A 3D Voice-Hearing Simulator Co-Created by Voice Hearers and Researchers: Preliminary Sound Quality Evaluation

Philippe-Aubert Gauthier *[1,2], **Kevin Zemmour**[3], **Sylvain Grignon**[3], **Bálint Demers**[4], **Catherine Lejeune**[1], **Sandrine Rousseau**[5], **Mouloud Boukala**[4], **Sofian Audry**[4], **Sylvio Arriola**[8], **Alain Berry**[6], **and Kevin Whittingstall**[7]
[1]École des arts visuels et médiatiques, UQAM, Montréal, Canada
[2]Centre for Interdisciplinary Research in Music, Media, and Technology (CIRMMT), McGill University, Montréal, Canada
[3]Département de psychiatrie, Faculté de médecine et des sciences de la santé, Université de Sherbrooke, Canada
[4]École des médias, UQAM, Montréal, Canada
[5]Association québécoise pour la réadaptation psychosociale, Québec, Canada
[6]Centre de recherche acoustique-signal-humain de l'Université de Sherbrooke, Université de Sherbrooke, Canada
[7]Département de radiologie diagnostique, Faculté de médecine et des sciences de la santé, Université de Sherbrooke, Canada
[8]École supérieure de Théâtre, UQAM, Montréal, Canada

1 Introduction

Voice hearing (VH) is an auditory hallucination that can touch a wide range of people, not only those with psychosis or schizophrenia. VH can be positive, negative, or anything in between and is distinctive in that the voices do not seem to only come from inside the head but also from the external world, the sound environment, or the body, sometime resulting in a sense of sound externalization. Poor support from professionals of mental health or social work is often reported by voice hearers. Leading to potential issues with service providers lacking proper experiential knowledge of VH. To address this issue, we co-created a 3D voice hearing simulator (3DV) within a participatory research paradigm with a multidisciplinary team of researchers and voice hearers from various fields (acoustics, acting, quantitative methods, neuroimaging, psychiatry, sound art). Several studies have assessed the impact of voice simulator on healthcare professionals, nursing students, etc. [1]. To the authors knowledge, the use of 3D audio or binaural sound and the co-creation approach are not reported in the literature. Since the aim of 3D audio is to recreate the experience of spatial hearing as in natural hearing [2], we explored 3D sound for immersion as described by voice hearers. In this work, since the 3DV should be a portable simulator for *in situ* listening and experiencing, we use basic binaural recording with binaural head or Ambisonics [3] to binaural conversion.

The objective of the project is the creation of an experiential knowledge sharing tool for the training of students in psychiatry or social work, since they both probably have to interact and provide service to voice hearers in their career. We here present our preliminary sound evaluation of the 3DV, including how binaural 3DV performed compared to stereo and to binaural decoding of Ambisonics sound.

2 Methods

2.1 Co-creation of the 3DV simulator

Because of the personal nature of VH, the 3DV should be designed for a specific population. Here, the tested 3DV is aimed at students in social work. It was designed as a pedagogical tool. For six working sessions (3 to 4 hours each), the team (including researchers and two students representing social work) and the voice-hearers have been defining script, keywords, rhythm, style, spatial composition, etc. We explored different VH: positive, negative, neutral, and command voices. These working sessions were non-structured in order to naturally discuss and elucidate the VH experience. We also used an iterative approach : from session to session, the actors/writers were proposing new scripts or demos for iterative validation with and feedback from the voice-hearers. These working sessions were gathering: (1) a recurring group of 2 to 5 voice hearers that freely discussed; (2) actors (one senior, two juniors directed by the senior); (3) researchers (acoustics, psychiatry, sound arts, etc.).

After these sessions, based on the validated 3DV script, the team conducted recordings at the CIRMMT (Fig. 1). For comparison purposes, multi-channel and synchronized records included: binaural with Numan artificial head, stereophonic with a cardioid pair, third-order Ambisonics with Zylia microphone. Next, these recording were cleaned to remove noises and assembled according to the script. This first version was then validated with voice hearers that gived feedabck. Modifications were asked, so new recording were done and combined to the first version. This second version was approved by all our co-researchers and the voice-hearers. The 3DV and working sessions were all done in French.

Figure 1 – Recording of the actors in a semi-anechoic room.

2.2 Participants

Nine social work students from UQAM (Canada) participated to the evaluation (sex: 7 women, 2 men ; gender: 7 women, 1 men, 1 genderqueer). They are considered as "naive listeners" (without experience in audio, confirmed in the form). Participants were informed about the study and provided written consent before participating.

*. gauthier.philippe-aubert@uqam.ca

2.3 Materials

The 3DV is a 45-minute sound file played on a portable media player using headphones (Sennheiser HD600). The volume was the same for each participant and it was adjusted to represent approximately 80% of perceived sound loudness of a normal discussion, as this was described by the voice hearers.

2.4 Procedure

In a room, participants were invited to complete the agreement form and were informed of the project. Next, they listened to the 3DV while going for a walk and perform their normal student activities (with the 3DV on). After listening, a qualitative evaluation was performed using an interview [4] on their general impression of the 3DV (not covered herein). Next, a quantitative listening test was done. Different versions (binaural, stereo, ambisonics to binaural) of the 3DV where compared and scaled from 0 to 100%. Adjectives (in French) were: sound quality, impact, immersion and envelopment, precision of localisation, externalization, distance, proximity. Each sound sample of 5 seconds for the adjectives were different while focusing on these key features.

3 Results

Preliminary quantitative results are in Tab. 1 with means in percent and standard-type (STD).

-2pt Besides, we had quantitative questions: "Do you see a clear difference between the sound samples ?", "How the rating was difficult ?", etc. From this, 5 from 9 replied that the differentiation was "difficult", one replied "very difficult", 3 replied "nor difficult, nor easy", and none replied "easy". Accordingly, the listening test was judged as a challenge, and therefore results must be interpreted carefully. This is also confirmed by the STD. The trend suggests that Ambisonics to binaural or stereophonic 3DV were the most well rated for all adjectives. Based on this trend, we think that binaural playback based on a binaural render of an Ambisonics microphone (spherical microphone array) is a better a avenue for future versions. Indeed, although it was slightly under-rated with respect to stereophonic sound for few adjectives, Ambisonics recording offer the possibility of spatial recomposition and also rendering for other devices such as Surround, loudspeaker arrays, etc. It would also be more appropriate for interactive binaural rendering using head-tracking for instance.

4 Discussion

Preliminary findings suggest that the 3DV is an effective tool for sharing the VH experience and can be used to increase experiential knowledge of VH among service providers. However, based on quantitative evaluation, the strong nature of the per se 3DV and the words used in the VH scenarios may make it difficult for participants to distinguish between different 3DV versions in terms of audio and immersion quality.

Although these results are limited to 9 participants, we expect more since the evaluation campaign is still active. Preliminary results from interviews suggest that the 3DV was efficient in sharing the VH experience. Still, we here focus on 3D audio and sound quality evaluation using the quantitative form. On this matter, by marked contrast with the qualitative insights, we think that it was difficult for participants to distinguish the sound quality (immersion, etc.) of different 3DV versions due to the strong and impactful nature of the 3DV and the words per se (stressful, annoying, distracting, etc.). Also, this might be explained by the fact, as mentionned, that all participants were self-identified "naive listeners." Besides, we must be careful in our interpretation, since the power of the preliminary data is very limited by its size. We can hardly conclude strongly on this matter from that data.

One limitation of the current data-set is that we only used social work students as participants. We also developed a 3DV for students to become physicians or psychiatrists. This second 3DV should be soon tested twith such students. A further limitation is the number of participants, our target of 30 participants should be reached in the future.

TABLE 1 – Results in terms of rating with respect to adjectives. Green, yellow and red indicate descending rating for each adjective.

	Binaural	Stereo	Ambi to bin
Quality	47.8% (32.7%)	60.2% (31.0%)	85.6% (19.4%)
Impact	53.4% (35.4%)	46.7% (39.7%)	74.8% (24.7%)
Immersion	51.3% (25.6%)	62.6% (25.2%)	61.1% (32.8%)
Precision	69.3% (35.7%)	66.8% (18.0%)	75.4 % (21.1%)
Externalization	35.0% (37.3%)	59.1% (28.7%)	50.1% (37.2%)
Distance	17.4% (12.7)	35.1% (26.4%)	20.8% (29.3%)
Proximity	56.0% (26.4%)	82.6% (24.2%)	57.1% (22.9%)

5 Conclusion

The co-creation and evaluation of a 3D voice hearing simulator were reported. It is designed to improve experiential knowledge and empathy for VH among service providers. Preliminary results suggested that the 3DV was efficient in sharing the VH experience for both voice hearers and participants. One of the success of the project was in the inclusion of voice hearers as non-traditional research collaborators with experiential knowledge in an horizontal and participatory structure. However, it was difficult for participants to distinguish the sound quality of different technological versions of the 3DV. To improve the immersion, interactive binaural audio with head-tracking or augmented reality may be necessary. For more personalized simulators, interactive and generative 3DV using context-dependent content and augmented reality is a potential avenue. This is planned future work.

Acknowledgments

Support: FRQ Audace. Ethics: UQAM (2023-4194).

References

[1] S. Patterson, N. Goulter, and T. Weaver. Hearing voices simulation : Process and outcomes of training. *J. Mental Health Training, Education and Practice*, 9 :46–58, 04 2014.

[2] J. Blauert. *Spatial Hearing*. MIT Press, 1996.

[3] J. Ahrens. *Analytic Methods of Sound Field Synthesis*. Springer, 2012.

[4] J.C. Kaufmann. L'entretien compréhensif. 1996.