



OPEN ACCESS

EDITED BY

Mehdi Abouzari,
University of California, Irvine, United States

REVIEWED BY

Sasan Dabiri,
Northern Ontario School of Medicine
University, Canada
Euyhyun Park,
Korea University, Republic of Korea

*CORRESPONDENCE

Melissa Ramírez,
✉ melissa.ramirez@th-koeln.de

RECEIVED 25 July 2024

ACCEPTED 24 October 2024

PUBLISHED 07 November 2024

CITATION

Ramírez M, Müller A, Arend JM, Himmelein H,
Rader T and Pörschmann C (2024) Speech-in-
noise testing in virtual reality.
Front. Virtual Real. 5:1470382.
doi: 10.3389/frvir.2024.1470382

COPYRIGHT

© 2024 Ramírez, Müller, Arend, Himmelein,
Rader and Pörschmann. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).
The use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Speech-in-noise testing in virtual reality

Melissa Ramírez^{1,2*}, Alexander Müller¹, Johannes M. Arend^{2,3},
Hendrik Himmelein^{1,2}, Tobias Rader⁴ and
Christoph Pörschmann¹

¹Institute of Computer and Communication Technology, TH Köln - University of Applied Sciences, Cologne, Germany, ²Audio Communication Group, Technische Universität Berlin, Berlin, Germany, ³Acoustics Lab, Department of Information and Communications Engineering, Aalto University, Espoo, Finland, ⁴Department of Otorhinolaryngology, Division of Audiology, University Hospital, Ludwig-Maximilians-University Munich (LMU), Munich, Germany

The potential of virtual reality (VR) in supporting hearing research and audiological care has long been recognized. While allowing the creation of experimental settings that closely resemble real-life scenarios and potentially leading to more ecologically valid results, VR could also support the current need for automated or remote assessment of auditory processing abilities in clinical settings. Understanding speech in competing noise is the most common complaint of patients with hearing difficulties, and the need to develop tools that can simplify speech-in-noise testing by reducing the time and resources required while improving the ecological validity of current assessment procedures is an area of great research interest. However, the use of VR for speech-in-noise testing has not yet been widely adopted because it is still unclear whether subjects respond to virtual stimuli the same way as they would in real-life settings. Using headphone-based binaural presentation, delivering visuals through head-mounted displays (HMDs), and using unsupervised (self-testing or remote) procedures are some aspects of virtualization that could potentially affect speech-in-noise measures, and the extent of this potential impact remains unclear. Before virtualization can be considered feasible, its effects on behavioral psychoacoustic measures must be understood. Thus, the ability to reproduce results from typical laboratory and clinical settings in VR environments is a major topic of current research. In this study, we sought to answer whether it is possible to reproduce results from a standard speech-in-noise test using state-of-the-art technology and commercially available VR peripherals. To this end, we compared the results of a well-established speech-in-noise test conducted in a conventional loudspeaker-based laboratory setting with those obtained in three different virtual environments. In each environment, we introduced one aspect of virtualization, i.e., virtual audio presentation in the first environment, HMD-based visuals with a visual anchor representing the target speaker in the second, and an alternative feedback- and scoring method allowing unsupervised testing in the last. Our results indicate that the speech-in-noise measures from the loudspeaker-based measurement and those from the virtual scenes were all statistically identical, suggesting that conducting speech-in-noise testing in state-of-the-art VR environments may be feasible even without experimenter supervision.

KEYWORDS

binaural hearing, speech reception thresholds, spatial release from masking, virtual reality, tele-audiology

1 Introduction

The cocktail party is a perfect metaphor for the auditory complexity of everyday life (Middlebrooks et al., 2017). Noisy classrooms, crowded restaurants, and busy offices are just a few examples of typical complex acoustic environments in which our auditory system demonstrates its ability to focus on signals of interest, such as the speech of a particular speaker in the presence of competing speech or background noise (Werner et al., 2012).

Speech intelligibility in noisy environments relies heavily on binaural processing, and the role of spatial hearing in this ability is well-established in the literature (Hawley et al., 2004). Speech intelligibility is enhanced when target speech and competing noise are spatially separated (Bronkhorst, 2000; Dirks and Wilson, 1969) compared to when they are colocated (Hess et al., 2018; Peng and Litovsky, 2022). This enhancement, known as spatial release from masking (SRM), can be measured as the difference in speech reception thresholds (SRTs) between the spatially separated and colocated noise conditions (Garadat et al., 2009; Hawley et al., 2004; Litovsky, 2005).

Previous research has consistently shown that audiograms alone are insufficient to predict speech understanding difficulties in noisy environments or to reveal a person's functional hearing ability in real-life listening scenarios (Ruggles et al., 2011; Strelcyk and Dau, 2009). There is a high incidence of hearing difficulties, especially in noisy environments, among patients who do not exhibit measurable hearing threshold loss. This includes individuals with subclinical hearing loss or supra-threshold listening disorders and those with auditory processing disorders (Bellis and Bellis, 2015; Beck, 2023). As a result, it has been recommended for over 50 years to include speech-in-noise testing in routine hearing evaluations for all patients (Carhart and Tillman, 1970), even those with pure-tone normal-hearing (NH) thresholds (Roup et al., 2021). Speech-in-noise testing provides a more comprehensive understanding of a patient's hearing abilities and facilitates the implementation of more effective treatment strategies. However, despite the availability of a wide range of accurate speech-in-noise tests (Bench et al., 1979; Cameron and Dillon, 2007; Killion et al., 2004; Nilsson et al., 1994; Niquette et al., 2003; Soli and Wong, 2008; Taylor, 2003), recent data indicate that speech-in-noise abilities are still not regularly tested in routine hearing evaluations (Beck, 2023; Mueller et al., 2023).

It is a matter of concern that less than 20% of hearing healthcare professionals include speech-in-noise testing in their routine hearing assessments. In most cases, when speech intelligibility measures are included, they are limited to SRTs in quiet (Beck, 2023). This is primarily attributed to time and resource limitations in the typical clinical practice, including a shortage of healthcare professionals, the unavailability of complex setups such as loudspeaker arrangements in large and acoustically treated rooms (Beck, 2023; Clark et al., 2017; Mueller, 2016; Mueller et al., 2023), and the perceived lack of *external validity* of some assessment procedures, i.e., the extent to which results are likely to generalize to conditions beyond those in which the data were collected (Beechey, 2002), also commonly known as *ecological validity* (Keidser et al., 2020).

In contrast, speech-in-noise abilities have been extensively studied in laboratory-based research contexts, showing that

several factors can affect speech intelligibility in NH and hearing-impaired listeners, including the spatial configuration of the sound sources, the acoustic properties of the listening environment, the type of masker (energetic or informational), and the spectral differences between the target and maskers, among others (Arbogast et al., 2005; Best et al., 2012; Bronkhorst, 2000; Kidd et al., 2005; Rader et al., 2013). However, laboratory-based studies often lack ecological validity because they are conducted in highly controlled environments that do not reflect real-life listening scenarios (Keidser, 2016). There is a need to improve ecological validity within behavioral hearing science. Current efforts focus on *realism*, i.e., the extent to which laboratory test conditions resemble those found in the everyday settings of interest (Beechey, 2002), and recent literature highlights the need to integrate perceptual variables that influence listening behavior in real-life scenarios into research paradigms and methods, such as the inclusion of visual information and the ability to make exploratory head movements (Keidser et al., 2020; Valzoghger, 2024).

Consequently, developing tools that can simplify speech-in-noise testing by reducing time and resource requirements (Jakien et al., 2017) while improving the ecological validity of current assessment procedures has become an area of great current research interest (Keidser, 2016; Keidser et al., 2020). With this study, we sought to answer whether speech-in-noise testing in virtual reality (VR) could be a viable solution to these challenges. Modern VR peripherals are affordable and portable devices that could improve clinical efficiency by allowing testing in any room, whether in a clinic or at home. In addition, the latest versions support standalone operation, further facilitating reproducibility and scalability of setups. VR technology has tremendous potential to support tele-audiology, improve the quality of care, and enhance the experience of patients and their families. However, the impact of virtualization on behavioral psychoacoustic measures must be investigated before it can be considered viable. To this end, we conducted a psychoacoustic study evaluating the ability to reproduce speech-in-noise outcomes from a conventional loudspeaker-based test setup using state-of-the-art technology and commercially available VR peripherals.

Virtualizing speech-in-noise testing involves significant modifications to the setups and procedures from the typical clinical practice. Therefore, to determine the potential impact of each of those changes on the test results, we started with a loudspeaker-based measurement setup, which we used as a baseline, and we gradually introduced different aspects of virtualization through three different virtual scenarios:

- a) In the first virtual scenario (VR1), we replaced the loudspeaker-based auditory presentation with headphone-based dynamic (i.e., motion-compensated) binaural rendering.

Previous research in multimodal perception has highlighted a strong link between binaural cues and self-motion, emphasizing that exploratory head movements play an essential role in spatial auditory perception (Grange and Culling, 2016; Gaveau et al., 2022). When head movements are not restricted, listeners tend to turn their heads to increase the target signal level in one ear. This instinctive response often results in an improved signal-to-noise ratio (SNR), leading to improved SRTs (Brimijoin et al., 2012; Kock,

1950). Despite this knowledge, current clinical and laboratory practice still uses headphone-based *static* binaural rendering for speech-in-noise testing. Although this approach was introduced to avoid the need for loudspeaker-based setups in acoustically treated rooms and is widely used, it has some limitations. Headphone-based static binaural rendering results in internalized sound images, i.e., they are perceived as being located inside the listener's head (Best et al., 2020; Brimijoin et al., 2013). Additionally, static binaural rendering causes the virtual location of the signals to move along with the listener's head movements, which does not mimic real-world listening conditions.

Combining headphone-based binaural rendering with head-tracking, i.e., *dynamic* binaural rendering, overcomes these limitations. It allows the auditory environment to be updated in real-time according to the subject's head movements, increasing both realism and externalization (Begault et al., 2001; Best et al., 2020). Thus, aiming to improve the naturalness of testing conditions, we did not limit head movements nor used static binaural rendering in this study. This enriches the complexity of the stimuli by making dynamic binaural cues available and allows people to behave more similarly as they would in the real world when performing the listening task (see Valzolgher (2024) for a comprehensive review).

- b) In the second scenario (VR2), we added (virtual) visual feedback with a visual anchor representing the target speaker in the virtual scene.

Vision is another modality influencing auditory spatial perception by aiding externalization and distance estimation (Best et al., 2020). The presentation of visual information congruent with the auditory environment supports the existence of an externalized sound source (Brimijoin, Boyd, and Akeroyd, 2013). Moreover, several psychophysical and neurophysiological studies have shown that auditory and visuospatial attention are linked such that when attention in one modality is focused on one location, attention in the other modality is also drawn there (Busse et al., 2005; Tiippana et al., 2011). This suggests that using a visual anchor at the location of the target speaker may aid listeners in directing their auditory attention to that location as well.

Although the graphics used in our implementation are still far from realistic, e.g., they do not include the speaker's facial expressions or other aspects that are highly relevant for speech understanding, such as lip movements (Helfer and Freyman, 2005; Yuan et al., 2021; Williams et al., 2023), we argue that the availability of the more reliable (albeit basic) visual information aids the brain in optimally calibrating the associations between auditory cues and spatial locations (Isaiah et al., 2014; Valzolgher et al., 2020). More importantly, we argue that a significant increase in the listeners' SRTs, when tested in this environment compared to VR1, would reveal an (undesired) effect of presenting visual feedback through a head-movement-display (HMD) for this application. For example, the choice of visuals could increase cognitive load, potentially resulting in poorer performance. See Methods and Discussion for more details.

- c) Last, in the third virtual scenario (VR3), we included an alternative feedback- and scoring method for unsupervised testing.

Speech intelligibility can be measured using an open- or closed-response set. With an open set (typical of clinical settings), the listener repeats aloud what they hear, and the tester rates these verbal responses as correct or incorrect. In a closed-set (forced-choice task), the listener chooses from a limited number of acceptable response alternatives. The response alternatives are usually presented visually (Buss et al., 2016). It is worth noting that closed sets generally result in reduced (better) SRTs. This is especially true when the response set contains few phonetically dissimilar alternatives (Buss et al., 2016; Miller et al., 1951). However, closed sets could facilitate testing without experimenter supervision, allowing self- or remote testing (Jakien et al., 2017) and supporting tele-audiology. In this setting, we tested whether allowing the participants to self-record their responses on the gamified HMD-based interface would affect the test scores relative to our baseline measure.

We invited three groups of NH subjects to participate in the study. Using a randomized mixed design, each group was tested by taking the German Hearing in Noise Test (HINT) (Joiko et al., 2021) in a loudspeaker-based setting (baseline condition) and in one of the three virtual scenarios (VR1, VR2, or VR3) presented via a head-mounted display (HMD) and headphones. The within-subjects conditions, i.e., baseline versus virtual, allowed us to evaluate the effect of using headphone-based dynamic binaural rendering with non-individual head-related transfer functions (HRTFs) compared to a conventional loudspeaker-based setup. The between-subjects conditions, i.e., the different virtual conditions, allowed us to measure the effect of introducing a visual anchor representing the target speaker in the virtual scene and introducing an alternative feedback- and scoring method for unsupervised testing.

2 Methods

2.1 Participants

Forty-five subjects aged 19 to 65 ($M = 31.5$ years, $Mdn = 26$ years, $\sigma = 13.43$ years) voluntarily participated in the study (without compensation). They were engineering students or colleagues from the TH Köln. They all reported no hearing complaints and no history of hearing loss or auditory processing disorders in a questionnaire completed prior to participation in the study. This was used as inclusion criteria. All subjects were native German speakers, and about 42% had previously participated in other listening experiments.

The study was designed following the principles of the Declaration of Helsinki (World Medical Association, 2013) and the guidelines of the local institutional review board of the Institute of Computer and Communication Technology at the TH Köln. All participants gave written informed consent for their voluntary participation in the study and the later publication of the results. All personal data and experimental results were collected, processed, and archived according to country-specific data protection regulations.

2.2 Setup and stimuli

The experiment took place in the sound-insulated anechoic chamber of the acoustics laboratory of the TH Köln, which has

dimensions of $4.5 \times 11.7 \times 2.3$ m (W×D×H), a lower cut-off frequency of about 200 Hz, and a background noise level of about 20 dB(A) SPL. We used three Genelec 8020D loudspeakers for the baseline measurement, i.e., the loudspeaker-based environment, and an Oculus Rift with a pair of Sennheiser HD600 headphones for the virtual environments. An RME Fireface UFXII interface connected to a PC controlled the loudspeakers and headphones.

We used the German hearing in noise test with a male target speaker, which includes twelve phonemically- and difficulty-matched 20-sentence lists and a spectrally matched masker noise. All sentences are four-to six-word-long simple sentences incorporating common nouns and verbs used at the elementary school level (Joiko et al., 2021).

2.3 Materials

2.3.1 Loudspeaker-based environment (baseline)

Three loudspeakers were placed at ear level and 1 m from the seated listener. They were placed at 0°, 90°, and 270° azimuth, i.e., front, left, and right directions, respectively, and were visible to the subjects. Participants were asked to repeat what they heard using the standard HINT protocol, using an open-set feedback method (Nilsson, Soli, and Sullivan, 1994; Soli and Wong, 2008; Joiko et al., 2021). The experimenter recorded their responses using a Python application developed specifically for this experiment. More details can be found in the Experimental procedure section.

2.3.2 VR environments

All virtual environments used headphone-based dynamic binaural rendering with head tracking. Sound sources were located at 0° and 90° (or) 270° azimuth, depending on the test condition, just like in the loudspeaker-based environment (see Experimental procedure section for more details).

We used the Unity wrapper for the 3D Tune-In Toolkit for dynamic binaural rendering because it is open-source, well-documented, and explicitly designed for hearing research (Cuevas-Rodríguez et al., 2019; Reyes-Lecuona and Picinali, 2022). For spatialization, we used a full-spherical HRTF set from a Neumann KU100 dummy head in SOFA format (Majdak et al., 2022), which was measured on a Lebedev grid with 2702 spatial sampling points in the far field (Bernschütz, 2013).

In addition, we applied a generic headphone compensation filter to the stimuli (target sentences and masker) to minimize the influence of the headphones used. The filter is based on twelve measurements (putting the headphones on and off the dummy head) to account for re-positioning variability and was designed by regularized inversion of the complex mean of the headphone transfer functions (Lindau and Brinkmann, 2012) using the implementation of Erbes et al. (2017).

a) VR1: Audio-only (AO)

There was no visual representation of the target speaker's location in this environment. Instead, a black screen was projected through the HMD (Figure 1). After the stimulus presentation, a visual icon and text appeared on the screen,

indicating to the subjects that it was time to repeat what they had heard. The experimenter recorded their responses in the system.

b) VR2: Audiovisual (AV)

We chose an open field as our visual environment because it closely mirrors the acoustic properties of the simulated auditory environment, which was anechoic (Figure 1). Rather than visually modeling the anechoic test room and displaying a “virtual experimenter,” we chose a simple virtual scene designed to make participants feel as if they were somewhere else so that they could “forget” that someone was sitting there recording what they said. This approach, as other researchers have reported, increases participant comfort (Murphy, 2017). Moreover, our goal is to assess the feasibility of conducting virtual and remote testing and, in the future, inside more complex, i.e., realistic scenarios, visually and auditorily. Therefore, we are interested in using visual scenes that look different from a research lab.

We used a robot avatar in front of the subject to represent the target speaker (located at 0° azimuth and 1 m distance). After the stimulus presentation, the subjects repeated aloud what they had heard, and the experimenter recorded their responses in the system. The Supplementary Material includes a short video illustrating some trials in this environment.

c) VR3: Audiovisual (AV) with word selection:

In this environment, otherwise identical to VR2, participants were asked to put together the target sentences word by word. The standard HINT procedure usually features an open set. However, being able to use a closed set may enable unsupervised testing. Thus, we presented five options for each word in a five-alternative forced-choice (5AFC) procedure.

Of these five options, only one was a correct word, while the other four were randomly selected alternatives from the sentence lists that were matched for length and capitalization, as capitalization is important in German for identifying words that are nouns. Participants had to choose each word to form the sentence, one at a time. Each decision about the current word was made before the alternatives for the next word were presented. Going back or changing previous answers was not possible (Figure 2). The Supplementary Material includes a short video illustrating some trials in this environment.

2.4 Experimental procedure

The HINT procedure includes four conditions where the target speech is always in front of the listener (0° azimuth). The noise (masker) is either at the same location, i.e., Noise Front (NF), shifted to the left (90° azimuth) or right (270° azimuth), i.e., Noise Left and Noise Right (NL and NR), or suppressed, i.e., Quiet (Q) (Joiko et al., 2021; Mönnich et al., 2023; Nilsson et al., 1994; Soli and Wong, 2008). The order in which the noise conditions were presented was randomized, changing each time a list of twenty sentences was completed.

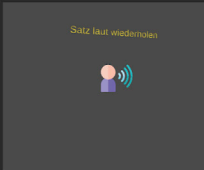


Test environment				
	Loudspeaker-based	Virtual		
Audio reproduction	Loudspeaker-based	Headphone-based dynamic binaural rendering		
Visual feedback	Real world	Virtual through head-mounted display		
				
Test condition	Baseline	VR1: Audio-only (AO)	VR2: Audiovisual (AV)	VR3: Audiovisual (AV) with word selection feedback
Response method	Subject repeats aloud what they understood Experimenter logs answers into the system			Word selection feedback system

FIGURE 1 Test environments.

Target sentence "Die Kinder essen Kuchen"			
Word selection pop-up screens			
1st word	2nd word	3rd word	4th word
Der	Mädchen	malen	Karten
Die	Männer	gehen	Kuchen
Das	Jungen	essen	Schnee
Sie	Kinder	haben	Puppen
Wir	Eltern	mögen	Futter

FIGURE 2 Example of different alternatives for a target sentence using the word selection system. The subjects' task was to form the target sentence one word at a time using a five-alternative forced-choice (5AFC) procedure. Of the five alternatives presented, only one was a correct word, while the other four were random alternatives matched for length and capitalization. The options for the first word were presented first, and participants had to select the correct word before being presented with the options for the next word. Once a word was selected, participants could not go back or change their answers.

Following the HINT procedure described by Soli and Wong (2008), speech and noise were initially presented at 65 dB(A) SPL measured in the free field at the listener's position (see Presentation level calibration section for more details). The SNR was automatically adjusted based on the subject's performance using

a 50% intelligibility criterion. It decreased if the subject repeated at least half of the sentence correctly, e.g., at least two words in a four-word sentence. Otherwise, it increased. The level of the masker remained constant, and the SNR was adjusted by increasing or decreasing the level of the target sentences. The procedure uses step

sizes of 4 dB for the first four sentences and 2 dB for the remaining 16 sentences (per list). SRTs are calculated by averaging the SNR over sentences 5-20 (including the SNR for a 21st sentence determined from the response to the 20th sentence) for each noise condition (NF, NR, NL, and Q), as described by Soli and Wong (2008).

Participants were randomly assigned to one of three groups. Each group underwent the HINT test in the loudspeaker-based environment and in one of the three different VR environments, as follows: *Group A* ($n = 15$, mean age of 33.4 years) was tested in the loudspeaker-based and VR1 environments, *group B* ($n = 15$, mean age of 29.5 years) was tested in the loudspeaker-based and VR2 environments, and *group C* ($n = 15$, mean age of 31.5 years) in the loudspeaker-based and VR3 environments.

The order in which subjects took the test in both environments (loudspeaker-based or virtual) was randomized across participants. Each participant completed four sentence lists (80 sentences) per test environment (160 sentences in total). No sentence or list was repeated per participant.

Before each test (loudspeaker-based or virtual), participants were informed that their task was to repeat aloud (or log into the system for VR3) what they heard. They were informed that they did not have to keep their head still (as in many laboratory-based listening experiments) and that dynamic binaural rendering was available in the headphone-based conditions. So they knew that they could move their head naturally if they wanted to, as in the loudspeaker-based condition. However, they were not given any verbal or written instructions about optimal head orientation strategies to preserve the undirected nature of the behavioral experiment concerning head movements.

After receiving instructions, participants had the opportunity to complete a practice run (5 sentences) in a random test condition (NL, NR, NF, or Q) to familiarize themselves with the test environment. This was followed by time to ask any questions they had before the test began. All subjects were given a 10–15 min break between tests.

Both tests took place in the same room (the anechoic chamber of the acoustics laboratory of the TH Köln). Participants sat in the same chair in the middle of the loudspeaker-based setup. The only difference between test conditions was whether the auditory presentation was loudspeaker-based or headphone-based and whether the subjects wore the HMD or not.

2.5 Presentation level calibration

First, we adjusted the presentation level for the loudspeaker-based condition. We played the masker noise on each loudspeaker and adjusted their level independently until the free-field sound pressure level at the listener's position was 65 dB(A). Then, for the headphone-based presentation, i.e., all VR conditions, we placed the dummy head (Neumann KU100) in the listener's position, played the same stimuli on the central loudspeaker (0° azimuth), and measured the electrical level produced at the dummy head's ears. Subsequently, the headphones were placed on the ears of the dummy head, and the same stimuli (noise signal from 0° azimuth) were played again via binaural rendering to adjust the headphone presentation level to the same electrical level.

2.6 Parameters and statistical analysis

Individuals may have different SRTs between the NL and NR test conditions. This difference may be due to (common) asymmetries between left and right hearing thresholds, cochlear function, neural processing, or head orientation strategies. Notably, in our study, these differences did not exceed 1 dB, as we tested only NH adults. Thus, to simplify the results' interpretation, we averaged each participant's SRTs from the NL and NR conditions. This resulted in one outcome measure for the spatially separated test conditions (NL and NR averaged) and one for the colocated noise condition (NF).

Following, we calculated the SRM as the difference between the SRTs in the spatially separated and colocated noise conditions (Equation 1).

$$SRM = \left(\frac{SRT_{NL} + SRT_{NR}}{2} \right) - SRT_{NF} \quad (1)$$

For statistical analysis, we performed a Bayesian repeated measures analysis of variance (ANOVA) with default priors (r scale fixed effects of .5, r scale random effects of 1) for SRM (which includes NL, NR, and NF noise conditions) and for the SRTs in quiet with the within-subjects factor environment [loudspeaker-based, virtual] and the between-subjects factor group [A, B, C] (Keyesers, et al., 2020; Rouder et al., 2012).

We performed posthoc testing through individual comparisons based on the default t-test with a Cauchy (0, $r = .707$) prior (Rouder et al., 2009; Wagenmakers, 2007) and corrected for multiple testing by fixing to .5 the prior probability that the null hypothesis holds across all comparisons (Westfall et al., 1997). All statistical analyses were performed using the Bayesian Repeated Measures ANOVA module of the Jamovi software package (Jamovi, 2022).

3 Results

3.1 SRTs in competing noise

Figure 3 shows the calculated SRTs as a function of the spatially separated and colocated noise conditions for all groups in both loudspeaker-based and virtual environments. The box plots show the individual SRTs per participant as points (with a horizontal offset for better readability).

The mean SRTs for the spatially separated noise conditions range from -14.6 to -14.0 dB SNR across groups and test environments, while the mean SRTs for the colocated noise conditions range from -6.0 to -5.7 dB SNR.

The variance in the results from the VR1 environment (Group A) is noticeably higher than the variance in the other groups. In particular, it is more pronounced towards lower (better) SRTs in the VR1 environment than in the loudspeaker-based baseline within the same group. This is the case for six of the fifteen participants in this group (40%) and holds for the same participants in both the spatially separated and the colocated noise conditions.

The effect may be due to the choice of visual feedback provided in the VR1 environment, as the auditory headphone-based presentation was the same in all other virtual environments, and this is not seen in the within-subjects comparison (against the loudspeaker-based baseline in the same group). One possible

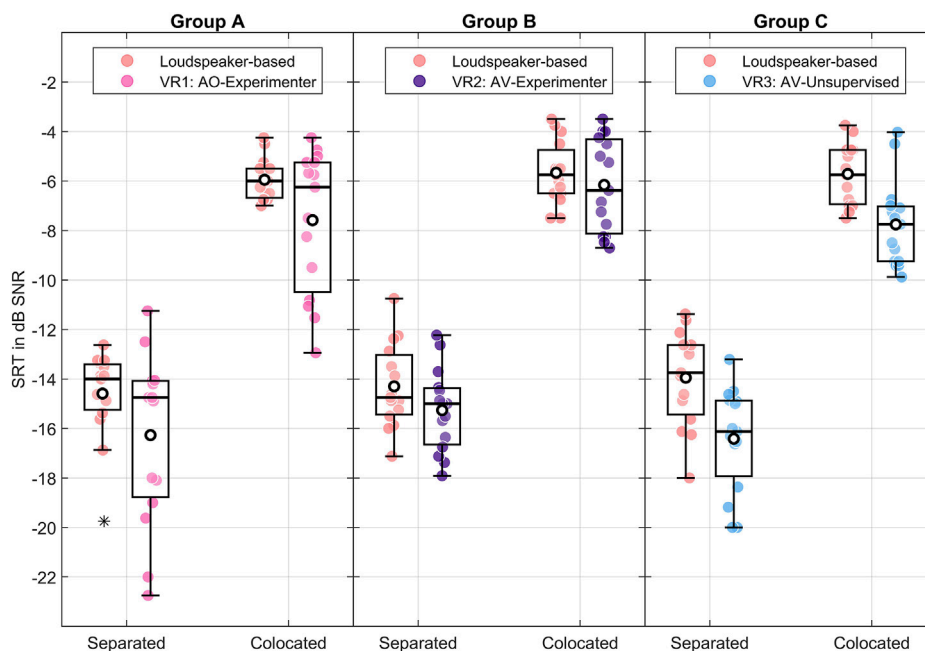


FIGURE 3 Speech reception thresholds (SRTs) as a function of noise condition (spatially separated and colocated) per test group and test environment. The individual SRTs per participant are shown as points per test condition. The boxes represent the (across participants) interquartile range (IQR), the means are shown as white points, and the medians are shown as solid black lines. The whiskers display $1.5 \times$ IQR below the 25th or above the 75th percentile, and asterisks indicate outliers beyond that range.

explanation is that the lack of visual feedback may have benefited some participants by allowing them to focus more on the auditory task, similar to closing their eyes and paying more attention to the auditory stimuli. This may have resulted in an advantage (compared to the loudspeaker-based measurement) where they could see the anechoic room, the loudspeaker array, and the experimenter, which may have distracted them and influenced their responses. However, as this effect appears to be homogeneous for both spatially separated and colocated noise conditions, it does not significantly affect our main outcome measure, the SRM values.

Supplementary Figure S1 shows the raw SRTs before averaging NL and NR conditions.

3.2 SRM

The data presented in Figure 4 shows the calculated SRM values for each subject per test environment and group. The trend lines at the bottom of the figure show individual performance trends across test environments. The line colors indicate an improvement (green) or deterioration (red) in SRM in the virtual environment compared to the baseline measurement in the loudspeaker-based environment.

The means of SRM across all groups and test environments range from 8.4 dB to 9.1 dB.

The Bayesian repeated measures ANOVA for SRM with default priors showed that the predictive performance $P(M | data)$ of the null model was higher than the predictive performance of all the rival models with and without each factor and their interaction (Table 1).

There is positive (moderate) evidence that the null model is more likely than the models including the factor

group ($BF_{01} = 4.72$), and strong evidence of the absence of an effect of both factors [environment + group] ($BF_{01} = 13.73$) and both factors and their interaction [environment + group + environment*group] ($BF_{01} = 70.33$). However, the null model is only 2.84 times more likely than those including the factor environment, and even though the data tends to prove the absence of an effect of environment, it is still inconclusive ($1/3 < BF_{01} < 3$) (Keyesers, et al., 2020).

The analysis of effects across all models, however, reveals evidence of the absence of an effect of environment, group, and their interaction (all $BF_{incl} < 1/3$) (Table 2), and a *post hoc* pairwise comparison confirms evidence for the absence of an effect of test environment in SRM ($BF_{01} = 3.84$) (Table 3).

Since the statistical analysis revealed no effect of group, we show the pooled data across groups in Figure 5 to provide a clearer visualization of the similarity between the SRM measures across test environments. The average SRM across subjects ($n = 45$) when tested in a conventional loudspeaker-based system was 8.5 dB. In comparison, the average SRM increased slightly to 8.8 dB when subjects were tested in virtual environments.

3.3 SRTs in quiet

Figure 6 displays the SRTs in quiet for each group and test environment in dB (A) SPL, i.e., free-field sound pressure level at the listener’s position. The means of SRTs in quiet across all groups and test environments range from 13.4 dB(A) to 16.2 dB(A) SPL.

The Bayesian repeated measures ANOVA for the SRTs in quiet revealed positive (moderate) evidence in favor of the absence of an

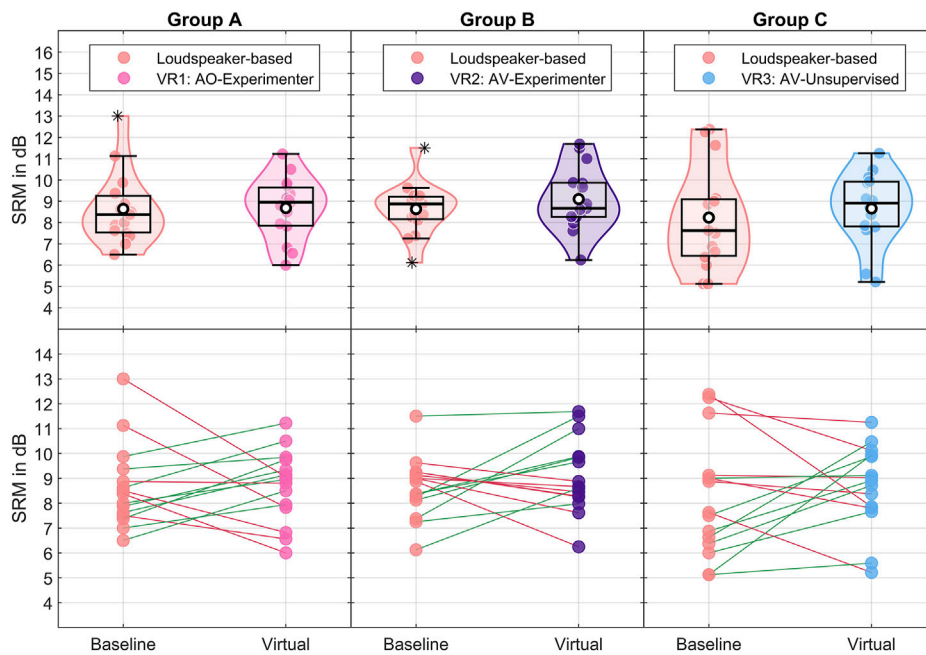


FIGURE 4 Spatial release from masking (SRM) as a function of test environment per group. The individual SRM values per participant are shown as points per test condition. Top: The boxes represent the (across participants) interquartile range (IQR). The means are shown as white points, and the medians are shown as solid black lines. The whiskers display 1.5 × IQR below the 25th or above the 75th percentile, and asterisks indicate outliers beyond that range. Bottom: Trend lines connect the results per participant. The color of the line indicates improved (in green) or deteriorated (in red) SRM in the virtual environment compared to the baseline measurement in the loudspeaker-based environment.

TABLE 1 Bayesian repeated measures ANOVA: Model Comparison - SRM.

Models	$P(M)$	$P(M data)$	BF_M	BF_{01}	error [%]
Null model	0.200	0.60490	6.1241	1.00	
Environment	0.200	0.21322	1.0840	2.84	0.924
Group	0.200	0.12802	0.5873	4.72	0.833
Environment + Group	0.200	0.04525	0.1896	13.37	1.271
Environment + Group + Environment * Group	0.200	0.00860	0.0347	70.33	1.727

TABLE 2 Analysis of effects - SRM.

Effects	$P(incl)$	$P(incl data)$	BF_{incl}
Environment	0.600	0.26707	0.2429
Group	0.600	0.18187	0.1482
Environment * Group	0.200	0.00860	0.0347

effect of environment ($BF_{01} = 3.665$) and in favor of the absence of an effect of both factors [environment + group] ($BF_{01} = 8.341$). However, for the models including the factor group and the full model (including both factors and their interaction), the evidence is too weak to be conclusive, i.e., there is an absence of evidence ($\frac{1}{3} < BF_{01} < 3$) (Table 4).

The analysis of effects across matched models provides further evidence that the null model is twelve times more likely than those

including the interaction between test environment and group ($BF_{incl} = 12.187$), confirms the absence of an effect of environment ($BF_{incl} = 0.271 < \frac{1}{3}$), but remains inconclusive regarding the factor group ($\frac{1}{3} < BF_{01} < 3$) (Table 5).

Post hoc pairwise comparisons confirm evidence for the absence of an effect of test environment ($BF_{01} = 4.78$) (Table 6) but remain inconclusive for all pairwise comparisons between groups ($\frac{1}{3} < BF_{01} < 3$) (Table 7).

4 Discussion

4.1 Comparing loudspeaker-based with virtual test environments

The means of all speech-in-noise measures from our study, in both loudspeaker-based and virtual environments, are consistent

TABLE 3 Posthoc comparison – Environment (SRM).

		Prior Odds	Posterior Odds	BF ₀₁	error [%]
Loudspeaker-based	Virtual	1.00	3.84	3.84	0.0486

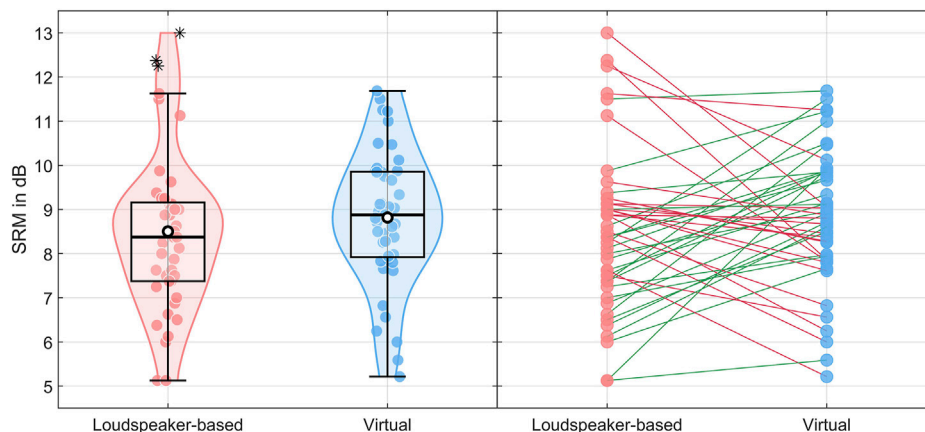


FIGURE 5 Spatial release from masking (SRM) as a function of test environment (pooled over groups). The individual SRM values per participant are shown as points per test condition. Left: The boxes represent the (across participants) interquartile range (IQR), the means are shown as white points, and the medians are shown as solid black lines. The whiskers display 1.5 × IQR below the 25th or above the 75th percentile, and asterisks indicate outliers beyond that range. Right: Trend lines connect the results per participant. The color of the line indicates improved (in green) or deteriorated (in red) SRM in the virtual environment compared to the baseline measurement in the loudspeaker-based environment.

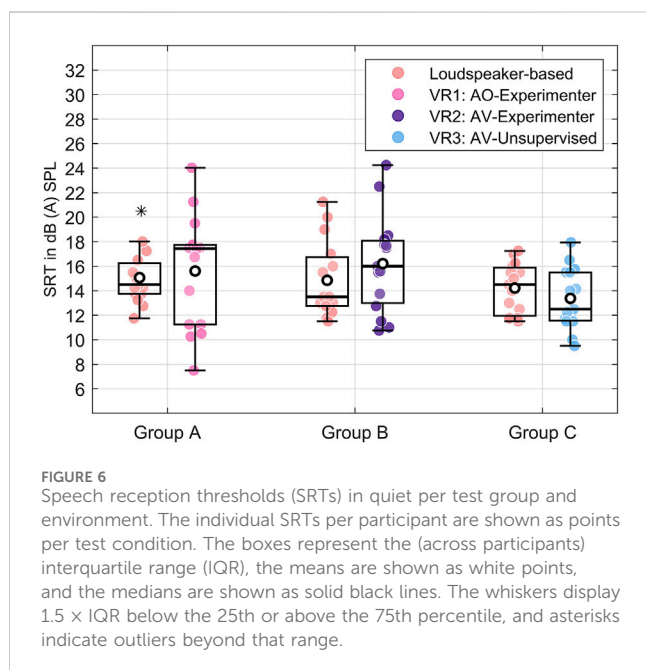


FIGURE 6 Speech reception thresholds (SRTs) in quiet per test group and environment. The individual SRTs per participant are shown as points per test condition. The boxes represent the (across participants) interquartile range (IQR), the means are shown as white points, and the medians are shown as solid black lines. The whiskers display 1.5 × IQR below the 25th or above the 75th percentile, and asterisks indicate outliers beyond that range.

with the norms reported in the literature for the German HINT (Figure 7) (Joiko et al., 2021; Mönnich et al., 2023).

Similarly, the average SRM from both loudspeaker-based and virtual environments align with the findings of previous studies using the same spatial configuration and masker type employed here (stationary noise at 90° azimuth) (Andersen et al., 2016; Beutelmann and Brand, 2006; Bronkhorst and Plomp, 1988; Cosentino et al.,

2014; Jelfs et al., 2011; Müller, 1992; Ozimek et al., 2013; Peissig and Kollmeier, 1997; Platte and vom Hövel, 1980; Plomp and Mimpen, 1981).

The validity of our baseline measurement was demonstrated by results consistent with those of previous studies (Figure 8). Furthermore, the similarity of the results obtained in the virtual environments with those of the baseline measurement for both outcome measures (SRM and SRTs in quiet) suggests that it may be feasible to replicate speech-in-noise results from conventional loudspeaker-based setups using portable and inexpensive VR peripherals.

It should be noted, however, that our results are limited to the experimental conditions tested: anechoic environment, with a single target speech source and a single energetic masker in spatially separated (90°) and collocated conditions.

With this preliminary study, we aimed to provide initial evidence that state-of-the-art technology can be used to reproduce results from conventional laboratory and clinical settings. The positive results of our study encourage further testing in more complex, i.e., realistic, listening environments, which may lead to more ecologically valid results. In particular, evaluating the ability to reproduce results from “real” reverberant conditions in virtual audiovisual environments is of great interest.

In this context, a significant body of research has shown that with current methods and technologies, such as those used in this study, it is possible to create virtual auditory environments indistinguishable from reality, even in reverberant listening conditions. Plausible and authentic virtual acoustic presentations are possible in both anechoic (Arend et al., 2021a; Weber et al., 2024) and reverberant conditions (Lindau and Weinzierl, 2012;

TABLE 4 Bayesian ANOVA: Model Comparison—SRTs in quiet.

Models	$P(M)$	$P(M data)$	BF_M	BF_{01}	error [%]
Null model	0.200	0.3029	1.738	1.000	
Environment	0.200	0.0827	0.360	3.665	1.09
Group	0.200	0.1356	0.627	2.234	1.43
Environment + Group	0.200	0.0363	0.151	8.341	1.11
Environment + Group + Environment * Group	0.200	0.4426	3.176	0.684	1.39

TABLE 5 Analysis of effects - SRTs in quiet.

Effects	$P(incl)$	$P(incl data)$	BF_{incl}
Environment	0.400	0.119	0.271
Group	0.400	0.172	0.446
Environment * Group	0.200	0.443	12.187

Brinkmann et al., 2017; Brinkmann et al., 2019; Arend et al., 2021b; Arend et al., 2024). However, as outlined in the introduction, exploiting multisensory integration is crucial to achieving this level of realism (Keidser et al., 2020). Many relevant technical aspects have to be considered, such as the use of real-time motion compensation, the use of matching visuals or matching the auditory environment perfectly to the visual real world, including appropriate descriptions of the source and receiver characteristics, e.g., source directivity and HRTFs, and correct headphone compensation filters, among others.

With this contribution, we include the Unity project (including all the virtual environments described here), which facilitates the reproducibility and extension of our experimental setup. Multiple sources can be easily added, and all stimuli (including audio and text) can be replaced using the regular file system. This means our setup can be easily transferred to different stimuli in different languages. The application also allows easy customization of various parameters, such as the number of noise conditions, lists, sentences, practice rounds, and adaptive step sizes. There is no need to modify the source code, as all customization can be done using the Unity Inspector interface.

4.2 On the use of headphone-based dynamic binaural rendering for speech-in-noise testing

To our knowledge, this study is the first to investigate speech-in-noise abilities using headphone-based dynamic binaural rendering with non-individual HRTFs. The role of spontaneous head movements in increasing the target signal level when speech intelligibility decreases, also known as the head orientation

benefit (HOB), has been extensively studied. Kock's work in 1950 was the first to demonstrate this phenomenon (Kock, 1950). Later, Grange and Culling (2016) investigated the benefits of head orientation away from the speech source in NH listeners. They analyzed spontaneous head orientations when listeners were presented with long speech clips of gradually decreasing SNR in an acoustically treated room. The speech was presented from a loudspeaker initially facing the listener, and the competing noise was presented from one of four other locations. In an undirected paradigm, they observed that listeners instinctively turned their heads away from the speech (between $\pm 10^\circ$ and $\pm 65^\circ$) in 56% of trials in response to increased intelligibility difficulties. They then observed that when subjects were explicitly instructed to perform head movements, all turned away from speech at lower SNRs and immediately reached head orientations associated with lower SRTs.

Similarly, Brimijoin et al. (2012) investigated head orientation strategies in a speech comprehension task in the presence of spatially separated competing noise. They found a clear tendency to orient approximately 60° away from the target, regardless of the position of the distractor signal, in listeners with large (>16 dB) hearing threshold differences between their left and right ears.

We did not log the head-tracking data in this study because spontaneous head movements and the resulting HOB have already been studied and demonstrated for a long time. However, as expected, we observed that participants' behavior regarding head orientations was consistent with the abovementioned findings.

The videos in the [Supplementary Material](#), recorded from the listener's perspective, exemplify the spontaneous use of head movements and illustrate the (very pronounced) level increase in one ear when head movements are exploited.

4.3 On the use of visual feedback

In addition to supporting auditory spatial perception by aiding externalization and distance estimation (Best et al., 2020), the use of visual cues in speech-in-noise testing paradigms may facilitate a better understanding of the mechanisms underlying auditory perception from a multisensory perspective and potentially lead to significant advances in hearing research. In clinical settings, using

TABLE 6 Posthoc comparison – Environment (SRTs in quiet).

		Prior Odds	Posterior Odds	BF_{01}	error [%]
Loudspeaker-based	Virtual	1.00	4.78	4.78	0.0539

TABLE 7 Posthoc comparisons - Groups (SRTs in quiet).

		Prior Odds	Posterior Odds	BF ₀₁	error [%]
A	B	1.70	4.67	2.742	0.01006
A	C	1.70	1.31	0.770	0.00928
B	C	1.70	4.36	2.560	0.01000

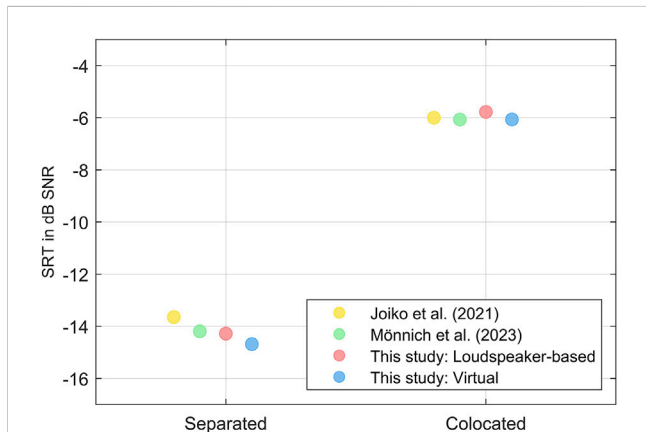


FIGURE 7 Mean speech reception thresholds (SRTs) for spatially separated and collocated noise conditions in this study in both loudspeaker-based and virtual test environments compared to the norms reported for the German HINT with male (Joiko et al., 2021) and female (Mönnich et al., 2023) speakers.

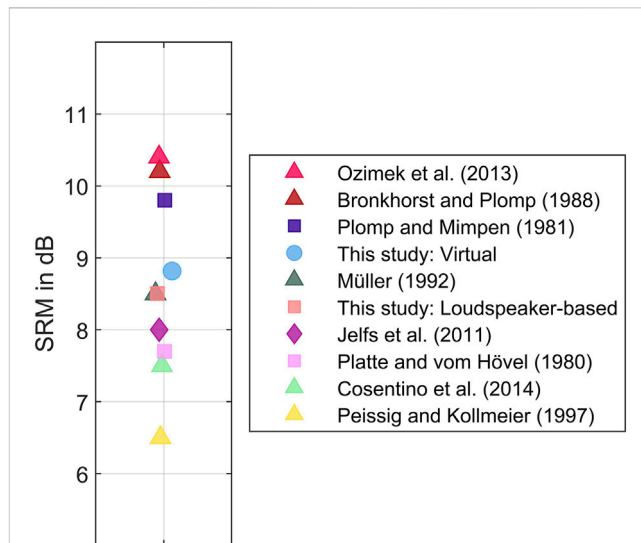


FIGURE 8 Mean spatial release from masking (SRM) in this study in both loudspeaker-based and virtual test environments compared to results from previous studies using the same spatial configuration and masker type (stationary noise at 90° azimuth) in different setups: Loudspeaker-based (squares), headphone-based with static binaural rendering (triangles), and headphone-based with dynamic binaural rendering (circle) - [Data other than those from this study are from Figure 8.7 in Culling and Lavandier (2021). The data point from Jelfs et al. (2011) corresponds to a model-based prediction for a stationary noise source at 90° azimuth and target speech in front using generic head-related transfer functions (diamond)].

appropriate visual cues can improve current assessment procedures' ecological validity and accuracy. For example, lip-reading has been shown to support speech intelligibility in noisy environments, and it is an aspect that is still overlooked in current speech-in-noise assessment methods (Helfer and Freyman, 2005; Williams et al., 2023; Yuan et al., 2021). In this study, we used a simple visual cue at the target speaker location without facial expressions or lip movements, and we found evidence of the absence of an effect of the visual feedback used in SRM.

Nevertheless, this may be different in more complex listening scenarios, for example, in cases where there is some uncertainty about the target speaker's location or in multi-speaker or cocktail party scenarios. There may be interactions between vision and auditory perception that require further investigation, and VR can play an important role in supporting auditory research in this regard. Further work should focus on understanding these potential interactions and their impact on speech intelligibility and listening effort.

4.4 On the use of unsupervised procedures

While currently available speech-in-noise tests are reliable, they require manual scoring by a clinician, which can be inconvenient in busy clinical settings. As a result, these tests are not widely used in routine hearing evaluations. The literature highlights the need for automated tests, which could allow testing while the patient is waiting in the clinic or remotely (Jakien et al., 2017). The use of closed-set tasks may facilitate unsupervised measurements.

To assess whether using a closed-set instead of the standard open-set procedure from the HINT would affect speech-in-noise measures, we incorporated the word selection feedback system into the VR3 test. We found evidence for the absence of an effect of group and environment on the SRM measure, suggesting that it is feasible to conduct speech-in-noise testing in VR, even with the unsupervised procedure introduced here. However, the evidence supporting our unsupervised procedure was too weak to be conclusive for SRTs in quiet.

Closed-set procedures may result in reduced (better) SRTs, mainly when the response set contains few phonetically dissimilar alternatives (Miller, Heise, and Lighten, 1951; Warzybok et al., 2015; Buss, Leibold, and Hall, 2016). Future research in remote or unsupervised speech-in-noise testing could explore alternative ways to design automated response systems. For example, Litovsky (2005) suggested adjusting task difficulty based on the listener's age. Their results indicated that 4AFC tasks were easier for adults than children, resulting in lower SRTs. Johnstone and Litovsky (2006) subsequently found significant differences in adult SRTs using 4AFC and 25AFC tasks as response methods, but only when speech was used as a masker, not when modulated noise was used, suggesting that appropriate response methods could be both population and stimuli-dependent.

Another attractive alternative may be to incorporate automatic speech recognition (ASR) into the virtual tests, preserving the open-set nature of the task while allowing for unsupervised testing. Ooster et al. (2023) proposed using an ASR for automatic response recording based on a time-delay neural network. They estimate an SRT deviation below 1.38 dB for 95% of users with this method, suggesting that robust unsupervised testing may be possible with

similar accuracy as with a human supervisor, even in noisy conditions and with altered or disordered speech from elderly severely hearing-impaired listeners and cochlear implant users.

4.5 Other remarks regarding virtualization

While VR offers potentially groundbreaking opportunities, many issues must be carefully considered before virtual testing can be used for individualized screening in clinical settings. Some of them are:

Our preliminary results represent only a small sample of NH adults. These results cannot be generalized to other populations. Follow-up studies should include large standardized samples, including patients with hearing loss, auditory processing disorders, and NH controls of different age groups in both conventional loudspeaker-based and VR conditions.

Although VR and its potential to serve children has been extensively researched, particularly in educational and medical settings, such as a tool for pain distraction, assessment of Attention Deficit/Hyperactivity Disorder (ADHD) and Autism Spectrum Disorder (ASD), and psychotherapy, among others, many questions remain about the potential impact of VR on children's development. It is still unclear whether we could use VR to assess speech-in-noise abilities in children. Further research should focus on creating controlled and safe environments that allow us to address these questions. This includes aspects such as appropriate exposure times, appropriate complexity of visual feedback, and considerations such as appropriate HRTF sets from a technical virtual acoustics perspective.

VR can cause motion sickness. In our study, participants did not report any adverse effects while using the HMD. However, this is a relevant aspect when using more complex visual feedback, as this may increase the likelihood of experiencing it.

Managing cognitive load is another major challenge in the design and use of VR. Excessive cognitive load can negatively affect the user experience, cause fatigue, and interfere with task performance. Therefore, future research should focus on understanding the relationship between increased realism in virtual environments, its associated cognitive load, and its potential impact on measures of auditory processing.

Recognizing and addressing the potential drawbacks of VR through research, innovation, and responsible use can help maximize its benefits while minimizing its risks. By promoting ethical and inclusive practices and fostering a balanced approach to VR adoption, we can harness this transformative technology in hearing research and audiological healthcare.

5 Conclusion

Our results suggest that conducting the HINT, a widely accepted and accurate speech-in-noise test, is feasible in state-of-the-art VR environments, even without experimenter supervision. We found no statistically significant differences between the SRM measures obtained in any of the VR environments tested and the loudspeaker-based setup used as a baseline. However, for the SRTs in quiet, the evidence was too weak to be conclusive. Nevertheless, as described in the Introduction,

speech-in-noise measures are considered to be more representative of a person's functional hearing ability in real-life listening scenarios than SRTs in quiet, and current literature and clinical guidelines encourage the use of speech-in-noise testing rather than measuring SRTs in quiet when a more comprehensive understanding of a patient's everyday hearing ability is desired.

Although our study was limited to an anechoic environment with a single target speech source and a single energetic masker (following the standard setup for the HINT), our findings pave the way for further research. Future studies should evaluate how these results generalize to more complex listening scenarios, such as those involving multiple speakers, greater visual complexity, and more diverse populations. Specifically, future work should investigate how these results apply to hard-of-hearing individuals, including patients with hearing loss and auditory processing disorders in different age groups.

Data availability statement

The original contributions presented in the study are included in the article/[Supplementary Material](#), further inquiries can be directed to the corresponding author. To promote and foster open-source and reproducible research, we have made the Unity project (including all the virtual environments described here) available at <https://github.com/AudioGroupCologne/HINT-VR> under a Creative Commons license CC BY-NC-SA 4.0. Supplemental material for this article is available online in <https://zenodo.org/records/13952337>.

Ethics statement

The studies involving humans were approved by Institute of Computer and Communication Technology at the TH Köln. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

MR: Conceptualization, Formal Analysis, Investigation, Methodology, Visualization, Writing–original draft, Writing–review and editing. AM: Data curation, Software, Writing–review and editing. JA: Funding acquisition, Supervision, Validation, Writing–review and editing. HH: Data curation, Writing–review and editing. TR: Resources, Validation, Writing–review and editing. CP: Funding acquisition, Project administration, Supervision, Validation, Writing–review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was sponsored by the German Federal Ministry of Education and Research BMBF (13FH6661A6-VIWER-S) and partly by the German Research Foundation (DFG WE 4057/21-1).

Acknowledgments

The authors express their gratitude to all the participants who took part in the study. They would also like to thank Tilman Brach and Johan Dasbach for their assistance in collecting the data. The authors would like to thank the handling editor and the reviewers for their constructive feedback, which significantly improved the quality of the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Andersen, A. H., de Haan, J. M., Tan, Z. H., and Jensen, J. (2016). Predicting the intelligibility of noisy and nonlinearly processed binaural speech. *IEEE/ACM Trans. Audio, Speech, Lang. Process.* 24 (11), 1908–1920. doi:10.1109/TASLP.2016.2588002
- Arbogast, T. L., Mason, C. R., and Kidd, G. (2005). The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.* 117 (4), 2169–2180. doi:10.1121/1.1861598
- Arend, J. M., Brinkmann, F., Ramirez, M., Scheer, C., and Weinzierl, S. (2024). “Auditory distance perception in a real and virtual walk-through environment,” in *Proceedings of the 50th DAGA*. Hannover.
- Arend, J. M., Ramirez, M., Liesefeld, H. R., and Pörschmann, C. (2021a). Do near-field cues enhance the plausibility of non-individual binaural rendering in a dynamic multimodal virtual acoustic scene? *Acta Acust.* 5 (3), 55. doi:10.1051/aacus/2021048
- Arend, J. M., Schissler, C., Klein, F., and Robinson, P. W. (2021b). Six-Degrees-of-Freedom parametric spatial audio based on one monaural room impulse response. *J. Audio Eng. Soc.* 69 (7/8), 557–575. doi:10.117743/jaes.2021.0009
- Beck, D. L. (2023). Speech-in-Noise testing: pivotal and rare. *Hear. J.* 76 (12), 28–32. doi:10.1097/01.HJ.0000997248.20295.53
- Beechey, T. (2002). Ecological validity, external validity, and mundane realism in hearing science. *Ear Hear.* 43 (5), 1395–1401. doi:10.1097/AUD.0000000000001202
- Begault, D. R., Wenzel, E. M., and Anderson, M. R. (2001). Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. *AES J. Audio Eng. Soc.* 49 (10), 904–916.
- Bellis, T. J., and Bellis, J. D. (2015). Central auditory processing disorders in children and adults. *Handb. Clin. Neurology* 129 (1954), 537–556. doi:10.1016/B978-0-444-62630-1.00030-5
- Bench, J., Kowal, Å., and Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *Br. J. Audiology* 13 (3), 108–112. doi:10.3109/03005367909078884
- Bernschütz, B. (2013). “A spherical far field HRIR HRTF compilation of the Neumann KU 100,” in *Proceedings of the 39th DAGA*, 592–595.
- Best, V., Baumgartner, R., Lavandier, M., Majdak, P., and Kopčo, N. (2020). Sound externalization: a review of recent research. *Trends Hear.* 24, 1–14. doi:10.1177/2331216520948390
- Best, V., Marrone, N., Mason, C. R., and Kidd, G., Jr. (2012). The influence of non-spatial factors on measures of spatial release from masking. *J. Acoust. Soc. Am.* 131 (4), 3103–3110. doi:10.1121/1.3693656
- Beutelmann, R., and Brand, T. (2006). Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.* 120 (1), 331–342. doi:10.1121/1.2202888
- Brimjoin, W. O., Boyd, A. W., and Akeroyd, M. A. (2013). The contribution of head movement to the externalization and internalization of sounds. *PLoS ONE* 8 (12), 830688–e83112. doi:10.1371/journal.pone.0083068
- Brimjoin, W. O., McShefferty, D., and Akeroyd, M. A. (2012). Undirected head movements of listeners with asymmetrical hearing impairment during a speech-in-noise task. *Hear. Res.* 283 (1–2), 162–168. doi:10.1016/j.heares.2011.10.009
- Brinkmann, F., Aspöck, L., Ackermann, D., Lepa, S., Vorländer, M., and Weinzierl, S. (2019). A round robin on room acoustical simulation and auralization. *J. Acoust. Soc. Am.* 145 (4), 2746–2760. doi:10.1121/1.5096178

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/frvir.2024.1470382/full#supplementary-material>

- Brinkmann, F., Lindau, A., and Weinzierl, S. (2017). On the authenticity of individual dynamic binaural synthesis. *J. Acoust. Soc. Am.* 142 (4), 1784–1795. doi:10.1121/1.5005606
- Bronkhorst, A. W. (2000). The cocktail party phenomenon: a review of research on speech intelligibility in multiple-talker conditions. *Acta Acustica united Acustica* 86 (1), 117–128.
- Bronkhorst, A. W., and Plomp, R. (1988). The effect of head-induced interaural time and level differences on speech intelligibility in noise. *J. Acoust. Soc. Am.* 83 (4), 1508–1516. doi:10.1121/1.395906
- Buss, E., Leibold, L. J., and Hall, J. W. (2016). Effect of response context and masker type on word recognition in school-age children and adults. *J. Acoust. Soc. Am.* 140 (2), 968–977. doi:10.1121/1.4960587
- Busse, L., Roberts, K. C., Crist, R. E., Weissman, D. H., and Woldorff, M. G. (2005). The spread of attention across modalities and space in a multisensory object. *Proc. Natl. Acad. Sci. U. S. A.* 102 (51), 18751–18756. doi:10.1073/pnas.0507704102
- Cameron, S., and Dillon, H. (2007). Development of the listening in spatialized noise-test (LISN-S). *Ear Hear.*, 28(2), pp. 196–211. doi:10.1097/AUD.0b013e318031267f
- Carhart, R., and Tillman, T. W. (1970). Interaction of competing speech signals with hearing losses. *Archives Otolaryngology* 91 (3), 273–279. doi:10.1001/archotol.1970.00770040379010
- Clark, J. G., Huff, C., and Earl, B. (2017). Clinical practice report card—Are we meeting best practice standards for adult hearing rehabilitation? *Audiol. Today* 29 (6), 15–25.
- Cosentino, S., Marquardt, T., McAlpine, D., Culling, J. F., and Falk, T. H. (2014). A model that predicts the binaural advantage to speech intelligibility from the mixed target and interferer signals. *J. Acoust. Soc. Am.* 135 (2), 796–807. doi:10.1121/1.4861239
- Cuevas-Rodríguez, M., Picinali, L., González-Toledo, D., Garre, C., de la Rubia-Cuevas, E., Molina-Tanco, L., et al. (2019). 3D Tune-In Toolkit: an open-source library for real-time binaural spatialisation. *PLoS ONE* 14 (3), e0211899. doi:10.1371/JOURNAL.PONE.0211899
- Culling, J. F., and Lavandier, M. (2021). “Binaural unmasking and spatial release from masking,” in *Binaural hearing. Springer handbook of auditory research*. Editor R. Y. Litovsky, (Cham: Springer), 209–241. doi:10.1007/978-3-030-57100-9_8
- Dirks, D. D., and Wilson, R. H. (1969). “The effect of spatially separated sound sources on speech intelligibility,” *J. speech Hear. Res.*, 12(1), pp. 5–38. doi:10.1044/jshr.1201.05
- Erbes, V., Geier, M., Wierstorf, H., and Spors, S. (2017). “Free database of low-frequency corrected head-related transfer functions and headphone compensation filters,” in *Proceedings of 142nd audio engineering society convention* 325, 1–5.
- Garadat, S. N., Litovsky, R. Y., Yu, G., and Zeng, F. G. (2009). Role of binaural hearing in speech intelligibility and spatial release from masking using vocoded speech. *J. Acoust. Soc. Am.* 126 (1522), 2522–2535. doi:10.1121/1.3238242
- Gaveau, V., Coudert, A., Salemm, R., Koun, E., Desoche, C., Truy, E., et al. (2022). Benefits of active listening during 3D sound localization. *Exp. Brain Res.* 240 (11), 2817–2833. doi:10.1007/s00221-022-06456-x
- Grange, J. A., and Culling, J. F. (2016). The benefit of head orientation to speech intelligibility in noise. *J. Acoust. Soc. Am.* 139 (2), 703–712. doi:10.1121/1.4941655
- Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (2004). The benefit of binaural hearing in a cocktail party: effect of location and type of interferer. *J. Acoust. Soc. Am.* 115 (2), 833–843. doi:10.1121/1.1639908
- Helfer, K. S., and Freyman, R. L. (2005). The role of visual speech cues in reducing energetic and informational masking. *J. Acoust. Soc. Am.* 117 (2), 842–849. doi:10.1121/1.1836832

- Hess, C. L., Misurelli, S. M., and Litovsky, R. Y. (2018). Spatial release from masking in 2-year-olds with normal hearing and with bilateral cochlear implants. *Trends Hear.* 22, 2331216518775567–13. doi:10.1177/2331216518775567
- Isaiah, A., Vongpaisal, T., King, A. J., and Hartley, D. E. H. (2014). Multisensory training improves auditory spatial processing following bilateral cochlear implantation. *J. Neurosci.* 34 (33), 11119–11130. doi:10.1523/JNEUROSCI.4767-13.2014
- Jakien, K. M., Kampel, S. D., Stansell, M. M., and Gallun, F. J. (2017). Validating a rapid, automated test of spatial release from masking. *Am. J. Audiology* 26 (4), 507–518. doi:10.1044/2017_AJA-17-0013
- Jamovi [Computer Software] (2022). "The Jamovi project."
- Jelfs, S., Culling, J. F., and Lavandier, M. (2011). Revision and validation of a binaural model for speech intelligibility in noise. *Hear. Res.* 275 (1–2), 96–104. doi:10.1016/j.heares.2010.12.005
- Johnstone, P. M., and Litovsky, R. Y. (2006). Effect of masker type and age on speech intelligibility and spatial release from masking in children and adults. *J. Acoust. Soc. Am.* 120 (4), 2177–2189. doi:10.1121/1.2225416
- Joiko, J., Bohnert, A., Strieth, S., Soli, S. D., and Rader, T. (2021). The German hearing in noise test. *Int. J. Audiology* 60 (110), 927–933. doi:10.1080/14992027.2020.1837969
- Keidser, G. (2016). Introduction to special issue: towards ecologically valid protocols for the assessment of hearing and hearing devices. *J. Am. Acad. Audiology* 27 (7), 502–503. doi:10.3766/jaaa.27.7.1
- Keidser, G., Naylor, G., Brungart, D. S., Caduff, A., Campos, J., Carlile, S., et al. (2020). The quest for ecological validity in hearing science: what it is, why it matters, and how to advance it. *Ear & Hear.* 41 (1), 5–19. doi:10.1097/AUD.0000000000000944
- Keyesers, C., Gazzola, V., and Wagenmakers, E. J. (2020). Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence. *Nat. Neurosci.* 23 (7), 788–799. doi:10.1038/s41593-020-0660-4
- Kidd, G., Mason, C. R., and Brughera, A. (2005). 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. doi:10.1121/1.4809166
- Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., and Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.* 116 (4), 2395–2405. doi:10.1121/1.1784440
- Kock, W. E. (1950). Binaural localization and masking. *J. Acoust. Soc. Am.* 22 (6), 801–804. doi:10.1121/1.1906692
- Lindau, A., and Brinkmann, F. (2012). Perceptual evaluation of head-phone compensation in binaural synthesis based on non-individual recordings. *J. Audio Eng. Soc.* 60 (1/2), 54–62.
- Lindau, A., and Weinzierl, S. (2012). Assessing the plausibility of virtual acoustic environments. *Acta Acustica united Acustica* 98 (5), 804–810. doi:10.3813/AAA.918562
- Litovsky, R. Y. (2005). Speech intelligibility and spatial release from masking in young children. *J. Acoust. Soc. Am.* 117 (5), 3091–3099. doi:10.1121/1.1873913
- Majdak, P., Zotter, F., Brinkmann, F., De Muyenke, J., Mihocic, M., and Noisternig, M. (2022). Spatially oriented format for acoustics 2.1: introduction and recent advances. *AES J. Audio Eng. Soc.* 70 (7–8), 565–584. doi:10.17743/jaes.2022.0026
- Middlebrooks, J. C., Simon, J. Z., Popper, A. N., and Fay, R. F. (2017). *Springer handbook of auditory research: the auditory system at the cocktail party*. Editor A. Press (Springer), 60. Available at: <https://doi.org/10.1007/978-3-319-51662-2>
- Miller, G. A., Heise, G. A., and Lighten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *J. Exp. Psychol.* 41(5), pp. 329–335. doi:10.1037/h0062491
- Mönnich, A.-L., Strieth, S., Bohnert, A., Ernst, B. P., and Rader, T. (2023). The German hearing in noise test with a female talker: development and comparison with German male speech test. *Eur. Archives Oto-Rhino-Laryngology* 280, 3157–3169. doi:10.1007/s00405-023-07820-5
- Mueller, H. G. (2016). *Signia expert series: speech-in-noise testing for selection and fitting of hearing aids: worth the effort?* Audiology Online. Available at: <https://www.audiologyonline.com/articles/signia-expert-series-speech-in-18336>.
- Mueller, H. G., Ricketts, T., and Hornsby, B. Y. G. (2023). *20Q: speech-in-noise testing - too useful to be ignored*. AudiologyOnline. Available at: <https://www.audiologyonline.com/articles/20q-speech-in-noise-testing-28760>.
- Müller, C. (1992). Perzeptive Analyse und Weiterentwicklung eines Reimtestverfahrens für die Sprachaudiometrie. *Univ. Göttingen*.
- Murphy, J. (2017). Virtual reality: the next frontier of audiology. *Hear. J.* 70, 24–27. doi:10.1097/01.HJ.0000525521.39398.8f
- Nilsson, M., Soli, S. D., and Sullivan, J. A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *J. Acoust. Soc. Am.* 95 (2), 1085–1099. doi:10.1121/1.408469
- Niquette, P., Arcaroli, J., Revit, L., Parkinson, A., Staller, S., Skinner, M., et al. (2003). "Development of the BKB-SIN test," in *Annual meeting of the American auditory society* (AZ: Scottsdale).
- Ooster, J., Tuschen, L., and Meyer, B. T. (2023). Self-conducted speech audiometry using automatic speech recognition: simulation results for listeners with hearing loss. *Comput. Speech Lang.* 78 (June 2021), 101447. doi:10.1016/j.csl.2022.101447
- Ozimek, E., Kociński, J., Kutzner, D., Sęk, A., and Wicher, A. (2013). Speech intelligibility for different spatial configurations of target speech and competing noise source in a horizontal and median plane. *Speech Commun.* 55 (10), 1021–1032. doi:10.1016/j.specom.2013.06.009
- Peissig, J., and Kollmeier, B. (1997). Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners. *J. Acoust. Soc. Am.* 101 (3), 1660–1670. doi:10.1121/1.418150
- Peng, Z. E., and Litovsky, R. Y. (2022). Novel approaches to measure spatial release from masking in children with bilateral cochlear implants. *Ear Hear.* 43 (1), 101–114. doi:10.1097/AUD.0000000000001080
- Platte, H. J., and vom Hövel, H. (1980). Zur deutung der ergebnisse von sprachverstaendlichkeitsmessungen mit stoerschall im freifeld. *Acta Acustica united Acustica* 45 (3), 139–150.
- Plomp, R., and Mimpen, A. M. (1981). Effect of the orientation of the speaker's head and the azimuth of a noise source on the speech-reception threshold for sentences. *Acustica* 48 (5), 325–328.
- Rader, T., Fastl, H., and Baumann, U. (2013). Speech perception with combined electric-acoustic stimulation and bilateral cochlear implants in a multisource noise field. *Ear Hear.* 34 (3), 324–332. doi:10.1097/AUD.0b013e318272f189
- Reyes-Lecuona, A., and Picinali, L. (2022). Unity wrapper for 3DTI. Available at: https://github.com/3DTTune-In/3dti_AudioToolkit_UnityWrapper.
- Rouder, J. N., Morey, R. D., Speckman, P. L., and Province, J. M. (2012). Default Bayes factors for ANOVA designs. *J. Math. Psychol.* 56 (5), 356–374. doi:10.1016/j.jmp.2012.08.001
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., and Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bull. & Rev.* 16 (2), 225–237. doi:10.3758/PBR.16.2.225
- Roup, C. M., Custer, A., and Powell, J. (2021). The relationship between self-perceived hearing ability and binaural speech-in-noise performance in adults with normal pure-tone hearing. *Perspectives* 6 (5), 1085–1096. doi:10.1044/2021_PERSP-21-00032
- Ruggles, D., Bharadwaj, H., and Shinn-Cunningham, B. G. (2011). Normal hearing is not enough to guarantee robust encoding of suprathreshold features important in everyday communication. *Proc. Natl. Acad. Sci. U. S. A.* 108 (37), 15516–15521. doi:10.1073/pnas.1108912108
- Soli, S. D., and Wong, L. L. N. (2008). Assessment of speech intelligibility in noise with the hearing in noise test. *Int. J. Audiology* 47 (6), 356–361. doi:10.1080/14992020801895136
- Strelcyk, O., and Dau, T. (2009). Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing. *J. Acoust. Soc. Am.* 125 (5), 3328–3345. doi:10.1121/1.3097469
- Taylor, B. (2003). Speech-in-noise tests: how and why to include them in your basic test battery. *Hear. J.* 56 (1), 40–46. doi:10.1097/01.HJ.0000293000.76300.ff
- Tiippana, K., Sams, M., and Puharinen, H. (2011). Sound location can influence audiovisual speech perception when spatial attention is manipulated. *Seeing Perceiving* 24 (1), 67–90. doi:10.1163/187847511X557308
- Valzolgher, C. (2024). Motor strategies: the role of active behavior in spatial hearing research. *Psychol. Rep.* 0 (0), 332941241260246–23. doi:10.1177/00332941241260246
- Valzolgher, C., Campus, C., Rabini, G., Gori, M., and Pavani, F. (2020). Updating spatial hearing abilities through multisensory and motor cues. *Cognition* 204 (November), 104409. doi:10.1016/j.cognition.2020.104409
- Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bull. & Rev.* 14 (5), 779–804. doi:10.3758/BF03194105
- Warzybok, A., Zokoll, M., Wardenga, N., Ozimek, E., Boboshko, M., and Kollmeier, B. (2015). Development of the Russian matrix sentence test. *Int. J. Audiology* 54 (November), 35–43. doi:10.3109/14992027.2015.1020969
- Weber, T., Lübeck, T., and Pörschmann, C. (2024). "Evaluating the influence of different generic head related transfer- functions on plausibility of binaural rendering," in *Fortschritte der Akustik - DAGA 2024* (Hannover: DEGA e.V. Berlin), 1–5.
- Werner, L. A., Fay, R. R., and Popper, A. N. (2012). "Human auditory development springer handbook of auditory research," New York, NY: Springer Springer Handbook of Auditory Research. doi:10.1007/978-1-4614-1421-6
- Westfall, P. H., Johnson, W. O., and Utts, J. M. (1997). A bayesian perspective on the bonferroni adjustment. *Biometrika* 84 (2), 419–427. doi:10.1093/biomet/84.2.419
- Williams, B. T., Viswanathan, N., and Brouwer, S. (2023). The effect of visual speech information on linguistic release from masking. *J. Acoust. Soc. Am.* 153 (1), 602–612. doi:10.1121/10.0016865
- World Medical Association (2013). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA. Revis. Ed.* 310 (20), 2191–2194. doi:10.1001/jama.2013.281053
- Yuan, Y., Lleo, Y., Daniel, R., and White, A. (2021). The impact of temporally coherent visual cues on speech perception in complex auditory environments. *Front. Neurosci.* 15 (June), 678029. doi:10.3389/frnins.2021.678029