INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2025, VOL. 71, NO. 4, PP. 1-6
Manuscript received May 29, 2025; revised September 2025. doi: 10.24425/ijet.2025.155462

Creation of immersive experience for headphone listening using directional room impulse responses for BRIR synthesis

Witold Mickiewicz, and Kaja Kosmenda

***Abstract*—This paper presents an algorithm for immersive processing of a multichannel recording for headphone listening. The material listened to with headphones should evoke impressions in the listener that are identical to those experienced when listening from a multi-speaker system. In order to allow the processing system to be adapted to the individual anatomical characteristics of the listener, an algorithm was developed, which is based on data in the form of directional room impulse responses acquired with an intensity probe, in the form of classical room pressure impulse responses at the excitation emitted by the individual loudspeakers of the listening system. The listening room characteristics recorded in this way are supplemented with data from the Head Related Transfer Function (HRTF) databases, which can be selected according to the listener's perception. The study compared the effects of an impulse response segmentation algorithm using publicly available HRTF averaging databases with the classic approach using individualized binaural room impulse responses (BRIR). Reference was also made to available binauralization algorithms using dummy head.**

***Keywords*—immersive sound; room impulse response; binaural technology; binaural room impulse response; sound intensity; headphones; externalization; vector acoustics**

I. Introduction

THE sound processing technology is developing at an extremely fast pace, offering increasingly sophisticated solutions for professionals and amateurs alike. Among the many modern techniques, immersive sound systems, which aim to transport the listener into the three-dimensional sound space of the reproduced space, increasing the realism of perception, are attracting particular attention [1, 2]. In this context, the term telepresence is increasingly being used. The topic of this article is therefore particularly relevant in the context of the increasing supply of content in immersive formats via telecommunication systems, the limited availability of multi-speaker immersive audio systems on the recipient side and the widespread use of headphones as the final channel for accessing audio content.

The possibility of experiencing sound immersion with headphone listening should be guaranteed by the increasingly widely available so-called two-channel binaural recordings or software rendering object-based immersive formats for headphone listening. Unfortunately, despite the assurances of the suppliers of ready-made recordings and/or rendering software, the listening experience offered is still imperfect, and its quality is judged differently by individual listeners.

A number of factors contribute to this, the most important of which are considered to be the individualized mechanism of spatial hearing [3, 4] and the influence of visual information on the perception of the surrounding sound space [5]. In the process of rendering immersive content into binaural format, errors resulting in poor sound externalization for headphone listening are primarily related to the use of HRTF databases non-specific to the listener [6] and the simulated, rather than real, image of sound reflections in the listening room [7]. This results in an inaccurate spatio-temporal distribution of the directions from which acoustic energy reaches the listener under real listening conditions.

This paper presents an algorithm for converting multichannel audio to a two-channel format, in order to guarantee, for headphone listening, an immersive audio experience that would be provided when listening to recordings in a good-class listening room equipped with a 7.1.4 loudspeaker system [8], which, at the current level of technology development, guarantees a satisfactory feeling of being surrounded or immersed by sound [9, 10]. To achieve this effect, the use of a convolution processor using binaural room impulse responses (BRIRs) [11] synthesized in a specific, novel way is proposed. To process the channel signals emitted by the individual loudspeakers of the 7.1.4 system, 11 versions of the BRIR corresponding to the emission of the measured forcing signal by each loudspeaker (excluding the subwoofer channel, which emits frequencies irrelevant to spatial localization [5]) have to be created.

Witold Mickiewicz and Kaja Kosmenda are with West Pomeranian University of Technology in Szczecin (e-mail: witold.mickiewicz@zut.edu.pl. kaja.kosmenda@zut.edu.pl).

II. OBJECTIVES

This paper presents an algorithm for immersive processing of a multi-channel recording for headphone listening based on the fusion of measurement data from two sensors: a pressure microphone and a sound intensity probe [12]. With these, the pressure and directional characteristics of the listening room were recorded. Appropriate segmentation of these data supplemented with data from the average HRTF databases enabled the synthesis of a set of impulse responses subsequently used in a convolution processor to create an immersive recording. The novelty of this approach lies in the use of a real rather than a simulated directional image of the acoustic energy propagation process in the room in combination with the non-individualized HRTF data bases. However, these bases are input parameters for the algorithm and can be selected from a number of available ones to best match the individual anatomical characteristics of the end user of the system. Ultimately, an individualized base can be used where the user has one [13].

This paper presents an algorithm for immersive multi-channel processing of a headphone listening recording based on data fusion from 3 different sources. In the proposed approach, there are some similarities to the methods described e.g. in [14, 15], but the main difference lies in the measurement sensors used and the number and location of sources. The first two are measurement data from two sensors: a pressure microphone and a sound intensity probe [12]. They are used to record the pressure impulse response of the room and its spatiotemporal directional characteristics. It allows the pressure response to be divided into time fragments (segments) in which acoustic energy reaches the listener from a specific direction. The individual segments are then convolved with HRTF data and assembled into continuous BRIRs, which are then used in the convolution processor to create a binaural representation of the multi-channel immersive recording. The novelty of this approach lies in the use of a real, rather than simulated, directional image of the sound energy propagation process in the room using sound intensity measurement in combination with any HRTF databases. These databases are input parameters for the algorithm and can be selected from many available to best match the individual anatomical features of the end user of the system. One can also use the individual HRTF database if it's available.

The evaluation of the algorithm's performance was carried out on the basis of direct comparisons of the impressions experienced by people from the test group. During the research, the sound field generated in the listening room by means of a loudspeaker system in the 7.1.4 configuration evoked an exemplary immersive sound impression in the listener and could be directly compared with the image presented by open headphones, which the tested listener wore all the time. In addition, by using a DAW, the prepared test signals can be seamlessly switched between speakers and headphones, so that the listener has no other clues as to the current sound source. It is also possible to dose visual signals by darkening the room or using acoustically transparent curtains to prevent the 7.1.4 loudspeaker equipment from being visually positioned.

III. PROPOSED METHOD

The acoustic signals that reach the eardrums and produce spatial sensations in real-life situations are the result of the interaction of acoustic waves emitted by sources distributed in 3D space with the boundary planes and obstacles in the listening room, and the interaction with the person himself: his head, pinnae and torso [16, 17, 18]. If these interactions can be modeled correctly, real-world effects can be simulated using a convolution processor in which the source signals are combined together with the impulse responses of the source-room-listener system shown above. These responses are Binaural Room Impulse Responses (BRIRs) and must be recorded by microphones placed in the listener's ears in the target listening room for all potential source locations.

The resulting *Phones* headphone signals for the left and right ears can be generated according to the relationships:

$$Phones_L = \sum_{i=1}^{N} (Source_i * BRIR_{L_i}) \tag{1}$$

$$Phones_R = \sum_{i=1}^{N} (Source_i * BRIR_{R_i}) \tag{2}$$

where: * - convolution, $Source_i$ - audio signal from given source, $BRIR_{Li}$, $BRIR_{Ri}$ - binaural impulse responses recorded in the listener's ears when the test signal is emitted from the source position $Source_i$

In this approach, the amount of source signals subjected to convolving and mixing is limited to speaker signals obtained, for example, from the output of a Dolby ATMOS renderer. This limits the N value to the number of speakers in the system, not the sound objects in the immersive input format. This further simplifies the processing of objects that change position. The research presented here used a 7.1.4 speaker system and the reference BRIR signals were recorded in the room shown in Figure 1 using the DPA 4560 core binaural microphone. The microphone itself and its location in the listener's ears are shown in Fig. 2. A logarithmic sine wave signal (frequency range: 20 Hz to 20 kHz, duration: 6 s) was used for all impulse response measurements [19].

Fig.1. Immersion sound laboratory of Faculty of Electrical Engineering of WPUT Szczecin.

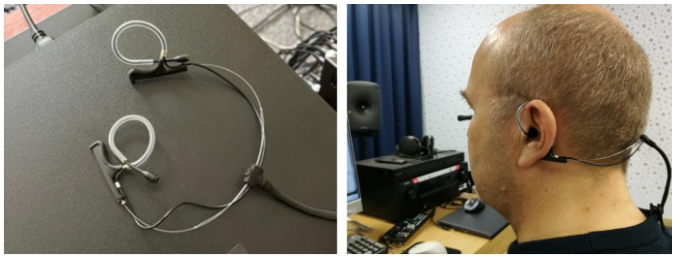


Fig.2. DPA 4560 core binaural microphone and its mounting on the listener.

A. BRIR synthesis - measurements

As mentioned earlier, the novelty of the presented approach (which distinguishes the presented method from the standard acquisition of BRIR responses described above) is the indirect synthesis of BRIR enabling separate acquisition of information about the location of the BRIR recording and the person who performs it. This allows the use of different HRTF databases in combination with measurement data obtained in a given room). The synthesis of the BRIR function will be performed on the basis of 3 independent components:

- actual directional sound intensity impulse responses of the room (SIIR);
- high quality omnidirectional room impulse responses (RIR);
- HRTF data set.

In the proposed approach one models the actual mechanism of BRIR formation when signals representing direct sound and reflections, which originate from different directions and are therefore filtered by HRTFs corresponding to different elevation and azimuth angles, interfere in the listener's ear. In order to synthesize BRIR using the proposed method, it is necessary to obtain information about the temporal evolution of directional acoustic energy in the room. In the proposed approach, we obtain them by sound intensity measurement. Microflown's commercial USP pressure - velocity (PU) probe was used for this purpose. It is a probe containing a pressurized microphone and three orthogonally arranged acoustic velocity sensors. The acoustic velocity is measured in this probe using the principle of hot wire anemometry. An enlarged image of the anemometer sensor and the location of individual sensors in the probe is shown in Fig. 3 [20]. The frequency range of the USP probe is from 20 Hz to 10 kHz.

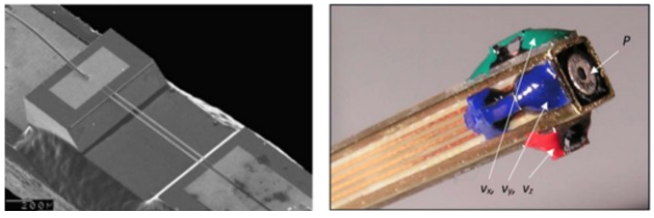


Fig.3. Wires of anemometry sound velocity sensor and construction of pressure-velocity intensity probe.

At this point, it should be noted that the probe used is expensive, which casts doubt on the economic aspect of the presented method. The authors used it only because of its accessibility, simplicity of use and the desire to test the idea of the proposed method. However, it should be mentioned that there are many more economical designs that could be used here, such as [21, 22, 23]. An example of a directional room impulse response that can be measured with an intensity probe is shown in Fig. 4. The individual rays indicate the directions of direct sound (thicker red ray) and individual reflections. The color scale allows you to reflect the temporal evolution. The direct sound corresponds to the red color and the last reflections to the blue color [24].

The next component to be measured is the classic omnidirectional room impulse response (RIR). Such a response is already obtained from the probe used, but the approach presented here uses a separate high-end omnidirectional microphone with recognized sound characteristics, a full audio frequency range and lower self-noise compared to the miniature electret microphone used in the Microflown probe (Schoeps MK5, frequency range 20 Hz - 28 kHz, equivalent noise floor A-weighted 12 dB). This has a direct impact on the final quality of the synthesized headphone signals. For comparison purposes, the BRIR responses were also recorded in the room using a dummy head Neumann KU100 head. The configuration of the sensor for measuring directional impulse responses and the artificial head is shown in Fig. 5.

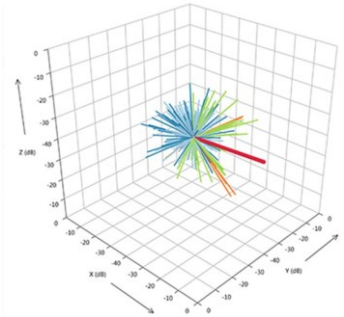
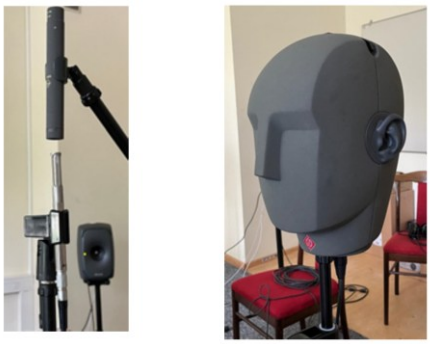


Fig.4. Directional sound intensity room impulse response in 3D. The arrival time is color coded from direct sound in red to latest reflections in dark blue.



a) b)

Fig.5. Sensors used in the study: a) pressure condenser microphone and PU-probe b) dummy head.

B. BRIR synthesis – segmentation algorithm

In the proposed BRIR synthesis method, the intensity response of the room is used to obtain information about the directions from which the acoustic energy reaches the listener's ears in subsequent moments of time. It was assumed that at certain intervals of time corresponding to more than one signal sample, the acoustic energy representing the discrete reflection arrives from one fixed direction. To identify these successive time intervals and directions, intensity averaging with a variable

width time window was used. The averaging process takes place over a period of time in which successive samples of the sound intensity signal have azimuth and elevation values that do not deviate from the spatial orientation of the sample considered at a given moment as a reference sample by more than the selected angular tolerance. The number of samples inside the window is therefore determined by the orientation of the first vector and the angular tolerance adopted. The averaged signal obtained by this algorithm preserves information about the reflections that occurred during the propagation of the sound. To speed up the above process, samples with too low amplitudes are treated as background noise and are pre-zeroed before the averaging process begins. The method used here consists in cutting off the signal below the experimentally selected threshold value. Fig. 6 shows the effect of the algorithm on the example of 2D data.

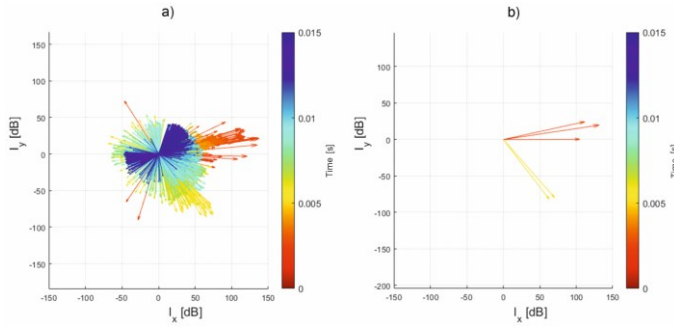


Fig.6. Determination of the main directions of acoustic energy arrival by averaging of the sound intensity impulse response: a) measured signal; b) signal after averaging.

The directional response signal of SIIR processed in this way was used to segment the RIR response. In this process, the pressure response of the RIR was divided into time segments corresponding to the fixed directions of arrival of the sound energy, and each segment was convolved with a two-channel HRIR function (the temporal form of the HRTF function) for a given elevation and azimuth value before being placed in the resulting BRIR.

Since the result of the convolution of the RIR segment with the HRIR is longer than the RIR segment, the folding of the resulting BRIR is done with overlaying, which naturally eliminates waveform discontinuities at the segment boundary and implements the crossfading process known in audio editing. The entire processing process is shown schematically in Fig.7.

IV. Experiment

To evaluate the performance of the algorithm, a subjective listening test was planned and performed in which a multichannel recording was alternately presented:

- through the loudspeaker system (the source designated in the table 1 as LS),
- through headphones without an externalising algorithm (the source designated in the table 1 as HP),
- through headphones using convolution with pre-recorded individual BRIR responses for a given listener (the source designated in the table 1 as HP-ALG1),
- through headphones using convolution with the non-individual BRIR responses recorded by the dummy head (the source designated in the table 1 as HP-ALG2),
- through headphones using convolution with synthesized BRIR responses according to the presented algorithm (the source designated in the table 1 as HP-ALG3).

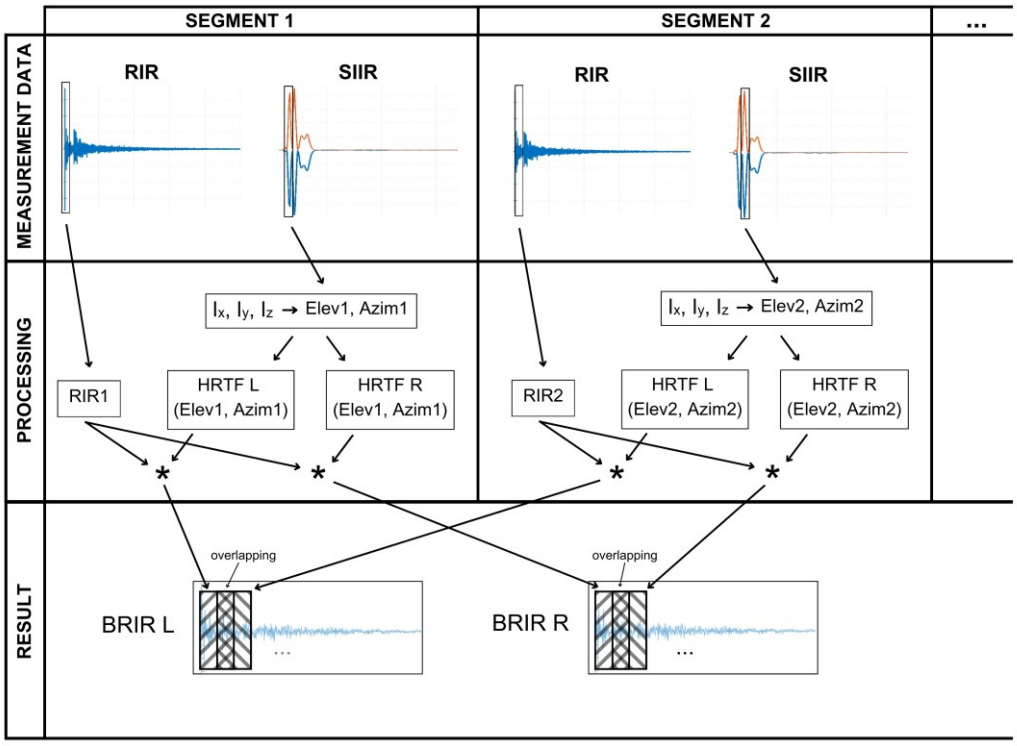


Fig.7. BRIR synthesis using SIIR, RIR and HRTF base (HRIR responses).

The signal processing block diagrams for the HP-ALG1, HP-ALG2 and HP-ALG3 algorithms are shown in Fig.8.

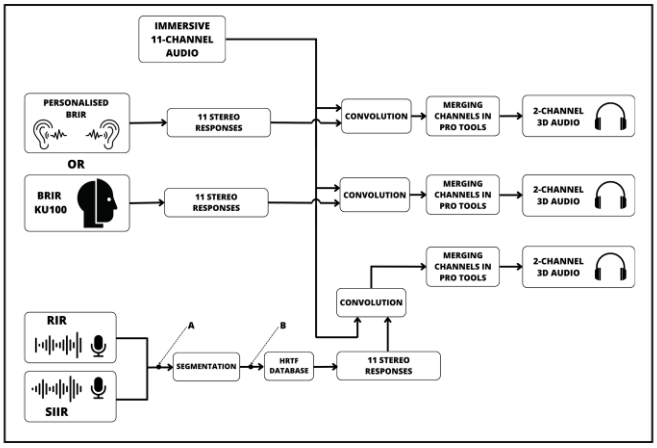


Fig.8. Types of processed headphone signals used in the subjective listening test.

Eight people took part in the test, three women and five men, aged between 20 and 26 years. The test subjects' task was to indicate what was the source of the immersive sound they heard. At this stage of the study, we were particularly interested in finding out whether the synthesized BRIR responses would ensure the disappearance of the internalization effect, which is present in unprocessed headphone signals but also in signals described as binaural and generated on the basis of HRTF applied only to the directions of direct sound arrival. Each listener was tested individually and occupied a position in the center of a circle with 7 loudspeakers in the surround layer. Throughout the test, the listener wore open-back headphones (model Sennheiser HD650) on their ears to allow both listening to the sound from the speakers and the headphones themselves. The details of the effect of wearing the headphones on the localization of the sound emitted from the speakers were not investigated at this stage. None of the listeners raised any concerns in this regard. The duration of the individual excerpts was about 10 s and the pauses about 2 s and was the same for all listeners. The entire test was emitted from within the DAW ProTools. An additional equalizer plug-in was applied to the headphone channels to introduce compensation for the change in timbre caused by the use of headphones. The starting point for the compensation settings was the frequency response of the HD 650 headphones available on the internet [25].

V. Results

The results of blind listening tests are shown in the table 1. At this stage of the research, the authors wanted to confirm whether the signals processed by the proposed algorithm would be able to evoke the impression of externalization and surround with sound identical or similar to the effects obtained with a multi-channel speaker system. As shown in the table, each of the algorithms used is able to evoke externalization impressions in the listener with similar effectiveness when compared directly with speaker listening. This initially proves the reasonableness of the created processing model and the correctness of its implementation. Closer conversations with individual participants also show that the use of the HP-ALG1 algorithm gives the best impression of immersion in the sound and is in line with modern knowledge about the importance of individual characteristics of the listener. On the other hand, the HP-ALG2 algorithm (dummy head), despite surprisingly good statistical results, was rated slightly worse in terms of immersion and the presented depth of the image – facts also known from the literature [4]. The proposed method (HP-ALG3) was evaluated at a similar level to HP-ALG2, which means that the processing process is correct and there is still untapped potential in terms of the ability to tune the algorithm to the individual user by being able to easily replace the HRTF database used. A statistical assessment of the effectiveness of such tuning will be the subject of further research. Finally, it is worth noting that of the subjects, 3 people had previous experience in listening to binaural content and, although they were not always confused, they confirmed the good processing quality of the proposed algorithm and its superiority over simple algorithms using only HRTF functions for direct audio processing.

Table I
Results of the subjective listening tests

No	Sex	Age	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
1.	M	22	HP	LS	HP	LS	HP	LS	HP	HP	HP	LS	HP	HP
2.	F	26	HP	LS	HP	HP	HP	LS	HP	HP	HP	LS	HP	LS
3.	F	23	HP	LS	HP	LS	HP	LS	HP	LS	LS	HP	LS	LS
4.	M	25	HP	LS	HP	LS	HP	LS	HP	LS	LS	HP	LS	LS
5.	M	20	HP	LS	HP	LS	HP	LS	HP	HP	LS	HP	LS	HP
6.	F	24	HP	LS	HP	HP	HP	LS	HP	LS	LS	HP	LS	LS
7.	M	23	HP	LS	HP	LS	HP	LS	HP	LS	LS	HP	LS	LS
8.	M	25	HP	LS	HP	HP	HP	LS	HP	LS	LS	HP	LS	HP
Actual	sound	source:	HP	LS	HP	HP ALG-1	HP	LS	HP	HP ALG-2	HP	LS	HP	HP ALG-3

VI. Conclusion

Analysis of the results of the research carried out in this way allows for a preliminary assessment of the usefulness of the designed sound processing technique, offering insight into potential directions for its further development and application. The research was aimed at providing new information and tools that can be used both in scientific research and in practical applications related to the creation of immersive audio experiences. As part of the preliminary research, the effects of the impulse response segmentation algorithm using a publicly available database of averaged HRTF responses were compared with the classic approach using individualized binaural impulse responses of the room and an artificial head. The results of the presented algorithm can be assessed as promising on the basis of initial subjective tests. In further work, a detailed statistical analysis will be carried out, especially with the use of several HRTF databases selected in accordance with the objective or subjective preferences of a given listener.

Acknowledgements

The authors would like to thank the volunteers who participated in the listening tests for their commitment and valuable comments during the discussions.

References

[1] Ed. A. Rogozinska, P. Geluso, “Immersive Sound. The Art and Science of Binaural and Multi-Channel Audio”, Routledge, Taylor & Francis, 2018

[2] E. Pfanzagl-Cardone, “The Art and Science of 3D Audio Recording”, Springer Cham, 2023, https://doi.org/10.1007/978-3-031-23046-2

[3] J.S. Bradley, G.A. Soulodre, “Objective measures of listener envelopment”, Journal of the Acoustical Society of America, vol. 98, pp. 2590–2597, 1995. https://doi.org/10.1121/1.413225

[4] J. Blauert, “Spatial Hearing The Psychophysics of Human Sound Localization”, The MIT Press, 1996. https://doi.org/10.7551/mitpress/6391.001.0001

[5] D. Reisberg, “Looking where you listen: visual cues and auditory attention”, Acta Psychologica, vol. 42, no. 4, pp. 331-341, 1978. https://doi.org/10.1016/0001-6918(78)90007-0

[6] W.S. Gan, S. Peksi, J. He, R. Ranjan, N. Hai, N.K. Chaudhary, “Personalized HRTF Measurement and 3D Audio Rendering for AR/VR Headsets”, AES 142nd Convention Paper, 2017.

[7] Ch. Pörschmann, P. Stade, J. Arend, “Binauralization of Omnidirectional Room Impulse Responses - Algorithm and Technical Evaluation”, in Proceedings of the 20th International Conference on Digital Audio Effects DAFx-17, Edinburgh, pp. 345-352, 2017.

[8] S. Agrawal, S. Bech, K. Barentsen, K. De Moor, S. Forchhammer, “Method for Subjective Assessment of Immersion in Audiovisual Experiences”, Journal of the Audio Engineering Society, vol. 69, pp. 656-671, 2021, https://doi.org/10.17743/jaes.2021.0013

[9] Kamekawa T, Marui A (2020) Evaluation of recording techniques for three-dimensional audio recordings: comparison of listening impressions based on difference between listening positions and three recording techniques. J Acoust Sci Tech 41:1

[10] Eaton C, Lee H (2022) Subjective evaluations of three-dimensional, surround and stereo loudspeaker reproductions using classical music recordings. Acoust Sci Tech 43(2):149–161

[11] W. Mickiewicz, J. Sawicki, “Spatial audio reproduction by headphones using binaural room impulse responses measured individually by the listener”, PAK, vol. 53, no 6, 2007. https://bibliotekanauki.pl/articles/156984

[12] F.J. Fahy, “Sound Intensity”, CRC Press, 1995.

[13] G.D. Romigh, “Individualized head-related transfer functions: efficient modeling and estimation from small sets of spatial samples”, presented at School of Electrical and Computer Engineering, Carnegie Mellon University, 2012

[14] M. Zaunschirm, M. Frank, F. Zotter, “BRIR synthesis using first-order microphone arrays”, AES 144nd Convention Paper, 2018. https://www.researchgate.net/publication/325392395_BRIR_synthesis_using_first-order_microphone_arrays

[15] P. Stade, J. Arend, Ch. Pörschmann, “A parametric model for the synthesis of binaural room impulse responses.” Proceedings of Meeting on Acoustics, vol. 30 no. 1, 2017, https://doi.org/10.1121/2.0000573

[16] D. Schröder, M. Vorländer, “RAVEN: A RealTime Framework for the Auralization of Interactive Virtual Environments,” Forum Acusticum, pp. 1541–1546, 2011.

[17] T. Lokki, “Physically-based Auralization - Design, Implementation, and Evaluation”, Ph.D. thesis, Helsinki University of Technology, Telecommunications Software and Multimedia Laboratory, report TML-A5, 2002, Available at http://lib.hut.fi/Diss/2002/isbn9512261588/.

[18] T. Lokki, H. Jaervelaeinen, “Subjective evaluation of auralization of physics-based room acoustics modeling”, in Proceedings of the 2001 International Conference on Auditory Display, Espoo, Finland, 2001 http://lib.tkk.fi/Diss/2002/isbn9512261588/article6.pdf

[19] P. Majdak, P. Balazs, B. Laback, “Multiple exponential sweep method for fast measurement of head related transfer functions”, Journal of the Audio Engineering Society, vol. 55, no. 7/8, pp. 623–637, 2007.

[20] H.-E. De Bree, “The Microflown: An acoustic particle velocity sensor”, Acoustics Australia, vol. 31, no. 3, pp. 91–94, 2003.

[21] J. Kotus, A. Czyżewski, B. Kostek, “3D Acoustic Field Intensity Probe Design and Measurements”, Archives of Acoustics, vol. 41, no. 4, pp. 701-711, 2016. http://dx.doi.org/10.1515/aoa-2016-0067

[22] W. Mickiewicz, M. Raczyński, “Modified pressure-pressure sound intensity measurement method and its application to loudspeaker set directivity assessment”, Metrology and Measurement Systems, vol. 27, no. 1, 2020.

[23] W. Mickiewicz, M. Raczyński, A. Parus, “Performance Analysis of Cost-Effective Miniature Microphone Sound Intensity 2D Probe”, Sensors, vol. 20, no. 1, 2020. http://dx.doi.org/10.3390/s20010271

[24] D. Protheroe, B. Guillemin, “3D impulse response measurements of spaces using an inexpensive microphone array”, presented at International Symposium on Room Acoustics, Toronto 2013. https://www.iris.co.nz/media/14459/ISRA2013.pdf

[25] https://www.sonarworks.com/blog/reviews/sennheiser-hd650-review#pros . Downloaded at 30.09.2024.