DOI: 10.61782/fa.2023.0301

EVALUATION OF A NEW VR-BASED HEARING DEVICE FINE-TUNING PROCEDURE

Gertjan Dingemanse[1*] **Maartje M.E. Hendrikse**[1] **André Goedegebure**[1]
[1] Department of Otorhinolaryngology and Head and Neck Surgery,
Erasmus Medical Center, Rotterdam, The Netherlands

ABSTRACT

When providing hearing devices to an individual, it is important to adjust the settings to fit their specific needs. This process, known as fine-tuning, is typically done in a quiet consultation room. However, this is not always reflective of the real-world situations that hearing-impaired individuals may encounter. To address this issue, a new fine-tuning procedure using virtual audiovisual environments has been developed. This method, which utilizes Virtual Reality technology, allows hearing care professionals to test different settings with hearing-device users in a variety of different situations. This way, professionals can ensure that the settings are adequate for the individual in various real-world scenarios and it is easier and faster for the users to compare different settings. Additionally, it is possible to fine-tune situation-specific settings in the automatic scene classifier present in many modern hearing devices in a matching listening situation. The effectiveness of this new fine-tuning procedure was evaluated through a study involving both hearing-aid and cochlear-implant users. Participants compared the fitting results of the standard and VR-based fine-tuning procedure over a period of two times two weeks and provided feedback through questionnaires and listening tests. Preliminary results of the cochlear implant users group are presented.

Keywords: *cochlear implants, fitting, virtual reality.*

1. INTRODUCTION

Hearing loss is a relatively common disease, at least in older people. Hearing aids (HA) can help to partially overcome the problems associated with hearing loss. For severe-to-profound hearing loss, hearing aids are less effective and cochlear implants (CI) are the treatment of choice. A CI consists of an electrode array that is inserted in the cochlea and a sound processor that transforms the incoming sound into electrical signals delivered to the auditory nerve by the electrodes in the cochlea. Both HAs and CIs have to be fitted to an individual patient, according to their hearing thresholds, sound perception, and individual needs or preferences, while taking into account the daily communication situations of the patient.

The fitting of HAs is based on the gain that is prescribed by a prescription procedure that aims to make speech intelligible and overall loudness comfortable. Several prescription methods are available, mainly based on the pure-tone thresholds [1]. Extensive research has resulted in several adjustments to the prescription rules, and current versions aim to give on average an appropriate gain for most hearing aids users. However, in clinical practice, fine-tuning is often needed to fit individual needs and preferences, after a trial period in the daily-life settings of the patient.

The fitting of CIs is based on psychometric measures: the perceptual threshold (T-level) and the most comfortable loudness level (M-level) are determined for each electrode. Therefore, subjective loudness is an important factor in CI-fitting. After a revalidation period and home trials, fine-tuning of the stimulation levels is needed.

Although hearing devices restore audibility of most sounds, more complex hearing functions like separating a sound from the background, remain difficult. Therefore, in addition, hearing devices need to aid the patient in filtering out the background noise in order to improve the ability to

---

[*] ***Corresponding author**: g.dingemanse@erasmusmc.nl.*

discriminate or understand the sounds or the speech that are important for the patient. For this purpose, advanced signal processing algorithms are nowadays included in most hearing devices, like noise reduction algorithms and advanced directional microphones [2]. These signal processing algorithms contain many parameters that can be adjusted according to the individual needs of the patient and the situation of use [3]. However, it is often not clear how to adjust these parameters in order to improve the effect of an algorithm on one's individual need. We call the combination of adjustable amplification (HA) or loudness (CI) settings and algorithm parameters the fitting parameters.

As the optimal fitting for one situation is not necessarily optimal for other acoustical situations, modern hearing devices automatically classify acoustic environments into multiple generic categories in order to select which signal processing algorithms are active and sometimes slightly adjust the amplification settings [4]. Such automatic environmental classifiers (AEC) usually allow a manual adjustment of the fitting parameters in each category. In general, automatic classifiers have some or all of the following categories: Quiet/calm, Speech (soft/loud), Speech in noise (moderate/loud; babble/car/other noise), Noise (moderate/loud; car/echo/machine/wind), Music. Manufacturers provide in their fitting software predefined settings for fitting parameters in each category, but little is known about how the classifier performs in the acoustical situations of the patient and how to optimize the parameters for these situations. It is known that AECs prioritize speech intelligibility, whereas some patients prefer a higher listening comfort, even if that leads to a slight decrement in intelligibility [5]. Being able to individualize the fitting parameters of the AEC-categories to achieve a good trade-off between intelligibility and comfort for each individual could improve the fitting process.

In current clinical settings, the fitting takes place in a quiet, often sound-treated, room, which is totally distinct from the noisy and reverberant situations that occur in daily life. The fitting is based on audiometric outcomes, like speech recognition scores. Also the input of a patient regarding the experiences in daily life is important and the clinician has to translate this input into a change of fitting parameters. However, it is not possible to try out the new settings immediately and the patient has to go home to try the new fitting. This often makes the fitting a process of trial and error, which is time-consuming. If a classifier available in the hearing device, often only one category is adjusted due to time constraints, and the settings of the other categories are derived from the adjusted category, according to internal rules of the fitting software. Given the limitations of the current clinical practice, the fitting process could be greatly improved if the clinician could test the adjustments in simulated relevant daily-life situations

With recent advances in virtual reality (VR) [6], we now have the ability to simulate daily-life situations in the clinic. We wanted to study whether a VR-environment can improve the fitting process by making it more efficient and whether this would result in better patient satisfaction. We designed a number of virtual audiovisual environments to simulate relevant daily-life listening situations and developed a procedure to incorporate this VR- technology in the fine-tuning process. This fine-tuning procedure is an addition to the standard clinical procedure. This way, clinicians and patients can try out adjustments of the fitting parameters immediately in relevant situations.

The aim of the study was to evaluate whether an additional VR fine-tuning results in a better final fit compared to the standard clinical practice in terms of preference, speech intelligibility in noise scores, perceived sound quality and overall satisfaction with the hearing device. Preliminary results obtained in a group of CI-users are presented.

2. METHODS

2.1 Study design

The study is a double-blind cross-over fundamental translational intervention study. Because of differences in hearing threshold and maximum speech recognition scores between participants, performance on listening tests can vary between participants. Therefore, a within-subject comparison and cross-over design is used. It is a fundamental translational study, because it is a first step in applying a new technique in the clinic.

Inclusion criteria were: age between 18 and 80 years, fluent in Dutch, a minimum free-field phoneme score of 65% at 65 dB SPL on a Dutch speech test [7], the hearing device has an automatic sound classifier.

During an initial visit, the devices were fitted or adjusted according to a standard clinical fitting protocol for setting of the T- and M-levels (given in clinical units, which represent constant charge). For all participants some fine-tuning was done if necessary, based on the first impression of the changed fitting, the preference, and the needs of the participant. Next, a new fine-tuning session was performed in the VR-setting, according to the VR fine-tuning procedure, as described in section 2.3. Participants tried both fits in random, blinded order, for two weeks each and rated their hearing experience with both fits. After four weeks, they returned to the clinic for listening tests (section 2.4).

2.2 Participants

Ten CI-users participated in this study (age (median: 61, range: 40-75) years; 4 male, 6 female). They were unilaterally implanted with an Advanced Bionics implant and wore a Naida M90 processor, since 16.5 (median) months. Six of them wore a hearing aid (Phonak Naida Link M) at the other ear, that was paired with the CI-processor.

2.3 VR Fine-tuning procedure

Several virtual relevant listening situations were created, in a sound-treated room. By playing realistic sound fields, using a ring of 12 loudspeakers (Genelec 8020), an acoustic environment was created. A matching animated visual environment was presented by a VR headset (HTC Vive Pro Eye). For details, see [8]. For this study three virtual environments were used: 1) a living room, with a 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°), some background noise, a TV, music and sounds from a vacuum cleaner, coffee machine, kitchen, doorbell, which all can be switched on and off; 2) a pub/restaurant, with a conversation between an adult (-27°) and a child (-53°) at a table, or a conversation between three talkers at a table (-27°, +44°, +77°), and background sound from several groups talking at different distances and music; 3) Street crossing, with various traffic to monitor, a railway crossing, and an ambulance.

The CI-processor has an AEC with categories: 'Calm situation', 'Speech in noise', 'Speech in loud noise', 'Comfort in noise'. All categories make use of the same map with T- and M-levels and only one setting of ClearVoice (noise reduction algorithm with settings Off/Weak/Moderate/Strong) and SoftVoice (noise reduction algorithm to increase soft voice intelligibility, settings On/Off) is possible. Microphone directionality (settings Omni to Ultrazoom, range 0 - 20) and SoundRelax (impulse noise reduction, settings Off/Weak/Moderate/ Strong, range 0 - 20) can be changed within each AEC-category. The fine-tuning by adjusting these fitting parameters was performed by experienced audiologists They adjusted the fitting parameters according to a participants' responses to standard questions regarding the intelligibility of speech, the audibility, clarity and naturalness of sounds, the loudness of sound, or the annoyance of loud sounds.

In the living room participants listened to the conversation in quiet and the AEC-category: 'Calm situation' was selected. Participants judged the intelligibility of the speech and its naturalness. M-levels were fine-tuned in response of this judgment by changing them in 3 frequency regions (low/mid/high frequencies). In addition, it was tested whether the SoftVoice algorithm contributes to the intelligibility. Next, the music was switched on and the AEC-category: 'Speech in noise' was selected. Microphone directionality was fine-tuned. Next the AEC-category: 'Comfort in noise' was selected and with noises of a vacuum cleaner, a coffee machine, and cutlery noises the settings of noise reduction algorithms ClearVoice and SoundRelax were fine-tuned.

In the pub/restaurant environment the AEC-category: 'Speech in noise' was selected for the conversation between an adult and a child. Finetuning options were higher M-levels for mid/high frequencies and a change in directionality. The default was the Ultrazoom setting for directionality. The AEC-category: 'Speech in loud noise' was selected for a conversation with three persons in loud noise. The overall loudness was fine-tuned by lowering of the M-levels overall.

In the street crossing environment the AEC-category: 'Comfort in noise' was selected and it was checked whether the traffic was audible and were detected on the right side. The directional microphone setting was fine-tuned. The overall loudness was fine-tuned by lowering of the M-levels overall for the sound of the railway crossing and the ambulance. For participants with a contralateral HA paired with the CI-processor, the settings of directionality and SoundRelax were also changed in the HA and the loudness balance between HA and CI was checked.

2.4 Outcome measures

The differences between M- and T-levels for the clinical fit and VR-fit were analyzed, as well as differences in preprocessing, such as the directionality of the microphone and strength of the impulse noise reduction.

Participants filled in the Speech, Spatial, and Qualities questionnaire [9] at the end of each trial period.

In the final session several tests are performed. During these tests, the hearing devices were connected to the fitting software, so that the appropriate automatic classifier program could be activated.

Speech intelligibility in noise was measured with the Dutch VU (Vrije Universiteit) female-spoken sentence material [10], coming from +/-45° at 70 dB SPL speech level. As diffuse restaurant noise, the mensa noise recording from Grimm & Hohmann [11] was used, which was filtered to resemble the long-term average speech spectrum of the VU sentences and rendered to a 12-loudspeaker layout. The level of the noise was adapted according to a stochastic approximation procedure to determine the speech reception threshold for 50% word score (SRT50) [12]. The size of the noise steps was 4·(0.5−*PC(n−1))* dB, in which *PC(n−1)* is the proportion correct words of the previous trial. First a training list was presented and next two lists of 13 sentences were used per condition. In one condition, the ‘speech in noise’ program from the automatic classifier of the clinical fit was used, in the other condition the one from the VR-fit. The order was randomized across participants.

Paired comparisons of both fits were used to determine the preference in four situations: the conversation in the living room in quiet (‘Calm situation’ program active); the conversation in the living room in noise (‘Speech in noise’ program active); the conversation between 3 persons in the pub in loud noise (‘Speech in loud noise’ program active); the street situation (‘Comfort in noise’ program active). The same fragment was played back for both fits (randomized order) and participants were then asked to indicate whether they had heard a difference between the two fragments, whether the first or second playback had their preference (or no preference), and whether this was a slight or pronounced preference. The preference rating was converted to a numeric score for the analysis: no difference or no preference = 0; slight preference = 1; pronounced preference = 2. The numeric scores were positive if the VR-fit had preference, and negative when the clinical fit had preference. The total score was the sum of the preferences in the four situations.

Finally, the final choice of the participants regarding which fit they wanted to keep was used as an outcome measure.

3. RESULTS

Fig. 1 shows the changes between M- and T-levels for each of the participants. On average the M-levels were higher in the VR-setting for all electrodes and most for electrodes 10 to 14, corresponding to the frequency range of 1.6 – 3.9kHz.

Fig. 2 shows the directionality was reduced in the AEC-category ‘calm situation’ for some participants and in the

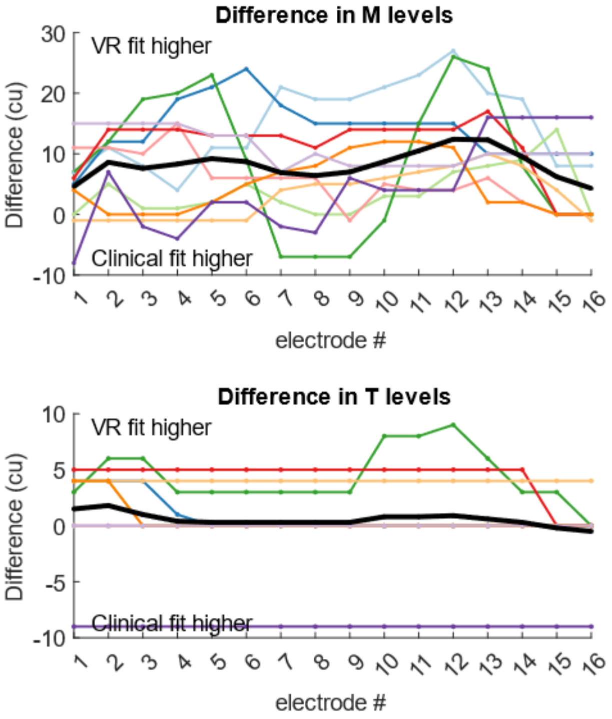

Figure 1. Differences in M-levels and T-levels between the VR-fitting and the clinical fitting for individual participants. The thick black lines show the average of the difference.

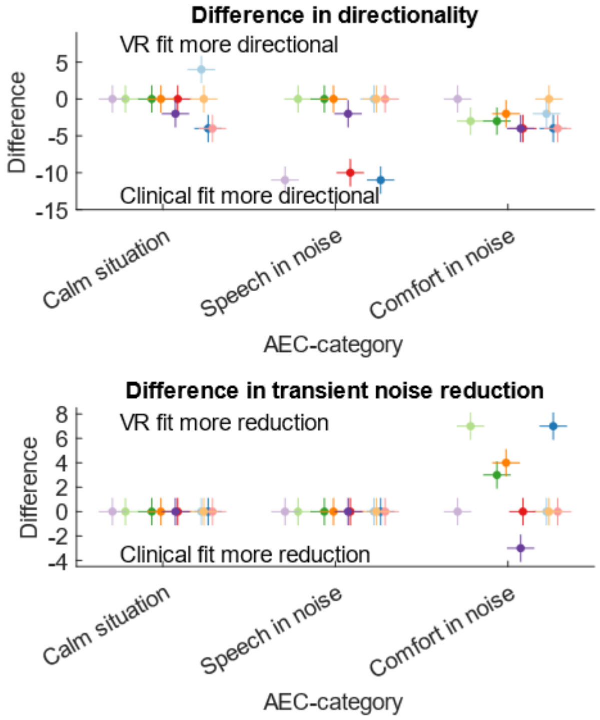

Figure 2. Differences in settings of microphone directionality (upper panel) and transient noise reduction SoundRelax (lower panel) between the VR-fitting and the clinical fitting for individual participants in different AEC-categories.

AEC-category 'Comfort in noise' for most participants. In AEC-category 'Speech in noise' the directionality was lowered for three participants. SoundRelax was only changed for AEC-category 'Comfort in noise' for several participants. SoftVoice was activated in 4 participants and was already switched on in the other participants in their clinical settings. ClearVoice was active in the clinical setting at 'Moderate' level in 9 participants, and it was 'Off' in one participant. For this participant ClearVoice was set to 'Moderate' in the VR-fitting. In two other participants it was set to the 'Weak' setting, and in one participant it was switched off.

Fig. 3 shows the final choice (VR-fit or clinical fit), the paired comparisons rating, changes of the SRT50, and the changes of the SSQ-scores. The gray areas in the figure show which difference is needed to reliably detect a difference between two scores at a 95% confidence level in an individual. These are calculated as 1.96·√2·SD of a single measurement and were based on [12] and [13].

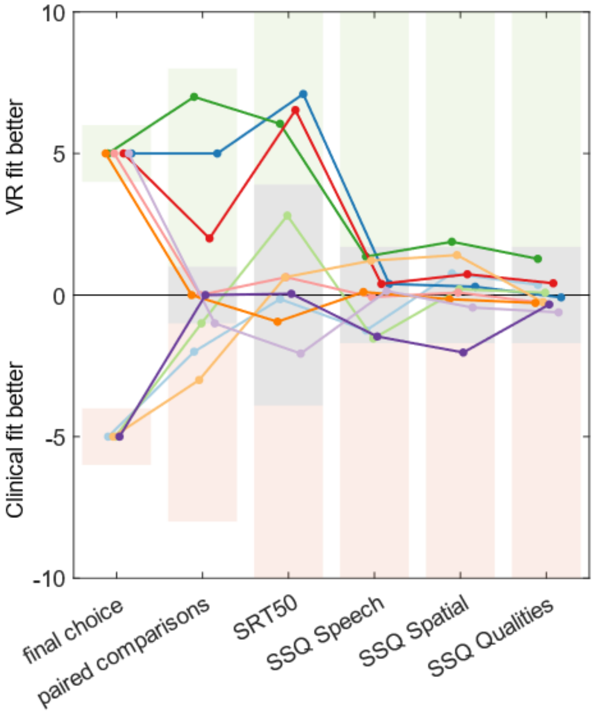

Figure 3. Differences in outcome measures between the VR-fitting and the clinical fitting for individual participants. The final score is VR-fit(+5) or clinical fit(-5), the paired comparison is the sum of the relative preference of four situations, the SRT50 is given in dB, and the maximum difference in SSQ subscales range from 0 to 10. The gray areas give estimated test-retest differences at the individual level (see text for details).

In three participants the VR-fit showed a significant improvement in SRT50 of more than 5 dB compared to the clinical fit. These participants preferred the VR-fit in the paired comparisons and chose the VR-fit as final program. Two subjects preferred the clinical fit and finally chose this fit, although no significant difference in SRT50 or SSQ was found. For the remaining five participants no significant differences were found, and three of them chose the VR-fit and the other two the clinical fit.

Because of the relatively small sample size in the preliminary data, results were only analyzed on individual level and not on group level.

4. DISCUSSION

The preliminary results of this study show that in part of the CI-users a significant benefit is obtained after the fine-tuning procedure in the VR-environments. This benefit is related to better speech recognition in noise scores. Eight out of ten participants rated the VR-fitting positive or neutral in the paired comparisons.

In most of the participants the M-levels were increased. The profile of M-levels adjustments over the electrodes is very different between subjects. An explanation for this observation is that clinical fitting procedures are partly based on loudness scaling with narrowband stimuli. Such stimuli may cause some annoyance and may result in lower, less optimal M-levels for some frequency regions. The VR-environment uses broadband sound, which is related to visual events, making these sounds more acceptable. In the VR-environment the M-levels can be raised for most subjects. However, the increase of M-levels did not result in better speech recognition as in most participants no significant difference in SRT50 is found nor in higher preference scores.

The perceived sound quality or naturalness of the sound may have influenced the final choice and the paired comparisons score. The four participants that chose the clinical fit at the end of the study had the smallest change in M-levels for the frequencies below 830Hz (electrodes 1-5). The M-levels are tilted towards the higher frequencies. This may have resulted in a less natural sound. It did not change the SRT50, nor the SSQ score (except the SSQ Spatial score in one of these patients) and only two of the four had a clear preference for the clinical fit in the pairwise comparisons. So, the differences between the clinical fit and VR-fit seem to be small here, and the VR-procedure did not make the fit much worse.

The SSQ scores appeared to be not very sensitive to changes in the fine-tuning of the CI. We expected a change in SSQ Speech for the participants with improved SRT50, but also these subjects did not report a better experience in the speech in noise situations the SSQ asks for. Even so, we expected an influence of the fine-tuning on the SSQ Qualities subscale, but we did not find any improvement. May be the fine-tuning steps are too small or may have some drawbacks in other aspects. For example, ClearVoice removes unwanted noise, but may also introduce some signal distortions [14].

The finetuning of multiple AEC-categories in various environments appeared to be very complex. As the fitting software did not allow a specific map of M- and T-levels for AEC-categories, these levels were changed several times in different environments. Given this limitation, we carefully designed the order of de environments. First, the calm situation with focus on intelligibility and naturalness, then followed by the various speech in noise situations. The large variation of M- and T-levels over participants and within participants reflects the multiple adjustments that were done in the different AEC-categories and different environments. Despite the used order in adjustments, the final M- and T-levels may not be best end result for all situations. Ideally, in the end the calm situation should be checked again, but this was not done in this study due to time constraints.

During the finetuning in the VR-environment and the paired comparisons, the same situations and sound tracks were used several times. This may have caused some habituation to the situation and a learning effect with respect to the recognition and intelligibility of the speech, making it more difficult to decide if a positive response of a participant was due to the finetuning or due to the habituation and learning effect. In future studies this aspect can be improved by using multiple fragments that were equal in perceptual aspects as intelligibility and sound quality ratings. However, this requirement is difficult to achieve and may be in conflict with the requirement that the VR-environment reflects the variation that is experienced in daily life.

VR-environments make it possible to simulate daily-life situations. However, they are just a small subsample of the largely varying situations in daily life. The question is whether the VR-environments are sufficiently representative and whether fine-tuning in the VR-environments results in a transfer to better experience en performance in other real life settings. In general, participants reported that they experienced the different environments as relevant and realistic. They appreciated this way of CI-fitting very much.

Despite this positive experiences, the results show that fine-tuning in a VR-environment not automatically results in better outcomes. Although eight participants rated the VR-fitting positive or neutral, two participants rated the clinical fitting better in the paired comparisons. These comparisons were done with exactly the same VR-environment and sound samples as used in the fine-tuning procedure. This shows that interactive fitting may be difficult for some participants. Fine-tuning requires that the participant can reflect on the situation and can describe their experience. The clinician that performs the fine-tuning task has to listen carefully, to ask questions to help the listeners to describe their listening experience. Furthermore, the clinician has to translate the response of the participant in adjustment of one or more of the many parameters. This ask for an experienced audiologist, but this experience does not guarantee that the right adjustments are made.

Despite this limitations, the VR-fitting resulted in a clear improvement for three participants, a positive or neutral preference rating for eight out of ten participants and no significant changes for the rest. This shows that the new VR fine-tuning procedure definitely has potential, but more data is needed before we can conclude that finetuning of a CI in a VR-environment is worth the effort.

5. ACKNOWLEDGMENTS

The authors would like to thank the CI-users participating in this study. 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).

6. REFERENCES

[1] G. Keidser, H. Dillon, L. Carter *et al.*, “NAL-NL2 Empirical Adjustments,” *Trends in Amplification,* vol. 16, no. 4, pp. 211-223, 2012.

[2] S. Launer, J. A. Zakis, and B. C. Moore, “Hearing aid signal processing,” *Hearing aids*, pp. 93-130, 2016.

[3] R. Sanchez-Lopez, M. Fereczkowski, F. Bianchi *et al.*, "Technical evaluation of hearing-aid fitting parameters for different auditory profiles."

[4] A. Yellamsetty, E. J. Ozmeral, R. A. Budinsky *et al.*, “A Comparison of Environment Classification Among Premium Hearing Instruments,” *Trends Hear,* vol. 25, pp. 2331216520980968, 2021.

[5] E. Übelacker, and J. Tchorz, “Untersuchung des Nutzens einer Programmwahlautomatik für

Hörgeräteträger," *Hörakustik,* vol. 1, pp. 8-11, 2015.

[6] M. M. E. Hendrikse, G. Llorach, V. Hohmann *et al.*, "Movement and Gaze Behavior in Virtual Audiovisual Listening Environments Resembling Everyday Life," *Trends Hear,* vol. 23, pp. 2331216519872362, 2019.

[7] A. J. Bosman, and G. F. Smoorenburg, "Intelligibility of Dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment," *Audiology,* vol. 34, no. 5, pp. 260-84, 1995.

[8] M. M. E. Hendrikse, G. Dingemanse, G. Grimm *et al.*, "Development of Virtual Reality scenes for clinical use with hearing device fine-tuning," in Forum Acusticum, Turin, 2023.

[9] S. Gatehouse, and W. Noble, "The Speech, Spatial and Qualities of Hearing Scale (SSQ)," *Int J Audiol,* vol. 43, no. 2, pp. 85-99, 2004.

[10] N. J. Versfeld, L. Daalder, J. M. Festen *et al.*, "Method for the selection of sentence materials for efficient measurement of the speech reception threshold," *J Acoust Soc Am,* vol. 107, no. 3, pp. 1671-84, 2000.

[11] G. Grimm, and V. Hohmann, "First Order Ambisonics field recordings for use in virtual acoustic environments in the context of audiology," *Zenodo*, 10.5281/zenodo.3588303, 2019.

[12] G. Dingemanse, and A. Goedegebure, "Efficient Adaptive Speech Reception Threshold Measurements Using Stochastic Approximation Algorithms," *Trends Hear,* vol. 23, pp. 2331216520919199, 2019.

[13] G. Singh, and M. Kathleen Pichora-Fuller, "Older adults' performance on the speech, spatial, and qualities of hearing scale (SSQ): Test-retest reliability and a comparison of interview and self-administration methods," *Int J Audiol,* vol. 49, no. 10, pp. 733-40, 2010.

[14] A. A. Kressner, T. May, and T. Dau, "Effect of Noise Reduction Gain Errors on Simulated Cochlear Implant Speech Intelligibility," *Trends Hear,* vol. 23, 2019.