# MODIFICATIONS OF THE SURROUNDING AUDITORY SPACE BY AUGMENTED REALITY AUDIO: INTRODUCTION TO WARPED ACOUSTIC REALITY

*Dominik Storek, Jan Stuchlik, Frantisek Rund*

Czech Technical University in Prague
Faculty of Electrical Engineering
Department of Radioelectronics
Technicka 2, Praha 6, 166 27, Czech Republic
storedom@fel.cvut.cz

## ABSTRACT

Augmented Reality Audio (ARA) is mostly employed in terms of adding virtual sound source into an existing auditory scene. Common goal of such system is to preserve the highest level of fidelity and natural character of both real and virtual components. This paper introduces another approach to the ARA systems based on intentional modification of the native parameters of both the real and the virtual auditory segments. The system employs binaural microphone-equipped earphones used as microphone-hear-through device, which are able to capture and immediately reproduce the auditory scene. A processing unit is included in the microphone-earphone signal path. The unit enables to apply various audio effects to both the captured and the virtual sound allowing warped perception of the auditory reality. The real-time processing algorithms were implemented in Pure Data environment. The proposed system provides immersive audio performance and has potential to be used in specific virtual reality applications, such as simulation of reality perception by people with mental disorders.

## 1. INTRODUCTION

The basic concept of Augmented Reality Audio (ARA) is already known for a few decades. However, its application possibilities are limited almost only by imagination of use. In general, the ARA concept is based on introducing virtual sounds source (often spatially positioned) to the surrounding real audio scene for artistic or interaction purposes [1]. Majority of the already presented systems is designed to enhance or extend the auditory scene by information carried by the virtual source, which involves its character and spatial position. The parameters of the real auditory scene are preserved simultaneously. As example, narrative comments for museums exposition [2], telecommunication improvements [3], tourist guide assistance [4], [5], or medial applications [6] employing ARA have been introduced so far. There exist several approaches, how to deliver the virtual source to the listener, e.g. mobile microphone-hear-through (mic-through) devices [7], or by bone-conducting system [8]. This paper utilizes the first option with the use of mic-through device, which consists of in-ear headphones (earphones) with embedded binaural microphones. This hardware arrangement allows the sound reaching both listener's ears to be binaural-recorded, processed, and reproduced in real-time conditions [9]. Embedded software processing offers various options of the sound adjustment. For our purposes, the binaural signal is modified beyond the edge of preserving the features of physical reality. The aim of the authors was to explore the perception of such modified auditory reality, especially in contrast to simultaneous presence of original physical reality perceived visually. Theoretical concept of this project is briefly described in the following Section 2, both hardware and software setups are described in Section 3. Perception of the modified reality and settings of the effect parameters are discussed in Section 4. Besides that listening to the modified acoustic reality is entertaining itself, it also has a potential of use in specific medical-based application (e.g. simulation of reality perception with mental disorders), as suggested in Section 5.



## 2. THEORETICAL CONCEPT

The main purpose of ARA system is to combine real and virtual acoustic objects together in real or pseudo-real environments. Proper function of the system is conditioned by real-time performance, which practically stands for ensuring *sufficiently* low latency. Modification of the acoustic reality parameters requires its previous capture. Since the frame of the surrounding acoustic space is determined by binaural representation of the scene, it is advisable to capture the signal at a place of ear canal entrance. This also requires to reduce the physical distance between the positions of capture and reproduction to minimum. The proposed system would also work with monaural representation of the surroundings, however, the final effect would lack spatial depth and fidelity. Therefore, binaural recording is necessary for proper performance of the augmented reality system. When captured, the signal parameters of the sound are subsequently modified by some of a wide range of optional audio effect algorithms. The captured signal (both processed or unprocessed) is referred to as *pseudo-reality* environment. Most impressive effect is achieved when the final virtual acoustic information is presented in contradiction with the real visual perception. For instance, when cathedral-like reverberation is added to real ordinary office environment, the auditory representation does not correspond to the visual perception of room size and space depth. This results in sensual confusion, which is very attractive to perceive (similar to e.g. optical illu-

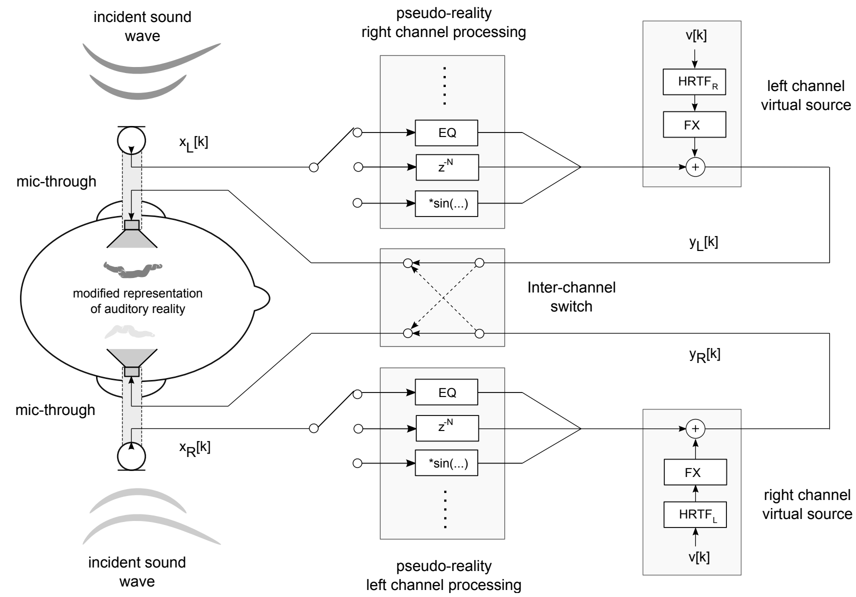


Figure 1: Auditory reality warping system. The signal is recorded by mic-through device, processed by a bank of tunable audio effects and mixed with positioned virtual source. Stereo sound is then presented to the listener by earphones embedded within the device.

sions). The issue of reproduction of the modified sound has to fulfill several requirements. First of all, the acoustic information needs to be captured with binaural cues included for perception of acoustical space depth and source arrangement [10]. Second, the reproduced sound needs to be separated from still existing *real* acoustic space (i.e. the information regarding the real unprocessed environment needs to be eliminated). Mixing of the real and the modified reality together would not deliver the exciting and immersive effect since the brain intends to focus on the real environment and prefer it to the virtual one. Third, the subject needs to be *sufficiently* free for at least head movements in order to preserve spatial perception supported by movement of the listener. In order to deliver the processed signal and simultaneously block the information regarding the real auditory environment, and also to allow partial freedom of the subject, in-ear headphones are promising. The system of previously described binaural microphones embedded within a pair of earphones already exists as commercial product and is easy to get.

The idea of the introduced system is resulting from a common design of the ARA systems, as already introduced in [11]. The concept of the system designed for our purposes is demonstrated in Fig. 1. The system consists of two parallel signal paths for treating both signal components of the captured stereo sound. At the beginning the incident sound is captured by miniature microphone attached on the in-ear headphone (the hear-through-microphone) and sent as signal $x_{L/R}[k]$ to the processing unit. The process of signal modification consists of two parts: application of selected audio effect and mix with added virtual sound, in Fig. 1 denoted as $v[k]$. The effects applied to the signal $x_{L/R}[k]$ are selected from a pre-determined bank of implemented algorithms. Theoretical concept of the effect unit also enables to arrange the effects to serial order. However, this may result in undesired high latency occurrence. After modification of the pseudo-reality parameters, virtual signal $v[x]$ is mixed with the original signal. This signal is a monaural file loaded from memory, thus both left and right channels are fed by the same input. The purpose of the virtual signal is to deliver additional information to the warped pseudo-reality environment, e.g. instructions for the listener, simulation of hallucination, etc. The signal $v[x]$ can be also virtually positioned in the pseudo-reality environment and also another effects processing is optional (FX block in Figure 1). Virtual positioning can be performed by processing of the signal by Head-Related Transfer Function (HRTF) [12] or by Differential HRTF (DHRTF) [13]. These transfer functions carry information of time and level differences in each ear in dependence on the source position. The location of the virtual signal can be easily adjusted by selection of appropriate (D)HRTF. Use of simple amplitude panning is also optional. In the standard ARA systems the FX block in the signal path is usually limited to introduce Room Impulse Response (RIR), which delivers *auralization* effect, thus the positioned signal is more successfully embedded into the corresponding pseudo-reality environment. This block of processing is also important for externaliztion of the source [14], [10]. The system proposed in this article takes into account much wider employment of audio effects. Both channels of the stereo signal $y_[k]$ are consequently delivered to the transducers of the earphones. Cross-channel routing as another audio effect (to hear someone in the opposite side than he visually appears) is also optional. Finally, the listener receives modified acoustic information representing the pseudo-reality environment with virtual sound present at specific location. There exist various approaches, how to implement the system described above. Our implementation is discussed in the following section.

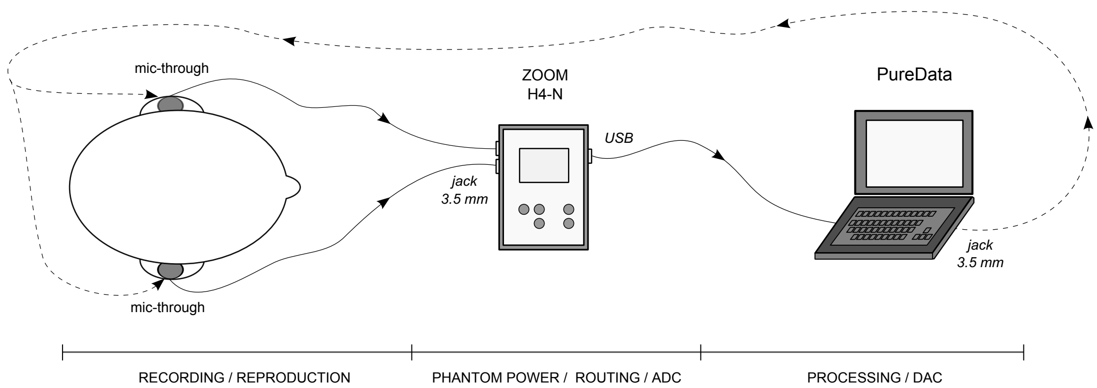


Figure 2: Demonstration of hardware implementation. Binaural earphones are powered by handy recorder, which also performs function of ADC. The recorder distributes the signal to the notebook with Pure Data via USB. The processed signal is returned in analog form.

## 3. TECHNICAL SETUP

Practical implementation of the system requires designing both hardware and software setup. The chosen configuration and corresponding issues are described in the subsections below.

### 3.1. Hardware configuration

The proposed ARA system is easy to set from the hardware point of view. For our purposes, microphone-hear-through binaural earphones Roland CS-10EM were chosen. Each earplug consists of electret condenser microphone with omni-directional characteristics and dynamic wide-band transducer with sufficiently flat transfer function. The microphones require power supply in range of 2-10 V, thus a handy recorder ZOOM H4n was employed in the path as a power source. The device is used as ADC (DAC also optional) and enables to record the binaural representation of the auditory scene for further purposes. The digitalized signal is then routed to notebook via USB. The processing algorithms consist of application of particular effects, mixing with the virtual positioned signal, converting by DAC, and routing back to earphone transducers. In this setup the performance of the system is limited by length of the cables and also by weight of the notebook and the recorder (e.g. too heavy for city exploration). Our hardware setup is demonstrated in Fig. 2.

### 3.2. Software implementation

For effectiveness of use, open source Pure Data (PD) [15] visual programming language was used as a platform to implement the algorithms. The PD enables intuitive control of the graphical pseudo-code arrangement of the signal blocks and also allows easy modifications and extensions of the existing program. The main principle of PD is based on a real-time distribution of data flow among graphical *objects* representing (and also performing) particular algorithms. An example of implementation of *frequency modulation* effect is demonstrated in Fig. 3. The signal path starts in the upper left corner by virtual analog-digital converter (ADC) and ends in the lower left corner by reverse digital-analog-converter (DAC), which defines the output. The signal is routed to delay blocks and afterwards used in an argument of $\cos(\cdot)$ function. Three slider (bars on the right side) are used to allow the user to modify the modulation parameters: *modulation speed*, *modulation depth*, and *modulation feedback*. The FM effect is used for demonstration purposes here. The system is capable of various performance (see further subsection 3.3 Effects).

Since the *raw* Pure Data interface may not appear to be suitable for technically non-skilled users, Graphical Interface for Pure Data (GrIPD) multi-platform was employed in the later stage. This extension of the original PD enables to create graphical user interface (GUI) for programs implemented in this environment.

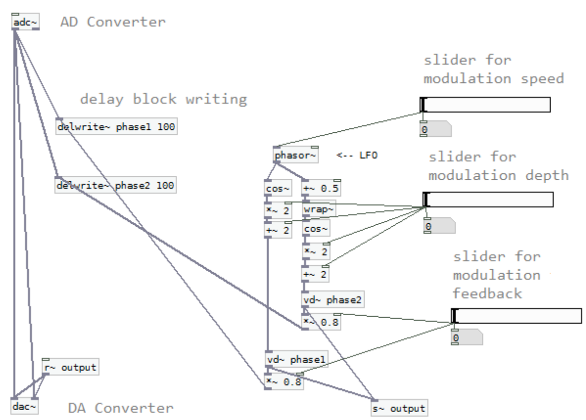


Figure 3: Demonstration of software implementation of a frequency modulation effect in Pure Data environment. The signal is routed within blocks, which are performing particular part of selected algorithm. The user can modify settings by sliders.

| RECOMMENDED EFFECTS | | |
|---|---|---|
| Effect | Settings | Description |
| Delay | *wet* sound delayed by 1-2 s | unable to read or talk coherently |
| Echoes | attenuated echoes after 0.3-0.8 s, 3 times repeated | introducing wider dimension |
| Reverbation | long (cathedral-like) or short (empty-room-like) reverb | different volume / space perception |
| Modulation | speed = 5 Hz, depth = 0.6, feedback = 0.6 | burbling of the sound objects |
| Pitch-shift | octave higher (smurf-like) or octave lower (demon-like) AM 1-127 Hz | resembles acceleration or decceleration of the world, respectively |
| Channel crossing | simply switching the channels | spatial confusion occurs |
| Overdrive | hard limiting for 0.7 of maximal amplitude | distorted boosting and compression |

Table 1: Summary of the effects suitable to be used within the ARA warping system.

### 3.3. Effects

The span of effect employment is more or less limited only by designer's imagination and system latency. As mentioned above, the most immersive impression of this auditory reality warping system is when it is producing contradiction within the visual and the hearing senses. Therefore, it is desirable to keep the effect performance within a limited range of values. The algorithms are supposed to be set in way to produce *enough* impressive reality warp while still preserving *enough* palpable real features to recognize the acoustic surrounding components. The issue of which effect is suitable to be implemented was resolved according to this assumption.

There exists a wide range of acoustic effects to be potentially employed in the system, moreover each effect can be then implemented by distinct algorithms with various performance. Very useful and inspiring overview of audio processing can be found e.g. in book [16]. From the exploratory character of this article, the following effects were taken in consideration for use within the auditory reality warping system. However, many other effect options can be employed. Note that all the effects are required to run continuously, thus an increasing sample buffering shall be avoided (e.g. occurring in time stretch). More detailed description of particular implementations is beyond the range of this article.

- **Filters**
  Application of standard filters is not interesting at all in this case, since high-passed or low-passed sound does not deliver any new information to the pseudo-reality environment. Considerable is *telephone bandpass* filtering to get coldness of the surroundings or application of time-varying filters like *wah-wah* to achieve impressive spectral movement and *breathing* of the sound scape.
- **Delays**
  Very efficient for use. Simple implementation of *delay line* with reasonable delay of approximately 1-2 seconds disables the listener to read a text (popular well known phenomenon) or even to talk coherently. Implementing of multiple echoes with reasonable feedback introduces very abstract dimension of the pseudo-reality. Artificially added *reverbation* delivers false cues for space dimensions (e.g. a church-like reverb in your office) and works perfectly.
- **Modulators**
  Introducing *Frequency Modulation* turns the surrounding objects into impressive gurgling sources, like inside of a fizzy drink. By means of *Amplitude Modulation* a pulsing vibrato-like sound is obtained. With proper settings of this effects even *pitch-shifting* is easy to achieve. When everything around is up-tuned or down-tuned *reasonably* (i.e. to keep the contours of the auditory environment), it also produces immersive result.
- **Nonlinear processing**
  This type of processing usually concerns dynamic-compression-like *gate* or *compressor*, which also does not introduces any new extension or disturbance into the binaural signal. However, implementation of hard limiting of the signal edge produces effect similar to *overdrive* or *fuzz*. This aggressively boost the sound into very sharp distorted signal.
- **Channel crossing**
  Probably the easiest effect to implement, however, very impressive in the result. Switching the left and right channel delivers reverse spatial perception: objects seen at one half-plane are heard in the opposite direction.

This section lists only the basic ideas to be implemented. Practically there is no limitation of implementing effective algorithm to achieve impressive results. Selected effects, which were implemented and tested by the date of submission of this article, are summarized with basic settings recommendation in Table 1 at the top of this page.

## 4. ISSUES & LIMITATIONS

The performance of the ARA warping system is impressive. However, there exist several limitations of the system, which cause some limitations. First of all, there is a leakage of the real signal (coming from the natural auditory scene) to the pseudo-reality environment [9]. This is caused by non-perfect fit of the earplugs in the listener's ears and results in mixing of both the real and the processed signals together. The earplugs need to be carefully adjusted in the ear, otherwise the brain inclines to prefer and target the real information, thus the whole concept of delivering warped reality is ruined. This problem may be solved by adjusting the size of the lining of the earplugs to individual requirements.

Second, the position of the hear-through microphones is not situated directly in the ear canal entrance and shifted more out of the pinna. Since the effect of pinna is reduced, it decreases externalization of the sources and spatial depth of the scene [17]. To avoid this unwanted effect a mic-through hardware of much smaller dimensions should be employed. However, this would also lead to limitation of physical parameters of the system regarding flatness and effective span of the transfer function. Another issue is the modified dynamics of the captured scene, i.e. the pseudo-reality environment. The natural character of *deep* space is reduced to more *flat* impression of the virtual scene. Probably some artificial dynamic expansion would be helpful in this case, however, this faces the edge of latency limitation. Since the signal path goes through multiple hardware devices and back, latency of approximately 140 ms occurs in the described implementations. Generally, this does not destroy the whole immersive spatial perception. However, the more quick response of the system is available, the better performance is provided. Useful advices on how to face general issues within latency and equalization of the system are introduced in [18].

## 5. RESULTS & DISCUSSION

This article introduced a system for modification of auditory reality based on Augmented Reality Audio concepts employing microphone-hear-through kit and effect processing unit. The system provides immersive perception of the modified reality determined by particular processing algorithm and also allows to add a virtual sound to the pseudo-reality environment. The system is designed to provide entertainment by combination of warped auditory reality and real visual information. The effect of fun interest is similar to watching own personality in a distorting mirror: the well known appearance is now drastically modified and one usually tries to explore different positions and mimics. Similar situation is reached with the warped audio reality system. However, the possibilities are not limited only to fun activities.

Several medical projects have introduced multimedia systems or static multimedia installations in order to simulate perception of reality by individuals with specific mental disorders. These projects intend to increase public awareness of their feelings and emotions within their daily routine tasks and inducing empathy to combat stereotypical responses. As an example, *schizophrenia* or *autism* are often presented by one's perception of false or distorted representation of the auditory (and also the visual) reality. For more detailed description of the projects simulating perception with the mentioned disorders see [19] and [20], respectively. Such projects generally present a short everyday situation (e.g. shopping in a market, visiting a hospital, child playing on playground, etc.) with corresponding visual and acoustic perception of the mentally impaired people (tactile information is also optional). In dependence on specific illnesses the presentation usually contains both *inner voices* (i.e. form of hallucination, when non-real character(s) is talking to the patient in various manners, or the patient is talking to himself) and modified auditory reality (echoes, distortions). These multimedia presentations are more-or-less based on fixed scenarios with no option for subject's interaction with the surroundings or allowance of own exploration of the *new* reality space. Therefore, this cinema-like concept turns the public observer into a role of passive consumer. Another approach is for instance more interactive *game*-based simulator of *sensory overload* symptoms of autistic child, which allows the user to move his character on playground with other children. For more details see [21]. However, even the game concept does not offer satisfactory interaction possibilities for the observer, since the environment is still presented by monitor and headphones.

Presentation of reality perception with specific mental illness can be significantly upgraded by our proposed system since it introduces both modifications of the existing reality and adding the virtual sound as an optional source of the *inner voice* talking to the subject. Our system would allow the observer to interact with the *real* environment and also free movement in the space with preserved warped auditory perception. Therefore, the presentation and impression of the mental illness would be much suggestive, when the subject turns from a passive observation to a real action and can interact with the surrounding environment. Our next effort will be put on designing and testing of several demo scenarios of such possible applications. Tuning the implementation issues is the next stage of the system development as well.

## 6. ACKNOWLEDGMENT

This work was supported by the Grant Agency of the Czech Technical University in Prague, grant No. SGS14/ 204/OHK3/ 3T/ 13. Thanks to Lucie Karlachova for voice recording cooperation.

## 7. REFERENCES

[1] A. Harma, J. Jakka, M. Tikander, M. Karjalainen, T. Lokki, and H. Nironen, "Techniques and applications of wearable augmented reality audio," in *Audio Engineering Society Convention 114*. Audio Engineering Society, 2003.

[2] A. Zimmermann and A. Lorenz, "Listen: a user-adaptive audio-augmented museum guide," *User Modeling and User-Adapted Interaction*, vol. 18, no. 5, pp. 389–416, 2008.

[3] H. Gamper and T. Lokki, "Audio augmented reality in telecommunication through virtual auditory display," *In Proceedings of the International Conference on Auditory Display (ICAD 2010)*, pp. 63–71, June 9–15 2010.

[4] J. M. Rozier, "Hear&there: An augmented reality system of linked audio," *In Proceedings of the International Conference on Auditory Display (ICAD 2000)*, pp. 63–67, 2000.

[5] B. Boren, M. Musick, J. Grossman, and A. Roginska, "I hear ny4d: Hybrid acoustic and augmented auditory display for urban soundscapes," in *Proceedings of the International Conference on Auditory Display (ICAD 2014)*, 2014.

[6] M. A. Livingston, W. F. Garrett, G. Hirota, M. C. Whitton, E. D. Pisano, H. Fuchs, *et al.*, "Technologies for augmented reality systems: realizing ultrasound-guided needle biopsies," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. ACM, 1996, pp. 439–446.

[7] R. Albrecht, T. Lokki, and L. Savioja, "A mobile augmented reality audio system with binaural microphones," in *Proceedings of Interacting with Sound Workshop: Exploring Context-Aware, Local and Social Audio Applications*. ACM, 2011, pp. 7–11.

[8] R. W. Lindeman, H. Noma, and P. G. de Barros, "Hear-through and mic-through augmented reality: Using bone conduction to display spatialized audio," in *Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*. IEEE Computer Society, 2007, pp. 1–4.

[9] A. Härmä, J. Jakka, M. Tikander, M. Karjalainen, T. Lokki, J. Hiipakka, and G. Lorho, "Augmented reality audio for mobile and wearable appliances," *Journal of the Audio Engineering Society*, vol. 52, no. 6, pp. 618–639, 2004.

[10] R. Gilkey and T. R. Anderson, *Binaural and spatial hearing in real and virtual environments*. Psychology Press, 2014.

[11] K. Lyons, M. Gandy, and T. Starner, "Guided by voices: An audio augmented reality system," *Proceedings of the International Conference on Auditory Display (ICAD 2000)*, 2000.

[12] J. Blauert, *The technology of binaural listening*. Springer Verlag, Berlin, 2013, eBook.

[13] D. Storek and F. Rund, "Differential head related transfer function as a new approach to virtual sound source positioning," in *Radioelektronika (RADIOELEKTRONIKA), 2012 22nd International Conference*. IEEE, 2012, pp. 1–4.

[14] W. M. Hartmann and A. Wittenberg, "On the externalization of sound images," *The Journal of the Acoustical Society of America*, vol. 99, no. 6, pp. 3678–3688, 1996.

[15] http://puredata.info/downloads/pure-data.

[16] U. Zölzer and X. Amatriain, *DAFX: digital audio effects - Second Edition*. Wiley Online Library, 2011, vol. 1.

[17] J. Goßmann, "Connecting virtual sound and physical space in audio-augmented environments," in *In Proceedings of the International Conference on Auditory Display (ICAD 2004)*, 2004.

[18] M. Karjalainen, V. Riikonen, and M. Tikander, "An augmented reality audio mixer and equalizer," in *Audio Engineering Society Convention 124*. Audio Engineering Society, 2008.

[19] D. L. Penn, J. D. Ivory, A. Judge, *et al.*, "The virtual doppelganger: Effects of a virtual reality simulator on perceptions of schizophrenia," *The Journal of nervous and mental disease*, vol. 198, no. 6, pp. 437–443, 2010.

[20] https://www.autismspeaks.org/news/news-item/animated-simulation-helps-you-experience-sensory-overload.

[21] http://arstechnica.com/gaming/2013/03/auti-sim-lets-you-experience-the-horror-of-sensory-overload/.

[22] N. Mariette, "Mitigation of binaural front-back confusions by body motion in audio augmented reality," in *Proceedings of the International Conference on Auditory Display (ICAD 2007)*, 2007.