TH Köln

Master thesis

Validating the efficacy of Speech Reception Tests in Virtual Reality

Alexander Müller

Supervised by
Prof. Dr. Christoph Pörschmann & M.Sc. Melissa Andrea Ramírez Caro

June 30, 2022

1 Introduction

There are several reasons for problems with auditory processing. The most common examples include Auditory Processing Disorder (APD) and negative side-effects of hearing aids or cochlea implants. These problems can diagnosed through specialized listening test applications like Listening in Spatialized Noise - Sentences Test (LiSN-S) [1] or Hearing in Noise Test (HINT) [2]. Furthermore, for some conditions even the mitigation through auditory training programs has been investigated [3][4]. These programs, however, often come with the requirement to be performed in the presence of schooled personnel and complex audio setups. This increases costs and efforts as well as decreases overall accessibility of such test and training programs. Especially given the circumstances of the Corona pandemic, the advantages of remote medical applications have become very clear. A lot of time can be saved on both the side of medical professionals and patients. Also the distribution of software and simple hardware is far easier than setting up a wide variety of locations for the treatment of specific conditions. This problem only grows larger for rare conditions and disorders, where there might be no specialist in range of affected people. Using Virtual Reality (VR) to implement auditory test and training applications might allow to transition existing programs to remote execution.

2 Method

To find possible issues originating from the translation of an auditory test procedure into VR a small study will be conducted. Participants will be adults without any known hearing impairments or APDs. The study will consists of two experiments that focus on the introduction of 3D audio rendering in VR and the addition of a non-supervised feedback system that omits the requirement for the presence of an experimenter.

For this purpose, the German HINT will be used as a foundation. Using the same speech and noise stimuli as in the original study, the test will be conducted in three different conditions. In the first experiment the HINT procedure will be performed in an anechoic chamber with a 3-way loudspeaker setup. The second condition will be performed using an Oculus Rift Head Mounted Display (HMD) with Sennheiser HD600 headphones. The audio spatialization is realized through the 3D Tune-In Toolkit (3DTI) Unity wrapper and the TH Köln KU100 Head Related Transfer Function (HRTF) set.

WE WILL USE THE BUILT-IN RIFT HEADPHONES INSTEAD

2.1 Test setup I: Loudspeaker

To recreate the original HINT experiment a python application has been developed. The program offers a simple **GUI!** (**GUI!**) that makes the usage for the experimenter easier and allows a detailed overview over all test parameters. At first the application requires a *username* to be set that identifies the current participant and the path to the test resources (the German HINT speech/noise stimuli). Afterwards a new screen presents the option of directly starting the test or entering the practice procedure. During both the actual test and the practice the experimenter has complete overview over the current noise condition (e.g. "noise left"), sentence list (1-10), Signal to Noise Ratio (SNR) and round (e.g. 7 of 20).

Once the test procedure is done, the **GUI!** presents a result screen that allows the experimenter to briefly check if the collected data is plausible and then stores all data in a **JSON!** (**JSON!**) file.

Packages To control the external audio interface the *sounddevice* and *soundfile* python packages have been used. The **GUI!** was developed based on *Tkinter* with the *customtkinter* wrapper package to provide a more contemporary look.

Calibration The initial playback levels have been measured using a CALIBDEVICE 1 located at the exact position where the participant will sit during the procedure. For this purpose the noise signal has been played from each of the three possible positions. As stated in the original HINT papers, the noise has been setup to be presented at 52 dB SPL while the speech will initially be played at 65 dB Sound Pressure Level (SPL). Since the speech stimuli are only presented from the "front" position only this loudspeaker has been used for calibration.

ALSO CONSIDER RMS MEASUREMENTS???

Adaptive SNR In order to change the SNR during the test procedure all speech stimuli had been normalized to a range of **RMS!** (**RMS!**) values. The test application only has to select the correct file for the requested SNR instead of processing the audio files in real-time or remotely controlling the audio interface. This approach offers the advantage of being able to verify the correct **RMS!** spacing beforehand and should also avoid unfavorable settings of the audio interface that might affect the noise floor (REFERENCE FOR THIS???). The biggest disadvantage from this approach is that the range of possible SNR values is limited through the pre-processed range of speech stimuli **RMS!** levels. However, through pre-testing it can easily be ensured that the provided range is sufficient for the purpose of this test. If the SNR limits are reached during the test procedure, the application will prompt a warning to the experimenter and will keep the current level.

All audio processing has been done using Audacity with the ***RMS!*** *measurement* and ***RMS!*** *normalize* plugins (ADD REFERENCES!).

2.2 Development

Feedback system To allow the procedure to be performed without the presence of an experimenter, a new system had to be added that could determine whether the participant has understood the last sentence or not. In the original HINT program this is done by the subject repeating what they've understood out loud and an experimenter comparing their reply with the actual sentence. However, even though this method would allow to rate each sentence on a word by word basis, the procedure only separated between either the majority of the sentence was understood or not. This rather broad division leaves some tolerance on the implementation of a new feedback system. A possible option would have been to follow the subjective feedback implemented in LiSN-S [1]. In this case, participant have been confronted with a forced 3-way choice consisting of the sentence was understood a) good, b) medium or c) bad. But it has to be noted that LiSN-S was not designed to be performed at home with completely unsupervised participants. To mitigate the issue of users not being honest, a randomized 5-way multiple choice system has been implemented for each word of the sentence. The wrong options are filtered to roughly match the length of the correct word and to consider capitalization and the start of the sentence (which is of course also capitalized). To make it impossible for the participant to recognize the correct sentence up front by eliminating grammatically incorrect or non-sensible options, only the options for the current word are shown and there is no possibility to go back and change a previous word-submission. Of course this system still has some weaknesses. If the participant understood a large part of the sentence it will be easy to eliminate a lot of the wrong options. However, since the system is only intended to determine whether the majority of the sentence has been understood this should not alter results strongly.

These problems could be addressed by multiple means. One option would be to manually design sensible and grammatically correct alternatives for each target sentence. Following this solution, it should be taken special care that the newly introduced wrong answers are still matching the complexity of the original HINT stimuli. An alternative would be to transition the test procedure over to nonsensical target sentences with a fixed grammatical structure (as done in Listening in Spatialized Noise (LiSN) & Learn [4]). This would allow to always make suiting proposals based on simple grammatical parsing. Another option would be to phonetically match the different alternatives for each questions as done by Salorio-Corbetto et al. [5].

The new feedback system will transition the HINT procedure from being an *open test* (without any limitations towards response options) towards a *closed test*. The most important advantage of a closed test is the omission of the requirement of an experimenter that evaluates the answers of the participant.

2.3 Subjects

3 Experiment I

4 Experiment II

5 Results

6 Discussion

7 Conclusion

Contents

Abbreviations

3DTI	3D Tune-In Toolkit
APD	Auditory Processing Disorder
CAPD	Central Auditory Processing Disorder
FIR	Finite Impulse Response
HMD	Head Mounted Display
HINT	Hearing in Noise Test
HRTF	Head Related Transfer Function
ILD	Interaural Level Difference
ITD	Interaural Time Difference
LiSN	Listening in Spatialized Noise
LiSN-S	Listening in Spatialized Noise - Sentences Test
SDK	Software Development Kit
SNR	Signal to Noise Ratio
SPD	Spatial Processing Disorder
SPL	Sound Pressure Level
SRM	Spatial Release from Masking
SRT	Speech Reception Threshold
VR	Virtual Reality

References

[1] Sharon Cameron and Harvey Dillon. “Development of the Listening in Spatialized Noise-Sentences Test (LISN-S)”. In: *Ear and hearing* 28 (Apr. 2007), pp. 196–211. DOI: 10.1097/AUD.0b013e318031267f.

[2] Michael Nilsson, Sigfrid Soli, and Jean Sullivan. “Development of the Hearing In Noise Test (HINT) for the measurement of speech reception thresholds in quiet and in noise”. In: *The Journal of the Acoustical Society of America* 95 (Mar. 1994), pp. 1085–99. DOI: 10.1121/1.408469.

[3] Richard Tyler et al. “Initial Development of a Spatially Separated Speech-in-Noise and Localization Training Program”. In: *Journal of the American Academy of Audiology* 21 (2010), pp. 390–403. DOI: 10.3766/jaaa.21.6.4.

[4] Sharon Cameron and Harvey Dillon. “Development and Evaluation of the LiSN & Learn Auditory Training Software for Deficit-Specific Remediation of Binaural Processing Deficits in Children: Preliminary Findings”. In: *Journal of the American Academy of Audiology* 22 (2011), pp. 678–696. DOI: 10.3766/jaaa.22.10.6.

[5] Marina Salorio-Corbetto et al. “Evaluating Spatial Hearing using a Dual-Task Approach in a Virtual Acoustics Environment”. In: *Frontiers in Neuroscience* (), p. 46.