Sound of Vision - spatial audio output and sonification approaches

M. Bujacz[1], K.Kropidlowski[1], G. Ivanica[2],
A. Moldoveanu[2], C. Saitis[3], A. Csapo[4], G. Wersenyi[4], S. Spagnol[5],
O. I. Johannesson[5], R. Unnborsson[5], M. Rotnicki[6], P. Witek[6]

[1] Institute of Electronics, Lodz University of Technology, Poland
[2] University POLITEHNICA of Bucharest, Romania
[3] ISI Foundation, Turin, Italy [4]Széchenyi István University, Hungary
[5] University of Iceland, Iceland
[6]Fundacja Instytut Rozwoju Regionalnego, Poland
michal.bujacz@p.lodz.pl

Abstract. The paper summarizes a number of audio-related studies conducted by the Sound of Vision consortium, which focuses on the construction of a new prototype electronic travel aid for the blind. Different solutions for spatial audio were compared by testing sound localization accuracy in a number of setups, comparing plain stereo panning with generic and individual HRTFs, as well as testing different types of stereo headphones vs custom designed quadrophonic proximaural headphones. A number of proposed sonification approaches were tested by sighted and blind volunteers for accuracy and efficiency in representing simple virtual environments.

Key words: electronic travel aid, spatial audio, HRTF, HRIR, sonification, sound model, sound synthesis

1 Introduction

With the XXI century advances in technology, such as embedded devices capable of real time image and audio processing, the possibilities of designing an electronic travel aid for the blind are greater than ever [1]. The main goal of the "Sound of Vision: natural sense of vision through acoustics and haptics" research project funded by the European Commission under the Horizon 2020 framework is to construct and test a wearable device that would convey an auditory and haptic representation of the surrounding environment to a visually impaired person. This paper presents some of the first year's results of the project in terms of audio-related research, especially spatial audio solutions and sonification models.

The overall concept of the Sound of Vision system is creation of an electronic aid for local navigation and obstacle avoidance, similar to a previous Naviton project [2]. The primary method of the environment sensing is stereovision (with possible data

fusion from other sensors, e.g. time of flight or accelerometers) [3,4]. The reconstructed 3D scene is processed (reconstruction, segmentation) and output about the detected (or recognized) obstacles is provided through auditory and haptic channels.

2 Spatial audio - state of the art

Wearable spatial audio technology focuses on the use of Head Related Transfer Functions that enable to artificially alter two channels in a stereo signal to simulate a sound wave's interaction with the human body, especially with the pinna, the head and the torso [5]. This is done by introducing a delay between the stereo channels called Interaural Time Delay (ITD) and filtering each of the channels with an HRTF filter, providing a frequency dependent Interaural Level Difference (ILD). The HRTF filters can be obtained by several methods:

- individual measurement for a specific listener (highest quality method) [4]
- utilizing a generic measurement for an acoustic mannequin (most common method) [6]
- selecting a similar HRTF set from a database of HRTFs from multiple listeners, either perceptually[7] or anthropometrically [8,9]
- modeling an individualized HRTF based on head and ear shape of the listener, either by sound wave simulations [10] or correlations between HRTF spectra and ear shapes of a large number of listeners [11,12].

The use of HRTFs alone is frequently insufficient for accurate sound spatialization. Further processing steps, such as headphone equalization, rendering reflections from the environment and head tracking can significantly improve localization accuracy and decrease the chances of common spatial audio problems, such as in-the-head localization or front-back confusions [13]. Also, a number of studies demonstrated that virtual sound localization accuracy can be significantly improved through training [14,15,16]

The idea of using several speakers on headphones has been used in commercial sets for emulating 5.1 or 7.1 sound systems, but we have found only one occurrence of attempted use for wearable spatial audio, and it was also in the context of an electronic travel aid [17].

3 Proposed spatial audio solutions

One of the basic assumptions for the output from the SoV device was that the generated sounds should be perceived as if they originated from the observed environment. This meant the inclusion of some form of spatial audio processing; however, another important requirement was that the sounds should not block natural environmental sounds, which are extremely important for a blind traveler.

Two main approaches were considered – filtering using head related transfer functions (HRTFs, both individualized and generic) with some type of headphones that do not cover the ear channels (e.g. bone conduction or open in-ear headphones) or a

custom solution with multiple proximaural speakers that would allow spatialization through amplitude panning.
Special software was prepared for testing of virtual source localization accuracy with various configurations of spatialization (plain stereo panning, generic and individual HRTFs) and output hardware (reference, bone conduction, in-ear, and custom multi-speaker headphones). Ten participants took part in the tests, localizing sound sources in a 5x7 grid spaced at 30° (azimuths from -90° to +90° and elevations: -60° to +60°) approximately 100 times in each possible configuration.

The results from the tests were that using individualized HRTFs provided no significant advantage over generic ones (perception of azimuth was actually slightly worse). High quality reference headphones provided the best spatial audio experience; however, at the cost of covering the ears. Bone conduction was a promising alternative; though it showed strongest variance between the test participants.

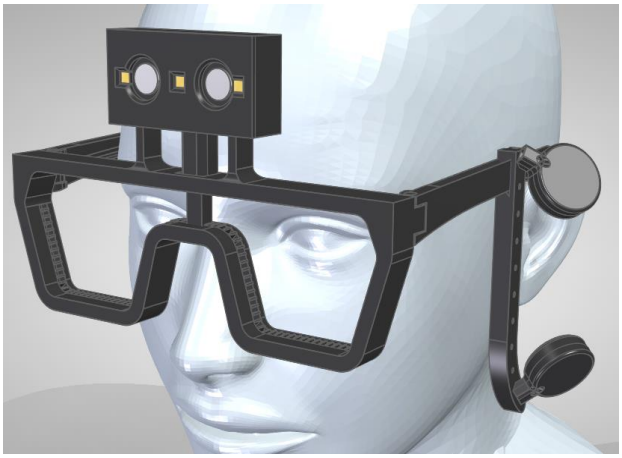

Fig. 1 Custom multi-speaker headphones and stereovision camera mount.

The custom headphones were designed and constructed as an alternative to spatial audio generated through the use of head related transfer functions (HRTFs). The 3D model of the headphones was prepared using Solid Edge ST7 software and manufactured using fused deposition modelling (FDM) on a Leapfrog Creator Dual Extruder 3D printer. The headphones included four speakers positioned above and below the ears, all slightly to the front. Amplitude panning was used to position a sound source both in the horizontal and vertical directions.

The tests of the custom quadrophonic headphones were conducted using a very similar procedure to the HRTF tests. The results were very promising, as both vertical and horizontal localization accuracy was on par with the high quality reference headphones, without the need for HRTF filtering.

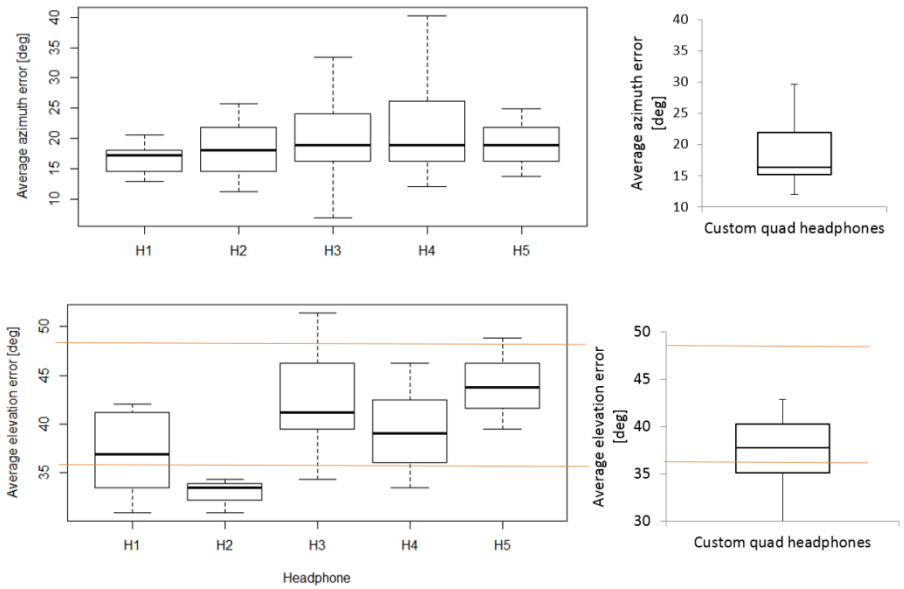

Fig. 2 Comparison of off-the-shelf headphones with individualized HRTFs with the custom quadrophonic headphones in terms of average azimuth (top) and elevation (bottom) localization error. H1 – AKG K612 Pro, H2 – Bose QC25, both high quality reference headphones, H3 – Aftershokz M3 bone conduction headphones, H4 – ear-Hero, H5 – Oticon P100 – in-ear air tube headphones. The red lines indicate predicted average errors if the replied either entirely randomly or always pointed to the central sound location.

4 Sonification – state of the art

Recent years have seen the emergence of many electronic systems aimed at aiding the blind [18]. Ranging from popular electronic range sensors [19], through smart phone apps[20,21], GPS or beacon-based navigation systems [22,23,24] to a number of research projects aimed at environmental imaging through sensory substitution and virtual acoustics spaces [2,25].

The sonification [26] in these devices can range from very simple, binary proximity alerts [19], through representing range using musical tones [27], to more complex synthesized sounds, such as clouds of spatially filtered impulses [25].

5 Proposed sonification approaches

Encoding the 3D visual scene through audio is considered the most important core functionality to be provided by the Sound of Vision project. Finding the most suitable

encoding method to provide valuable information about the environment surrounding the user through sound is, thus, a pivotal and challenging task.

The authors decided to explore a number of competing alternatives, dubbed "sound models" in order to identify the most promising sonification solution.

All the tested sonification methods based on a number of common assumptions about the output of the image processing module:

- the observed 3D scene fragment is 90° wide and 5m in depth
- the 3D scene can be roughly divided into individual objects, described by height and width
- these objects may or may not be continuously tracked from frame to frame
- the objects may be classified as belonging to specific categories (e.g. walls, stairs)

Approaches using the scene depth directly or basing on simplified occupation grids were also considered, however not included in the tests.
All the proposed models also used a similar basic virtual sound source positioning – encoding direction with generic HRTF filtering and distance with loudness.

The four tested sound models were:
Model 1 - **Objects as loudspeakers** – This model treats each object in the frontal hemisphere as an independent virtual sound source that continuously emits impact sounds, as if the user was striking the white cane on it. The pitch and timbre of the sound resulting from the impact are considered dependent on the object's width and category. The distance between object and user is coded into sound level and repetition rate: the closer the object, the higher the sound level and the more frequent the sound, just like in parking systems.
Model 2 – **Time and Frequency division multiplexed scene rendering -** The visual field was divided into three 30° wide regions. If a region was empty this was signified by a quiet heartbeat sound. For every object in a region a glissandi was played using granular synthesis such that: the wider the object, the slower and further down the musical scale the glissando went, the taller the object, the coarser was the sound texture.
Model 3 **– Depth scanning** - A virtual "scanning plane", a surface parallel to the camera view that moves away from the observer through the scene. As the surface intersects scene elements, sound sources originating from the places of intersection are released. Sources correspond to object parameters (distance to loudness and pitch, width to duration, category to instrument type). The model distinguishes two categories of objects – walls (any object with a sufficiently large surface area) and generic obstacles. This model has been previously successfully implemented in the Naviton prototype [2].
Model 4 – **Horizontal sweep** - A popular approach in many sonification studies (e.g. Navbelt [27]) sometimes referred to as a "piano scan" [28]. The method basically translates distance to pitch in several directions from the observer, making this method suitable for use with unprocessed depth maps or occupancy grids.

The first phase of tests was performed by 10 sighted volunteers, the second phase by 6 blind participants. The tests included such tasks as identifying the position of several obstacles, picking "the odd one out" (an obstacle of different size than others) and choosing a safe route to turn. Some of the scenes were static and in some

the observer moved at a constant walking speed. The metrics included reaction times, accuracy, and subjective opinions gathered in surveys with test participants.
Unfortunately, the results of these first tests were inconclusive. In most tasks of the models showed a very significant advantage over the others in accuracy or response times. The testers also had mixed preferences as to the nature of the sounds used. One important conclusion was that testers complained of the lack of continuity of the sounds, i.e. the cyclic auditory "snapshots" of the environment common to all the tested models were a poor method of observing dynamic scene changes.

6 Testing procedures

The tests of the Sound of Vision are conducted with participation of visually impaired testers – blind and partially-sighted, as these groups of end-users are the best experts regarding how the SOV solution is going to meet their needs. What is important in the case of Sound Of Vision, testing is to be interspersed with training, in an approach sometimes called "gamification" [29] – users progress to more difficult tests after completing simpler ones with sufficient efficiency.

So far the tests with blind participants focused on computer simulations for the purpose of selection of sounds for the auditory representation of objects in the environment; however some data was also gathered in real environments, observing travel speeds and patterns while also attempting to collect EEG and other biometric signals [30].

Further testing, introduced by training phases, will include use of virtual environment scenarios, then controlled laboratory environments (e.g. with cardboard obstacles) to finally move on to real-world environment and scenarios (under the care of orientation and mobility specialists). Testers in this phase are going to use the SOV prototype to navigate various indoor and outdoor paths. This will give final confirmation of compliance of the SOV solution with the needs of visually impaired persons.

7 Conclusions and future work

The conclusions from the spatial audio tests led us away from HRTF-based solutions towards the idea of the custom multi-speaker headphones. Currently two additional aspects are being tested – whether the use of additional speakers (4 per ear) can enhance the perception of spatial audio, e.g. by improving the perception of distance or accuracy of localization, and whether the use of HRTFs in conjunction with the custom headphones produces a significant difference than plain amplitude panning in virtual sound source perception.
From the sonification modes the depth scan shows most promise, though it is clear different approaches seem preferred in different scenarios and further testing is necessary.

Acknowledgment. This work received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 643636 "Sound of Vision".

References

1. Strumiłło, P.: Elektroniczne systemy nawigacji osobistej dla niewidomych i słabowidzących [Electronic personal navigation systems for the blind and visually impaired], PŁ, Wydział Elektrotechniki, Elektroniki, Informatyki i Automatyki, Łódź (2012)
2. Bujacz, M., Skulimowski, P., Strumiłło, P.: Naviton - a prototype mobility aid for auditory presentation of 3D scenes. In: Journal of Audio Engineering Society, Vol 60, No. 9, pp. 696-708 (2012)
3. Skulimowski, P., Strumillo, P.: Obstacle Localization in 3D Scenes from Stereoscopic Sequences. In: 15th European Signal Processing Conference EUSIPCO 2007, Poznan, Poland (2007)
4. Owczarek, M., Skulimowski, P., Strumillo, P.: Sound of Vision – 3D scene reconstruction from stereo vision in an electronic travel aid for the visually impaired. In: Computers Helping People with Special Needs of the series Lecture Notes in Computer Science (2016) (in press)
5. Dobrucki, A., Plaskota, P.; Pruchnicki, P.; Pec, M.; Bujacz, M., Strumiłło, P.: Measurement System for Personalized Head-Related Transfer Functions and Its Verification by Virtual Source Localiza-tion Trials with Visually Impaired and Sighted Individuals. In: 58(9) Journal of Audio Engineering Society, pp. 724-738 (2010)
6. Gardner, W. G. and Martin, K. D. HRTF measurements of a KEMAR. J. Acoust. Soc. Am, 97(6):3907-3908, (1995)
7. Middlebrooks, J. C., Macpherson, E. A. and Onsan, Z. A.: Psychophysical customization of directional transfer functions for virtual sound localization, J. Acoust. Soc.Am., vol. 108(6), pp. 3088-3091, (2000)
8. Geronazzo, M.; Spagnol, S.; Bedin, A. & Avanzini, F. Enhancing Vertical Localization with Image-Guided Selection of Non-Individual Head-Related Transfer Functions. Proc. IEEE Int. Conf. Acoust. Speech Signal Process. ICASSP 2014, pp. 4496-4500 (2014)
9. Zotkin, D. N., Hwang, J., Duraiswami, R., Davis, L. S.: HRTF Personalization Using Anthropomentric Measurements, Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'03), pp. 157-160, (2003)
10. Dobrucki, A., Plaskota, P.: Computational modelling of head-related transfer function, Archives of Acoustics, vol. 32 (2007)
11. Spagnol, S., Avanzini, F.: Frequency Estimation of the First Pinna Notch in Head-Related Transfer Functions with a Linear Anthropometric Model. Proc. 18th Int. Conf. Digital Audio Effects (DAFx-15), 231-236. (2015)
12. D. Trapenskas, N. Frenne, Ö. Johansson, Relationship between HRTF's and anthropometric data, The 29th International Congress and Exhibition on Noise Control Engineering, (2000)
13. Wightman, F. L., Kistler, D. J.: Binaural and Spatial Hearing in Real and Virtual Environments, Ch. Factors Affecting the Relative Salience of Sound Localization Cues. In: Lawrence Erlbaum Associates, pp. 1-24, Mahwah, New Jersey (1997)
14. Bălan, O., Moldoveanu, A., Moldoveanu, F. and Morar, A.: Experiments on Training the Human Localization Abilities. In Proceedings of the 10th International Scientific Conference eLearning and Software for Education-Bucharest. (2014)

15. Bălan, O., Moldoveanu, A., Butean, A., Moldoveanu, F., Negoi, I.: Comparative research on sound localization accuracy in the free-field and virtual auditory displays. In The 11th eLearning and Software for Education Conference - eLSE 2015. (2015)
16. Bălan, O., Moldoveanu, A., Moldoveanu, F., Negoi, I.: The Role of Perceptual Feedback Training on Sound Localization Accuracy in Audio Experiments. In Proceedings of The 11th International Scientific Conference eLearning and software for Education. (2015)
17. Vitek, S., Klima, M., Husnik, L., Spirk, D.: New Possibilities for Blind People Navigation. In: IEEE 2011 International Conference on Applied Electronics (AE), Plisen (2011)
18. Hersh, M., Johnson, M.: Assistive technology for visually impaired and blind people. Springer, London (2008)
19. Farcy, R., Bellik, Y., Locomotion assistance for the blind.[in:] Universal Access and Assistive Technology, Keates S., Langdom P., Clarkson P., Robinson P. [Eds.], pp. 277–284, Springer. (2002)
20. Manduchi, R., Coughlan, J., Ivanchenko, V.: Search Strategies of Visually Impaired Persons Using a Camera Phone Wayfinding System. In: Computers Helping People with Special Needs Vol. 5105 of the series Lecture Notes in Computer Science, pp. 1135-1140 (2008)
21. Matusiak, K., Skulimowski, Skulimowski, P., Strumillo, P.: A Mobile Phone Application for Recognizing Objects as a Personal Aid for the Visually Impaired Users in: HUMAN-COMPUTER SYSTEMS INTERACTION: BACKGROUNDS AND APPLICATIONS 3, Book Series: Advances in Intelligent Systems and Computing, vol 300, pp 201-212 (2014)
22. Skulimowski, P., Korbel, P., Wawrzyniak, P.: POI Explorer - A Sonified Mobile Application Aiding the Visually Impaired in Urban Navigation Proc. of FedCSIS, ACSIS-Annals of Computer Science and Information Systems, Volume: 2, pp: 969-976, (2014)
23. Ferreira, E.J., Navmetro: Preliminary Study Application of Usability Assessment Methods, Human Factors in Design, (2013)
24. Mayerhofer, B., Pressl, B., Wieser, M.: ODILIA - A Mobility Concept for the Visually Impaired. In Computers Helping People with Special Needs Vol. 5105 of the series Lecture Notes in Computer Science, pp. 1109-1116 (2008)
25. González-Mora, J., Rodríguez-Hernández, A., Rodríguez-Ramos, L.: Development of a new space perception system for blind people, based on the creation of a virtual acoustic space. In: Engineering Applications of Bio-Inspired Artificial Neural Networks, pp. 321-330, Springer, Berlin/Heidelberg (1999)
26. Hermann, T., Hunt. A., Neuhoff, J.G.: The Sonification Handbook. Logos Verlag Berlin (2011)
27. Shoval, S.,Borenstein J., Koren, Y.:Auditory guidance with the Navbelt - a computerized travel aid for the blind. In: 28(3) IEEE Transactions on Systems, Man, and Cybernetics.pp. 459-467 (1998)
28. Bujacz, M., Strumiłło, P.: Stereophonic representation of virtual 3D scenes - a simulated mobility aid for the blind. In: Dobrucki, A., Petrovsky, A., Skarbek, W. (eds.) New Trends in Audio and Video, vol. 1, pp. 157-162 (2006)
29. Balan, O., Moldoveanu, A., Moldoveanu, F., Dascalu, M.I.: Audio games- a novel approach towards effective learning in the case of visually-impaired people. In: Proc 7th Int. Conference of Education, Research and Innovation, p. 7, Seville, Spain (2014)
30. Saitis, C., Kalimeri, K.: Identifying Urban Mobility Challenges for the Visually Impaired with Mobile Monitoring of Multimodal Biosignals. In: Antona, M., Stephanidis, C. (eds.) Universal Access in Human-Computer Interaction - 10th International Conference, UAHCI 2016, Held as Part of HCI International 2016, Toronto, ON, Canada, July 17-22, 2016, Proceedings. Springer-Verlag, Berlin (2016) (in press)