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

July 22, 2022

Contents

Abstract

Central Auditory Processing Disorder (CAPD) is described by the **ASHA!** (**ASHA!**) as a condition, which "may lead to or be associated with difficulties in higher order language, learning, and communication functions" without being caused by actual hearing loss or inabilities [**ASHA**]. Focusing on the aspect of spatial hearing Cameron and Dillon established the term Spatial Processing Disorder (SPD) and designed the Listening in Spatialized Noise (LiSN) & Learn auditory training software to improve binaural processing abilities of affected children [**LiSN-dev**].

Based on this foundation a new training program shall be created as an open-source project with Virtual Reality (VR) support inside the *Unity* game development engine. Apart from offering a free alternative to the follow up product of the original LiSN & Learn software (*Soundstorm)*[1], this project shall also make initial steps towards evaluating whether *Unity* is a viable platform for auditory training applications. In further iterations potential improvements to the concept may be reviewed such as the inclusion of head tracking.

[1] Reference: `https://www.soundstorm.app/`

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 VR to implement auditory test and training applications might allow to transition existing programs to remote execution.

1.1 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.

1.2 Motivation

2 Theoretical foundations

2.1 Spatial hearing

Cocktail-party-effect

Spatial Release from Masking (SRM)

2.2 3D audio rendering

Head Related Transfer Functions (HRTF)

Simulaiton of sound propagation effects

2.3 Virtual Reality - Design considerations

Multisensory information

Usage of generic HRTFs

2.4 Auditory test procedures

Hearing in Noise Test (HINT)

German HINT

LiSN-S

Probably more!

3 Approach

3.1 HINT analysis

3.2 VR HINT - concept

4 Study design

Test conditions

Test setup

4.1 Test setup I: Loudspeaker

To recreate the original HINT experiment a Python (version 3.9) 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. Alternatively, the same functionality could have also been achieved using MATLAB in conjunction with the *playrec* utility. But in contrast to MATLAB, Python has the advantage of being free to use without any license requirements.

Application Overview At startup the application presents a simple setup screen that requires a *username* for the participant and the path[2] to the test stimuli to be set. Once both requirements have been fulfilled the test can be initialized and a new screen is presented in the **GUI!**.

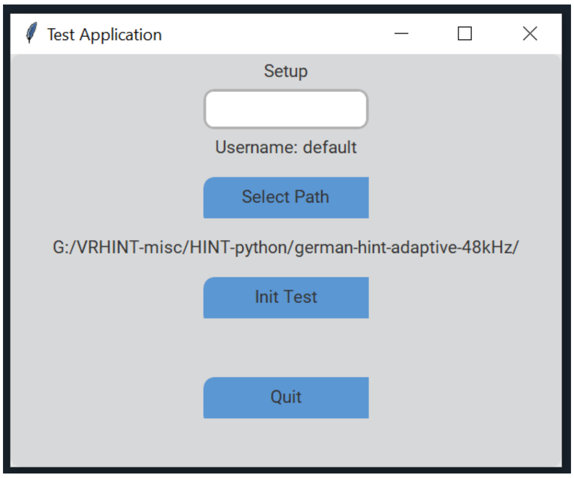

Figure 1 Python HINT GUI - Setup

The *Test Overview* screen allows to either access the *practice mode* or directly start the *test 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).

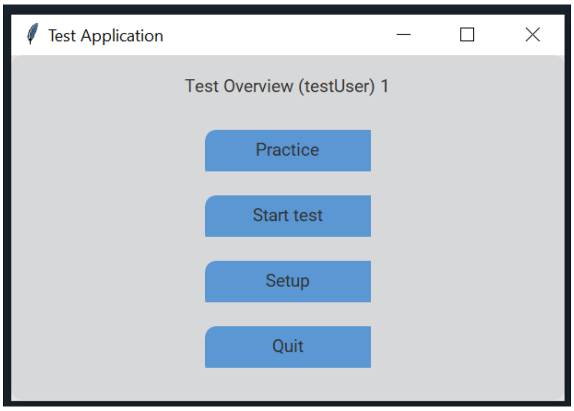

Figure 2 Python HINT GUI - Test Overview

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 JavaScript Object Notaion (JSON) file.

Packages To control the external audio interface the *sounddevice* [5] 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.

[2] The application will perform a simply sanity check on the path via trying to load a noise file from the given directory.

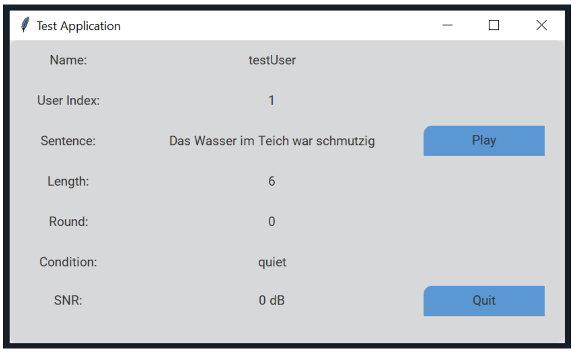

Figure 3 Python HINT GUI - Practice

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 Root Mean Square (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!).

4.2 Test setup II: VR HINT

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 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. [6].

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.

5 Study design

To evaluate the efficacy of VR peripherals to perform Spatial Release from Masking (SRM) tests as remote medical applications a small study will be conducted. In this study two main aspects will be investigated using two separate experiments:

1. The efficacy of audio source virtualization in VR applications using non-individualized HRTFs

2. The effects of of non-supervised feedback systems

Both experiments will be based on the HINT procedure and use the German HINT speech and noise stimuli provided by Rader et. al.

Setup	noiseLeft (db RMS)	noiseFront (db RMS)	noiseRight (db RMS)
Loudspeaker	-33.2	-32.7	-33.0
VR HD600 (1)	-31.2	-32.7	-31.2
VR HD600 (2)	-32.1	-33.4	-31.7
VR HD600 (3)	-31.6	-33.2	-31.4
VR HD600 (4)	-31.5	-32.7	-30.9
VR HD600 (5)	-31.6	-32.8	-30.9

Table 1 VR calibration - RMS measurements

5.1 Setup and Calibration

All experiments will be conducted in the anechoic chamber at TH cologne. For the louspeaker based conditions three Genelec XYZ speakers. These will be controlled using an *RME UFXII* interface.

The speakers have been placed at 0, 90 and 270 degrees azimuth around the listeners positions at a distance of 1 m. The height of the speakers has been set to 1.2m, which resembles the elevation of the ears of a listener of normal height while seated. During the experiment the seat height will be adjusted according to the height of the participant.

Loudspeaker calibration At the position of the listener an *Ono Sokki* Intergrating Sound Level Meter has been placed. Using the interface control software each speaker has been calibrated to 65 dBA while playing back the noise signal.

VR calibaration To get a reference value a *Neumann KU100* dummy head has been placed at the listeners location. The microphone signals are picked up using an *Edirol UA-25EX* audio interface and feed into a *Reaper* session running the *TBProAudio dpMeter4* plugin. Using this plugin the integrated RMS values have been noted for the noise playback from all three positions. Afterwards a pair of *Senneheiser HD600* headphones has been connected to the *RME UFXII* and set as default output device for the VR HINT application. In a demo scene the noise signal has been played from the three virtual positions and the level of the virtual audio mixer has been adjusted to achieve the same RMS levels as in the loudspeaker condition. Since a deviation between the front and left / right positions of 1.5 dB could be measured, it was decided to calibrate the VR condition for the front channel, which leads to slightly increase RMS levels for the side channels compared to the loudspeaker condition.

These measurements have been repeated several times with headphone readjustements in between to mitigate effects of headphone misplacement.

5.2 Subjects

Subjects for both experiments will be adults with - self proclaimed - regular hearing abilities.

5.3 Experiment I

The first experiment will focus on possible deviations in Speech Reception Threshold (SRT)s based on SRM due to the virtual spatialization of audio sources. For this purposes participants will perform the HINT procedure in both a reference condition and in a VR application. The reference condition will be performed using loudspeakers in an anechoic chamber located at the locations defined by the HINT experiment. For this purpose speakers will be located at 0, 90 and 270 degrees azimuth around the participant in a distance of 1 m, with the center of the speakers matching the elevation of the listeners ears. By this means the SRM of participants can be determined from their SRTs without any deviations caused by non-individualized HRTFs or similar means.

The second condition will be run in a VR-based implementation of the HINT experiment. The application has been developed using the Unity engine and is presentend using an *Oculus Rift* HMD and a pair of *Sennheiser HD600* headphones. All stimuli have been filtered in advance to equalize the frequency response curve of the headphones. Spatialization will be performed by the 3DTI framework using a set of HRTFs recorded using a *Neumann KU100* dummy head.

In both cases the participant will be advised to look straight forward, facing the *front* position in the experiment. Only small head movements will be permitted. If the participants turns their head too far, the experiment will be stopped. This deviation from other HINT experiments where the headphone-based approach does not consider head movements at all is planned, since the head-tracking capabilities of VR peripherals might prove advantageous to more accurately reproduce real-world listening scenarios.

Since only ten sentence-lists (excluding the two practice lists) are available to us and to avoid exhaustion of the participants to affect the results, the test procedure will be shortened. Instead of testing ten sentence-lists at once the experiment will separate them into two pairs of five lists for each condition. This still allows to include every noise condition in both conditions (*quiet*, *noiseLeft*, *noiseRight*, *noiseFront*), while keeping the duration of the experiment at a manageable level. As in the original experiments the participant will repeat the sentences out loud and the experimenter will enter the number of correct words into the test system, which will then automatically adjust the SNR according to the listeners performance.

5.4 Experiment II

While the first experiment focuses on the aspect of audio spatialization the second one will investigate the effects a non-supervised feedback system might introduce. In this case both test conditions will be performed in VR. As in experiment I the procedure will be split into two parts, both consisting of five test lists. One half will be performed using the established feedback method of repeating the understood sentences out loud and an experimenter entering the amount of correct words into the system. For the other half a 5-way forced-choice word-proposal system will be used. After each sentence participants have

Q	F	R	L
F	L	Q	R
L	R	F	Q
R	Q	L	F

Figure 4 Conditions - Latin Squares
Q: quiet F: noiseFront
R: noiseRight L: nosieLeft

1	2	10	3	9	4	8	5	7	6
2	3	1	4	10	5	9	6	8	7
3	4	2	5	1	6	10	7	9	8
4	5	3	6	2	7	1	8	10	9
5	6	4	7	3	8	2	9	1	10
6	7	5	8	4	9	3	10	2	1
7	8	6	9	5	10	4	1	3	2
8	9	7	10	6	1	5	2	4	3
9	10	8	1	7	2	6	3	5	4
10	1	9	2	8	3	7	4	6	5

Figure 5 Sentence lists - Latin Squares

to recreate the target sentence as good as possible on a word-by-word basis of randomly selected alternatives. To avoid participants to easily guess the correct sentence by context or grammar only suggestions for one position at a time will be shown with no option of going back and adjusting a previous selection after reading the suggestions of for the following words.

5.5 Counterbalancing

To avoid order effects in the test procedure, an automated latin-squares based counterbalancing systems has been implemented. The latin-square matrices have been pre-generated and imported into the test systems as .csv files. At the initialization of the procedure the applications check the amount of files in the result storage to determine the start vector for the counterbalancing matrices. This start index and the resulting list and conditions order is made visible to the experimenter to add another control instance. Figures 4 and 5 show the latin-squares matrices used.

1. Latin Squares
2. First / Second test
3. UserIndex tracking

6 Results

Experiment I

Experiment II

7 Discussion

8 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
JSON	JavaScript Object Notaion
LiSN	Listening in Spatialized Noise
LiSN-S	Listening in Spatialized Noise - Sentences Test
RMS	Root Mean Square
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] Spatial Audio. *Webpage*. 2022. URL: https://github.com/spatialaudio/python-sounddevice.

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