HINT-VR: A child-appropriate application for the self-assessment of spatial hearing abilities in VR

Alexander Müller[1], Melissa Ramírez[1,2], Johannes M. Arend[2], Tobias Rader[3], Christoph Pörschmann[1]
[1] *Institute of Computer and Communication Technology, TH Köln – University of Applied Sciences, D-50679 Cologne, Germany*
[2] *Audio Communication Group, TU Berlin, D-10587 Berlin, Germany*
[3] *Division of Audiology, Department of Otorhinolaryngology, University Hospital, LMU Munich, D-81377 Munich, Germany*
Corresponding author: melissa.ramirez@th-koeln.de

Introduction

The cocktail party is the perfect metaphor for everyday life's auditory scene complexity[1]. Noisy classrooms, crowded restaurants, and busy offices are just a few examples of acoustically challenging situations in which the normal auditory system demonstrates a remarkable ability to parse complex scenes.

Successful auditory scene analysis relies on the listeners' ability to form different perceptual images, i.e., auditory objects. Forming an auditory object requires grouping the multiple sound components that belong to a particular source and segregating those components from those of other sources. These primitive perceptual processes facilitate our ability to focus on signals of interest, such as the speech from a particular talker in the presence of competing speech, background noise, and room reflections, among others.

The auditory mechanisms enabling listeners to accomplish these tasks involve binaural processing to a large degree[2]. The role of spatial hearing in effective speech intelligibility in noise is well established in the literature. Many studies have shown that speech intelligibility improves when target speech and competing sounds are spatially separated[3–5] compared to when they are colocated[6–8]. This improvement is commonly known as spatial release from masking (SRM), and it is measured as the difference in speech reception thresholds (SRTs) between collocated and spatially separated noise conditions.

The extent of the spatial benefit, i.e., the effect of SRM, depends on several factors, such as the degree of spatial separation, the acoustic properties of the listening environment, or the masker's characteristics[9,10]. Moreover, hearing loss and auditory processing disorders (APDs) can significantly impact some listeners' ability to localize sound sources and achieve SRM[11].

Hence, recent research focuses on disentangling those effects on cocktail party processing, not only for normal hearing listeners but also for patients with hearing difficulties, such as people with hearing loss, APDs, cochlear implant, and hearing aid users. Speech audiometry is the most commonly used procedure in clinical settings and laboratories to assess these difficulties, evaluating speech intelligibility in the presence of competing speech or noise. Good examples are the Listening in Spatialized Noise–Sentence Test (LiSN-S) developed by Cameron & Dillon[12] and the well-established Hearing in Noise Test (HINT) introduced by Nilsson et al.[13].

Understanding speech in challenging acoustical environments is particularly important for school-aged children. With high reverberation and low signal-to-noise ratios (SNRs), classrooms are far from ideal listening conditions. Children with speech understanding difficulties face significant disadvantages in their development regarding social and educational aspects. Fortunately, the previously mentioned test procedures have proven to be efficient tools for promptly assessing and diagnosing difficulties related to spatial hearing. Moreover, research on spatial auditory rehabilitation, such as Cameron and Dillon's work around the LiSN & Learn project[14,15], has shown that the effects of APDs, such as spatial processing disorders, can be moderated significantly through auditory training programs.

Nevertheless, these services are still unavailable for an important part of the world's population. Many traditional audiometric testing and training setups require elaborate calibration, costly equipment, or acoustically treated rooms. Likewise, with the shortage of trained healthcare professionals, infrastructure, and resource limitations, even essential hearing health services are still inaccessible for specific populations. Thus, developing valid and efficient test and training procedures that can be performed with inexpensive consumer-grade hardware with simple setups is of great research interest nowadays.

Modern-day virtual reality (VR) peripherals may be very suitable for this use case since they support spatial audio, are portable, and are relatively cheap. Besides, the newest versions support standalone operation without additional computing hardware, facilitating the reproducibility and scalability of the setups. Combining these advantages with some adjustments to the procedures so they can be performed without supervision, at-home conduction of both hearing tests and auditory training programs may become feasible.

We present HINT-VR, a child-appropriate application for assessing spatial hearing abilities in VR that does not require an experimenter's supervision. HINT-VR is an open-source tool that can support research on VR-based spatial auditory testing and rehabilitation. For its evaluation, we analyze the efficacy of binaural rendering with non-individual HRTFs for assessing spatial hearing abilities through the well-known HINT procedure. Moreover, we investigate and discuss other relevant items like the effect of using an autonomous feedback system, i.e., a non-supervised method that does not require an experimenter, and the effect of using an HMD for presenting visual stimuli in the test.

Methods

The HINT-VR was developed using the Unity game engine, and the virtual acoustic scene simulates free field conditions (Figure 1). We used the 3D Tune-In Toolkit (3DTI) Unity wrapper[16,17] with far-field measured non-individual HRTFs (Neumann KU100 dummy head)[18] for the audio spatialization.

The test program contains a login screen where the participant can be registered. Moreover, the main menu offers access to a dedicated calibration scene and the HINT test procedure. All the relevant test data is collected in an extendable JSON format and stored locally. To offer greater flexibility for different study designs, most parameters, such as the number of noise conditions or practice rounds, can easily be adjusted using the Unity Inspector interface without changing the source code.

Figure 1: HINT-VR - open field virtual environment.

Furthermore, all stimuli (text and audio) can be replaced easily by using the regular file system, e.g., to perform HINT with different stimuli, such as in other languages.

We used the stimuli from the German hearing in noise test featuring a male target speaker[19]. The procedure features four noise conditions. The target speech is always presented from the front (0° azimuth), and the noise (distractor) is located either in the same position, i.e., Noise Front (NF), shifted 90° to the left or the right sides, i.e., Noise Left and Noise Right (NL and NR), or is disabled, i.e., Quiet (Q). The noise condition presentation order is randomly selected and changes every time a list is completed (1 list = 20 sentences).

At the start of each condition, speech and noise are presented at 65 dBA (SPL). The SNR is adjusted automatically according to the participant's performance based on a 50% intelligibility criterion. If less than 50% of the sentence is understood correctly, the SNR is increased by 2 dB. Otherwise, it is decreased by the same amount. The SRT is calculated as the average SNR over all sentences on each noise condition (NF, NR, NL, and Q). The SRM is defined as the difference in SRTs between the collocated and the spatially separated noise conditions.

HINT-VR supports two different user feedback methods. First, the traditional way, i.e., the participant repeats what they have understood aloud to the experimenter, who records the user's answers in the system. Second, a new system based on random word proposals, i.e., the autonomous feedback system (AFS). With the AFS, the user's task is to form the sentence word by word. The user has five options for each word, one correct and four random alternatives matched in length and capitalization.

The words conforming each sentence must be selected one by one in order. The player must decide on the current word before the alternatives for the next word are shown. It is not possible to go back or change previous answers.

Figure 2. HINT-VR: Autonomous feedback system. At first, five options from the first-word container are shown to the user. Once a selection has been made, they are replaced with the entries from the second-word container, and so on.

Additionally, HINT-VR offers two different virtual test environments to investigate the influence of the HMD-based visual scene representation on the subjects' performance. The first is a simple open field with avatars representing the sound sources (talker and distractor - Figure 1). The second alternative is a simple, uniformly colored squared room. Hereafter, the dark room.

Evaluation

The HINT-VR was evaluated through a listening experiment in the anechoic room at the TH Köln. The study participants performed the HINT procedure in a loudspeaker-based setup used as a reference and using the VR application developed for this study. The test setup order (loudspeaker-based or in VR) and noise conditions (NF, NR, NL, or Q) was counterbalanced between subjects using a Latin-squares-based approach. Each participant performed five of the ten sentence lists in the loudspeaker-based setup and the remaining five in the VR setup.

The reference (loudspeaker-based) condition was implemented and controlled as an independent Python application. It provides a simple GUI that offers the experimenter an easy and reliable way of controlling and monitoring the test procedure. We used three loudspeakers (Genelec 8020D) located at 0°, 90°, and 270° azimuth around the listener at a distance of 1m. The setup was calibrated to 65 dBA (SPL) using an Ono Sokki LA-1240 Sound Level Meter placed at the listener's position.

For the HINT-VR, a pair of Sennheiser HD600 was used for audio playback. Headphone equalization was performed for all audio files in the test using a generic headphone compensation filter. The filter was applied to the stimuli to minimize the influence of the employed headphones. The filter is based on twelve measurements (putting the headphones on and off the Neumann KU100 dummy head) to account for re-positioning variability and was designed by regularized inversion of the complex mean of the headphone transfer functions[20] using the implementation by Erbes et al.[21]. Moreover, for the headphone-based level calibration, we measured the RMS levels of the calibrated loudspeaker setup using the dummy head (Neumann KU100) and adjusted the headphone playback level to match that level. An RME Fireface UFXII interface at 48 kHz was used to drive both setups.

In the study, different aspects of the new VR implementation were investigated separately through three groups of participants. First, two groups were used to evaluate the effect of the feedback system. Group 1 provided feedback by repeating what they understood aloud in both test setups (loudspeaker-based and VR). Group 2 repeated out loud in the loudspeaker-based setup but used the autonomous feedback word-selection system in HINT-VR. Second, to investigate if using HMD-based matching 3D graphics (visual stimuli with avatars to represent the sound sources) affected the subjects' performance, Group 3 performed the VR test in the dark room.

The listening experiment is completed, and the results align with values reported in the literature, such as the German HINT[19,22] and other HINT studies[23] for both conditions, loudspeaker-based reference setup, and HINT-VR. The statistical analysis showed no significant differences in SRM between the loudspeaker-based and the VR test setups, using both the traditional and the proposed autonomous feedback system. These results suggest that the German HINT can be performed reliably in VR and at home without an experimenter's supervision. Regarding the visual feedback, in line with previous research, the results of this study confirm that matching visuals aid sound source localization[24,25]. Consequently, the group performing the test in the virtual environment with avatars representing the sound sources showed improved SRMs compared with the group performing the test in the empty dark room (with auditive but not visual stimuli). A publication with a more extensive elaboration on the study results will follow.

License and Access

Acknowledgments

This work was sponsored by the German Federal Ministry of Education and Research BMBF (13FH666IA6-VIWER-S).

Literature

1. Middlebrooks, J. C., Simon, J. Z., Popper, A. N. & Fay, R. R. *Springer Handbook of Auditory Research: The Auditory System at the Cocktail Party*. (Springer, 2017). doi:10.1007/978-3-319-51662-2.
2. Werner, L. A., Fay, R. R. & Popper, A. N. *Human Auditory Development. Springer Handbook of Auditory Research* (Springer, 2012). doi:10.1007/978-1-4614-1421-6.
3. Dirks, D. D. & Wilson, R. H. The effect of spatially separated sound sources on speech intelligibility. *Journal of speech and hearing research* **12**, 5–38 (1969).
4. Bronkhorst, A. W. The Cocktail Party Phenomenon : A Review of Research on Speech Intelligibility in Multiple-Talker Conditions. *Acta Acustica united with Acustica* **86**, 117–128 (2000).
5. Hawley, M. L., Litovsky, R. Y. & Culling, J. F. The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer. *The Journal of the Acoustical Society of America* **115**, 833–843 (2004).
6. Hess, C. L., Misurelli, S. M. & Litovsky, R. Y. Spatial Release From Masking in 2-Year-Olds With Normal Hearing and With Bilateral Cochlear Implants. **22**, 1–13 (2018).
7. Peng, Z. E. & Litovsky, R. Y. Novel Approaches to Measure Spatial Release from Masking in Children with Bilateral Cochlear Implants. *Ear and Hearing* **43**, 101–114 (2022).
8. Garadat, S. N., Litovsky, R. Y., Yu, G. & Zeng, F.-G. Role of binaural hearing in speech intelligibility and spatial release from masking using vocoded speech. *The Journal of the Acoustical Society of America* **126**, (2009).
9. Kidd, G., Mason, C. R., Brughera, A. & Hartmann, W. M. The Role of Reverberation in Release from Masking Due to Spatial Separation of Sources for Speech Identification. *Acta Acustica united with Acustica* **91**, 526–536 (2005).
10. Best, V., Marrone, N., Mason, C. R. & Kidd, G. The influence of non-spatial factors on measures of spatial release from masking. *The Journal of the Acoustical Society of America* **131**, 3103–3110 (2012).
11. Arbogast, T. L., Mason, C. R. & Kidd, G. The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America* **117**, 2169–2180 (2005).
12. Cameron, S. & Dillon, H. Development of the Listening in Spatialized Noise-Sentences Test (LISN-S). *Ear and Hearing* **28**, 196–211 (2007).
13. Nilsson, M., Soli, S. D. & Sullivan, J. A. Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise. *The Journal of the Acoustical Society of America* **95**, 1085–1099 (1994).
14. Cameron, S. & Dillon, H. Development and Evaluation of the LiSN & Learn Auditory Training Software for Deficit-Specific Remediation of Binaural Processing Deficits in Children: Preliminary Findings. *Journal of the American Academy of Audiology* **22**, 678–696 (2011).
15. Cameron, S., & Dillon, H. Remediation of spatial processing issues in central auditory processing disorder. in *Handbook of central auditory processing disorder: comprehensive intervention Vol. 2* (ed. Chermak Gail, F. M.) 201–224 (Plural Publishing, 2014).
16. Cuevas-Rodríguez, M. *et al.* 3D Tune-In Toolkit: An open-source library for real-time binaural spatialisation. *PLoS ONE* **14**, (2019).
17. Reyes-Lecuona, A. & Picinali, L. Unity Wrapper for 3DTI. https://github.com/3DTune-In/3dti_AudioToolkit_UnityWrapper (2022).
18. Bernschütz, B. A Spherical Far Field HRIR HRTF Compilation of the Neumann KU 100. in *Proceedings of the 39th DAGA* 592–595 (2013).
19. Joiko, J., Bohnert, A., Strieth, S., Soli, S. D. & Rader, T. The German hearing in noise test. *International Journal of Audiology* **60**, 927–933 (2021).
20. Lindau, A. & Brinkmann, F. Perceptual evaluation of head-phone compensation in binaural synthesis based on non-individual recordings. *The Journal of the Audio Engineering Society* **60**, (2012).
21. Erbes, V., Wierstorf, H., Geier, M. & Spors, S. Free Database of Low-Frequency Corrected Head-Related Transfer Functions and Headphone Compensation Filters. in *Proceedings of 142nd Audio Engineering Society Convention, e-Brief 325* 1–5 (2017).
22. Mönnich, A.-L., Strieth, S., Bohnert, A., Ernst, B. P. & Rader, T. The German hearing in noise test with a female talker: development and comparison with German male speech test. *European Archives of Oto-Rhino-Laryngology* (2023) doi:10.1007/s00405-023-07820-5.
23. Soli, S. D. & Wong, L. L. N. Assessment of speech intelligibility in noise with the hearing in noise test. *International Journal of Audiology* **47**, 356–361 (2008).
24. Maravita, A., Spence, C. & Driver, J. Multisensory integration and the body schema: Close to hand and within reach. *Current Biology* **13**, 531–539 (2003).
25. Valzolgher, C. *et al.* The impact of a visual spatial frame on real sound-source localization in virtual reality. *Current Research in Behavioral Sciences* **1**, 100003 (2020).