forum acusticum 2023

DEVELOPMENT OF VIRTUAL REALITY SCENES FOR CLINICAL USE WITH HEARING DEVICE FINE-TUNING

Maartje M. E. Hendrikse[1*] **Gertjan Dingemanse**[1] **Giso Grimm**[2]
Volker Hohmann[2] **André Goedegebure**[1]
[1] Department of Otorhinolaryngology and Head and Neck Surgery, Erasmus Medical Center, Rotterdam, The Netherlands
[2] Auditory Signal Processing and Cluster of Excellence "Hearing4all", Department of Medical Physics and Acoustics, University of Oldenburg, Oldenburg, Germany

ABSTRACT

Virtual reality (VR) scenes were developed for a new clinical approach to fine-tune hearing devices. This new VR-based fine-tuning procedure allows hearing care professionals to try out various settings with hearing-device users in different scenes, so that individualized and optimized settings can be achieved for dedicated situations. The focus of this contribution is on the development of the VR scenes. We explain how we selected the different VR scenes, and what choices were made when implementing the VR scenes. Moreover, we present the results of a technical evaluation with two different hearing aids, showing how the VR scenes are classified by the automatic scene classifiers of two hearing aid brands.

Keywords: *hearing aids, cochlear implants, fitting, VR*

1. INTRODUCTION

The current standard clinical practice for fitting hearing devices usually takes place in a quiet, sound-treated room, which does not reflect the noisy and reverberant situations that hearing-impaired individuals encounter in everyday life. As a result, the settings of the hearing devices may be suboptimal and require further adjustments, which can be inefficient. Additionally, the clinical practice does not fully utilize the potential of modern hearing devices, which can automatically classify acoustic environments into multiple generic categories and adjust amplification and noise reduction settings accordingly. During clinical practice, typically only the basic amplification settings are adjusted, and the gain adjustments and noise reduction settings for each of the categories remain at the default values determined by the manufacturer.

The fitting process could be improved by fine-tuning the settings of each automatic environmental classifier (AEC) category in relevant everyday situations, which would allow a direct comparison of settings, leading to better individualized and optimized settings. Such a procedure would supplement, not replace, the current clinical practice and is hypothesized to lead to a more efficient fitting process, better final fit, and higher patient satisfaction.

To achieve this, we developed VR scenes to simulate everyday situations in the clinic. Using VR technology, the acoustic scene can be controlled, which allows playback of the same scene for easier comparison between settings. It is important to include both an acoustic and a visual scene in the VR scenes. Visual scenes increase immersion [1] in the VR scene and are important for speech intelligibility, as information can be extracted from mouth movements [2], face movements [3], and gestures [4]. Visual cues can also guide spatial attention [5] and affect self-motion [6, 7]. While a previous study has published an extendable set of complex audiovisual environments for hearing research [8], we found considerable effort was still needed to turn them into VR scenes applicable for hearing device fine-tuning in the clinic. The aim of this work is to give insight into the choices that need to be made in the implementation of VR scenes. Section 2 of this work focuses on the research question of which scenes are important for hearing device fine-tuning. Section 3 focuses on how these scenes can be

[*] ***Corresponding author***: m.hendrikse@erasmusmc.nl.

implemented in VR. Section 4 describes the technical evaluation that was carried out with two hearing aid brands to determine whether the automatic scene classifiers classified the VR scenes as the intended category. The VR scenes developed in this work are published [9] as an application that is ready-to-use but less modifiable than the set of audiovisual environments in [8].

2. SELECTION OF SCENES

To find scenes that are important for use in the hearing device fine-tuning process, first the patient perspective is considered. We wanted the scenes to represent everyday situations where hearing-impaired listeners experience problems and want to improve their hearing. Second, the technical perspective is considered. For the fine-tuning, it is important to keep in mind what possibilities different hearing devices have to adjust the amplification and noise reduction settings. Finally, we compare these two perspectives to the common sound scenarios framework proposed by Wolters et al. [10] and discuss how this leads to the final selection of scenes.

2.1 Patient perspective

The patient perspective was investigated in a group of 240 patients who visited the audiology department in the Erasmus MC in 2020 with hearing loss and/or tinnitus complaints and an indication for a hearing aid or bone conduction device. Patients were asked to complete an online survey before their visit to inform the audiologists about their hearing problems and individual situation. One of the questions asked patients to describe 3-5 goals they would like to achieve with their new hearing devices. For this study, the responses were categorized, and the numbers of responses within the categories were counted. The responses varied in detail, and aspects such as intent, location and circumstances were separated to structure the counting.

Following group conversations better and understanding better in one-on-one conversations were mentioned most often as hearing goals, by 53% and 43% of the patients, respectively. For the one-on-one conversations, most patients specifically mentioned they wanted to understand their partner better. Hearing better when watching TV (mentioned by 27% of the patients), listening to music (23%), and making (video) calls (17%) were also mentioned often. Specific locations where at least 5% of the patients wanted to hear better, are: in a meeting (19%), at work (15%), at a party with family/friends (13%), at home (11%), in a classroom (7%), in the car (5%), outdoors (5%), and in a restaurant (5%). Specific circumstances under which at least 5% of the patients wanted to hear better, are: in (babble) noise (16%), when the sound is coming from the side or from a distance (13%), when listening to children (11%), and when the speech is soft/whispered (5%). Furthermore, comments related to spatial hearing were made, most often about improving participation in traffic (27%), but also about wanting to hear which direction sounds come from in general (12%). Many patients also named hearing goals not necessarily related to specific scenes. Some individuals only mentioned they wanted to "hear better" (18%). Not having to ask for people to speak up or repeat their words (24%), and reducing tinnitus complaints (18%) and fatigue (12%) were also mentioned often. Finally, some comments were made about perceiving less annoyance from loud or unpleasant sounds (14%), and detecting alarm sounds better (5%).

2.2 Technical perspective

Most modern hearing devices have AECs that detect specific scenes and optimize the sound processing to enhance the listening experience for the user. This way, the appropriate noise processing can be applied in every situation, for example the use of wind noise reduction for outdoor situations. These classifiers are limited to a small number of categories, and manufacturers provide several options to adapt the sound processing strategies applied in each category. Most brands have 'quiet', 'speech', 'speech in noise', 'noise', and 'music' categories in their AEC. Furthermore, some brands differentiate between speech in noise and speech in loud noise, and include special noise situations in their AEC, such as wind noise, noise when driving in a car, machine noise, and reverberation.

2.3 Final selection

We combined the listening tasks with the different locations that were mentioned by the patients, and checked whether the scenes were sufficiently different from each other in terms of noise and other aspects. We also checked whether the scenes we distinguished would fit in one of the seven listening tasks defined in the common sound scenarios framework (CSSF) [10]: speech communication between two persons, speech communication between more than two persons, speech communication through media device, focused listening to live sounds, focused listening through media device, monitoring surroundings, and passive listening. Moreover, for each of the most common AEC categories ('quiet', 'speech', 'speech in noise', 'noise', and 'music') there should be a matching scene in the selection. Furthermore, it would be good to cover the special noise

situations in the classifiers. Patients also mentioned problems with listening to children, speakers from a distance or side, annoying sounds, and alarm sounds, which can be included in the scenes. In some scenes, hearing-impaired listeners would normally make use of streaming to their hearing devices with/without external microphone. Hearing devices have a separate streaming program for this, which may also need fine-tuning. It would be possible to stream the sound of a virtual sound source to the hearing devices using a TV connector of the same brand. Based on these considerations, the following scenes are selected:

1. At home, listening to a conversation between two persons in quiet. CSSF: speech communication between more than two persons. AEC: speech.
2. At home, listening to a conversation between two persons with soft background noise from kitchen appliances and radio. Meanwhile, the doorbell rings and someone yells from the kitchen. CSSF: speech communication between more than two persons. AEC: speech in noise.
3. At home, watching TV (with/without streaming). CSSF: focused listening through media device. AEC: speech (or streaming program).
4. At home, listening to music. CSSF: focused listening through media device. AEC: music.
5. At home, noise from the vacuum cleaner, the coffee machine and cutlery. CSSF: passive listening. AEC: (machine) noise.
6. In a pub, listening to a conversation between an adult and a child while sitting at a table with loud babble noise and music. CSSF: speech communication between more than two persons. AEC: speech in loud noise.
7. In a pub, listening to a conversation between three persons while sitting at a table with loud babble noise and music. CSSF: speech communication between more than two persons. AEC: speech in loud noise.
8. Standing at a street intersection where traffic is passing by. CSSF: monitoring surroundings. AEC: quiet/noise.
9. Standing at a street intersection where an ambulance with siren is passing by. CSSF: monitoring surroundings. AEC: noise.
10. Standing at a railway crossing where a train is passing by. CSSF: monitoring surroundings. AEC: noise.
11. In a meeting room, sitting at a table while a meeting is taking place. CSSF: speech communication between more than two persons. AEC: speech.
12. In a classroom, listening to the teacher in babble noise. CSSF: focused listening to live sounds. AEC: speech in noise/speech in loud noise.
13. In a car, listening to a conversation between the other passengers. CSSF: speech communication between more than two persons. AEC: speech in (car) noise.

The final selection does not include wind noise and the task 'speech communication through a device', because they are impractical to implement in VR. Instead of interacting with the VR user, pre-recorded conversations are used. Because of this lack of interaction, the situations with speech communication could also be regarded as focused listening. In all scenes with conversations, VR users are required to switch their attention between two or three talkers, which is considered 'speech communication between more than two persons' in the CSSF. The task 'speech communication between two persons' does not require attention shifts and was not included, because this would result in a monologue to the VR-user when there is no interaction. There is only interaction between the hearing care professional and the patient. The VR scenes facilitated this by providing a virtual representation of the hearing care professional, as explained in the next section.

3. IMPLEMENTATION

Three virtual environments were created to accommodate the living room, pub, and outdoor situations (situations 1-10 in section 2.3). Situations 11-13 were not implemented yet due to time constraints. The resulting VR scenes are described in section 3.1, and in sections 3.2 and 3.3 it is explained how the virtual acoustic and visual scenes were implemented.

3.1 Description of VR scenes

A description and images of the VR scenes are provided in the following. The online manual of the VR scenes [9] includes plots of the sound level at the listener position and azimuth relative to the listener position of all point noise sources, as well as measurements of the acoustic properties.

3.1.1 The VR-environment 'Living room'

The scenes within this environment (Figure 1) are:

1. Conversation between 2 persons in quiet: conversation between a woman sitting in a chair in front of the listener to the left (+39°), and a man sitting on the sofa next to the listener to the right (-87°). Ambient outdoor noise (birds) is present from an open door.
2. Conversation between 2 persons in noise: same as scene 1, but there is also background noise from a radio and noise of the dishwasher and cooker hood from the kitchen. While playing this scene, the sound

of the doorbell ringing and someone calling from the kitchen can be played at random time points.

3. Watching TV: the same ambient noise as in scene 1 is used. The TV sound is played from two locations to the left and right of the screen, and some distortion was added to simulate stereo TV loudspeakers of low/medium quality.
4. Listening to music: listening to music on the radio. The same ambient noise as in scene 1 is used. Three different music styles can be selected: classical (orchestra) music, instrumental jazz, or pop music with vocals. The music is played from the center of the two TV stereo positions.
5. Annoying sounds:
   a. Vacuum cleaner: the sound of a robot vacuum cleaner switching on and vacuuming. Background noise consists of the ambient noise, and noise from a dishwasher and cooker hood in the kitchen.
   b. Kitchen noises: someone putting cutlery in a drawer and turning on the coffee machine. The listener position switches to the kitchen. The background noise consists of the ambient noise, and noise of the dishwasher and cooker hood from the kitchen.

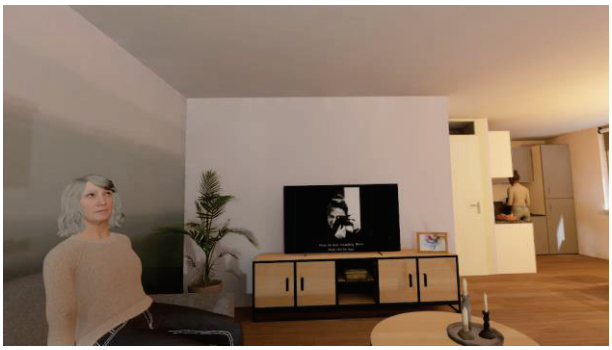

Figure 1: Visual representation of living room.

3.1.2 VR-environment 'Pub'

This environment (Figure 2) contains the scenes:

6. Conversation with child: Listening to a dialogue between adult and child at a table (-27° and -53°). As background noise, there is babble noise as well as conversation snippets from nearby tables. Moreover, there is music playing. Further away, at the bar, is the sound from glasses clinging and noises from the beer tap. Finally, at some time points there is the sound of chairs moving on the wooden floor. The babble noise was generated with ~16 persons speaking simultaneously, and the levels were chosen so that the resulting SNR is about +3 dBA.
7. Conversation between 3 persons in (loud) babble noise: Listening to a conversation between three talkers at a table (-27°, +44°, +77°). The background noise is similar to the scene 6, except for the babble noise and sound levels. The babble noise was generated with ~54 persons speaking simultaneously, and the resulting SNR is about 0 dBA.

Figure 2: Visual representation of pub.

3.1.3 VR-environment 'Street'

The scenes in this environment (Figure 3) are:

Figure 3: Visual representation of street crossing.

8. Monitoring traffic: the listener is standing at the corner of a street crossing where traffic is passing by (cars, pickup truck, bus, ambulance without siren). Moreover, a bicycle is approaching and ringing the bell, people are passing by from behind (footsteps and talking), and a mother with a baby in a pram is passing by (mother singing, baby babbling) and riding over an empty soda can. Finally, a car making an emergency stop, and a train passing by at a railway crossing further away.
9. Ambulance: this scene is a fragment of scene 8 (31 s to 54 s), but in this scene the rescue car has its siren on when it is passing by at the street crossing.
10. Railway crossing: this scene is a fragment of scene 8 (34 s to 86 s), but the listener position switches to the

railway crossing. A first order ambisonics recording made at a railway crossing is played back as diffuse sound. The warning sound of the railway crossing starts, the barriers close, a train passes by, then the warning sound stops and the barriers open.

3.2 Implementation of virtual acoustic scenes

To implement the virtual acoustic scenes, the room acoustic properties of the room or space where the scene takes place need to be simulated, and content matching the listening task and environment is needed. Appropriate sound levels have to be selected for the sources, and the scene needs to have an appropriate SNR. To implement the virtual acoustic scenes developed in this work, we used TASCAR [11], an acoustic simulation and auralization software package. To simulate the reverberation in the virtual acoustic scenes in a computationally efficient way that sounds perceptually plausible, first-order reflections of the sound sources on the floor, walls, ceiling, and nearby tabletops were rendered using an image source model, and late reverberation was rendered using a simple feedback delay network based on Schroeder [12] and Rocchesso & Smith [13]. The parameters of the feedback delay network were optimized to resemble the room IRs measured in the real rooms as described in [9]. TASCAR can use different spatial audio reproduction schemes for a loudspeaker ring; here horizontal higher order Ambisonics with embedded decoder was used (5th order with our setup of 12 loudspeakers). For all target and noise sources containing speech, audio recordings in Dutch were used, since the VR scenes are used in the Netherlands. All sound files were either recorded by the authors or have licenses that allow sharing, as listed in [9]. Music was downloaded from Opsound (no longer available) and the Free Music Archive, some other sound files from Freesound. Speech material from IFA Dialog Video Corpus [14] and Nederlandse Dialectenbank [15] was used.

For the virtual living room, we designed a new virtual environment, because we wanted a living room with open kitchen so that kitchen noises could be presented. The virtual living room is based on an existing room, in which the room IR was measured as described in [9]. Ambient noise was recorded in the existing room using a Zoom H3-VR first-order Ambisonics recorder at the listener position with the balcony door open, so that birds could be heard. The B-format recording was played back as diffuse noise in the virtual acoustic scene. A movie with Creative Commons license was used as TV signal [16].

The virtual pub is based on the pub presented in [8, 17]. Diffuse babble noise fitting this environment was generated by calculating the sum of the convolutions of the recorded IRs from the pub [17] with conversation snippets. The positions further away from the listener were used to generate the diffuse babble noise [P and S positions to first order ambisonics microphone at R1; see 17]. A distracting conversation was positioned at the table closer to the listener at the back, with separate sound source positions for each talker. To record the target and distracting conversations, the Zoom PodTrak P4 recording device was used, which could record from up to four AKG C 417 PP Lavalier microphones simultaneously on separate channels, while playing back the diffuse babble noise and microphone recordings via Sennheiser HD 599 headphones to the talkers. This way, the uttered speech matched the background noise, as talkers increase their vocal effort in noise [18]. The talkers' faces were filmed using webcams placed in front of them and recorded simultaneously using OBS Studio. The audio recording of the webcams was used to time-align the videos with the microphone recordings.

The virtual street was based on the acoustic street scene published in [19]. Some modifications were made, to the listening positions, the language of speech fragments, and a train passing at a railway crossing was added. The sound of a train passing at a railway crossing was recorded using a Zoom H3-VR first-order Ambisonics recorder. In scene 8 and 9, the railway crossing was far away, and the omni channel of the B-format recording was used as a point source. In scene 10, the listener position was close to the railway crossing, as when making the recording, and the B-format recording was rendered as diffuse sound. Because the street scene is outdoors, no late reverberation was included, only first-order reflections on the ground.

To allow communication with the hearing care professional, a virtual sound source was positioned in front of the listener and connected to a microphone. When this source is unmuted, the other sources are reduced in gain by 30 dB, to make communication in the noisy scenes easier.

To find appropriate sound levels for the different listening situations, the SNR values and noise RMS levels reported in [20] for conversation in different types of noise were taken as a reference. These reported values represent the mean of a range of sound levels that can occur in real life, so different sound levels may also be realistic. Wu et al. [21] show how large this range of realistic SNRs is, and both the sound levels of the VR scenes and the values from [20] fall within this range (see [9]).

3.3 Implementation of virtual visual scenes

The virtual visual scenes of course have to match with the acoustic scenes. For the virtual scene, an arrangement of 3D

objects is needed, which needs to be rendered by a game engine to be able to present it with a VR headset. Besides the geometry, 3D objects also need textures and materials to look realistic. The most complex 3D objects are the virtual characters, that also need to be animated. Characters that are further away from the user require less detail. Characters that are closer and speaking to the user require more detail and facial animations. When the 3D objects needed for a virtual visual scene are collected, they need to be imported in a game engine and arranged spatially. In the game engine, animations and lighting can be added as well as logic to start/stop the animations and control other aspects of the scene, and the scene is rendered (including shadows) so that it can be viewed with a VR headset.

When using a VR headset, cybersickness, related to motion sickness, can occur when the user's perception of self-motion does not match the user's visual perception. Symptoms can include nausea, oculomotor problems, and disorientation [22]. Risk factors for cybersickness are movements forced upon the users and a high motion-to-photon latency. Game engines usually provide guidelines to keep the motion-to-photon latency low and framerate high in VR applications [23, 24].

In this work, we used Unreal Engine to make the virtual visual scenes. Virtual characters that were further away from the user were made with MakeHuman, and for the virtual characters close to the user, Unreal's MetaHumans were used. Facial animations for the MetaHumans were generated using Faceware Studio, based on the video recordings of the conversations. The face of the hearing care professional can be recorded with a webcam, and Faceware Studio can send the facial animation data to a virtual character in the virtual scene, which appears when the microphone is unmuted. It is unclear how much information such facial animations provide for speechreading, but not as much as a video of a real person, although these techniques are developing rapidly.

For the pub scene, the 3D model of the room from Grimm et al. [17] could be used, but the lighting, shadows, virtual characters with animations, and logic to control the scene had to be added. For the street scene, the floor plan and motion tracks from [19] were used, and the 3D objects were added. The living room scene was newly developed.

To control the VR scenes, a GUI was programmed in Matlab. The GUI sends Open Sound Control (OSC) messages to TASCAR and Unreal in order to start the playback and the animations for each scene simultaneously.

4. TECHNICAL EVALUATION

We know that the AECs of hearing devices differ a lot between brands, and that they are still being developed and do not always give the right result [25, 26]. However, when using the virtual audiovisual scenes for fine-tuning it is important that the AEC would select the category that is being fine-tuned in that scene, otherwise the settings would become active in different real-world situations, which may lead to sub-optimal performance.

To check which AEC category was selected in the VR scenes, the procedure proposed by Husstedt et al. [26] was used. Using this procedure, first the acoustic output of binaurally fitted hearing aids was measured in one of the VR scenes as a reference, using in-ear microphones (Affinity Compact) on a dummy head. Next, a marker was placed in one of the AEC categories, by reducing the gain for all frequency bands in that category by 12 dB. Then, another recording was made when the setting with the marker was active. This was done for all AEC categories in all VR scenes, using Phonak Audéo M90 and ReSound ONE 961 hearing aids. Finally, each recording with a marker was compared to the corresponding reference recording to check for differences. In order to compare the recordings, the RMS differences in dB in all 1/3 octave bands between 500 Hz and 8 kHz were computed in 1 s time windows. If the mean difference per time window was >2 dB, the AEC category with marker was likely active.

The results showed that in most cases, the expected AEC category as listed in section 2.3 was active most of the time, after some initialization period. Some minor changes were made to the background noise in the VR scenes where this was not the case (the description in section 3.1 is after final changes). Phonak's AEC only selected the 'music' category in scene 4 when there was no background noise from the dishwasher and cooker hood. ReSound's AEC did not have a 'music' category. Moreover, the babble noise in scene 7 was changed to be less fluctuating (more speech snippets used) and louder so that Phonak's AEC selected the 'speech in loud noise' category. With the more realistic sound levels according to literature, as in scene 6, Phonak's AEC selected 'speech in noise'. ReSound's AEC switched between 'speech in noise (moderate/loud)' and 'noise (moderate/loud)' in scenes 6 and 7, selecting 'noise' more often in scene 7 (lower SNR).

5. DISCUSSION

In this work, we described what steps need to be taken to implement VR scenes for hearing device fine-tuning. It shows that considerable effort was needed to further

develop the virtual environments as they were published in [8] to VR scenes that can be used for hearing device fine-tuning in the clinic. We therefore decided to publish the VR scenes as an application that is ready-to-use. Unfortunately, due to licensing, this also means that the VR scenes are not as easily modifiable. Modifications may be necessary to implement all selected situations as VR scenes, and to make versions in other languages that can also be used with other VR headsets.

The developed VR scenes are a new tool for hearing device fine-tuning. With this tool, it would be possible to let patients try out different settings in different situations, so that settings can be compared more easily, and individualized and optimized settings can be achieved. It could also be a great tool for counselling patients about their listening behavior. An ongoing study is exploring a fine-tuning procedure using the developed VR scenes, with preliminary results discussed in [27].

We demonstrated that it is possible to develop VR scenes for listening situations that are important for hearing device fine-tuning and that are recognized by the AECs. However, this implementation is just one realization, and we do not yet know how representative the implementation is for similar listening situations that occur in daily life. In general we get positive comments of the hearing-users about the recognizability of the acoustic scenes. Still, a major drawback is the implementation of facial movements of the speakers. We are aware that the facial animations of the animated characters are not good enough to allow speechreading. Hearing-aid users and cochlear-implant users experiencing the VR scenes in the abovementioned study commented that it was confronting for them to realize how much they rely on speechreading. Some also commented that it was a positive effect that they could not use speechreading, because this forced them to focus more on the sound, but it would be more realistic if speechreading were possible. Moreover, the hearing-device users commented that the conversations were more difficult to follow for them than would normally be the case in such a situation, because the talkers did not take into account their hearing impairment and were talking too fast and unclear. This occurred because pre-recorded conversations were used instead of real interaction. Real interaction is possible and took place when the hearing care professional was talking with the hearing-device users. However, we opted for pre-recorded conversations so that multi-talker conversations would be possible, as well as playback of the same fragment. Repetition of the same fragment allows for an easier comparison, but it is also a complication that learning effects may make the fragment more understandable when repeating it more often.

Thus, the developed VR scenes are an important first step, and the best we could achieve with the current state of technology, equipment, and reasonable effort. However, more research and further development is needed regarding the need for more realism, especially of the facial animations, the need for multiple versions of the VR scenes with slightly different sound levels and content, and the need for real interaction with the VR user.

6. AVAILABILITY

The VR scenes and a manual describing how to install and use them are available under the CC-BY-NC-SA 4.0 license on Zenodo: https://doi.org/10.5281/zenodo.7092789 [9]. The manual also provides plots of the sound level at the listener position and azimuth relative to the listener position of all point sound sources and measurements of the acoustic properties of the VR scenes.

7. ACKNOWLEDGMENTS

The authors would like to thank Allart Knoop for making the dummy head for the technical evaluation. The authors would also like to thank the hearing-aid manufacturers for providing the devices for testing. This research was funded by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 101028117 (VR-FIT).

8. REFERENCES

[1] S. Jung and R. W. Lindeman, "Perspective: Does realism improve presence in vr? suggesting a model and metric for vr experience evaluation," *Frontiers in Virtual Reality,* vol. 2, p. 693327, 2021.

[2] A. Macleod and Q. Summerfield, "Quantifying the contribution of vision to speech perception in noise," *British Journal of Audiology,* vol. 21, pp. 131-141, 1987, doi: 10.3109/03005368709077786.

[3] K. W. Grant, "The effect of speechreading on masked detection thresholds for filtered speech," *The Journal of the Acoustical Society of America,* vol. 109, pp. 2272-2275, 2001, doi: 10.1121/1.1362687.

[4] L. Drijvers and A. Özyürek, "Visual Context Enhanced: The Joint Contribution of Iconic Gestures and Visible Speech to Degraded Speech Comprehension," *Journal of Speech Language and Hearing Research,* vol. 60, p. 212, 2017, doi: 10.1044/2016_JSLHR-H-16-0101.

[5] K. Jokinen, H. Furukawa, M. Nishida, and S. Yamamoto, "Gaze and turn-taking behavior in casual

conversational interactions," *ACM Transactions on Interactive Intelligent Systems (TiiS),* vol. 3, pp. 1-30, 2013.
[6] M. M. E. Hendrikse, G. Llorach, G. Grimm, and V. Hohmann, "Influence of visual cues on head and eye movements during listening tasks in multi-talker audiovisual environments with animated characters," *Speech Communication,* vol. 101, pp. 70-84, 2018, doi: 10.1016/j.specom.2018.05.008.
[7] M. Hartwig, V. Hohmann, and G. Grimm, "Speaking with avatars-influence of social interaction on movement behavior in interactive hearing experiments," in *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2021: IEEE, pp. 94-98, doi: 10.1109/VRW52623.2021.0002.
[8] S. Van De Par *et al.*, "Auditory-visual scenes for hearing research," *Acta Acustica,* vol. 6, p. 55, 2022.
[9] M. M. E. Hendrikse, G. Dingemanse, and A. Goedegebure, "Virtual audiovisual scenes for hearing device fine-tuning," *Zenodo,* 2022, doi: 10.5281/zenodo.7092789.
[10] F. Wolters, K. Smeds, E. Schmidt, E. K. Christensen, and C. Norup, "Common Sound Scenarios: A Context-Driven Categorization of Everyday Sound Environments for Application in Hearing-Device Research," *Journal of the American Academy of Audiology,* vol. 27, pp. 527-540, 2016, doi: 10.3766/jaaa.15105.
[11] G. Grimm, J. Luberadzka, and V. Hohmann, "A Toolbox for Rendering Virtual Acoustic Environments in the Context of Audiology," *Acta Acustica united with Acustica,* vol. 105, pp. 566-578, 2019, doi: 10.3813/AAA.919337.
[12] M. R. Schroeder, "Natural sounding artificial reverberation," *Journal of the audio engineering society,* vol. 10, no. 3, pp. 219-223, 1962.
[13] D. Rocchesso and J. O. Smith, "Circulant and elliptic feedback delay networks for artificial reverberation," *IEEE Transactions on Speech and Audio Processing,* vol. 5, no. 1, pp. 51-63, 1997.
[14] R. van Son, W. Wesseling, E. P. Sanders, and H. Heuvel, "The IFADV corpus: A free dialog video corpus," *LREC,* pp. 501-508, 2008.
[15] Nederlandse Dialectenbank [Online] Available: https://www.meertens.knaw.nl/ndb/
[16] B. Kommerij, "Flick Radio," ed, 2007.
[17] G. Grimm, M. M. E. Hendrikse, and V. Hohmann, "Pub environment," 2021, doi: 10.5281/zenodo.5534258.
[18] Y. Lu and M. Cooke, "Speech production modifications produced by competing talkers, babble, and stationary noise," *The Journal of the Acoustical Society of America,* vol. 124, no. 5, pp. 3261-3275, 2008, doi: 10.1121/1.2990705.
[19] G. Grimm, "Virtual acoustic street environment," *Zenodo,* November 29 2022, doi: 10.5281/zenodo.7355483.
[20] K. Smeds, F. Wolters, and M. Rung, "Estimation of Signal-to-Noise Ratios in Realistic Sound Scenarios," *Journal of the American Academy of Audiology,* vol. 26, pp. 183-196, 2015, doi: 10.3766/jaaa.26.2.7.
[21] Y. H. Wu, E. Stangl, O. Chipara, S. S. Hasan, A. Welhaven, and J. Oleson, "Characteristics of Real-World Signal to Noise Ratios and Speech Listening Situations of Older Adults With Mild to Moderate Hearing Loss," *Ear Hear,* vol. 39, no. 2, pp. 293-304, Mar/Apr 2018, doi: 10.1097/AUD.0000000000000486.
[22] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal, "Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness," *The international journal of aviation psychology,* vol. 3, no. 3, pp. 203-220, 1993.
[23] "Virtual Reality Best Practices | Unreal Engine Documentation." Epic Games. https://docs.unrealengine.com/4.26/en-US/SharingAndReleasing/XRDevelopment/VR/DevelopVR/ContentSetup/ (accessed 2022).
[24] "VR best practices - Unity Learn." Unity Technologies. https://learn.unity.com/tutorial/vr-best-practice (accessed 2022).
[25] A. Yellamsetty, E. J. Ozmeral, R. A. Budinsky, and D. A. Eddins, "A Comparison of Environment Classification Among Premium Hearing Instruments," (in eng), *Trends Hear,* vol. 25, January 2021, doi: 10.1177/2331216520980968.
[26] H. Husstedt, S. Wollermann, and J. Tchorz, "A method to analyse and test the automatic selection of hearing aid programs," in *International Symposium on Auditory and Audiological Research*, Nyborg, Denmark, 2017, vol. 6, pp. 143-150.
[27] G. Dingemanse, M. M. E. Hendrikse, and A. Goedegebure, "Evaluation of a new VR-based hearing device fine-tuning procedure," presented at the Forum Acusticum, Turin, Italy, 2023.