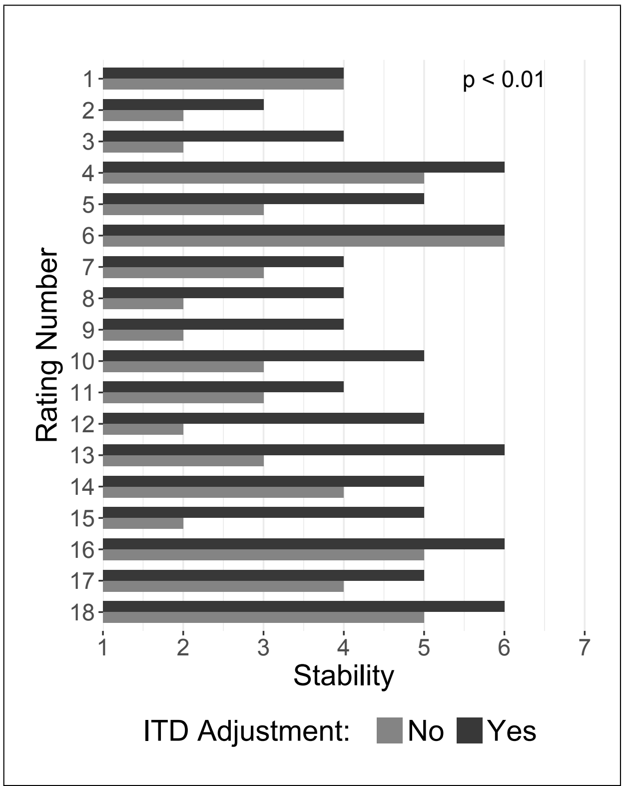

Figure 3. Preliminary results regarding the perceived sound source stability of a single sound source before and after individual ITD adjustment. The ratings were obtained from 6 particpants for which 3 HRTFs were selected based similar anthropometrical parameters.

5. Conclusion

The motivation and goals of BINCI have been outlined together with the tools and technologies implemented in the project. Special focus was given to the topic of HRTF individualization, where a brief overview of different approaches was presented and the proposed HRTF individualization procedure was briefly described. Moreover, the status of current evaluations regarding the performance of BINCI's spatial audio production tools and ongoing experiments validating the advantages of an HRTF individualization have been addressed. It is worth mentioning that the relevance and potential for spatial audio production tools can be confirmed from the feedback given by the user groups. Similarly, benefits from a HRTF selection and adjustment procedure can be seen from preliminary listening test results, although further relevant aspects and improvements towards a better binaural experience and the minimization of localization errors should be considered in future investigations.

Acknowledgement

The authors would like to acknowledge the productive and cooperative work of all partners involved in the BINCI project. Special recognition is given to Andreas Herweg for his prototype implementations and the execution of current validation experiments. 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] A. Garriga, M. E. Fuenmayor, M. Caballero: Binaural Tools for 3D Audio Production at Home. Proc. NEM Summit 2017.

[2] A. Kruh-Elendt et. al.: Tools for the production of spatial audio within BINCI. Proc. DAGA 2018.

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

[4] A. Meshram et. al.: P-HRTF: Efficient personalized HRTF computation for high-fidelity spatial sound. Proc. ISMAR 2014, 53-61.

[5] P. Majdak, W. Kreuzer, R. Baumgartner: Head-Related-Transfer-Functions aus Fotos. Proc. DAGA 2018.

[6] D. Zotkin et. al.: HRTF Personalization Using Anthropometric Measurements. Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics 2003, 157-160.

[7] P. Bilinski et. al.: HRTF Magnitude Synthesis via Sparse Representation of Anthropometric Features. Proc. ICASSP 2014, 4468-4472.

[8] R. Sottek, K. Genuit: Physical modeling of individual HRTFs (Head Related Transfer Functions). Proc. DAGA 1999.

[9] V. R. Algazi, R. O. Duda: Approximating the head-related transfer function using simple geometric models of the head and torso. The Journal of the Acoustical Society of America 112, 5 (2002) 2053-2064.

[10] J. He, R. Ranjan, W. S. Gan: Fast Continuous HRTF Acquisition With Unconstrained Movements of Human Subjects. Proc. ICASSP 2016, 321-325.

[11] B. Masiero et. al.: Design of a Fast Individual HRTF Measurement System. Proc. DAGA 2012.

[12] R. Bomhardt, M. Lins, J. Fels: HRTF Individualization: Investigations on the Influence of Adapted Binaural Parameters on the Localization. Proc. DAGA 2016, 81-82.

[13] R. Bomhardt, J. Fels: Analytical Interaural Time Difference Model for the Individualization of Arbitrary Head-Related Impulse Responses. Proc. of the 137th AES Convention 2014.

[14] A. Lindau, J. Estrella, S. Weinzierl: Individualization of Dynamic Binaural Synthesis by Real Time Manipulation of the ITD. Proc. of the 128th AES Convention 2010.

[15] B. Seeber, H. Fastl: Subjective Selection of Non-Individual Head-Related Transfer Functions. Proc. International Conference on Auditory Display 2003, 259-262.

[16] Y. Iwaya: Individualization of Head-Related Transfer Functions With Tournament-Style Listening Test: Listening With Other's Ears. Acoust. Sci. & Tech. 27, 6 (2006) 340-343.

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