Representing inner voices in virtual reality environments

Audio Engineering Society

Convention Paper 10639

Presented at the 154th Convention
2023 May 13–15, Espoo, Helsinki, Finland

Representing Inner Voices in Virtual Reality Environments

Kuura Parkkola[1], Thomas McKenzie[2], Jukka Häkkinen[3], and Ville Pulkki[1]

[1]*Acoustics Lab, Department of Information and Communications Engineering, Aalto University, Espoo, Finland*
[2]*Acoustics and Audio Group, Reid School of Music, University of Edinburgh, United Kingdom*
[3]*Visual Cognition Research Group, Department of Psychology and Logopedics, University of Helsinki, Finland*

Correspondence should be addressed to Kuura Parkkola (parkkola.kuura@gmail.com)

ABSTRACT

The inner auditory experience comprises various sounds which, rather than originating from sources in their environment, form as a result of internal processes within the brain of an observer. Examples of such sounds are, for instance, verbal thoughts and auditory hallucinations. Traditional audiovisual media representations of inner voices have tended to focus on impact and storytelling, rather than aiming to reflect a true-to-life experience. In virtual reality (VR) environments, where plausibility is favoured over this hyper-real sound design, a question remains on the best ways to recreate realistic, and on the other hand, entertaining inner and imagined voices via head-tracked headphones and spatial audio tools. This paper first presents a questionnaire which has been completed by 70 participants on their own experience of inner voices. Next, the results of the questionnaire are used to inform a VR experiment, whereby different methods to render inner voices are compared. This is conducted using a short film created for this project. Results show that people mostly expect realism from the rendering of inner voices and auditory hallucinations when the focus is on believability. People's expectations for inner voice did not change considerably in an entertainment context, whereas for hallucinations, exaggerated reverberation was preferred.

1 Introduction

The subjective auditory experience of an individual comprises numerous stimuli that originate not only from external sources in their environment but also from internal processes within the nervous system and brain [1]. Such sounds include, for instance, hallucinations and inner monologue. Internally generated sounds are henceforth referred to in this paper as inner voices. The psychology behind inner voices has been studied extensively in the literature, and the development of verbal thought has been researched since the early 20th century by psychologists such as L. Vygotsky and J. Piaget, whereas more recent studies have investigated the neurological mechanisms behind subvocal speech as well as its characteristics [2, 1, 3]. Hallucinations are a well-researched but not comprehensively understood topic. Over the previous decades, many potential mechanics which provide a sound basis for the phenomenon [4, 5] have been published.

While the psychology and neurology of inner voices is well understood, their reproduction is not. Films, video games, and various other media often present inner voices in particular ways that best fit the artistic vision of the work. These representations, however, typically rely on well-established industry conventions rather than scientific research. Realism is defined in this paper as authenticity to real life, whereas hyper-realism is an exaggeration and inauthentic, in order to

increase engagement and storytelling narratives. For example, film dialogue is expected to sound clear and crisp and similar sound effects are chosen for specific types of events. While these hyper-realistic approaches are immersive in more traditional media, virtual reality (VR) experiences engage with a spectator in a fundamentally different way by making the experience more dynamic and interactive. Therefore, the representation of inner voices in VR requires investigation.

This paper studies both realistic (plausible) and entertaining (engaging) reproduction of inner voices in VR environments and how prioritising either changes the experience. The aim is to answer the question of how sounds like the perceived sound of one's own thoughts, auditory hallucinations, and other such sounds should be represented, and by extension, how should they be rendered in practice using spatial audio techniques. This is approached via a two-stage study comprising a questionnaire and a listening test. The questionnaire aims to understand people's experiences and expectations regarding inner voices. The purpose of the listening test is to then test a set of inner experience simulations derived from the questionnaire results on a group of test subjects.

2 Inner Voices in the Brain

The human auditory system can be examined on three layers. A physical layer captures sound vibrations, a neurological layer transmits and processes them, and a cognitive layer perceives and understands them. The auditory system is also connected to speech production through a number of physical and neurological paths. Physical feedback results from sound vibrations created through speech being received as sensory inputs through the ears [6]. This phenomenon defines some of the characteristics, such as the effects of bone conduction, that people associate with their own voice. The neurological connections are important for the semantic understanding of language and various other cognitive phenomena. One such connection is the inner simulation of motor actions. As the premotor and motor cortices interact with muscles, copies of these stimuli are sent as feedback into the brain areas that typically receive relevant physical and somatosensory feedback [4]. This behaviour allows the brain to track its performance. Research has shown that these *efference copies* are sent regardless of whether physical movements are made [7]; therefore, speech is carried back to the auditory regions of the brain in both vocal and subvocal speech. This could be one factor contributing to the formation of inner speech.

The key difference between inner speech and auditory vocal hallucinations (AVH) is agency. Inner speech holds a sense of agency which is believed to enable the brain to detect language inputs originating from within. This is thought to be connected to efference copies [8, 5]. In hallucinations, the sense of agency is lost, which causes confusion in the brain since it can no longer accurately detect internally generated stimuli. AVHs are difficult to discern from reality [9].

While neurology can provide a context for how inner voices form, psychology helps to understand their perceptual features. One theory explaining human consciousness is the multi-component model of working memory [2]. The theory identifies three primary processes of consciousness: the central executive and its subsystems, the visuospatial scratchpad, and the phonological loop. The phonological loop itself comprises two components: a temporary store which can hold on to verbal content for a few seconds, and an articulatory rehearsal process which feeds and refreshes the store. The articulatory process is thought to be linked to vocal and subvocal speech.

The characteristics of inner voices have been approached from many directions. One influential theory is Vygotsky's theory on the development of language and thought [10], which argues that subvocal speech is an internalised version of voiced speech, often serving the same purpose as the egocentric speech, customary to small children. Studies done by J.B. Watson also found that children and adults employ similar techniques in problem-solving, with a distinction that children often go through their thinking process outloud, whereas adults think subvocally [11]. In a 2011 study, 380 people were interviewed on their use of inner speech [3]. The most common interactions were found to revolve around one's appearance and state of affairs, such as finances, stress and future, the planning of actions and future conversations, various problem-solving tasks, and self-regulation.

3 Questionnaire on Inner Auditory Experience

In the first phase of this work, a three-part questionnaire was created to collect statistics on people's inner

experiences. The questionnaire focused on three areas: inner speech, auditory hallucinations, and sound in dreams.

The topics of the questionnaire were chosen based on a number of expected characteristics: inner speech in VR should sound similar to the perceived qualities of one's own voice, with bone conduction present and little-to-no reverberation. The voice should be perceived inside the head. Auditory hallucinations should be difficult to tonally discern from physical sources, except when aiming for entertainment, where the separation should be easier to make. These assumptions were derived from background research and authors' introspection.

3.1 Questions

The goal of the first section was to acquire first-hand experiences of inner voices including ones that had not been thought of when designing the questionnaire, and was therefore worded to be open ended and not suggestive. The section focused on two types of inner voices: inner speech and imaginary hallucination-like sounds. The ordering of the sections was chosen to minimise bias by allowing the participants to freely think, before asking more direct questions. The questions are listed below:

- If inner monologue is a part of your thinking, how would you describe your inner monologue?
- Have you ever experienced any form of an auditory hallucination? If so, how would you describe it?
- Can you recall any dreams you have recently had, can you describe how different things sounded like in your dream?

The second section comprised three questions aiming to more directly validate specific assumptions. Each question had a selection of predefined options assumed to be the most common ones, but open responses were also allowed. The questions are presented below:

- Is your inner voice located... In the centre of your head? / In front of you? / Above you? / Other (specify).
- When you think about something someone else has said or might say, do you sense the voice as... Your own voice? / Their voice? / Other (specify).
- When you think about something you have heard, do you sense the sound... In the ambience of the space you were in? / In the ambience of the space you are currently in? / Without any ambience? / Other (specify).

The data collected in the first and second sections were difficult to use directly as inputs for concrete sound design parameters. In the third section, therefore, a selection of parameters was defined based on introspection while designing the survey. Here, values for each parameter were collected with sliders ranging from 0 - 10, each end corresponding to one of the edge cases for the parameter, as listed below:

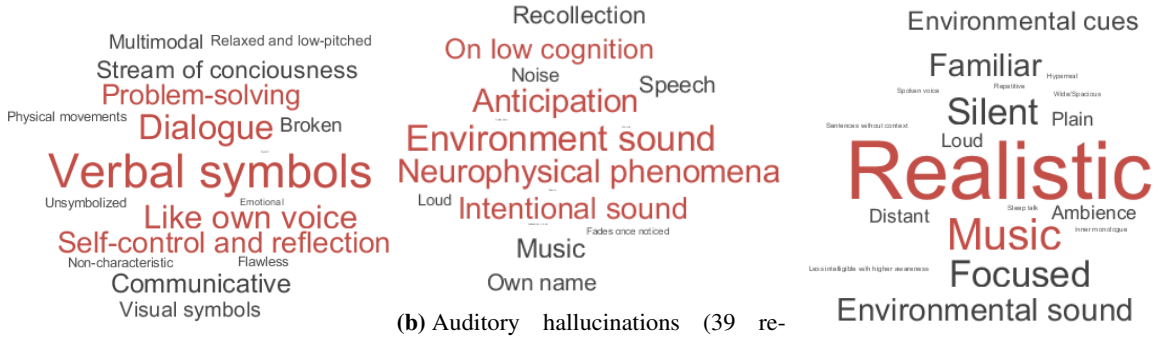
- Loudness - quiet to loud.
- Depth - boomy to thin.
- Spatial characteristics - dry to reverberant.
- Tonality - tonal to whispery.
- Location - inside to outside.

The participants repeated this part for each of three contexts: their experience of their own inner speech; their expectation for the tone of simulated inner speech; and their expectation for the voices of imagined hallucination-like characters.

3.2 Results

Responses were gathered over two months from 70 people of diverse backgrounds with ages ranging from 22 to 76 (median age of 32). Some participants did not answer all of the questions and some answers did not contain useful information, so the number of valid responses is disclosed alongside the data. With the wide range of ages among the participants, some minor age-related bias may be present in the results, particularly in the third stage of the questionnaire, which discussed relatively recent technology.

The first section produced open responses which had to be quantified before they could be analysed further. First, the responses were inspected for recurring features which were used to define a set of labels. The responses were then processed again marking each response with relevant labels. The labels and their relative frequencies are displayed in a word cloud in Fig. 1.



(a) Inner speech (65 responses)

(b) Auditory hallucinations (39 responses)

(c) Sounds in dreams (46 responses)

Fig. 1: Word cloud representations of the questionnaire results.

Many contributors characterised their inner speech as similar to their own voiced speech. Some of the subjects noted, however, that their largely verbal thoughts also contained visual and unsymbolised elements. Only a few people described the qualities of their inner speech, but the ones who did mentioned features such as lack of speech defects, being formed of incomplete sentences, and causing physical movements of the mouth, like when speaking out loud. Some specified the quality of the voice as relaxed and low-pitched. The most commonly reported uses of subvocal speech were inner dialogues preparing for social interactions, self-control, and self-reflection. Many subjects also reported using their inner monologue for problem-solving and explaining complex concepts for themselves. This is in keeping with the results of [3].

Characteristics of auditory hallucinations were reported by far fewer people compared to inner speech. Many responses described various neurophysiological phenomena, such as tinnitus, which have little relevance to inner voices. Environmental sounds, such as door creaks, footsteps, and police sirens were also commonly reported. Intentional sounds, including, for example, speech, music, and calls to one's own name, were less frequent but not rare. Speech and music were reportedly experienced in lowered states of awareness and high-stress situations, and environmental sounds were commonly paired with anxiety or as 'deja vu'-type phenomena.

Many participants reported not remembering their dreams or the sound in them. These sounds were mostly characterised as similar to any sound heard while awake. In some accounts, the sound would be extremely focused such that only the focal sound would be perceived with little to no background ambience. The most common sounds recalled from dreams were spoken words and musical tones.

Fig. 2: Heatmap of perceived inner voice location (70 responses).

The next part of the questionnaire featured more targeted questions focusing on the location, agencies, and acoustic features of inner voices. Firstly, Fig. 2 presents a heatmap of the locations in the head where inner speech was perceived to originate from. Most subjects considered their inner speech to come from the centre of their head (47 people), though some also stated the back of the head (5 people); either above the head (4 people) or inside the head, but towards its crown (4 people); near the eyes and mouth area (3 people); or all over in an indefinite location (7 people).

When recalling past conversations, most people recalled the experience of the other person in the conver-

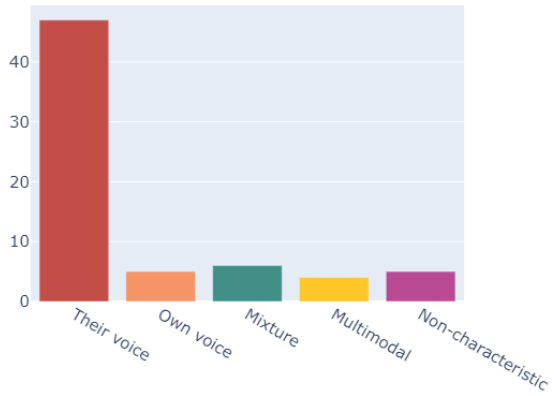


Fig. 3: The voice perceived in memories of previously had conversations (67 valid responses).

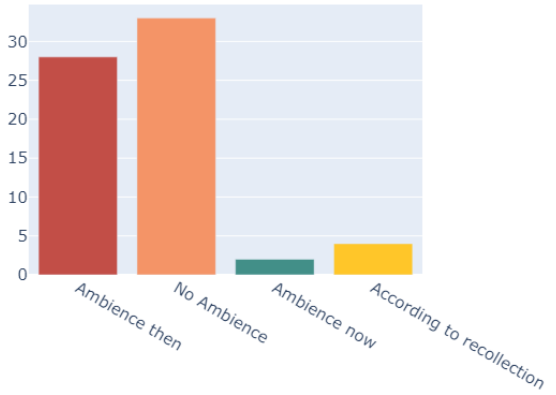


Fig. 4: The presence of acoustical characteristics and other ambience features in memories of past events (67 valid responses). Then stands for the time of the event, now refers to the time of completing the questionnaire, and according to recollection refers to only noticeable ambiences recalled.

sation speaking in that person's voice. Some people did, however, report how context and the topic of conversation play a role. A breakdown of the responses is displayed in Fig. 3. According to some responses, for example, if a person had contributed to an idea originally introduced by someone else, the voice would be different. Some people also felt the voice of the interlocutor to be their own voice, or an impersonal generic voice.

Room acoustic phenomena such as reverberation and elements of ambience was a divisive subject. While most participants recalled no ambience, the number of people who would recall the characteristics of spaces did not fall far behind. Some participants also felt that their recollection was stronger if the acoustics were distinctive or had significance in the memory. The responses are shown in Fig. 4.

The third section of the questionnaire collected parameters for the rendering of inner speech and other inner voices in VR. The results are displayed as violin plots in Fig. 5, whereby the representation shows both the distribution of the responses and the key statistics of the data. Inner speech was considered to be not particularly loud or quiet. The character of the inner speech was not particularly boomy nor thin, which can be interpreted as lacking the effects of bone conduction. Inner speech should have very little reverberation or spatial cues, and the voice should originate from within the head. The voice should resemble a normal speaking voice with a clear fundamental in favour of a more whisper-like tone. Nearly identical results were also reported for the expectations regarding the reproduction of the inner speech in a VR experience.

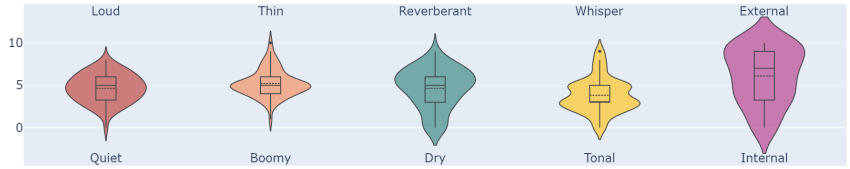
The participants' expectations for the characteristics of imaginary characters followed similar patterns to those of inner speech. The spatial characteristics are defined contrarily, however, with an emphasis placed on reverberation and the auralisation of the characters. The pseudo-acoustic space in which the character is located is expected to be reverberant, and the character's voice should not localise inside the head.

4 Rendering Inner Voices in Virtual Reality

The questionnaire identified several disagreements with initial assumptions. For some parameters, the results also varied significantly, making definitive conclusions difficult to draw. The primary goal in the second phase was to resolve the aforementioned uncertainties. The listening test approached the problem by deriving sound design parameters for an average participant based on the median values provided by the third section of the questionnaire. The integral questions were then tested with a number of alternate soundtracks to find which parameters best match the test subjects' experiences and expectations in terms of realism and entertainment.

(a) Reported characteristics of inner speech (70, 69, 69, 68, and 68 responses respectively)

(b) Expected characteristics of inner speech in VR (65 responses for each category)



(c) Expected characteristics of imaginary characters in VR (63, 62, 62, 63, and 63 responses respectively)

Fig. 5: Experienced and expected parameters of two types of inner voices, one's inner speech, and imaginary characters.

The source material for the listening test was a VR short film produced by YLE and Aalto University as a part of the Human Optimised Extended Reality (HUMOR) project. The film is set in prison, and the characters in the story include two (real) prisoners, an (imaginary) angel and demon, and the lead character himself who is also a prisoner. The main character uses his vocal and subvocal speech to communicate with the real and imaginary characters.

4.1 Test Design

Three questions were left unanswered in the first phase of the study. Firstly, the subjects did not report experiencing nor expecting their inner speech to carry over the effects of bone conduction. Secondly, the participants did not have a clear consensus on whether imaginary characters should be experienced internally or externally. Finally, for both, inner speech and imaginary characters, participants did not unambiguously agree on whether the sources should be reverberant.

To test these observations, four test scenes were created. The first two focused on inner speech, and the next two on imaginary characters. In each scene, a baseline render and two alternative renders were AB tested in three steps. The first scene evaluated the preferred amount of simulated bone conduction [12], the second scene compared realistic and synthetic reverberation on the inner voice, the third scene compared internal, external and close proximity source positions of the imaginary characters, and the last scene tested realistic and synthetic reverberation of the imaginary characters' speech. In each step the test subject was first asked to select the soundtrack they felt the most realistic, and then the one they felt was most entertaining. The order of comparisons of the soundtracks was randomised.

The test was conducted in a quiet conference room on an Oculus Rift S headset with Sony WH-1000XM3 headphones using a test environment implemented in Cycling 74 Max. The subject had three buttons at their disposal: A and B buttons which allowed them to switch on the fly between the two soundtracks currently under comparison, and a selector button that logged the preferred soundtrack. The test operator switched between scenes and test steps.

To promote intuitiveness and reduce the risk of overanalysing, the test subjects were not told what they should be listening for in each scene. This created

possible ambiguity, however, since an untrained ear might have difficulty in hearing differences between the soundtracks. Since the focus of the test was on the subjective and partially subconscious experience of the test subject, this ambiguity was considered a part of the experiment; thus, the test subjects were further instructed to answer using intuition. To understand the reasoning behind a participant's responses, the test operator discussed the scenes with the participants after the test had finished.

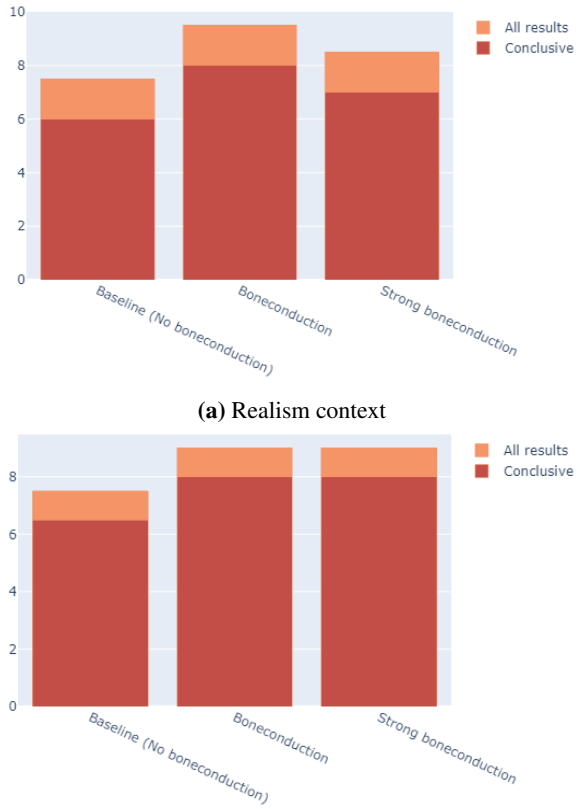
The soundtrack variations for the listening test were compiled from three types of source material. In-location Ambisonic recordings were captured as the video was taken. The same recording setup was used to separately capture the lead character's external speech. Mono recordings of dialogue were captured with lavalier microphones worn by the non-imaginary characters. The inner speech of the lead character and the voices of the imaginary characters were recorded in a studio environment onto mono tracks. The soundtracks were mixed in Steinberg Cubase. The resulting mixes were in Ambisonic format, and rendered to headphones using the Sparta VST plugin HO-DirAC binaural decoder[1] [13, 14]. The Oculus fed head orientation data to Max, which allowed for dynamic compensation of head rotations.

4.2 Results

The listening test was completed by 17 people with backgrounds mostly in acoustics and engineering, and ages ranging from 19 to 32 (median age of 24). In most cases the responses were unambiguous with one of the soundtracks in a scene selected more times than the others. On some occasions, a participant chose a different render in each of the three steps, thus not affecting the distribution of the responses. These ambiguous results are labelled here as inconclusive. The ratio between conclusive and inconclusive results is used to judge uncertainty since these responses display either indecisiveness or an inability to hear differences in the audio material.

The first scene considered the amount of simulated bone conduction expected by the listener in the inner speech of the lead character. The results are presented in Fig. 6. Many subjects reported that the differences

[1] https://leomccormack.github.io/sparta-site/docs/plugins/hodirac-suite/#binaural



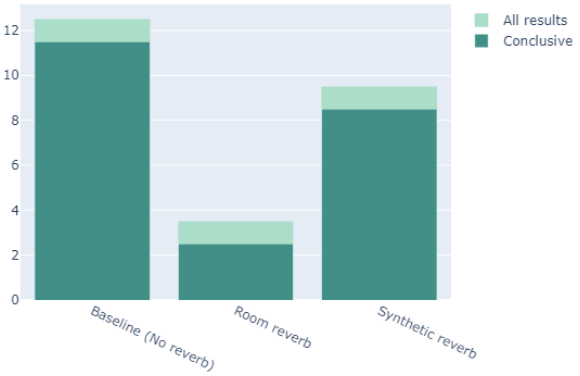
(a) Realism context

(b) Entertainment context

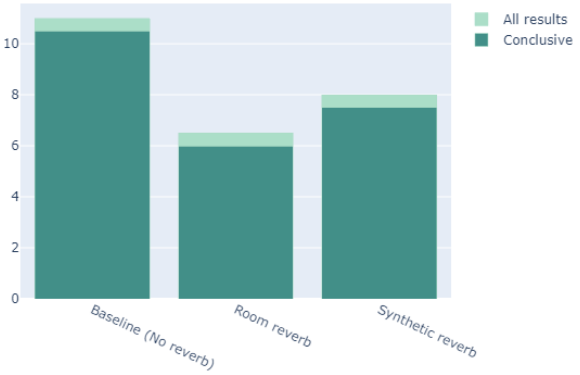
Fig. 6: Preferred rendering of bone conduction in inner speech.

between conditions were difficult to notice. In these cases the test subjects were instructed to make the selection based on their intuition. The similarity between the three options increases uncertainty in the results. The responses show a slight skew towards the soundtrack having a small amount of simulated bone conduction when looking for realism and an expectation for light to notable bone conduction in the entertainment context.

The second scene focused on the reverberation applied to the inner speech of the main character. The results displayed in Fig. 7 show that people expect their inner speech to stand out acoustically from the actual room in the case of realism; in the case of entertainment, the results are similar but less pronounced. It should be noted that in the second scene, a few people mentioned difficulty discerning the synthetic reverberation from the baseline. Some also pointed out that from the perspective of the spectator, the VR space seemed larger

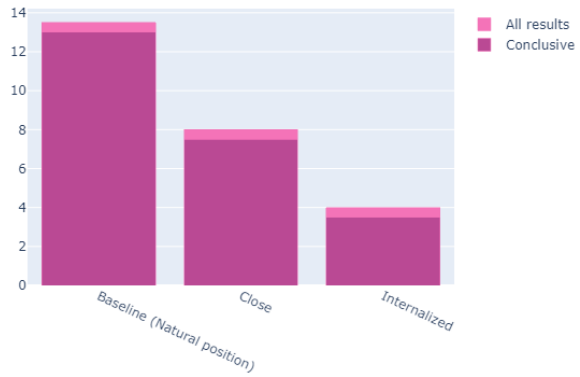


(a) Realism context

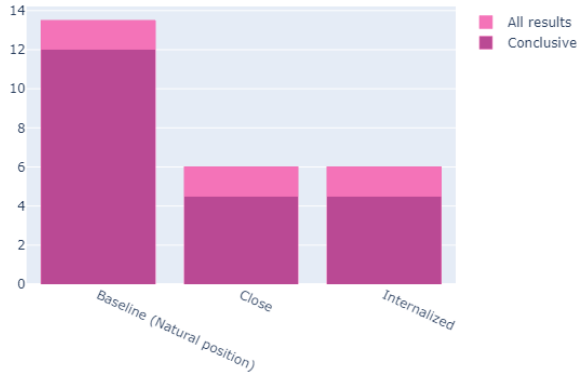


(b) Entertainment context

Fig. 7: Preferred reverberation of inner speech.



(a) Realism context



(b) Entertainment context

Fig. 8: Preferred location of imaginary characters.

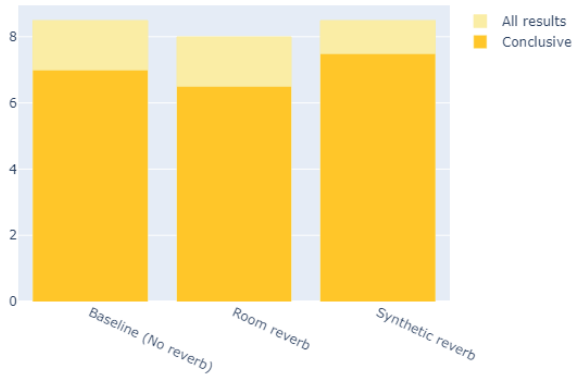
than it was in reality, causing the room reverberation measured in-situ to sound unnatural.

The results for the third scene are presented in Fig. 8. This asked whether the voices of imaginary external characters should be rendered inside or outside the head. Most participants considered a render with the voices placed in their natural position in the room as the most realistic and entertaining. The voices placed inside the listener's head were not considered realistic, however, in the context of entertainment, the ratio between the internalised voices and the voices rendered very close became equal.

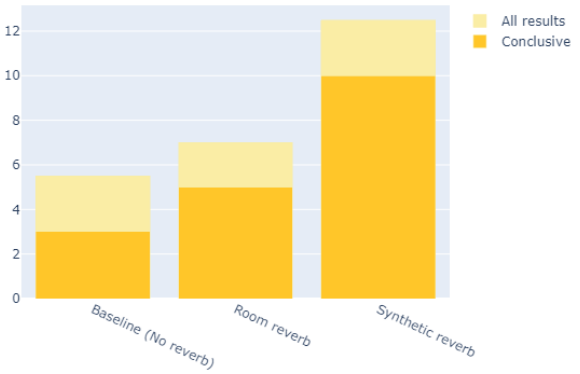
The final scene in the listening test examined the effect of reverberation in the voices of imaginary characters. The results are shown in Fig. 9. In terms of realism, choices were near-equal for all three options. Based on discussions with the test subjects, just as in the second scene, the room reverberation did not precisely match the perspective of the spectator, causing a sensory conflict. Imaginary characters, or hallucinations, were not familiar to many of the participants either. The expectations towards entertaining reproduction showed considerably more explicit results. In this context, the synthetic reverberation was voted to best match people's expectations.

5 Discussion

According to the initial assumptions of this study, the expected characteristics of inner speech should be close to the somatosensory response of speech, and hallucinations as realistic as possible. For the most part, these assumptions were confirmed by the questionnaire; the questions on the presence of bone conduction and the localisation of imaginary characters were left ambiguous, however. The listening test confirmed the initial assumption that hallucinations are in fact expected to be rendered outside the viewer's head. While it also seems



(a) Realism context



(b) Entertainment context

Fig. 9: Preferred reverberation of imaginary characters.

that the effects of bone conduction are expected in VR simulations of inner speech, the results do not provide a strong enough indication for definitive claims to be made and further investigation is therefore required.

The listening test results show notable variation in both reverberation-related questions. Both questions produced clear differences in the context of realism and entertainment. In the case of inner speech, dry non-reverberant speech was deemed the most realistic and entertaining. Some participants mentioned after the test that they had not chosen the room reverberation as realistic because, from the spectator's perspective, the space sounded smaller than it looked.

Auditory hallucinations are not an approachable topic for most people. An indication of this is the nearly 50% reduction in the number of questionnaire responses to the hallucination questions compared to those of inner speech. The same is also evident in the listening test, where participants disagreed on the acoustic characteristics of realistic hallucinations. On the other hand, participants were able to give more conclusive data on the entertainment characteristics of hallucinations. This suggests that fewer people personally experience auditory hallucinations, but have a clear idea of how they should be perceived in a film setting nonetheless. The clear peak on the synthetic reverberation suggests that voices of the imaginary characters should be rendered in some space, but the space should be easily discernible from the actual room. Many of the participants also confirmed this idea in the post-experiment interview.

6 Summary

This paper has investigated people's perception of their inner voices, namely, inner speech and auditory hallucinations, to find out how these sounds should be rendered using spatial audio techniques in virtual reality (VR) in either a realistic or entertaining way.

A questionnaire collected experiences of inner speech, auditory hallucinations, and sound in dreams in the participants' own words, as well as the perceived location of the inner speech, and recollection of agency and ambience via semi-open multiple-choice questions. Parameter data for the rendering of inner speech and auditory hallucinations was collected for use in the second stage of the study.

The parameters were evaluated in a listening test using a short VR film, whereby participants viewed the film in VR, whilst switching between different possible rendering options, before rating which was deemed the most realistic and entertaining. Results of the listening test suggest that people expect their inner speech in VR to include some effects of bone conduction. They also expect imaginary characters to localise to their perceived positions outside the head. The inner speech was expected to either have no reverberation or a large synthetic one. The subjects did not agree on what type of reverberation was expected for imaginary characters to be realistic. For the case of entertainment, the consensus was, however, that a synthetic reverberation that gives the characters space but isolates them acoustically from the actual environment sounded better.

Possible future avenues for research are the use of special effects to improve the realism or entertainment of VR video. The questionnaire gave inspiration for a

selection of commonly mentioned events, such as the sound in dreams fading away as one wakes up. Through a further listening test, it could be tested whether effects used in films or inspired by actual inner experiences improve immersion and the feeling of realism of VR media. A problem with inner speech representation in VR is that the voice actor heard by the spectator is not the voice of the viewer themselves. Since the experience of inner speech is closely tied to one's own voice identity, an unfamiliar voice is difficult to identify with. A research topic considered at an early stage of the study presented the idea of leveraging recent deep fake technologies to render the inner speech of a character with their own voice simulated by artificial intelligence.

7 Acknowledgements

The authors would like to thank Rébecca Kleinberger for her insights, and YLE for producing the VR film.

References

[1] Heavey, C. and Hurlburt, R., "The phenomena of inner experience," *Consciousness and cognition*, 17(3), pp. 798–810, 2008, doi:10.1016/j.concog.2007.12.006.

[2] Baddeley, A. D. and Hitch, G., "Working Memory," volume 8 of *Psychology of Learning and Motivation*, pp. 47–89, Academic Press, 1974, doi:10.1016/S0079-7421(08)60452-1.

[3] Morin, A., Uttl, B., and Hamper, B., "Self-reported frequency, content, and functions of inner speech," *Procedia-Social and Behavioral Sciences*, 30, pp. 1714–1718, 2011, doi:10.1016/j.sbspro.2011.10.331.

[4] Wolpert, D. M., Ghahramani, Z., and Jordan, M. I., "An Internal Model for Sensorimotor Integration," *Science*, 269(5232), pp. 1880–1882, 1995, doi:10.1126/science.7569931.

[5] Raij, T. T., Valkonen-Korhonen, M., Holi, M., Therman, S., Lehtonen, J., and Hari, R., "Reality of auditory verbal hallucinations," *Brain*, 132(11), pp. 2994–3001, 2009, doi:10.1093/brain/awp186.

[6] v. Békésy, G., "The structure of the middle ear and the hearing of one's own voice by bone conduction," *The Journal of the Acoustical Society of America*, 21(3), pp. 217–232, 1949, doi:10.1121/1.1906501.

[7] Tian, X. and Poeppel, D., "Mental imagery of speech: Linking motor and perceptual systems through internal simulation and estimation," *Frontiers in human neuroscience*, 6, p. 314, 2012, doi:10.3389/fnhum.2012.00314.

[8] Haggard, P. and Clark, S., "Intentional action: Conscious experience and neural prediction," *Consciousness and cognition*, 12(4), pp. 695–707, 2003, doi:10.1016/s1053-8100(03)00052-7.

[9] Dudley, R., Aynsworth, C., Mosimann, U., Taylor, J., Smailes, D., Collerton, D., McCarthy-Jones, S., and Urwyler, P., "A comparison of visual hallucinations across disorders," *Psychiatry research*, 272, pp. 86–92, 2019, doi:10.1016/j.psychres.2018.12.052.

[10] Vygotsky, L., *Thought and Language, revised and expanded edition*, MIT Press, Cambridge, MA, USA, 2012.

[11] Watson, J. B., *Psychology: From the standpoint of a behaviorist*, JB Lippincott, Philadelphia, PA, USA, 1919.

[12] Won, S. Y. and Berger, J., "Estimating transfer function from air to bone conduction using singing voice," in *Proceedings of the 2005 International Computer Music Conference*, International Computer Music Association, Barcelona, Spain, 2005.

[13] Politis, A., McCormack, L., and Pulkki, V., "Enhancement of Ambisonic binaural reproduction using directional audio coding with optimal adaptive mixing," in *IEEE Workshop on Applications of Sig. Proc. to Audio and Acoustics*, pp. 379–383, 2017, doi:10.1109/WASPAA.2017.8170059.

[14] McCormack, L. and Politis, A., "SPARTA and COMPASS: Real-time implementations of linear and parametric spatial audio reproduction and processing methods," in *Proceedings of the AES Int. Conf. on Immersive and Interactive Audio*, volume 2019-March, pp. 1–12, 2019.