PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

The use of inertial measurement units in virtual reality systems for auralization applications

Kristian JAMBROŠIĆ[1]; Miljenko KRHEN[1]; Marko HORVAT[1]; Tin OBERMAN[2]

[1] University of Zagreb, Faculty of Electrical Engineering and Computing, Unska 3, Zagreb, Croatia

[2] University of Zagreb, Faculty of Architecture, Fra Andrije Kačića Miošića 26, Zagreb, Croatia

ABSTRACT

Auralisation finds increasingly more important applications in everyday life with the development and broad availability of virtual reality technologies and systems. This paper presents a new research project focused on auralisation implementation in different hardware systems, mainly by using binaural technology with an added head tracking system. Besides commercially available VR systems, the possibility of using inexpensive inertia measurement sensors connected to simple embedded systems will be investigated as an alternative auralisation system for adding natural aural experience in VR applications. In addition, the possibility of using widely available smartphones with already integrated movement and direction sensors for auralisation purposes using a binaural system will be examined. The listed platforms will be measured and compared considering their precision, latency and quality of the virtual sound field using practical auralisation examples in laboratory conditions, and in real life situation in-situ. Measurement plans for sensor parameters and characteristics, such as delay, sensitivity and accuracy, are presented in the paper for both IMU sensors connected to an embedded system and sensors integrated in smartphones.

Keywords: Virtual Reality, Auralization, Inertial measurement units

1. INTRODUCTION

The range of possible indoor and outdoor noise perception scenarios is so wide that novel techniques and methods must be designed to predict and model the variety of acoustical situations under a controlled experimental setup. It has been shown that the combination of virtual acoustic reality and computer-based acoustic simulation allows the study of sound in any environment, offering multiple research possibilities for urban designers, architects and acoustical engineers (1).

Auralisation is a growing research field within applied acoustics. It was defined in the 1990s, as the process of rendering audible the sound field of a source in both indoor and outdoor environments, using physical or mathematical modelling. This rendering is done by simulating the binaural listening experience at a given position in the modelled space (2). This concept has been further developed, having evolved into a research field known as Virtual Acoustics (3). Virtual Acoustics is part of a broader domain called Virtual Reality (VR), which unites sound with vision, as well as with tactile sensation (to a smaller extent) into the immersive experience.

The discipline of auralisation integrates methods of physics and acoustic engineering with psychoacoustics, auditory cognition and electroacoustic sound reproduction. The interest in auralisation has been enhanced in recent years, due to the advances in areas such as spatial sound reproduction, sound field analysis, and numerical methods. However, there are still many unsolved problems, including: modelling real sources for these environments, limitations on methods of computing the acoustic parameters for annoyance assessment, lack of evaluation benchmarks, uncertainty evaluation of auralisation models, etc. (4). The importance of auralisation and virtual acoustics has been recognized by governmental organizations and national funding bodies in Europe. They have funded several research projects related to auralisation and virtual acoustics.

[1] kristian.jambrosic@fer.hr

2. AURALIZATION TECHNOLOGIES

Currently, auralisation techniques are penetrating three main fields: the virtual world of computer games and smartphone applications, walk-through modes of architectural design programs, and the design of acoustically sensitive spaces. More applications are expected for predictive orientation to the auditory system, indoor-outdoor acoustics measurement, effects of noise exposure and potential social impacts, involving: urban acoustic design, social interaction, social and health impact, the creation of quieter environments enhancing speech perception, etc. (5, 6).

The scope of auralisation and audio in virtual reality is an inter- and trans-disciplinary field, so the usual channels for interchanging information are not effective. Contributions are needed from audio and acoustic engineers, architects and urban designers, computer and social scientists, physiologists, cognitive scientists, etc. New research in auralisation should focus on the application of innovative technologies for monitoring acoustic, psychoacoustic, and psychological parameters in urban areas, and on modelling these spaces with the aim of developing new solutions for indoor/outdoor noise-related problems. The combination of new acoustic modelling methods and immersive audio rendering techniques are expected to provide new auralisation-based tools for the assessment of acoustic impact in diverse environments. Although one could argue that in-situ measurements are the most realistic testing tool, many non-auditory factors can influence in-situ evaluations that can be avoided in laboratory experiments, such as excessive low or high temperature, strong daylight or night conditions, smell, etc. (7, 8).

Binaural hearing is essential for realistic auditory perception because it presumes the ability of the human hearing to perceive spatially separated sound sources, which is quite different from the perception of simple mono sound recordings. Head-related transfer functions (HRTF) are essential for simulating spatial hearing, and they are the basic tool for a wide range of virtual auditory display applications (guiding systems, game and mobile sound, room acoustic design, etc.). Although many datasets of HRTFs of acoustic mannequins already exist, different methods have been developed for measuring personal HRTFs, even in an inexpensive and/or fast manner (9).

2.1 IMU sensors

The binaural method of auralisation is used more and more in Virtual Reality (VR) systems and Augmented Reality (AR) systems (10). Dedicated commercial binaural systems with head movement tracking are quite expensive today because they are not produced in large batches, and require several sensors and a computer unit used exclusively for this task. At the same time, there are many experimental systems developed in recent years that are successfully used for head tracking applications in the process of auralisation. Such systems include ultrasonic trackers (usually 6 Degrees of Freedom, 6-DOF), simpler 1-DOF ITD systems for horizontal plane tracking, electromagnetic devices with 6-DOF and 3-DOF capable sensors, optical analogue devices with 2-DOF and 1-DOF, optical digital systems with 3-DOF and 2-DOF, inertial systems with 6-DOF (with and without optical reference), and one- or two-camera face-tracking systems with 6-DOF (11). All these sensors and systems are not equally suitable for auralisation purposes, but there is no detailed comparison of their advantages and disadvantages for binaural synthesis. Not all sensors are necessarily expensive; some of the systems are low-cost solutions connected to well-known embedded computer systems such as Arduino or Raspberry Pi (12). In all systems and technologies, it is essential that the latency of the head position measurement is read with a sufficient frequency, and that the data transfer and data processing does not take too much time. Otherwise, the feeling of transparency and non-existing lag is ruined for the listener who is experiencing the virtual auditory scene while turning their head (13). The common name used for many of these sensors is the Inertial Measurement Unit (IMU).

In recent years, a technology that integrates all functionalities in terms of motion sensors has become widely available – smartphones. Nowadays, they can be found in virtually every home, and their number surpasses the number of people using them. Smartphones are already equipped with several sensors that can be readily used for real-life calculation of their position and orientation in space, such as accelerometers, magnetometers and gyroscopes, all built in space- and money-saving MEMS technology, as shown in Figure 1.

An accelerometer is a sensor for measuring the acceleration, e.g. the change of velocity over time. In smartphones, it is built as a MEMS sensor (Micro Electro Mechanical System). A gyroscope is a device that consists of a spinning disc, which can rotate freely around its axis, regardless of the orientation of the housing frame. In smartphones, the gyroscope is again designed in MEMS technology. Vibration structures are used. A magnetometer measures the magnetic field in space, the compass being its most primitive version. The magnitude of the magnetic field is often required, along with the direction of the magnetic field. For this reason, the Hall Effect is used. The Hall sensors are too big to fit into an average smartphone. Therefore, the magnetometer is also made in MEMS technology. It uses the action of Lorentz forces on the MEMS element, which causes a conductor with current flowing through it to move when if finds itself in a magnetic field.

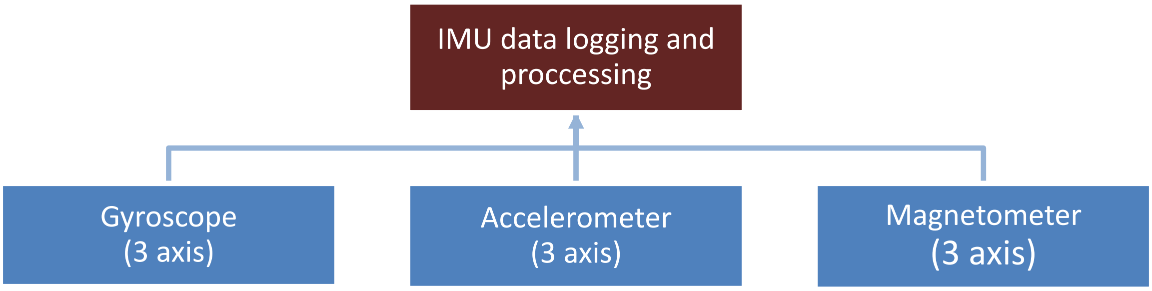

Figure 1 – Typical sensor types found in IMUs used for head tracking in auralization

Smartphones are a good choice for head movement detection if they are firmly fastened to the head(phones). Since they integrate all required sensors, the only task is to extract the positional data and transfer them to the computer that performs the binaural synthesis. At the same time, smartphones are already computers with enough processing power to render binaural signals and scenes on the fly. There can be many applications of these systems, such as multi-party conference systems, pedestrian navigation (12), or even mapping the sound field with the built-in or external microphones for any acoustic measurement (14).

Many more applications of auralization exist, and many more are coming in near future. This is especially true for music production that desperately searches for a way-out from the obsolete stereo system that has been around for over 60 years. Moreover, there has been no real breakthrough with the home-cinema surround sound systems as a mere upgrade to the old 2-channel stereo audio systems (15). Health application that uses augmented reality for visually disabled persons is also a new field for auralization (16). For people with the need of wearing hearing aids, auralization can be used for many augmented reality applications (17).

3. RECENT RESEARCH IN THE FIELD

In order to understand the auralization process, one must understand the principles of room acoustic measurements and computer simulation. This was also one of our research topics where ray-tracing parameters were harmonized in order to achieve realistic room impulse responses in computer simulations, essential to experiencing an auralized space as realistic (18). Auralization of closed spaces is essential for many applications, such as making an aural experience of any new or reconstructed room for architects, investors, historians, nonprofessionals, but also for visually impaired people who rely completely on their hearing rather than on vision.

Two studies showed how the interior acoustic design of spaces directly influences the way we experience the size of the room and our position in it, the so-called self-localization experience (19, 20). This is achieved by choosing various kinds of wall, floor and ceiling materials with different sound absorption and sound scattering characteristics. The idea was to make some guidelines for architects in order to support visually impaired and blind persons entering public buildings.

Another very recent research topic was connected to the use of auralisation as a part of a holistic, augmented reality experience. Sound is used here as a mechanism for path following and trajectory tracking of remotely operated vehicles. Using this technique, the visual load for the operators is reduced. Using more senses together gives the operator a more harmonized cognitive load when being exposed to tasks that require a long-lasting mental concentration (21).

Finally, a pilot project on how to use head tracking systems in virtual audio environments for a real-time binaural synthesis. This is again essential for receiving a well-balanced and realistic auditory experience of a simulated reality, or when simply playing back a recorded sound event that already happened in the past, but we want to experience it now (22).

4. RESEARCH METHODOLOGY

This project proposal is a continuation of previous research done by the research team. The main research outcomes expected in the field of auralisation are the precisely defined limitations and constraints connected to various auralisation systems. Firstly, limitations of head tracking of the listeners in order to achieve a seemingly transparent aural experience that completely immerses the listener in the virtual reality surrounding, at least in terms of hearing. Secondly, the comparison of the binaural systems with implemented head tracking with available loudspeaker-based auralisation technology with dedicated auralisation software, such as Ambisonics system, or similar.

The main difficulty of audio technologies that are sensitive to head movement today is the speed of head tracking, i.e. the latency of the system when freely moving the head and looking around the virtual surrounding. This means that any latency larger than approximately 20 ms is experienced as a "drag" of the sound field behind the movement. This is no noted issue for commercial VR systems, but they are certainly not affordable for a wide group of potential beneficiaries and users. Therefore, the usability and optimization of using inexpensive IMU units connected to embedded systems as an interface towards computers rendering binaural scenes is expected in this project, as well as a comparison with commercially available systems.

The final goal is to compare these two systems with smartphones that generally possess all the required computational power and positioning sensors, but are not widely used for auralisation purposes. Smartphones are very favorable because of their existence in virtually every home, and certainly for the great majority of people today. This is already shown in the usage of cheap VR system (like Google Cardboard) where the visual part of VR immersion can be achieved to a certain extent, but not the aural one, at least not uniting it with head tracking which is essential for good immersion and proper auralisation, as mentioned before.

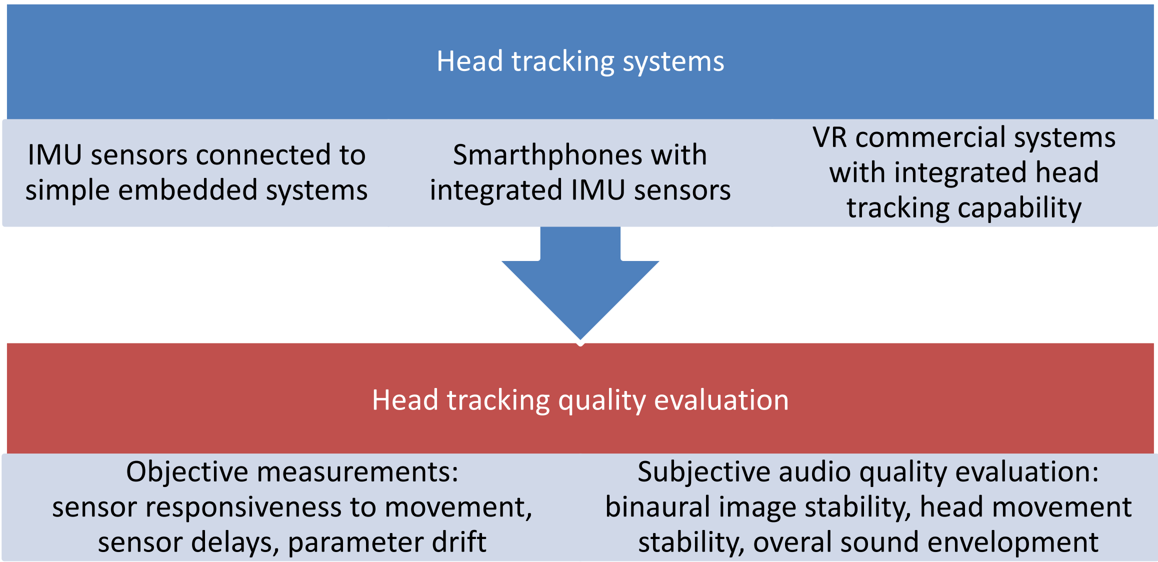

Figure 2 – Basic schematics of the research plan

A lot of research time will be dedicated to measurements of various parameters of the sensors. This includes measurement of the precision of absolute and relative position and orientation, their pitch, roll and yaw values, and their responsiveness to movements depending on the speed of movement, travelled distance and time offset, and parameter drift. Existing measurement equipment available at the Faculty of Electrical Engineering and Computing, and the Auralization Laboratory in particular, will be used for precise measurements of these parameters. Measurement data will be statistically analyzed and repeatability and uncertainty of the measured data will be calculated. The general overview of the research plan is shown in Figure 2.

Another key objective is auralisation by computer simulation and numerical methods. For this, several levels of auralisation algorithms will be used – from simple coding of binaural synthesis using mathematical convolution (for stationary and time-varying signals) to dedicated software for auralisation of virtual sound environments using position data from sensors and advanced algorithms for binaural synthesis. All this will be done for the available hardware platforms, and when possible, for multiple sets of HRTFs. The goal is to compare the speed of execution, the latency introduced by the software itself, and all audible artefacts that occur because of filtering, averaging, and other data processing. Such processing is required because HRTFs are always recorded in discrete angular positions, while binaural synthesis must provide a smooth aural experience when the head of listeners is in motion.

The final key objective is to test and compare all hardware and software systems. For this, several aspects are needed, comparing measurable delays and sound localization precision, where team members will evaluate the precision and overall impression of the systems regarding several measurable parameters and using semantic difference scale in order to obtain usable statistics. This includes the evaluation of direction (azimuth and elevation), overall satisfaction with the rendered auditory scenes, the feeling of involvement in the scene, sound envelopment, spaciousness, lag, artefacts in sound coloration and distortion, etc. The evaluation of the auditory quality in the VR systems can be added with the visual scene which changes the overall evaluation, as known from literature. In the end, auralisation of binaural synthesis with head tracking will be compared with simple binaural recordings without head tracking ability, and finally with auralisation using loudspeakers in systems such as Ambisonics. All these systems will be compared using the same semantic difference questionnaires in order to obtain a relevant conclusion on the quality of the system.

Since auralisation with smartphones is a mobile system used in-situ, in many different surrounding outside laboratories, the final field tests will be done. All parameters of auralisation quality will be evaluated by the same group, i.e. the team members of the project. Now they will be exposed to other type of surroundings, including noisy environments where many sound sources emit noise at the same time, thus changing the overall impression of the auralisation quality and auralisation immersions. Both evaluations – in laboratory and in the field, will be compared, and the usability of the proposed system for field applications will be assessed.

5. EXPECTED RESULTS

In terms of general impact of this project, the outcomes can directly influence a broad community that extends far beyond the interested scientific community. VR technologies are currently being developed in an accelerated pace, and this project focuses on the use of smartphones with all embedded sensors for real-time binaural auralisation with many potential applications. It is reasonable to expect that the outcomes will stimulate a lot of follow-up research and development in the field of new games for smartphones, new possibilities for visually impaired users, improved intelligibility for video and teleconference systems (due to spatial distinction of speaker positions), etc. Real-time auralisation of indoor and outdoor spaces is expected to become more and more important to architects, urban designers, urban planners, decision makers, etc. The possibility will be opened for participative sensing and evaluation of environmental sounds and soundscapes by a large group of people, thus potentially participating in decision making, mobile television within the VR systems including both audio and video immersion, etc.

Thus, the group of potential beneficiaries from the project outcomes goes far beyond the scientists and researchers in the field. As the research is interdisciplinary by nature, any positive development in the field opens new possibilities of application to new groups of beneficiaries, from small children to adults, from educational purposes to leisure applications, from single assisted aids for disabled to participative application that are used by many.

The impact of the proposed project can be notable and important, having in mind the potential usability of auralisation for a wide group of people, i.e. the owners of smartphones. This could involve many participants in all kinds of subjective assessment tasks, first coming to mind the question of noise annoyance level depending on the soundscape, and VR systems are an ideal platform for this. Participation of citizens in policy-making and evaluation increases involvement in reaching a sustainable European society. It can also be the basis for creating European societies of citizens, which in turn can help strengthen social cohesion. Focusing on the local living environment and on noise can be a deliberate choice. In contrast to other environmental components, environmental sounds can be easily heard and appraised by people. It is discussed on many internet forums, and can become a point of debate: certain sounds are loved by some, and hated by others.

6. PILOT STUDY MEASUREMENTS

Measurements of the quality of head tracking systems have already started. A simplified system has been used built out of an Android-based smartphone mounted on top of headphones, thus serving as a head tracking device, an open access software for collecting head movement data, and a PC running a Matlab-based application for real-time binaural synthesis based on the listener's head direction. For smartphone (head) position measurements, the following IMU sensor data collection apps were used: Wireless IMU and freePIE. All used applications are released under ISC and GNU licenses (22).

6.1 Wireless IMU, freePIE and opentrack

Wireless IMU sends data from sensors using UDP (User Datagram Protocol) to the computer, Figure 3 left. The program supports the collection of data from three sensors: accelerometer, gyroscope and magnetometer. The UDP packages are sent as numerical values. The speed of sending packages can be set from the slowest 380 ms to the fastest 5 ms. The best results were achieved with 20 ms because at 5 ms, only every other magnetometer value was sent. The UDP packages are read by a Matlab function, and the collected data was further used for calculating the position of the smartphone, e.g. the head orientation.

A screenshot of freePIE (Programmable Input Emulator) is shown in Figure 3 middle. It also collects sensor data from the smartphone and sends it to the computer. The freePIE sends packages with a speed of 20 ms to 3 ms. The packages can be collected on the computer using the opentrack program that processes the data and sends them to Matlab for computing the binaural audio signal.

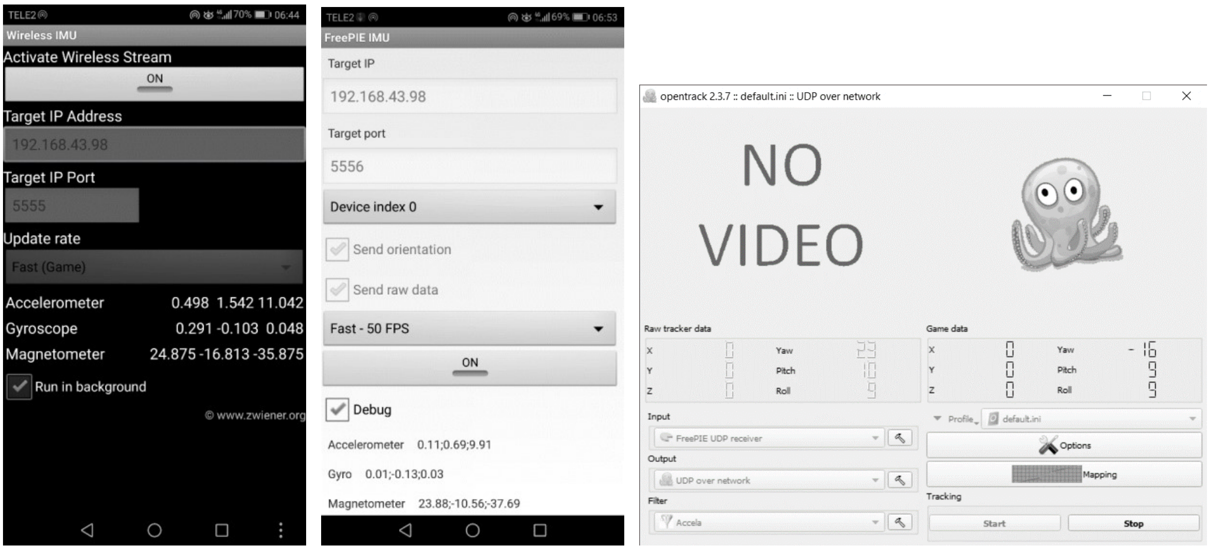

Figure 3. Wireless IMU interface (left), freePIE interface (middle) and opentrack interface (right).

The opentrack program is used for head tracking and sending the data to video games, flight simulators, etc., Figure 3 right. It supports PointTracker, Oculus Rift, Arduino, Android, and other protocols. Depending on the input types, the data are processed to calculate yaw, pitch and roll of the device. These parameters can be sent to specific software (e.g. Microsoft Flight Simulator), they can emulate some devices (joystick or mouse), or can be sent further as UDP packages. The opentrack sends a package every 100 ms.

Binaural synthesis is done in Matlab using a code written by the authors for this purpose (22). The first step is to choose the required HRIR's for the binaural synthesis. The CIPIC HRTF database was used.

6.2 Head tracking quality evaluation

For testing the time required for the execution of the code, a 5-second pure sine signal with 44.1 kHz sampling frequency was used, generated with MATLAB. The average speed is shown in Table 1. From the measurements, it can be concluded that freePIE has a shorter processing time by about 20 ms. This is not unexpected because the UDP packages already include information about the orientation, so additional processing is not needed. In contrary, the concept of Wireless IMU slows down the execution of the synthesis because of additional processing time required to calculate the position parameters.

The testing of the subjective quality of head tracking was done using short music excerpts sampled at a frequency of 44.1 kHz. During the testing, the overlap percentage was varied, the size of the frame, the refresh rate of the UDP packages and the speed of head movement. It was found that the sound artefacts became inaudible at approx. 15 % overlap for music excerpts.

Table 1. Average execution time.

	Wireless IMU [ms]	freePie [ms]
Average time	5140.3	5121.9

It became clear that the main advantage of the Wireless IMU application is that there are no problems in making a full circle, e.g. crossing the azimuth 180°. This is due to the way that only the relative change of orientation is coded in the application. At the same time, the freePIE algorithm tracks the head movement much better. Even at fast head movements there are no delays in the sound image. The problem occurs only at sharp turns of the head. Then it becomes obvious that the sound field needs some time to position in its new direction.

7. CONCLUSIONS

This project will focus on research of auralisation implementation in different hardware systems, mainly by using binaural technology with an added head tracking system. The focus will be on head tracking systems, namely to investigate the possibilities, limitations, shortcomings and delay of three very different hardware platforms that can be used for auralisation in the binaural field. These platforms are 1. simple IMU units connected to inexpensive embedded platforms, 2. smartphones that already have integrated all needed sensors, and the required computational power, and 3. commercial virtual reality systems.

It is expected that the project will contribute to the scientific community, but also to general community and wide beneficiary groups. By showing the possibility to use auralisation with already widely available devices, a number of new applications can be expected in many different fields.

The research is highly interdisciplinary, so any positive development in the field will lead to new possibilities of application by new groups of beneficiaries.

ACKNOWLEDGEMENTS

The authors acknowledge financial support by the Croatian Science Foundation, (HRZZ IP-2018-01-6308, "Audio Technologies in Virtual Reality Systems for Auralization Applications (AUTAURA)".

REFERENCES

1. Savioja, L, Huopaniemi, J, Lokki, T, Väänänen, R. Creating Interactive Virtual Acoustic Environments, J Audio Engineering Soc. 1999;47(10):675-705.
2. Vorländer, M. Auralization - Fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality. Berlin, Springer-Verlag; 2007.
3. Lokki, T, Savioja, L. Virtual Acoustics, Handbook of Signal Processing in Acoustics (Ed. Havelock, D, Kuwano, S, Vorländer, M). New York Springer Verlag; 2008, p. 761-771.
4. Rumsey, F. Sonification, assistive listening, and soundscapes, J Audio Engineering Soc. 2017;65(7/8):652-656.
5. Altman, M, Krauss, K, Susal, J, Tsingos, N. Immersive Audio for VR, Audio Engineering Society Conference on Audio for Virtual and Augmented Reality, Los Angeles, 2016.
6. Rumsey, F. Binaural audio and virtual acoustics, J Audio Engineering Soc. 2017;65(6):524-528.
7. Rychatrikova, M, Vermeir, G, Vorländer, M. Laboratory Listening Tests in Building and Room Acoustics, European Symposium on Harmonization of European Sound Insulation Descriptors and Classification Standards, Florence 2010. p. 1-8.
8. Rychtarikova M, Van den Bogaert T, Vermeir G, Wouters J. Binaural sound source localization in real and virtual rooms, J Audio Engineering Soc., 2009;57(4):205-220.
9. Reijniers, J, Partoens, B, Peremans, H. DIY measurement of your personal HRTF at home: low-cost,

fast and validated, 143rd Audio Engineering Society Convention, Convention e-Brief 399, New York, 2017.
10. Park, J. W. Audio Localization Method for VR Application, 143rd Audio Engineering Society Convention, Convention e-Brief 400, New York, 2017.
11. Hess, W, Head-Tracking Techniques for Virtual Acoustics Applications, 133rd Audio Engineering Society Convention, Paper 8782, San Francisco, 2012.
12. Väänänen, R, Vesa, S, Hämäläinen, M. Testing the User Experience of an Augmented Reality Headset and 3D Audio-Guided Pedestrian Navigation, Audio Engineering Society 55th International Conference, Helsinki, 2014.
13. Stitt, P, Hendrickx, E, Messonnier, J-C, Katz, B. The Influence of Head Tracking Latency on Binaural Rendering in Simple and Complex Sound Scenes, 140th Audio Engineering Society Convention, Paper 9591, Paris, 2016.
14. Hawley, S H, Alegre, S, Yon, B. Sound Fields Forever: Mapping sound fields via positionaware smartphones, 143rd Audio Engineering Society Convention, Convention e-Brief 401, New York, 2017.
15. Faller, C, Menzer, F, Tournery, C. Binaural Audio with Relative and Pseudo Head Tracking, 138th Audio Engineering Society Convention, Paper 9223, Warsaw, 2015.
16. Kolarik, A J, Cirstea, S, Pardhan, S, Moore, B C J. A summary of research investigating echolocation abilities of blind and sighted humans, Hearing Research, 2014;310:60-68.
17. Oreinos, Ch, Buchholz, J M. Objective analysis of ambisonics for hearing aid applications: Effect of listener's head, room reverberation, and directional microphones. J. Acoust Soc Am, 2015;137(6): 3447-3465.
18. Plašć, V, Jambrošić, K, Domitrović, H. Influence of Ray Tracing Parameter Variation on Simulated Room Impulse Responses, Proc Euronoise 2012, p. 1099-1104.
19. Horvat, M, Jambrošić, K, Francetić, J, Domitrović, H, Rychtarikova, M, Chmelik, V. On the Ability of Normal Sighted Persons to Assess Room Size and Position Inside the Room Based on its Acoustic Response, Akustika, 2015;24:2-8.
20. Chmelik, V, Francetić, J, Jambrošić, K, Horvat, M, Domitrović, H, Rychtarikova, M, Glorieux, C. Perception of room size and ability of self- localisation in a virtual environment: headphone experiment, Proc Forum Acusticum 2014. p. 1-5.
21. Vasilijević, A, Jambrošić, K, Vukić, Z. Teleoperated path following and trajectory tracking of unmanned vehicles using spatial auditory guidance system, Applied acoustics, 2018;129(1): 72-85.
22. Trepšić, M, Jambrošić, K. Virtual Audio Environment Synthesis Based on Headtracking Systems, Proc ATF 2017 5th Conference on Applied Technology, 2017. p. 239-245.