

# Acoustic Realism of Clinical Speech-in-Noise Testing: Parameter Ranges of Speech-Likeness, Interaural Coherence, and Interaural Differences

Trends in Hearing

Volume 29: 1–11

© The Author(s) 2025

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/23312165251336625

journals.sagepub.com/home/tia



S Theo Goverts<sup>1</sup> , Virginia Best<sup>2</sup> , Julia Bouwmeester<sup>1</sup>, Cas Smits<sup>3</sup> and H Steven Colburn<sup>4</sup>

## Abstract

Speech-in-noise testing is a valuable component of audiological examination that can provide estimates of a listener's ability to communicate in their everyday life. It has long been recognized, however, that the acoustics of real-world environments are complex and variable and not well represented by a typical clinical test setup. The first aim of this study was to quantify real-world environments in terms of several acoustic parameters that may be relevant for speech understanding (namely speech-likeness, interaural coherence, and interaural time and level differences). Earlier acoustic analyses of binaural recordings in natural environments were extended to binaural re-creations of natural environments that included conversational speech embedded in recorded backgrounds and allowed a systematic manipulation of signal-to-noise ratio. The second aim of the study was to examine these same parameters in typical clinical speech-in-noise tests and consider the “acoustic realism” of such tests. We confirmed that the parameter spaces of natural environments are poorly covered by those of the most commonly used clinical test with one frontal loudspeaker. We also demonstrated that a simple variation of the clinical test, which uses two spatially separated loudspeakers to present speech and noise, leads to better coverage of the parameter spaces of natural environments. Overall, the results provide a framework for characterizing different listening environments that may guide future efforts to increase the real-world relevance of clinical speech-in-noise testing.

## Keywords

hearing assessment, ecological validity, acoustic realism, binaural recordings, natural environments

Received: May 5, 2024; revised: March 26, 2025; accepted: April 5, 2025

## Introduction

The goal of auditory rehabilitation is to improve functioning in everyday life for individuals who live with hearing loss. To diagnose hearing loss, and to prescribe and assess the success of auditory rehabilitation, clinicians often rely on behavioral tests of functional auditory abilities (such as speech understanding) using the available clinical infrastructure. It has long been recognized that clinical hearing tests are a simplification of real-world listening situations, and that this may explain why clinical measures sometimes do not show good correspondence with a patient's real-world experience (e.g., Badajoz-Davila & Buchholz, 2021; Cord et al., 2004; Gnewikow et al., 2009). To combat this disconnect, there is a lot of interest in developing clinical tests that are more relevant or have higher “ecological

validity” (e.g., Keidser et al., 2020). It should be noted that this term actually merges two related but distinct

<sup>1</sup>Department of Otolaryngology-Head and Neck Surgery, Section Ear & Hearing, Amsterdam Public Health Research Institute, Amsterdam UMC Location Vrije Universiteit, Amsterdam, the Netherlands

<sup>2</sup>Department of Speech, Language and Hearing Sciences, Boston University, Boston, MA, USA

<sup>3</sup>Department of Otolaryngology-Head and Neck Surgery, Amsterdam Public Health Research Institute, Amsterdam UMC Location University of Amsterdam, Amsterdam, the Netherlands

<sup>4</sup>Department of Biomedical Engineering, Boston University, Boston, MA, USA

### Corresponding Author:

S Theo Goverts, Department of Otolaryngology-Head and Neck Surgery, Audiological Center, Amsterdam, the Netherlands.

Email: st.goverts@amsterdamumc.nl



elements: realism and external validity (Beechey, 2022; Keidser et al., 2020, 2022). Realism refers to the similarity of relevant parameters in the test setting to those in natural situations, whereas external validity refers to the correspondence between the results of a certain test and performance in natural situations.

For speech-in-noise (SIN) testing, several broad domains have been identified that may modulate ecological validity, including the linguistic content of the speech stimuli (e.g., Best et al., 2016); the social context in which stimuli are presented (e.g., Carlile & Keidser, 2020); and the cognitive demands of a particular task (e.g., Lunner et al., 2020). In the current study we focus on a domain that we call “acoustic realism”, which motivates our own investigations and has received increasing attention in the literature in recent years. For example, several studies have aimed to measure the range of realistic signal-to-noise ratios (SNRs) that listeners encounter when communicating in their daily life (e.g., Mansour et al., 2021a; Smeds et al., 2015; Weisser & Buchholz, 2019; Wu et al., 2018) and have argued that SIN testing that targets these SNRs may lead to more relevant outcome measures. Other studies have investigated the consequences of introducing more realistic background noises into SIN tests (e.g., Best et al., 2015; Mansour et al., 2021b), considering factors such as amplitude modulation, informational masking, spatialization, and reverberation. These studies demonstrate that variations in realism can affect outcome measures and even the conclusions one might draw from SIN testing.

Goverts and Colburn (2020) approached acoustic realism by documenting the acoustics of natural environments. They obtained binaural recordings from the ears of human subjects actively immersed in eight everyday environments (Home, Restaurant, City Walk, City Talk, City Bike Ride, Station, Train, and Bus). From those recordings, they extracted a number of acoustic parameters that related to the spatial characteristics of the natural environments, including interaural coherence (IC), interaural time differences (ITDs), and interaural level differences (ILDs). This analysis showed that natural environments result in a wide range of interaural parameters as measured at the ears, but also that only a small proportion of measurements show high IC. This implies that highly reliable information in realistic scenarios is relatively sparse, and indeed there are many demonstrations in the literature of binaural mechanisms designed to take advantage of this fact (e.g., Dietz et al., 2013; Stecker et al., 2021). Goverts and Colburn also quantified “speech-likeness” (SpLi) at the left and right ears, using a modulation-spectrum-based metric for which high values indicate a high similarity to speech. This metric provides a convenient estimate of the presence of speech in a noisy signal, and is related to the SNR, but can be calculated non-intrusively (i.e., without access to the signal and the noise separately). SpLi results showed large variability within and between environments, and revealed prevalent

across-ear asymmetries, highlighting that there is often an acoustically better ear for speech reception that can change from moment to moment.

In the current study, our first goal was to further characterize the acoustic properties of real-world environments, especially those in which people typically communicate. The approach closely followed that of Goverts and Colburn (2020), but provided additional insights. For example, because the recordings of Goverts and Colburn (2020) were conducted in real-world settings, they contained speech and other sounds in rather uncontrolled combinations. In the current study, we used realistic simulations that afforded us much more control. Specifically, we chose an existing set of recordings of six everyday environments in which communication often takes place (Office, Church, Living Room, Café, Dinner Party, and Food Court), and combined them with conversational speech materials that are intended for those environments. As the waveforms for both environments and conversational sentences were available separately, it was possible to vary the SNR, and thus calculate SpLi, IC, ITD and ILD as a function of SNR.

The second goal of the study was to consider and quantify the acoustic realism of clinical SIN testing using the same analyses we applied to the real-world environments. It is important to point out that SIN testing is not consistently included as a part of regular clinical audiological care (e.g., Billings et al., 2024; Fitzgerald et al., 2023; Parmar et al., 2022; Vermiglio et al., 2019). When it is included, there are variations in its format across different clinics and countries. SIN testing may be performed over headphones to assess ear-specific impacts of hearing loss, or it may be performed using an external loudspeaker to assess a patient’s functional speech recognition capacity with or without hearing-assistive devices. In the latter case, a typical test setup uses co-located speech and noise stimuli played from a single loudspeaker. Arrangements using two or more loudspeakers are common in clinical research studies (e.g., Corbin et al., 2021; D’Onofrio et al., 2020; Griffin et al., 2019; Vermeire & Van de Heyning, 2009) and in certain clinical settings for sound localization testing (e.g., Ramirez et al., 2024) but to date such setups are not in widespread use for SIN testing in clinical care. Beyond the physical test setup, there are also many variations in the stimuli that may be considered for SIN testing. Target stimuli may be words, digit triplets, sentences, or stories (Griffin et al., 2020; Kollmeier et al., 2015; Nilsson et al., 1994; Plomp & Mimpen, 1979; Smits et al., 2013; Versfeld et al., 2000; Wilson et al., 2007; see also reviews in Billings et al., 2024 and Sanchez et al., 2022). The interfering signal is typically a stationary speech-spectrum noise, but may also be fluctuating noise, babble, reversed speech, or speech (e.g., Versfeld et al., 2021; Wilson et al., 2007). Most commonly, SIN performance is quantified according to an adaptive procedure that estimates the SNR where 50% of the speech items are recognized correctly (Levitt, 1971; Nilsson et al., 1994; Plomp & Mimpen,

1979). Variants of test procedure use different target points in the adaptive procedure, or use fixed SNRs (e.g., Dambha et al., 2022; Herbert et al., 2023).

In the current study we took one example of a clinical SIN test, which was the basic single-loudspeaker test used in Dutch audiology clinics (with the loudspeaker at 0° azimuth). This test uses sentences embedded in stationary speech shaped noise, and an adaptive procedure targeting 50%. We made binaural recordings using microphones placed in the ears of a listener completing this test, to facilitate a direct comparison with the real-world recordings and simulations described above. Our aim here was to quantify the parameter spaces for the clinical test and understand which parameters deviate most from the real-world environments. In addition, we wanted to understand how much these parameter spaces change (or how much acoustic realism may be improved) by simply adding a second loudspeaker to the setup. To that end, we also implemented and made recordings during a simple two-loudspeaker variation of the Dutch test with the speech and noise at  $\pm 45^\circ$  azimuth. We chose a two-loudspeaker version deliberately given that it is a highly feasible arrangement to set up in a typical clinic; most audiometric booths are large enough to house two loudspeakers in this configuration and have two audio channels readily available.

To address these goals, the approach of the present study was to:

1. Extend the analyses of SpLi, IC, ITD, and ILD from Goverts and Colburn (2020) to six realistic communication-focused environments and to binaural recordings of two clinical SIN tests;
2. Explore the effect of SNR on SpLi, IC, ITD, and ILD in the communication-focused environments;
3. Quantitatively compare parameter spaces of SpLi, IC, ITD, and ILD for the realistic environments and for the clinical SIN tests.

## Methods

### Binaural Stimuli

**Binaural Recordings of Natural Environments** (“*Binaural Recordings*”). Binaural recordings made in eight natural environments were used as described in Goverts and Colburn (2020). The purpose of that earlier study was to have a representative, though far from complete, set of environments that individuals with hearing loss are active in. The choice of the specific environments was inspired by interviews with specialists involved in research or in clinical care or who live with hearing loss. The environments included Home, Restaurant, City Walk, City Talk, City Bike Ride, Station, Train, and Bus. In-ear recordings were made using two commercially available microphones (Sound Professionals MS-TFB-2) and a digital recorder (Olympus linear PCM recorder, Model LS-11) at a 44.1-kHz sample rate. The microphones

were placed in the concha. Note that the subjects who wore the microphones in their ears were not just observers but actively participated in the environments.

**Binaural Re-Creations of Communication Situations** (“*Binaural Re-Creations*”). The binaural re-creations made use of six natural environments from the Ambisonic Recordings of Typical Environments (ARTE) database (Weisser et al., 2019). These environments included Office, Church, Living room, Café, Dinner Party, and Food Court. The ARTE database contains 64-channel recordings in the environments that were converted for binaural playback as described by Weisser et al. (2019). For each environment, those recordings can be combined with “clean” sentences extracted from conversations held by real talkers conversing while listening to the recordings over headphones (see Miles et al., 2020). We followed the approach of Miles et al. (2022), which was to spatialize the sentences using head-related transfer functions measured in each environment using an acoustic manikin and a source at a distance of 1 m directly in front. A unique realistic SNR was chosen for each environment (−6.6, −3.0, −1.3, 1.6, 3.0, 6.2 dB for the Office, Church, Living Room, Café, Dinner Party, and Food Court, respectively) based on the estimates made by Weisser and Buchholz (2019). The analyses presented in the current study made use of 4 lists of 16 target sentences spoken by a single female talker (see Miles et al., 2020). Furthermore, for each environment, mixtures were generated with SNRs varying from −30 dB through +30 dB in 5 dB steps. We note that the binaural stimuli in the condition with SNR of −30 dB are dominated by the background sounds, whereas the binaural stimuli in the condition with SNR of +30 dB are dominated by the conversational speech.

**Binaural Recordings of “Clinical SIN Tests”.** Using the same setup described above, binaural recordings were made in an audiological clinic at Amsterdam UMC during the administration of two SIN tests using the stimuli of Versfeld et al. (2000). The first was the commonly used test in which the target speech (Dutch sentences spoken by a female talker) and the steady state masking noise (female Long-Term Average Speech Shaped noise; LTASS) are co-located, presented via one loudspeaker at approximately 1.5 m in front of the listener. This test is referred to as the “clinical SIN test.” The second test was a two-loudspeaker variation based on the test described by Boymans et al. (2009). The test uses two loudspeakers positioned at  $\pm 45^\circ$  at 1.5 m distance from the listener, which present the target speech and masking noise. In our variation, two adaptive tracks were interleaved. One corresponded to a female talker on the right presented with steady state male LTASS noise on the left, and the other corresponded to a male talker on the left presented with steady state female LTASS noise on the right. This test is referred to as the “spatialized clinical SIN test.” A volunteer with normal hearing completed two runs of each of these two tests. Initially, the regular adaptive procedure was followed, with the listener

providing verbal responses and all responses (correct/incorrect) were noted. The actual recording was made in the second run, where we replicated this procedure, but without the listener responding, to avoid the listener's own voice appearing in the recordings and influencing the acoustic analyses. The listener was instructed to make no head movements, though the head was not fixed.

### Calculation of Acoustic Parameters

SpLi was calculated using the approach of Goverts and Colburn (2020), and further details can be found there. Briefly, signals were processed into non-overlapping 1-second time-slices (with 0.5-seconds cosine-shaped rise and fall times) and bandpass filters were applied resulting in six, one-octave frequency bands centered at 250, 500, 1,000, 2,000, 4,000, and 8,000 Hz. For each frequency band the envelope was extracted and the discrete Fourier transform of this envelope was calculated. The modulation energy for each frequency band was determined in each of the six one-octave modulation bands centered at 4, 8, 16, 32, 64, and 128 Hz. Then for each spectral band the speech-to-reverberation modulation energy ratio (SRMR) was calculated as the ratio between the sum of energies present at low modulation frequencies and the sum of energies present at high modulation frequencies. This modulation-strength measure was developed and used by Falk and colleagues (e.g., Cosentino et al., 2014; Falk et al., 2010) in their studies of speech quality and intelligibility. It is based on the assumption that low modulation frequencies are related primarily to speech, and high modulation frequencies are related primarily to noise or reverberation. Goverts and Colburn (2020) adapted this measure to estimate SpLi in dB,  $\text{SpLi}(f_i) = 10 \times \log_{10}(\text{SRMR}(f_i))$ , and then calculated a broadband SpLi using a speech-intelligibility-index-weighted sum of the  $\text{SpLi}(f_i)$  values. The resulting SpLi values vary from about -2 dB for a stationary masker to 10 dB for natural clean speech and 12.5 dB for well-pronounced clean recorded speech materials used in SIN tests (Goverts & Colburn, 2020). SpLi was calculated for both left and right ears.

The interaural parameters were also calculated using the approach of Goverts and Colburn (2020), and further details can be found there. The left and right signals were filtered through a bank of quarter-octave-bandwidth filters, and selected frequencies were analyzed with respect to their interaural differences. In the results reported here, following Goverts and Colburn (2020), one low-frequency band (centered at 500 Hz) and one higher frequency band (centered at 2,000 Hz) were chosen. For the band at 2,000 Hz, only the interaural information contained within the envelopes of the 2,000-Hz waveforms was examined. In all cases, the bilateral signals were analyzed in 300-ms time-slices including with 5-ms, cosine-shaped rise and fall times. Successive analysis windows were shifted by 20-ms and thus partially overlapping. The window duration was somewhat arbitrary, but

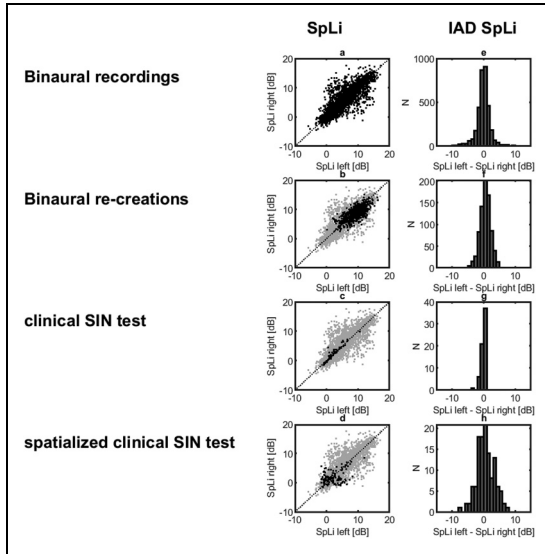
followed that used in Goverts and Colburn (2020) and was broadly in line with window sizes used in previous studies that have considered spectrotemporal “glimpses” in speech (e.g., Conroy et al., 2020) or have estimated binaural integration windows under different conditions (e.g., Dietz et al., 2013; Eurich & Dietz, 2023; Reed et al., 2016).

For each time window, we computed the cross-correlation function using the left and right waveforms. The cross-correlation was normalized by the product of the root-mean-square amplitudes, giving a maximum value of unity for waveforms that differ only in a pure delay and/or a fixed scale factor. The IC was defined as the maximum normalized correlation. The ITD was defined as the time shift for which the cross-correlation function reached its maximum value. Windows with an ITD larger than 1 ms were not included in the ITD distributions. Moreover, if the IC for a given window was less than 0.5, then that window was not included in the ITD distributions. The ILD was defined as the difference in dB between the two ears for each window. Whereas Goverts and Colburn (2020) analyzed IC, ITD and ILD in both 500 and 2,000 Hz bands, in the current study we considered only the IC and ITD at 500 Hz, and the ILD at 2,000 Hz, according to the frequency range generally considered to be most relevant for each kind of cue (e.g., Hartmann, 2021).

## Results

### Parameter Spaces of Speech-Likeness

Figure 1 displays the parameter spaces and left-right asymmetries for SpLi. The rows present results for the binaural recordings in the eight natural environments (panels a, e), the binaural re-creations at their realistic SNRs (panels b, f), the clinical SIN test (panels c, g), and the spatialized clinical SIN test (panels d, h). In the first column SpLi values for the right ear are plotted against SpLi values for the left ear (in black). For reference, the parameter spaces for the binaural recordings of panel a are replotted in gray in the other panels. For the binaural recordings, the SpLi values range from -6 to 18 dB. For the binaural re-creations, the distribution of SpLi values was similar but with a slightly reduced range (0–16 dB); the absence of extremely low values in the binaural re-creations is because speech was always present and at a sufficiently high SNR to support communication. The clinical SIN test shows a restricted distribution of SpLi values that is weighted towards low values (-1 to 11 dB) and includes very few asymmetries. The spatialized clinical SIN test has a similar range (-2 to 12 dB) but shows more asymmetry as a result of the laterally placed targets and maskers. The overall low values of SpLi observed in both versions of the clinical test reflect the dominance of negative SNRs in the adaptive tracking procedure, as is typically observed. The second column of Figure 1 displays histograms of the left-right ear asymmetries in SpLi. This representation shows



**Figure 1.** Parameter Spaces and Left-Right Asymmetries for SpLi. The Rows Present Analyses of Binaural Recordings in Eight Natural Environments (Panels a, e), in the Binaural Re-creations (Panels b, f), During the Clinical SIN Test (Panels c, g) and During the Spatialized Clinical SIN Test (Panels d, h). The First Column Presents Parameter Spaces for SpLi. In Each Panel, SpLi Values for the Right Ear are Plotted Against SpLi Values for the Left Ear. The Distribution from Panel a is Plotted in Gray in the Other Panels to Facilitate Comparisons to the Earlier Study. The Second Column Presents Interaural SpLi Differences to Illustrate the Amount of Asymmetry.

that the distribution of asymmetries is comparable across the binaural recordings and the binaural re-creations. The clinical SIN test has a very tight distribution close to zero, but the spatialized version has a broader distribution comparable to that of the natural environments. Statistical testing ( $F$  tests with Bonferroni correction) showed significant differences between all distributions ( $P < .01$ ).

### Parameter Spaces of Interaural Coherence, Interaural Time Differences, and Interaural Level Differences

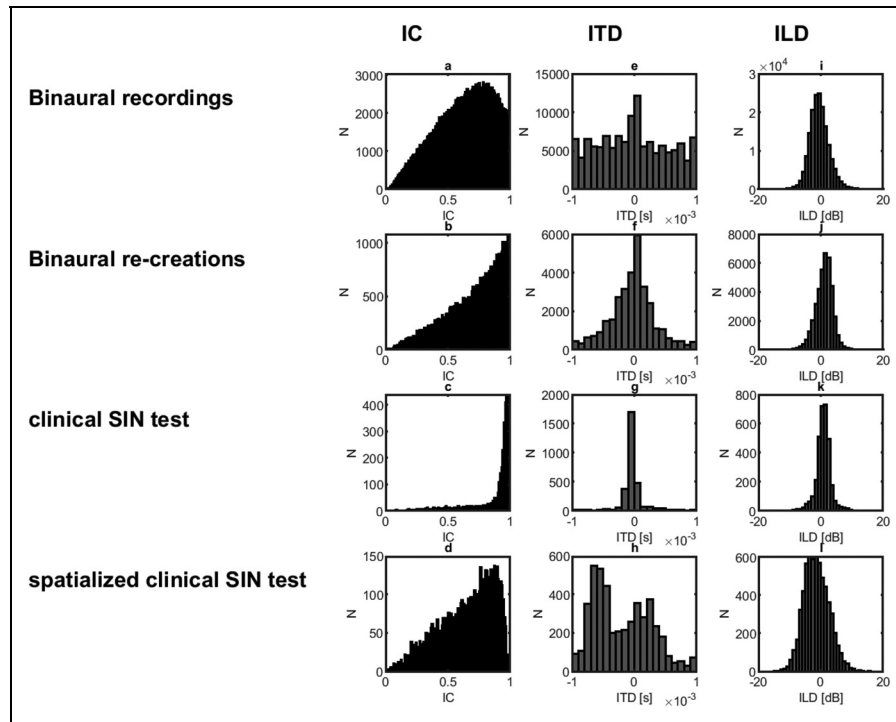
Figure 2 displays the parameter spaces for the IC at 500 Hz, ITD at 500 Hz, and ILD 2,000 Hz in the first, second and third column, respectively. Comparing panels a and b, it appears that the binaural recordings, which were collected during a variety of activities, have a stronger representation of low IC values than the binaural re-creations. The effect of the frontal target talker at a relatively close distance in the binaural re-creations gives a shift towards high IC values, although the binaural recordings include segments with the microphone wearer's own voice, which likely explains the strong representation of IC values at 1 (Pohlhausen et al., 2022). The clinical SIN test shows a highly skewed distribution of IC values near unity correlation (panel c), whereas the distribution for the spatialized clinical SIN test (panel d) has a broader distribution that much more closely matches that of

the natural environments. Statistical testing ( $F$  tests with Bonferroni correction) showed significant differences between all distributions ( $P < .01$ ), except for the difference between the binaural re-creations and the spatialized clinical test ( $P = .24$ ). The second column of Figure 2 displays the parameter spaces for the ITD at 500 Hz. Comparing panels e and f, it appears that the binaural recordings have a stronger representation of extreme ITDs than the binaural re-creations. The effect of the frontal target talker at a relatively close distance in the binaural re-creations is visible in stronger representation of ITD values around 0 ms. The clinical SIN test shows an even stronger representation of ITD values near 0 ms (panel g), whereas the distribution for the spatialized clinical SIN test (panel h) has a range of values that more closely matches that of the natural environments. Statistical testing ( $F$  tests with Bonferroni correction) showed significant differences between all distributions ( $P < .01$ ), except for the difference between the binaural re-creations and the spatialized clinical test ( $P = .013$ ).

The third column of Figure 2 displays the parameter spaces for the ILD at 2,000 Hz. Comparing panels i and j, it appears that the different sets of binaural recordings have a similar distribution of ILD. The effect of the frontal target talker at a relatively close distance in the binaural re-creations is partly visible in a more centrally weighted representation. The clinical SIN test also shows a representation of ILD values that is focused around 0 dB (panel k), whereas the distribution for the spatialized clinical SIN test (panel l) has a broader distribution that is even broader than that of the natural recordings and the binaural re-creations. Statistical testing ( $F$  tests with Bonferroni correction) showed significant differences between all distributions ( $P < .01$ ).

### Parameter Variations as a Function of SNR

Figure 3 presents several key interaural measures as a function of SNR for the binaural re-creations. For comparison, on the right side of each panel, values for the binaural recordings and the two clinical SIN tests are presented. Figure 3a shows interquartile ranges of the interaural SpLi differences (IQR-ISD), capturing the width of the distributions shown in Figure 1e–h. There seems to be no general pattern of dependence of this parameter on SNR that holds across all environments. For most of the binaural recordings, the IQR-ISD agrees well with that of the binaural re-creations at their ecological SNRs, although the City Walk and City Talk show rather high values, consistent with the fact that these are environments with dominant talkers to one side. The clinical SIN test shows a very reduced IQR-ISD, whereas the spatialized version produces rather high IQR-ISD values, in line with the more asymmetric binaural recordings. Figure 3b shows median values for IC, capturing the central tendency of the distributions shown in Figure 2a–d. The median IC shows a strong dependence on SNR for all environments. At the low-SNR end, the differences across environments reflect



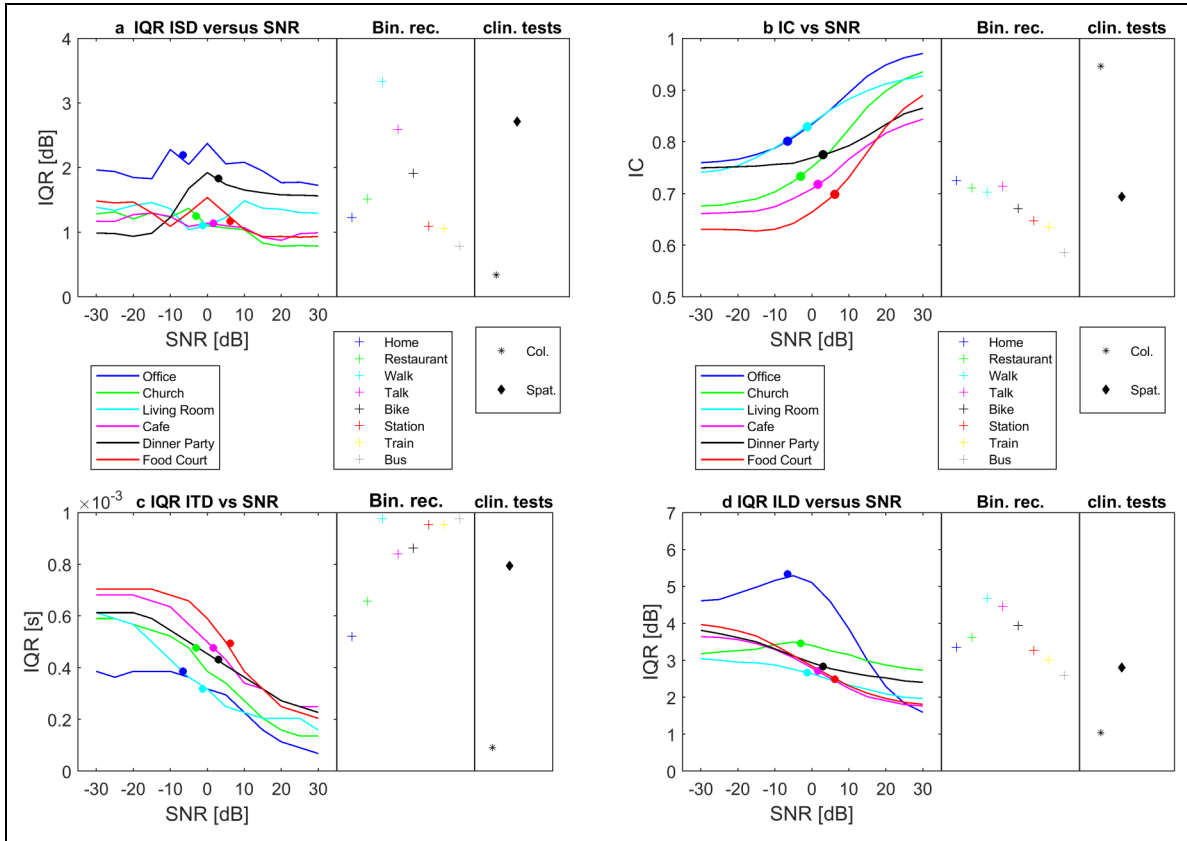
**Figure 2.** Parameter Spaces for the IC at 500 Hz, ITD at 500 Hz and ILD at 2,000 Hz. The Rows Present Analyses of Binaural Recordings in Eight Natural Environments (Panels a, e, i), in the Binaural Re-creations (Panels b, f, j), During the Clinical SIN Test (Panels c, g, k) and During the Spatialized Clinical SIN Test (Panels d, h, l). Histograms of IC Values, ITD Values, and ILD Values are Presented in the First, Second, and Third Column, Respectively.

differences in the competing low-frequency sounds, their spatial configuration and the reverberation present. At the high-SNR end, the differences across environments reflect differences in how the acoustics of the room or environment impact the target talker. The highest median IC values, over the range of SNRs, are found for the Office and Living Room, whereas the lowest values are found for the Food Court and Café. For most of the binaural recordings the median IC agrees well with that of the binaural re-creations at their ecological SNRs, although the Station, Train, and Bus recordings show relatively low values. This reflects the fact that these are noisy environments in which communication was typically not occurring and there was no one dominant target source. The clinical SIN test shows very high IC values, corresponding to the binaural re-creations at +30 dB, whereas the spatialized version has a much-reduced median IC that matches well the IC for the binaural recordings and the binaural re-creations at their ecological SNRs. Figure 3c shows interquartile ranges of ITDs (IQR-ITD), capturing the width of the distributions shown in Figure 2e–h. The IQR-ITD shows a strong dependence on SNR for all environments. At the low-SNR end, the differences across environments reflect differences in the competing low-frequency sounds, their spatial configuration and the reverberation present. For most of the binaural recordings the IQR-ITDs are higher than those of the binaural re-creations, reflecting the fact that they are not always dominated by a centrally located talker. The clinical

SIN test shows a very low IQR-ITD, similar to the binaural re-creations at +30 dB, whereas the spatialized version has an IQR-ITD that agrees rather well with the binaural recordings. Figure 3d shows interquartile ranges of ILDs (IQR-ILD), capturing the width of the distributions shown in Figure 2i–l. The IQR-ILD does not depend strongly on SNR. For most of the binaural recordings, the IQR-ILD agrees well with that of the binaural re-creations. Again, the City Walk and City Talk show relatively high IQR-ILDs consistent with strongly asymmetric sources. The clinical SIN test shows a very low IQR-ILD compared to that of the binaural recordings and binaural re-creations, even lower than the binaural re-creations at +30 dB SNR. The spatialized version has an IQR-ILD that agrees rather well with that of the binaural recordings and the binaural re-creations at their ecological SNRs.

## Discussion

The first major goal of the current study was to characterize the acoustic properties of real-world environments in which people typically communicate, with an emphasis on those properties that are relevant for binaural hearing. To do so we extended the results of Goverts and Colburn (2020), who analyzed a broad range of everyday environments, to binaural re-creations of natural environments that included conversational speech embedded in real-world backgrounds. Applying a similar



**Figure 3.** Key Interaural Measures as a Function of SNR for the Binaural Re-creations. For each environment, the value of the Interaural Measures at its specific ecological SNR is indicated by a colored dot. For Comparison, on the Right Side of Each Panel, Values for the Binaural Recordings and the Two Clinical SIN Tests are Presented. Shown are Interquartile Ranges of the Interaural SpLi Differences (IQR-ISD; Panel a), Median Values of IC (Panel b), Interquartile Ranges of ITDs (IQR-ITD; Panel c), and Interquartile Ranges of ILDs (IQR-ILD, panel d).

analysis to the new binaural re-creations produced broadly consistent results, but some key differences were noted. In general, the range of SpLi values was smaller, and the range of interaural SpLi differences was smaller. Furthermore, the IC distribution for the re-creations had a greater fraction of higher values and the ITD and ILD ranges were smaller. These differences can largely be understood by the specific origin of the two sets of binaural stimuli. The binaural recordings cover a more diverse range of situations, including some that are not dominated by speech. Furthermore, in the binaural recordings the dominant source could move relative to the subject, leading to more/larger asymmetries than for the binaural re-creations, in which the dominant speech source was always close to and directly in front of the subject. Considering all of the binaural stimuli, everyday life can be considered as a condition with highly variable amounts of speech information, consistent with previous studies that estimated a wide range of SNRs in realistic environments (e.g., Mansour et al., 2021a; Smeds et al., 2015; Weisser & Buchholz, 2019; Wu et al., 2018). Our results also emphasize that everyday environments can be highly asymmetric, show interaural coherences varying from perfect (unity) to diffuse, and have binaural properties spanning a wide range of values.

One key advantage of the binaural re-creations over the real-world recordings is that they make it possible to directly explore the impact of SNR on the parameter spaces of communication-focused scenarios. Our data show that while many of the parameters we examined are sensitive to SNR, there are also differences in the parameters across environments for a given SNR. For example, the IC varies quite dramatically across environments, at a given SNR. The Office represents an environment with very little reverberation and showed a consistently high IC, suggesting that it offers good opportunities for extracting spatial information. The other environments are more reverberant and influence the IC in different ways depending on the size and characteristics of the room. For example, the Food Court is a large room with highly diffuse late reverberation that leads to low IC values particularly at low SNRs where the background sounds dominate. On the other hand, the Café is a small room with strong early reflections that interact with the target and reduce the IC even at the high-SNR end.

The second major goal of the current study was to measure the parameter spaces of clinical SIN tests in relation to realistic environments, to identify and quantify specific ways in which the clinical tests lack acoustic realism. For this



comparison we considered both a typical single-loudspeaker test, which was expected to show very poor acoustic realism, but also explored how the picture changes for a simple variation on the clinical test, which uses two separated loudspeakers to present the speech and noise. We found that the parameter spaces of everyday environments are rather poorly covered by that of the clinical SIN test. The clinical SIN test includes only low SpLi values, which are only observed in the binaural re-creations for unnaturally low SNRs, or in the natural recordings that are not intended for speech communication. The clinical SIN test also displays very little asymmetry in SpLi, relatively limited variation in ITD and ILD, and very high IC values. Broadly speaking, the starkly different parameter spaces we observed for the clinical SIN test relative to the natural environments are not surprising and are well explained by the strong symmetry of the single-loudspeaker test. A concern for clinical practice is that this configuration produces performance measures that mainly reflect the capacities of the patient's better ear and will not reveal effects related to the poorer ear. This makes it difficult to assess symmetric losses, and may lead to over-estimates of real-world performance for this population (who may have to rely on their poorer ear in certain real-world situations). Moreover, the single-loudspeaker configuration does not require the utilization of binaural differences, and thus cannot reveal any specific deficits related to binaural processing (Bronkhorst & Plomp, 1990). Finally, the unnaturally high IC values in the clinical test avoid any assessment of the processing of stimuli with low IC. Thus, the clinical test may over-estimate how listeners will cope in real-world situations with multiple interferers and strong reverberation.

Encouragingly, we have demonstrated that it is relatively simple with one additional loudspeaker to introduce asymmetry, reduce IC, increase the range of ITDs and ILDs, and arrive at more realistic parameter spaces. The parameter spaces we observed for the spatialized clinical SIN test were generally well-matched to the parameter spaces of the realistic environments. Of course, other spatial configurations of a two-loudspeaker test will also lead to an improved coverage of the parameter spaces relative to the co-located clinical test, as will larger loudspeaker arrays. Moreover, there are certain characteristics of real-world environments that can only be captured with a larger number of sources. For example, a "symmetric" three-talker configuration is very commonly used to minimize acoustic effects related to the head shadow and focus instead on binaural processing (Cameron et al., 2009; Marrone et al., 2008; Srinivasan et al., 2020). Configurations with three or more talkers are also believed to maximize "informational masking" (Freyman et al., 2004; Simpson & Cooke, 2005) and thus capture the ability of listeners to overcome this particularly disruptive form of interference. However, the added value in terms of acoustic realism of adding a second loudspeaker to a one-loudspeaker setup is likely to be greater than going from two to three (or more) loudspeakers. From a practical perspective, two audio

channels and two loudspeakers are relatively common in clinical settings and thus this configuration would be very feasible to implement widely in clinics (or countries) not currently doing so. It would be useful to make more extensive recordings of clinical tests with different loudspeaker configurations, materials, and hearing-impaired participants, to fully explore the range of parameter spaces that can be achieved.

Finally, it should be noted that, while we have explored just a few acoustic parameters, and focused on those relevant to binaural hearing, there are many more parameters that could be taken into account in the search for acoustic realism in clinical SIN testing (e.g., temporal dynamics and spectral characteristics of the background, variations in loudness, etc.). Furthermore, there are many non-acoustic factors that may affect the realism of a speech communication situation and even interact with the factors we have explored (e.g., visual information, linguistic content; see e.g., Keidser et al., 2020; Wu et al., 2018). An important next step is to determine the extent to which different deviations from realism have a meaningful impact on outcomes and on our ability to estimate an individual's functional abilities in the real world.


## Conclusions


Everyday situations in which people communicate vary widely in their binaural properties and are often highly asymmetric. These properties are not captured well by commonly used clinical speech tests that are limited to a single loudspeaker. Simple and clinically feasible test setups using as few as two loudspeakers can provide a much better coverage of the acoustic parameters encountered in daily life, which may be important for the assessment of functional SIN abilities, particularly in patients with asymmetric losses or binaural processing deficits.

## Acknowledgements

The authors are grateful to Jörg Buchholz for sharing materials and code for the binaural re-creations, and for many helpful and inspiring discussions about this topic.

## ORCID iD

Theo Goverts  <https://orcid.org/0000-0002-6887-5909>

Virginia Best  <https://orcid.org/0000-0002-5535-5736>

Cas Smits  <https://orcid.org/0000-0003-1617-1253>

## Statements and Declarations

### Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: VB was supported by the National Institutes of Health (Award No. DC015760).



### Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Data Availability

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

### References

- Badajoz-Davila, J., & Buchholz, J. M. (2021). Effect of test realism on speech-in-noise outcomes in bilateral cochlear implant users. *Ear Hear*, 42(6), 1687–1698. <https://doi.org/10.1097/AUD.0000000000001061>
- Beechey, T. (2022). Ecological validity, external validity, and mundane realism in hearing science. *Ear and Hearing*, 43(5), 1395–1401. <https://doi.org/10.1097/AUD.0000000000001202>
- Best, V., Keidser, G., Buchholz, J. M., & Freeston, K. (2015). An examination of speech reception thresholds measured in a simulated reverberant cafeteria environment. *International Journal of Audiology*, 54(10), 682–690. <https://doi.org/10.3109/14992027.2015.1028656>
- Best, V., Keidser, G., Freeston, K., & Buchholz, J. M. (2016). A dynamic speech comprehension test for assessing real-world listening ability. *Journal of the American Academy of Audiology*, 27(7), 515–526. <https://doi.org/10.3766/jaaa.15089>
- Billings, C. J., Olsen, T. M., Charney, L., Madsen, B. M., & Holmes, C. E. (2024). Speech-in-noise testing: An Introduction for audiologists. *Seminars in Hearing*, 45(1), 55–82. <https://doi.org/10.1055/s-0043-1770155>
- Boymans, M., Goverts, S. T., Kramer, S. E., Festen, J. M., & Dreschler, W. A. (2009). Candidacy for bilateral hearing aids: A retrospective multicenter study. *Journal of Speech Language and Hearing Research*, 52(1), 130–140. [https://doi.org/10.1044/1092-4388\(2008/07-0120](https://doi.org/10.1044/1092-4388(2008/07-0120)
- Bronkhorst, A. W., & Plomp, R. (1990). A clinical test for the assessment of binaural speech perception in noise. *International Journal of Audiology*, 29(5), 275–285. <https://doi.org/10.3109/00206099009072858>
- Cameron, S., Brown, D., Keith, R., Martin, J., Watson, C., & Dillon, H. (2009). Development of the north American listening in spatialized noise-sentences test (NA LiSN-S): Sentence equivalence, normative data, and test-retest reliability studies. *Journal of the American Academy of Audiology*, 20(2), 128–146. <https://doi.org/10.3766/jaaa.20.2.6>
- Carlile, S., & Keidser, G. (2020). Conversational interaction is the brain in action: Implications for the evaluation of hearing and hearing interventions. *Ear and Hearing*, 41(Suppl. 1), 56S–67S. <https://doi.org/10.1097/AUD.0000000000000939>
- Conroy, C., Best, V., Jennings, T. R., & Kidd, G. (2020). The importance of processing resolution in “ideal time-frequency segregation” of masked speech and the implications for predicting speech intelligibility. *Journal of the Acoustical Society of America*, 147(3), 1648–1660. <https://doi.org/10.1121/10.0000893>
- Corbin, N. E., Buss, E., & Leibold, L. J. (2021). Spatial hearing and functional auditory skills in children with unilateral hearing loss. *Journal of Speech, Language, and Hearing Research*, 64(11), 4495–4512. [https://doi.org/10.1044/2021\\_JSLHR-20-00081](https://doi.org/10.1044/2021_JSLHR-20-00081)
- Cord, M. T., Surr, R. K., Walden, B. E., & Dyrland, O. (2004). Relationship between laboratory measures of directional advantage and everyday success with directional microphone hearing aids. *Journal of the American Academy of Audiology*, 15(5), 353–364. <https://doi.org/10.3766/jaaa.15.5.3>
- Cosentino, S., Marquardt, T., McAlpine, D., Culling, J., & Falk, T. H. (2014). A model that predicts the binaural advantage to speech intelligibility from the mixed target and interferer signals. *Journal of the Acoustical Society of America*, 135(2), 796–807. <https://doi.org/10.1121/1.4861239>
- Dambha, T., Swanepoel, W., Mahomed-Asmail, F., De Sousa, K. C., Graham, M. A., & Smits, C. (2022). Improving the efficiency of the digits-in-noise hearing screening test: A comparison between four different test procedures. *Journal of Speech, Language, and Hearing Research*, 65(1), 378–391. [https://doi.org/10.1044/2021\\_JSLHR-21-00159](https://doi.org/10.1044/2021_JSLHR-21-00159)
- Dietz, M., Marquardt, T., Salminen, N. H., & McAlpine, D. (2013). Emphasis of spatial cues in the temporal fine structure during the rising segments of amplitude-modulated sounds. *Proceedings of the National Academy of Sciences*, 110(37), 15151–15156. <https://doi.org/10.1073/pnas.1309712110>
- D’Onofrio, K., Richards, V., & Gifford, R. (2020). Spatial release from informational and energetic masking in bimodal and bilateral cochlear implant users. *Journal of Speech, Language, and Hearing Research*, 63(11), 3816–3833. [https://doi.org/10.1044/2020\\_JSLHR-20-00044](https://doi.org/10.1044/2020_JSLHR-20-00044)
- Eurich, B., & Dietz, M. (2023). Fast binaural processing but sluggish masker representation reconfiguration. *Journal of the Acoustical Society of America*, 154(3), 1862–1870. <https://doi.org/10.1121/10.0021072>
- Falk, T. H., Zheng, C., & Chan, W.-Y. (2010). A non-intrusive quality and intelligibility measure of reverberant and dereverberated speech. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(7), 1766–1774. <https://doi.org/10.1109/TASL.2010.2052247>
- Fitzgerald, M. B., Gianakas, S. P., Qian, Z. J., Losorelli, S., & Swanson, A. C. (2023). Preliminary guidelines for replacing word-recognition in quiet with speech in noise assessment in the routine audiologic test battery. *Ear and Hearing*, 44(6), 1548–1561. <https://doi.org/10.1097/AUD.0000000000001409>
- Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *Journal of the Acoustical Society of America*, 115(5 Pt 1), 2246–2256. <https://doi.org/10.1121/1.1689343>
- Gnewikow, D., Ricketts, T., Bratt, G. W., & Mutchler, L. C. (2009). Real-world benefit from directional microphone hearing aids. *The Journal of Rehabilitation Research and Development*, 46(5), 603–618. <https://doi.org/10.1682/jrrd.2007.03.0052>

- Goverts, S. T., & Colburn, H. S. (2020). Binaural recordings in natural acoustic environments: Estimates of speech-likeness and interaural parameters. *Trends in Hearing*, 24. <https://doi.org/10.1177/2331216520972858>
- Griffin, A. M., Poissant, S. F., & Freyman, R. L. (2019). Speech-in-noise and quality-of-life measures in school-aged children with normal hearing and with unilateral hearing loss. *Ear and Hearing*, 40(4), 887–904. <https://doi.org/10.1097/AUD.0000000000000667>
- Griffin, A. M., Poissant, S. F., & Freyman, R. L. (2020). Auditory comprehension in school-aged children with normal hearing and with unilateral hearing loss. *Language, Speech, and Hearing Services in Schools*, 51(1), 29–41. [https://doi.org/10.1044/2019\\_LSHSS-0CHL-19-0020](https://doi.org/10.1044/2019_LSHSS-0CHL-19-0020)
- Hartmann, W. M. (2021). Localization and lateralization of sound. In R. Y. Litovsky, M. J. Goupell, R. R. Fay, & A. N. Popper (Eds.), *Binaural hearing* (pp. 9–45). Springer.
- Herbert, N., Keller, M., Derleth, P., Kühnel, V., & Strelcyk, O. (2023). Optimised adaptive procedures and analysis methods for conducting speech-in-noise tests. *International Journal of Audiology*, 62(8), 776–786. <https://doi.org/10.1080/14992027.2022.2087112>.
- Keidser, G., Naylor, G., Brungart, D. S., Caduff, A., Campos, J., Carlile, S., Carpenter, M. G., Grimm, G., Hohmann, V., Holube, I., Launer, S., Lunner, T., Mehra, R., Rapport, F., Slaney, M., & Smeds, K. (2020). The quest for ecological validity in hearing science: What it is, why it matters, and how to advance it. *Ear and Hearing*, 41(Suppl. 1), 5S–19S. <https://doi.org/10.1097/AUD.0000000000000944>
- Keidser, G., Naylor, G., Brungart, D. S., Caduff, A., Campos, J., Carlile, S., Carpenter, M. G., Grimm, G., Hohmann, V., Holube, I., Launer, S., Lunner, T., Mehra, R., Rapport, F., Slaney, M., & Smeds, K. (2022). Comment on the point of view “ecological validity, external validity and mundane realism in hearing science”. *Ear and Hearing*, 43(5), 1601–1602. <https://doi.org/10.1097/AUD.0000000000001241>
- Kollmeier, B., Warzybok, A., Hochmuth, S., Zokoll, M. A., Uslar, V., Brand, T., & Wagener, K. C. (2015). The multilingual matrix test: Principles, applications, and comparison across languages: A review. *International Journal of Audiology*, 54(Suppl. 2), 3–16. <https://doi.org/10.3109/14992027.2015.1020971>
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, 49(2), 467–477. <https://doi.org/10.1121/1.1912375>
- Lunner, T., Alickovic, E., Graversen, C., Ng, E. H. N., Wendt, D., & Keidser, G. (2020). Three new outcome measures that tap into cognitive processes required for real-life communication. *Ear and Hearing*, 41(Suppl. 1), 39S–47S. <https://doi.org/10.1097/AUD.0000000000000941>
- Mansour, N., Marschall, M., May, T., Westermann, A., & Dau, T. (2021a). A method for realistic, conversational signal-to-noise ratio estimation. *Journal of the Acoustical Society of America*, 149(3), 1559–1566. <https://doi.org/10.1121/10.0003626>
- Mansour, N., Marschall, M., May, T., Westermann, A., & Dau, T. (2021b). Speech intelligibility in a realistic virtual sound environment. *Journal of the Acoustical Society of America*, 149(4), 2791–2801. <https://doi.org/10.1121/10.0004779>
- Marrone, N., Mason, C. R., & Kidd, G. Jr. (2008). The effects of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms. *Journal of the Acoustical Society of America*, 124(5), 3064–3075. <https://doi.org/10.1121/1.2980441>
- Miles, K. M., Beechey, T., Best, V., & Buchholz, J. M. (2022). Measuring speech intelligibility and hearing-aid benefit using everyday conversational sentences in real-world environments. *Frontiers in Neuroscience*, 16. <https://doi.org/10.3389/fnins.2022.789565>
- Miles, K. M., Keidser, G., Freeston, K., Beechey, T., Best, V., & Buchholz, J. M. (2020). Development of the everyday conversational sentences in noise test. *Journal of the Acoustical Society of America*, 147(3), 1562–1576. <https://doi.org/10.1121/10.0000780>
- Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America*, 95(2), 1085–1099. <https://doi.org/10.1121/1.408469>
- Parmar, B. J., Rajasingam, S. L., Bizley, J. K., & Vickers, D. A. (2022). Factors affecting the use of speech testing in adult audiology. *American Journal of Audiology*, 31(3), 528–540. [https://doi.org/10.1044/2022\\_AJA-21-00233](https://doi.org/10.1044/2022_AJA-21-00233)
- Plomp, R., & Mimpen, A. M. (1979). Improving the reliability of testing the speech reception threshold for sentences. *International Journal of Audiology*, 18(1), 43–52. <https://doi.org/10.3109/00206097909072618>
- Pohlhausen, J., Holube, I., & Bitzer, J. (2022). Near-ear sound pressure level distribution in everyday life considering the user's own voice and privacy. *Acta Acustica*, 6, 11. Article 40. <https://doi.org/10.1051/aacus/2022035>
- Ramirez, M., Arend, J. M., von Gablenz, P., Liesefeld, H. R., & Pörschmann, C. (2024). Toward sound localization testing in virtual reality to aid in the screening of auditory processing disorders. *Trends in Hearing*, 28. <https://doi.org/10.1177/23312165241235463>
- Reed, D. K., Dietz, M., Josupeit, A., & van de Par, S. (2016). Lateralization of stimuli with alternating interaural time differences: The role of monaural envelope cues. *Journal of the Acoustical Society of America*, 139(1), 30–40. <https://doi.org/10.1121/1.4938018>
- Sanchez, V. A., Arnold, M. L., Moore, D. R., Clavier, O., & Abrams, H. B. (2022). Speech-in-noise testing: Innovative applications for pediatric patients, underrepresented populations, fitness for duty, clinical trials, and remote services. *Journal of the Acoustical Society of America*, 152(4), 2336–2356. <https://doi.org/10.1121/10.0014418>
- Simpson, S. A., & Cooke, M. (2005). Consonant identification in N-talker babble is a nonmonotonic function of N. *Journal of the Acoustical Society of America*, 118(5), 2775–2778. <https://doi.org/10.1121/1.2062650>
- Smeds, K., Wolters, F., & Rung, M. (2015). Estimation of signal-to-noise ratios in realistic sound scenarios. *Journal of the American Academy of Audiology*, 26(2), 183–196. <https://doi.org/10.3766/jaaa.26.2.7>

- Smits, C., Goverts, S. T., & Festen, J. M. (2013). The digits-in-noise test: Assessing auditory speech recognition abilities in noise. *Journal of the Acoustical Society of America*, 133(3), 1693–1706. <https://doi.org/10.1121/1.4789933>
- Srinivasan, N. K., Holtz, A., & Gallun, F. J. (2020). Comparing spatial release from masking using traditional methods and portable automated rapid testing iPad app. *American Journal of Audiology*, 29(4), 907–915. <https://doi.org/10.1044/2020>
- Stecker, G. C., Bernstein, L. R., & Brown, A. D. (2021). Binaural hearing with temporally complex signals. In R. Y. Litovsky, M. J. Goupell, R. R. Fay, & A. N. Popper (Eds.), *Binaural hearing* (pp. 145–180). Springer.
- Vermeire, K., & Van de Heyning, P. (2009). Binaural hearing after cochlear implantation in subjects with unilateral sensorineural deafness and tinnitus. *Audiology and Neurotology*, 14(3), 163–171. <https://doi.org/10.1159/000171478>
- Vermiglio, A. J., Herring, C. C., Heeke, P., Post, C. E., & Fang, X. (2019). Sentence recognition in steady-state speech-shaped noise versus four-talker babble. *Journal of the American Academy of Audiology*, 30(1), 54–65. <https://doi.org/10.3766/jaaa.17083>
- Versfeld, N. J., Daalder, L., Festen, J. M., & Houtgast, T. (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *Journal of the Acoustical Society of America*, 107(3), 1671–1684. <https://doi.org/10.1121/1.428451>
- Versfeld, N. J., Lie, S., Kramer, S. E., & Zekveld, A. A. (2021). Informational masking with speech-on-speech intelligibility: Pupil response and time-course of learning. *Journal of the Acoustical Society of America*, 149(4), 2353–2366. <https://doi.org/10.1121/10.0003952>
- Weisser, A., & Buchholz, J. M. (2019). Conversational speech levels and signal-to-noise ratios in realistic acoustic conditions. *Journal of the Acoustical Society of America*, 145(1), 349–360. <https://doi.org/doi10.1121/1.5087567>
- Weisser, A., Buchholz, J. M., Oreinos, C., Badajoz-Davila, J., Galloway, J., Beechey, T., & Keidser, G. (2019). The ambisonic recordings of typical environments (ARTE) database. *Acta Acustica United with Acustica*, 105(4), 695–713. <https://doi.org/10.3813/AAA.919349>
- Wilson, R. H., McArdle, R. A., & Smith, S. L. (2007). An evaluation of the BKB-SIN, HINT, QuickSIN, and WIN materials on listeners with normal hearing and listeners with hearing loss. *Journal of Speech, Language, and Hearing Research*, 50(4), 844–856. [https://doi.org/10.1044/1092-4388\(2007/059\)](https://doi.org/10.1044/1092-4388(2007/059))
- Wu, Y. H., Stangl, E., Chipara, O., Hasan, S. S., Welhaven, A., & Oleson, J. (2018). Characteristics of real-world signal to noise ratios and speech listening situations of older adults with mild to moderate hearing loss. *Ear and Hearing*, 39(2), 293–304. <https://doi.org/10.1097/AUD.00000000000000486>