On the Usability of Directional Information through Bone Conduction Headphones in a Virtual Reality Environment

György Wersényi
Department of Telecommunications
Széchenyi István University
Győr, Hungary
wersenyi@sze.hu

Ádám Csapó
Department of Computer Science
Széchenyi István University
Győr, Hungary
adam.csapo@gmail.com

***Abstract*—Current Virtual and Augmented Reality solutions offer spatial audio rendering and navigation in real-time using 3D scenarios. The MaxWhere framework allows users to define 3D spaces integrating 3D objects and 2D content for various tasks, mostly focusing on education, training and rehabilitation. As part of an international research project targeting the visually impaired, bone conduction headsets were tested in the virtual reality (VR) scene to test usability of the headphones in virtual environments for assistive technologies. This paper presents results of a localization task restricted to the horizontal plane, as well as the evaluation of the capabilities of the applied rendering and programming methods.**

Keywords—directional information, bone conduction, headphones, virtual reality, MaxWhere

I. INTRODUCTION

In the past years, the study of spatial audio has expanded beyond the implementation of classic localization tasks over speakers and headphones. Today, spatial audio is often used with alternative playback systems (bone conduction, multi-speaker headphones etc.) and to investigate and stimulate higher-level cognitive functions (broadening capabilities of the visually impaired, military applications, simulators etc.) [1].

Furthermore, VR and AR applications are increasingly using spatialized audio not only for gaming and simulators, but also in the areas of teaching, rehabilitation, training and more. State-of-the-art solutions offer easy-to-use programming APIs, fast processing and thus, can reach a wider target audience than ever before.

The Internet of Digital Reality has introduced many aspects of the higher-level connections between cognitive entities and digital realities using VR interfaces, which brings to the forefront new challenges in a variety of fields including human factors and modeling of cognitive capabilities [2]. Traditional measurements and tests for the evaluation of spatial audio solutions can be moved into this space, especially since audio/video devices and headsets will play a significant role during everyday use in the future. Given the broadening of the user audience, it is all the more important to understand how less traditional devices can provide useful information to untrained users in unfamiliar VR spaces.

A. *Virtual Reality Applications and Scenarios*

Traditionally, VR has often been regarded as an environment for flashy 3D simulations that supports gaming and (primarily social) entertainment applications. Recently, however the VR scene has expanded considerably, with the appearance of a plethora of web-based platforms (including JanusWeb, Mozilla Hubs and Spatial.io) that allow users to upload, display and share 2D content besides create 3D objects and avatars. Fig.1. shows examples created in MaxWhere.

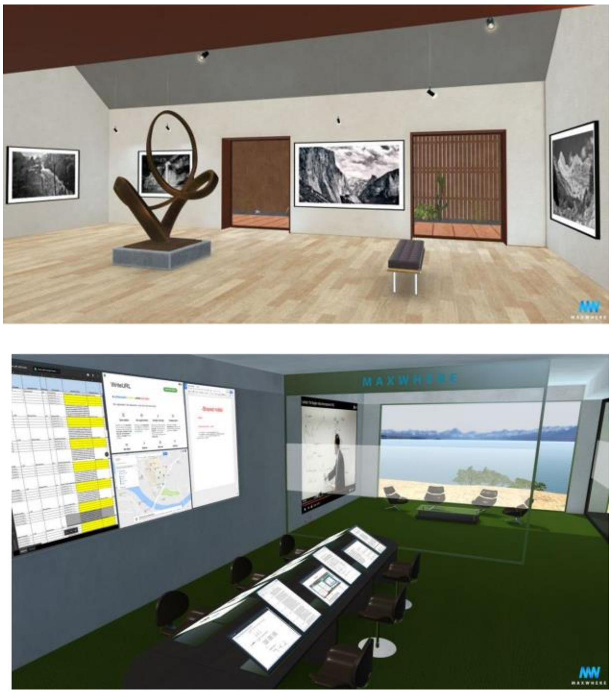

Fig.1 Two examples showing 3D objects and smartboards in MaxWhere.

The MaxWhere VR platform is an extensible desktop VR solution that was created primarily for productivity, showcasing and VR-learning applications [3, 4]. MaxWhere includes a cloud service for publishing and sharing VR spaces (which are VR apps in their own right and can be downloaded with suitable user privileges). The spaces integrate both interactive 3D objects and 2D display panels (called smartboards) which feature a Chromium-based browser engine, thus supporting images, PDF documents, audio and video besides (static and dynamic) web documents. Each space is represented inside the runtime as a document tree (referred to as the Where Object Model – or WOM – by analogy to the classic Document Object Model) that is dynamic and hence programmable in real-time.

Based on its generality, extensibility and versatile capabilities for integrating all areas of users' digital life, a platform like MaxWhere can be considered as a key framework for implementing digital realities [2]. From a more narrow viewpoint, it is also a suitable testbed environment for usability experiments centered around unconventional interaction devices.

B. Bone Conduction

Standard audio headphones cover the ears of the listener and thus may impair the perception of ambient sounds. Bone Conduction (BC) headphones offer a possible alternative, but traditionally their use is limited to monaural applications due to the high propagation speed of sound in the human skull. These devices are advantageous for mobile guidance for blind people (travel aids), during outdoor sport activities where environmental sounds are a safety issue and in VR/AR applications (military simulators, gaming).

In case of the visually impaired, applications require an audio-based user interface as well. Special focus in the VR section could be training, where different scenarios (2D and 3D spaces) can be created. The project Sound of Vision was aimed at developing a wearable travel aid for the visually impaired 5, 6]. Although the prototype includes a multi-speaker headphone device, bone conduction and alternative headphones were also taken into consideration.

Hearing thresholds are usually different by using BC type equipment and even a lowering of 10–20 dB in thresholds can be observed when skin penetration is performed [7, 8]. Skull locations are highly sensitive to bone conduction auditory signal reception [9]. The condyle has the lowest mean threshold for all signals followed by the jaw angle, mastoid and vertex. The condyle would be the most effective location based on sensitivity, however, higher audio output levels can make other locations effective as well.

Walker et al showed that stereo bone-conduction headsets can be used to provide a limited amount of interaural isolation [10]. Results suggest that reliable spatial separation and externalized spatial audio is possible with bone-conduction headsets, but that they probably cannot be used to lateralize signals to extreme left or right apparent locations.

Spatial simulation in the horizontal plane and vertical displacement simulated with adjusting the tone's pitch can result in better localization performance even using BC (such as used in [11] and [12] or in the ActiVis project [13]). High-pass and low-pass filtering of sounds is a useful tool to increase the number of correct answers in vertical localization tasks without HRTF filtering [14]. HRTF filtering is meant to be for standard headphone use. However, it is an interesting question, how they contribute in case of BC headsets [15].

Analyses of the transmission characteristics show an uneven frequency response of bone conduction headphones compared to conventional headphones or speakers.

Listening to distracting sounds over bone conduction may still disrupt a listener's awareness of their auditory environment [16]. As expected, participants have greater localization error in the distractor-present conditions, due to masking effects, and was more important for narrow-band targets.

Mixing real and computer-generated audio for AR applications can use some kind of Hear-Through and Mic-Through audio as well. This may include BC devices or headphones using microphones, usually used for noise cancelling (active noise cancellation, ANC). The goal is to leave the ear canal open or at least, letting environmental noise and sounds passing through. Mic-Through AR allows audio signals to be mixed with computer-generated audio in the computer, and delivered both to the user over headphones.

BC headphones was tested also for the military, where headphones are the standard presentation devices for radio communication [17]. This particular study tested the feasibility of a multi-channel bone conduction system by measuring the localizability of spatialized auditory stimuli presented through a pair of bone conduction vibrators. Localization performance was found to be nearly identical for BC and traditional audio systems, indicating that bone conduction systems can be effectively used for displaying spatial information.

This paper presents first results of selected BC devices. Aim of the study was to create a versatile, easy-to-use headphone testing VR scenario, with focus on currently available, newly designed low-cost BC headsets. Thus, testing is aimed at examining the VR environment for such purposes and the devices itself.

II. MEASUREMENT SETUP

In the experiment 20 volunteers participated. Subjects were seated in front of the computer screen in a silent laboratory room wearing a BC headphone.

Fig.2. The AudioBone and the AfterShokz headphones.

Fig.2. shows two different type of devices that were used: the AudioBone (the same that was used in [18], and AfterShokz (wired version) [19]. They were selected as possible equipment for visually impaired and for scenarios where the ears can not be covered during operation. The AudioBone was used already in other experiments and the

AfterShokz is one of the most popular device nowadays (usually for sport activities). Both are cost-efficient and can be placed on the jaw bone instead of behind the ears (as former models). Transfer characteristics are not supplied by the vendors and were not measured (due to the special requirements of measuring BC headphones). For everyday use e.g. during navigation tasks having limited directional information anyway this is not of great importance.

The localization task included identification of sound sources in the horizontal plane using 3 (0°, ±90°), 5 (0°, ±90°, ±45°) and 7 (0°, ±90°, ±60°, ±30°) locations in the frontal hemisphere. Stereo panning was used for directional rendering (Fig.3.). Short, iconic samples of white noise, 1 kHz sine and female speech were used for stimuli.

The scene was created in MaxWhere, and the audio feedback was implemented using the ToneJS library - a library for sound rendering, allowing 2D horizontal or vertical displacement as well as a 3D spatialization of sounds.

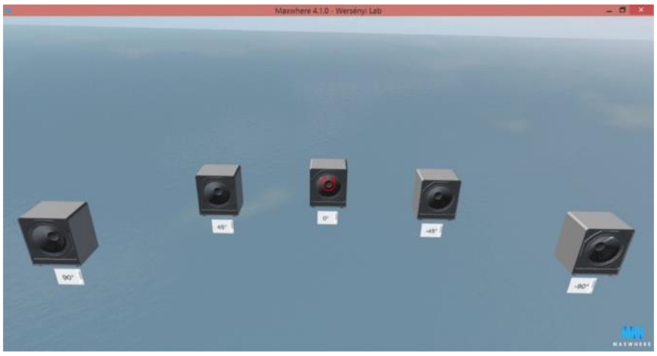

Fig. 3. Screenshot of the headphone testing application in MaxWhere. In this test scenario, users have to identify five sound sources in a 45-degree spatial resolution in the frontal hemisphere.

Every stimulus was played back once (non-looped) in randomized order over the source locations. One measurement round included stimulus presentation at every location three times (unknown to the subjects), thus, 9 answers for 3x1, 15 answers for 5x1 and 21 answers for 7x1 were collected. Total time for the game and error rates (%) were recorded.

III. RESULTS AND DISCUSSION

A. Evaluation of the listening test

The experiment allows comparison of different

- types of BC equipment (3),
- types of stimuli (3),
- resolutions, number of sources (3),
- and presentation method: looped or single (2).

Testing all possibilities would result in 54 different scenarios, thus, only a limited number of tests were carried out. Using only two wired BC devices with different users and non-looped presentation reduces scenarios to 9/subject allowing indicative results to be collected about the effect of *resolution* (number of sources) and *sound types*. Including other parameters, such as looping or more devices would have increased the number of possibilities (see future works).

We can support former findings about the effect of placement of the transducers on the head, as subjective experience and experienced transmission quality varied upon the location (even swallowing can move the transducer on the skull bone and left/right symmetry is hard to maintain).

Comparing directions showed that 3 directions (center, left and right) can be determined with 100% accuracy independent of stimulus (9/9 points). In case of 5 source locations, 45-degree directions are often confused with the side and center at the beginning, but subjects learn quickly the difference during the test sessions (12-14 out of 15 points). If we have 7 sources, locations of 60 degrees tend to be confused with 90 degrees, and 30 degree locations with the center direction (15-17 out of 21 points). As expected, the more directions the more difficult the task becomes.

Results are affected by the fact, in which order the random generator activates sound sources: if neighbor sources are presented after each other, discrimination is easier by comparing and remembering than in case of contralateral presentation, where it is more difficult to identify sources correctly. Furthermore, experience (effect of training) can also bias results: subjects' performance increased after the first test runs. This needs to be investigated more in detail with repeated measurements designed focusing on learning effects.

Comparing stimuli, best results (number of correct hits) can be achieved with the female voice, followed by noise and sinus. This ranking was expected, as results of listening tests (localization tasks) using familiar natural sounds (especially human speech) are in general better than unnatural sounds, such as noises or artificial test signals. Furthermore, bandwidth contributes to better results: signals having larger bandwidth are easier to localize.

B. Evaluation of the VR scenario

The developed VR space was deemed useful by the authors and the test subjects alike for the evaluation of spatialized sound production capabilities. Given that the space integrated all functionality relevant to the tests – including test subject data input, test configuration, as well as implementation of tests and recording of test performance – using a standard JSON format – the space served as a compact environment with all the relevant tools available.

IV. CONCLUSIONS

This paper presented first results of a listening test with different bone conduction headphones using panned spatial audio in the horizontal plane in a VR environment. Results indicate correct localization using three directions, and decreased localization performance with 5 and 7 sources, suggesting that 5 source locations can be used with high accuracy (above 85%). Training and placement of the BC device on the head is critical.

The VR environment created in MaxWhere was found to be appropriate for audio listening tests as well. The framework offers easy programming and inclusion of different scenarios, data collection and evaluation in the same space. Testing of Digital Reality applications including VR scenes and connecting devices can contribute to adequate use and information transfer using Internet of Digital Reality [20].

V. Future Work

Future work includes introducing more variables and statistical analysis (ANOVA) of results, implementing distractor sounds in the 3D space, testing the usability of the VR space for various auditory training (including serious game solutions for crowdsourcing) and real-time experiments with visually impaired.

Acknowledgment

This research was supported by the Digital Development Center in the national framework GINOP-3.1.1-VEKOP-15-2016-00001 “Promotion and support of cooperations between educational institutions and ICT enterprises”.

References

[1] B. F. G. Katz, and L. Picinali, “Spatial audio applied to research with the blind,” In P. Strumillo (ed.): Advances in Sound Localization, InTech Publishing, pp. 225–250, 2011.

[2] P. Baranyi, Á. Csapó, T. Budai, and Gy. Wersényi, “Introducing the Concept of Internet of Digital Reality - Part I,” *Acta Polytechnica Hungarica*, vol. 18, no. 7, pp. 225–240, 2021., 2021.

[3] Á. Csapó, I. Horváth, P. Galambos, and P. Baranyi, “VR as a medium of communication: from memory palaces to comprehensive memory management,” in Proc. 9th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), IEEE, 2018.

[4] B. Berki, “Experiencing the Sense of Presence within an Educational Desktop Virtual Reality,” Acta Polytechnica Hungarica, vol. 17., no. 2., pp. 255-265, 2020.

[5] Á. Csapó, Gy. Wersényi, H. Nagy, and T. Stockman, “A survey of assistive technologies and applications for blind users on mobile platforms: a review and foundation for research,” J. on Multimodal User Interfaces, vol. 9, no.4., pp. 275-286, 2015.

[6] https://soundofvision.net/

[7] B. Håkansson, A. Tjellström, and U. Rosenhall, “Hearing Thresholds with Direct Bone Conduction Versus Conventional Bone Conduction,” Journal Scandinavian Audiology, vol. 13, no.1, pp. 3-13, 1984.

[8] B. N. Walker, and R. M. Stanley, “Thresholds of Audibility for Bone-Conduction Headsets,” in Proc. of ICAD 05, Limerick, Ireland, pp. 218-222, 2005.

[9] M. McBride, T. Letowski, and P. Tran, “Bone conduction reception: Head sensitivity mapping,” Journal Ergonomics, vol. 51, no. 5., pp. 702-718, 2008.

[10] B. N. Walker, R. M. Stanley, N. Iyer, B. D. Simpson, and D. S. Brungart, “Evaluation of Bone-Conduction Headsets for Use in Multitalker Communication Environments,” in Proc. of the Human Factors and Ergonomics Society Annual Meeting, Orlando, USA, vol. 49, no. 17, pp. 1615-1619, 2005.

[11] K. Kaga, M. Setou, and M. Nakmura, “Bone-Conducted Sound Lateralization Of Interaural Time Difference And Interaural Intensity Difference In Children And A Young Adult With Bilateral Microtia And Atresia Of The Ears,” Acta Oto-Laryngologica, vol. 121, no.2, pp. 274–277, 2001.

[12] A. Liu, “Sound Enhancement and Source Localization with Bone Conduction,” Senior Project CPSC 490, Yale University, 6 pages, 2018. https://zoo.cs.yale.edu/classes/cs490/17-18b/liu.alan.all68/

[13] J. C. Lock, I. D. Gilchrist, G. Cielniak, and N. Bellotto, “Bone-Conduction Audio Interface to Guide People with Visual Impairments,” In: Wang G., El Saddik A., Lai X., Martinez Perez G., Choo KK. (eds) Smart City and Informatization. iSCI 2019. Communications in Computer and Information Science, vol. 1122, Springer, Singapore, pp-542-553, 2019.

[14] Gy. Wersényi, “Localization in a Head-Related Transfer Function-based virtual audio synthesis using additional high-pass and low-pass filtering of sound sources,” Acoustical Science and Technology, vol. 28, no. 4, pp. 244-250, 2007.

[15] T. M. Voong, and M. Oehler, “Auditory Spatial Perception Using Bone Conduction Headphones along with Fitted Head Related Transfer Functions,” in Proc. of 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Osaka, Japan, pp. 1211-1212., 2019.

[16] K. R. May, and B. N. Walker, “The effects of distractor sounds presented through bone conduction headphones on the localization of critical environmental sounds,” Applied Ergonomics, vol. 61, no. 5, pp. 144-158, 2017.

[17] J. A. McDonald, P. P. Henry, and T. R. Letowsky, “Spatial audio through a bone conduction interface,” International Journal of Audiology, vol. 45, no. 10, pp. 595-599, 2006.

[18] R. W. Lindeman, H. Noma, P. G. de Barros, “Hear-Through and Mic-Through Augmented Reality: Using Bone Conduction to Display Spatialized Audio,” in Proc. of the International Symposium on Mixed and Augmented Reality (ISMAR), Nara, Japan, pp. 4, 2007.

[19] https://aftershokz.com/

[20] Gy. Wersényi, Á. Csapó, T. Budai, and P. Baranyi, “Internet of Digital Reality: Infrastructural Background - Part II,” *Acta Polytechnica Hungarica*, vol. 18, no. 8, pp. 91–104, 2021.