



# Speech-in-Noise Testing: An Introduction for Audiologists

**Curtis J. Billings, Ph.D.,<sup>1,2</sup> Tessa M. Olsen, B.S.,<sup>1</sup> Lauren Charney, Au.D.,<sup>2</sup> Brandon M. Madsen, Au.D.,<sup>2,3</sup> and Corrie E. Holmes, Au.D.<sup>1</sup>**

## ABSTRACT

Speech-in-noise testing has been proposed as a useful part of the audiometric test battery dating back to the earliest years of the field of audiology. Many speech-in-noise tests have been developed and used to varying degrees. However, multiple barriers have prevented speech-in-noise testing from being used widely in the clinic. The purpose of this article is to provide a resource to audiologists and other hearing health professionals who want to know (1) what tests are available for use, (2) the rationale behind specific tests, and (3) important considerations when selecting one or more tests to use clinically. In addition, data are presented for four speech-in-noise tests with the purpose of comparing outcomes as a function of age and hearing status. The four tests (QuickSIN, Words in Noise [WIN], Listening in Spatialized Noise–Sentences [LiSN-S], and Coordinate Response Measure [CRM]) were completed by 30 individuals from three groups: 10 young adults with normal hearing, 10 older adults with normal hearing, and 10 older adults with hearing loss. The results suggest that, despite significant differences in performance between groups, group overlap was present such that some individuals from one group performed similar to some individuals of other groups; therefore, individual performance was more important than associated group. When selecting an appropriate speech-in-noise

<sup>1</sup>Department of Communication Sciences and Disorders, Idaho State University, Pocatello, Idaho; <sup>2</sup>VA RR&D National Center for Rehabilitative Auditory Research, VA Portland Health Care System, Portland, Oregon; <sup>3</sup>Hearing Systems Section, Department of Health Technology, Technical University of Denmark, Kongens Lyngby, Denmark.

Address for correspondence: Curtis J. Billings, Ph.D., VA RR&D National Center for Rehabilitative Auditory Research, VA Portland Health Care System, Portland, OR 97239 (e-mail: [curtis.billings2@va.gov](mailto:curtis.billings2@va.gov)).

VA Rehabilitation Research for Hearing and Balance Disorders; Guest Editors, Dawn Konrad-Martin, Ph.D. and Michelle E. Hungerford, Au.D.

Semin Hear 2024;45:55–82. © 2023. The Author(s).

This is an open access article published by Thieme under the terms of the Creative Commons Attribution-NonDerivative-NonCommercial-License, permitting copying and reproduction so long as the original work is given appropriate credit. Contents may not be used for commercial purposes, or adapted, remixed, transformed or built upon. (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Thieme Medical Publishers, Inc., 333 Seventh Avenue, 18th Floor, New York, NY 10001, USA

DOI: <https://doi.org/10.1055/s-0043-1770155>.

ISSN 0734-0451.

test to use clinically, audiologists should carefully consider the purpose of their testing and the type of information they desire as an outcome. A quick-resource table and appendix is provided to aid audiologists and other health professionals in their selection of an appropriate speech-in-noise test.

**KEYWORDS:** speech-in-noise testing, hearing loss, aging, background noise, audiology, audiometry, QuickSIN, Words in Noise, Listening in Spatialized Noise-Sentences, Coordinate Response Measure

Successful communication in complex listening environments is difficult for most of us, but especially for those with whom audiologists interact on a daily basis. Foremost among these individuals are those who are hard of hearing, but others might include individuals who are older and individuals with other conditions that impact the auditory system or other higher-order systems. While there are many factors that contribute to a complex listening environment, this article focuses on the contribution of background noise.

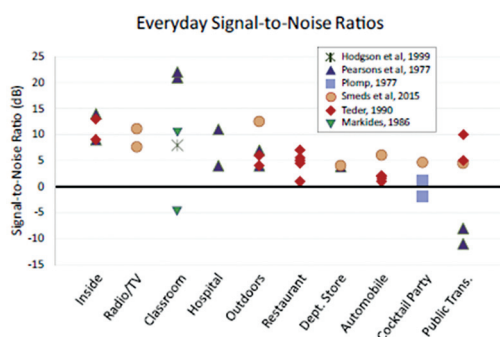
### Listening in Background Noise

When it comes to communication, difficulty understanding speech in the presence of background noise is the primary complaint among those treated for hearing loss (Kochkin 2010). Communicating in background noise is a universal challenge because there is so much noise in the world around us. For decades, researchers have tried to understand the extent of background

noise in common communication settings. Plomp (1977) and Pearsons et al. (1977) were some of the first to attempt to measure the levels of signals and noise in everyday listening environments to determine signal-to-noise ratios (SNRs) typical in communication. Fig. 1 illustrates data from several studies published over the past 45 years that measured SNRs present in everyday listening environments (Hodgson 1999; Markides 1986; Pearsons et al. 1977; Plomp 1977; Smeds et al. 2015; Teder 1990). Methods of measurement varied across studies, and there can be a large degree of variability even within one setting depending on the conditions that are present; however, results are generally consistent across studies with SNRs mostly between 0 and 15 dB across a range of settings.

### Historical Perspective on Speech-in-Noise Testing

Difficulties understanding speech in background noise have been studied for decades,



**Figure 1** Signal-to-noise ratios (SNRs) measured in everyday life. Measurements across several studies dating from 1977 to 2015 reveal general trends regarding typical SNRs in each environment, although some environments (e.g., classroom, public transportation) demonstrate more variability than others. (Figure used with permission of Hearing Research [see Billings & Madsen 2018, for details of measured settings]).

dating back to the beginnings of the field of audiology (Carhart 1946; Miller 1947; Cherry 1953). With respect to measuring success with hearing aids, Carhart (1946) identified understanding speech in noise as one of the main dimensions needing further exploration. Miller (1947), and later Cherry (1953), explored the effect of different competing maskers (i.e., different types of noise and different sources of noise) on speech understanding. Similarly, it has been suggested since the early days of audiology that speech-in-noise testing is important to include in audiological testing (Carhart 1946; Davis et al. 1946; Hardy 1950). Despite its consistent recommendation by professional associations and researchers alike (e.g., APSO 2021; Davidson et al. 2021, the prevalence of regular speech-in-noise testing by hearing professionals remains below 50% (ASHA 2019; Mueller 2010; Strom 2006; Clark et al. 2017).

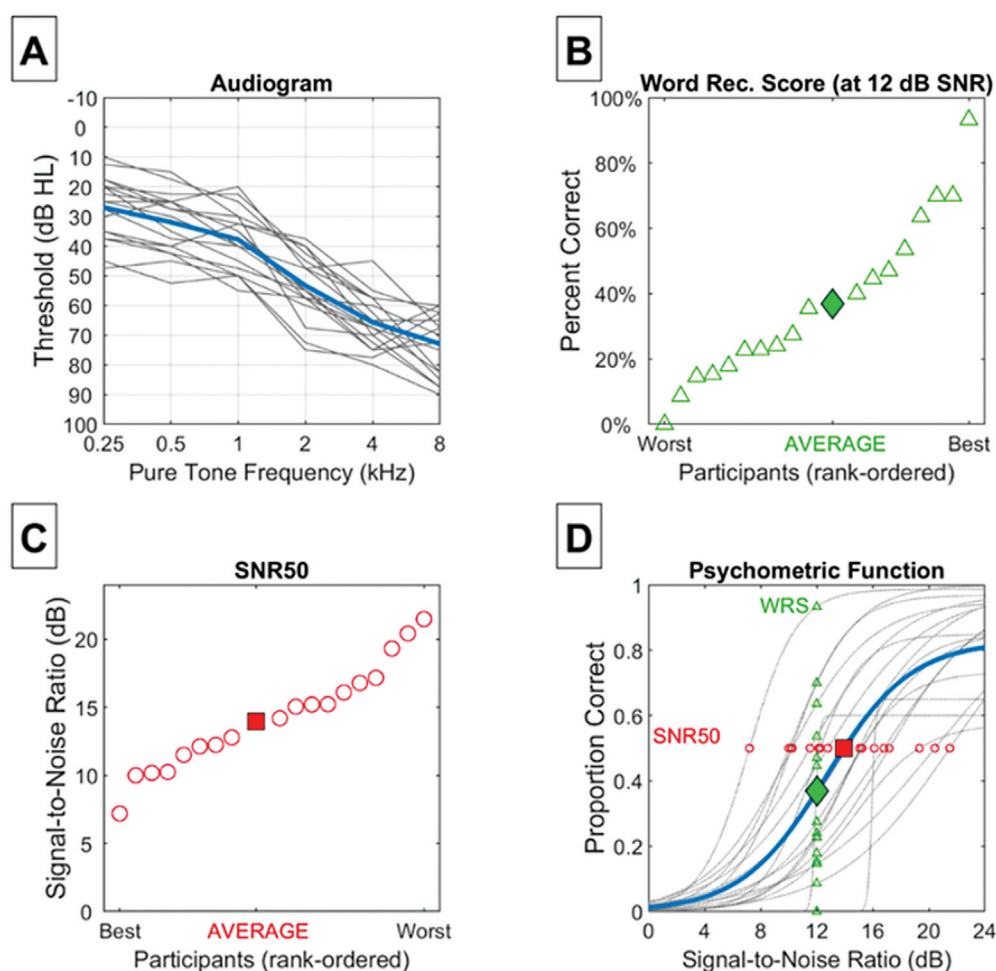
The fundamental challenge of helping patients understand speech in the presence of competing maskers continues to be a critical task for audiology. More than 70 years ago, Hardy (1950) introduced two different categories of hearing difficulty. The first is the “louder please” category. Individuals with this type of difficulty do well if the volume of the signal can be increased. The second is the “I can’t understand you” category. These individuals may need increased volume, but they also have a perceptual impairment involving auditory distortion that results in listening difficulties. Other researchers have since proposed similar categories (Carhart 1951; Stephens 1976). In 1978, in a seminal paper, Plomp formalized these categories in terms of “attenuation” versus “distortion” problems, with the latter being especially pronounced in noisy environments. He proposed that hearing losses in the *attenuation* category simply decrease a person’s *sensitivity* to sound (i.e., raise one’s audibility threshold), while losses in the *distortion* category cause a degradation in the *fidelity* of a person’s perception of sounds even when they are well above threshold. In most listeners, both attenuation and distortion-related factors contribute to speech-in-noise performance, resulting in considerable variability.

## Why Should I Test Speech in Noise?

An emblematic feature of speech-in-noise perception is the wide range of variability in performance across individuals, even when all are of similar age and hearing status. For example, Fig. 2A shows audiograms for 18 individuals who are over 65 years of age; despite the similarities across individuals (i.e., bilateral sensorineural hearing loss in the normal/mild sloping to moderately severe/severe range), understanding words in background babble noise at a given SNR (12 dB in this case) varies from ~0 to 95% (Fig. 2B). The large range of variability is also apparent in SNR50s, or the SNR at which an individual achieves 50% intelligibility (Fig. 2C). Both the percent correct score at a given SNR and the SNR at a given percent correct can be derived from the more complete psychometric function (Fig. 2D), which in this case represents the performance of the individuals tested across a range of SNRs.

The wide range of variability in understanding speech in background noise presents a challenge for audiologists. For example, two patients may present with similar pure-tone thresholds and speech-in-quiet understanding but very different speech-in-noise understanding. Given that pure-tone testing and speech-in-quiet testing are completed in an optimal listening situation, it is not surprising that these measures do not adequately characterize the listening difficulties experienced by patients in background noise. In a group of 3,430 Veterans, Wilson (2011) demonstrated the relationship between pure-tone average (PTA), word recognition in quiet, and word recognition in noise using the Words in Noise (WIN) test. While these measures were correlated, there was a wide range of performance on WIN scores even among individuals with similar PTAs or performance in quiet. As many as 70% of those tested had word recognition scores  $\geq 80\%$  in quiet, whereas only 7% of the group demonstrated normal WIN performance ( $\leq 6$  dB; Wilson 2011). Wilson suggested that speech-in-noise testing “puts substantial pressure on the auditory system and should be considered as the ‘stress test’ of auditory function” (p. 418).

Another challenge for audiologists is determining which speech-in-noise test to use.



**Figure 2** Variability in speech-in-noise listening performance. Performance on the WIN for 18 individuals over the age of 65 years with symmetrical hearing loss. Panel A shows the average (thick blue line) and individual (thin gray lines) pure-tone thresholds. Panels B and C reveal the group-mean (diamond/square) and individual (triangles/circles) percent correct and SNR50 scores, illustrating the wide range of variability across individuals. Panel D shows the modeled group (thick blue line) and individual (thin gray lines) psychometric functions.

Several different tests have been developed over the years; however, the rationale for using one test instead of another is not always clear. Interestingly, the most commonly used of these tests is reportedly the QuickSIN (Clark et al. 2017; Strom 2006; Mealings et al. 2020; Mueller 2003), perhaps because of its ease of use and the length of time that it has been available to audiologists. The purpose of this article is to introduce audiologists to the variety of tests that are available to them as well as the underlying rationale for each, and to discuss important factors that should be considered when selec-

ting a test to use in their audiometric test battery. Data will also be presented that compare performance across a variety of speech-in-noise tests within the same group of individuals who vary in age and hearing status.

### How Is Speech Understanding in Noise Measured?

Speech understanding in noise is most often characterized either in terms of percent correct at a given SNR or, conversely, as the SNR needed to achieve a specific percent correct

(most commonly the SNR50, i.e., the SNR needed to understand 50% of the signal) as shown in Fig. 2 (B and C, respectively). Some tests go a step further and compare the SNR score of an individual to a group of normal-hearing individuals, resulting in an “SNR loss” value that is similar to the conversion of dB SPL to dB HL on the audiogram (i.e., a group of normal-hearing individuals are used to provide a reference performance level). The main motivation that has been advanced for using *SNR loss* rather than *SNR50* is that SNR loss is likely to be more comparable within an individual across tests than SNR50 is, with the latter being extremely test-specific. Comparing SNR50 across tests with the purpose of tracking individual changes in performance should be done with caution as differences may be more reflective of signal or noise differences rather than individual performance difference. If the purpose is to track changes in performance over time, then it may be advisable to normalize the results from different tests in some way. Theoretically, using SNR loss should work to subtract out the differences in mean performance between tests, but the conversion procedure does not address any differences that might exist in the width of the performance spread across tests; so, one would still expect definite limits to the degree of across-test generalizability. It is noteworthy that the SNR50, or 50%-point (also known as the speech reception threshold in noise [SRTN] or speech reception threshold [SRT] in the speech-in-noise literature, but not to be confused with the audiometric SRT, which is typically the detection threshold of spondees in quiet), has historically been the convention for characterizing speech-in-noise performance.

It is important to consider the methodology of administration of speech-in-noise tests that are available to audiologists. Each test's methodology is unique but can be divided into three categories: fixed, adaptive, or progressive. A *fixed* protocol uses a constant SNR (e.g., +5 dB) for the entirety of the test and the outcome measure is a percent correct score. An *adaptive* protocol changes the SNR of a given condition according to how the participant performs and is designed to “bracket” a particular predetermined level of performance (most often target-

ing 50% correct), with the response variable being the SNR that, on average, results in performance closest to the target level. Finally, a progressive protocol gradually changes the SNR (usually in either an increasing only or decreasing only direction) using step sizes and numbers of trials that are independent of how the participant performs, and typically the SNR50 is derived or the conversion to SNR loss is made. One partial exception to this performance-independence is that some progressive tests employ a stopping rule such that if the individual performs below a certain criterion at a given SNR (e.g., 0 of 5 correct) then the testing can be terminated and no poorer SNRs are presented, effectively reducing test time (in which case all SNRs lower than the one that triggered the stopping criterion would also be assumed to have a score of 0).

Another important consideration is how the SNR for a given test is changed throughout the test. As an example, the QuickSIN holds the signal level constant and adjusts the noise level to create different SNRs, whereas in contrast, the WIN holds the noise level constant and adjusts the signal level to manipulate SNR. For most speech-in-noise tests, the presentation level should be high enough to ensure audibility; however, it is critical to carefully consider how audibility of the signal and audibility of the noise may impact speech-in-noise testing for an individual given their specific hearing loss.

### What Is the Best Test for You?

There are many speech-in-noise tests that are available to audiologists. Table 1 represents basic information for the common tests available to audiologists. A more thorough reference is provided as **Appendix A**, which compiles additional information for each test about the purpose, materials, administration, scoring, norms if available, and key references. Of course, the selection of a given test will be dependent on many factors. For example, the type of noise, the type of signal, and other factors may be important to consider.

### EFFECTS OF SIGNAL TYPE

There is a range of speech signal types, including nonsense syllables, words, sentences,

Table 1 Current speech-in-noise tests that are frequently used or discussed in the literature

	Acronym	Target Speech	Noise	Protocol	Use	Time (min)	Purchase Source	Approximate Cost
Acceptable Noise Level	ANL	Running speech Female	Multi-talker babble	Adaptive	Pre-hearing aid fitting Estimate noise tolerance	5-10	Interacoustics www.interacoustics.com	Included with equipment purchase \$155
Arizona Biomedical Sentences Recognition	AzBio	AzBio sentences 2 Female 2 Male	Auditec four-talker babble	Fixed	CI evaluation CI post-op monitoring	5-7	Auditory Potential www.auditorypotential.com	
Bamford Kowal Bench Sentence Test	BKB-SIN	BKB sentences Male	Auditec four-talker babble	Progressive	CI evaluation CI post-op monitoring APD evaluation When QuickSIN is too difficult	5-7	Etymotic Research, Inc. Auditec, Inc.	\$215
Connected Speech Test	CST	Recorded passages	Six-talker babble	Fixed	Pre- post hearing aid fitting	<10	https://harlmemphis.org/connected-speech-test-cst/ http://auditory.org/mhonarc/2012/msg00034.html	Free
Coordinate Response Measure	CRM	Ready [CALL SIGN] go to [COLOR] [NUMBER] now BKB Sentences	Multi-talker babble	Adaptive or Progressive	Analyze spatial hearing			Free
Hearing in Noise Test	HINT		Speech spectrum noise	Adaptive	Measure SNR threshold while avoiding ceiling & floor effects Determine benefit of directional mics	5-10	Owned by Interacoustics; not currently available to clinicians	
Listening in Spatialized Noise - Sentences	LiSN-S	LiSN-S sentences Female	Children stories	Adaptive	APD evaluation or specifically, spatial processing disorder	<15	Sound Scouts https://www.soundscouts.com/apd/	Monthly subscription
Quick Speech in Noise	Quick-SIN	IEEE sentences Female	Auditec four-talker babble	Progressive	Pre- Post- hearing aid fitting May indicate cognitive processes	1/list	Etymotic Research, Inc. Auditec, Inc.	\$176AOS
Revised Speech in Noise	R-SPIN	R-SPIN Sentences 1 Female 1 Male	Multi-talker babble	Fixed	Determine use of context clues		https://ahs.illinois.edu	Free
Speech Reception in Noise Test	SPRINT	NU-6 words	Six-talker babble	Fixed	Aid in determination of H3 active-duty members with retention, reclassification, or separation	10	Richard Wilson, Ph.D.: Richard.H.Wilson@asu.edu or wilsonr1943@gmail.com	\$50 donation
Words in Noise	WIN	NU-6 words Female	Causey six-talker babble	Progressive	Pre- hearing aid fitting Minimize cognitive influences	2.5	Richard Wilson, Ph.D.:Richard.H.Wilson@asu.edu or wilsonr1943@gmail.com	\$50 donation



running monologue, and conversation. Syllables typically increase the number of scoring opportunities, allowing for less variability and more precise measurements. Syllable- or word-level tests may also give the clinician information about specific phoneme errors while reducing top-down processing effects (Billings et al. 2016). Within each broad type, there are also many other potential considerations to account for, such as the properties of the target talker (e.g., pitch, speed, or dialect characteristics), the length and familiarity of the target speech, the amount of contextual information, and so on. Different signal types will require contributions from different levels and processes of the auditory system. In some cases, audiologists may wish to limit higher-order contributions such as cognitive processes (e.g., working memory, cognitive processing speed, sustained attention) by administering syllable- or word-level tests such as the WIN. At other times, it may be important to use signal types that allow for more cognitive contribution as with sentence-level tests such as the QuickSIN. The assumption is that speech-in-noise perception with simpler signals, such as syllables or words, will be driven by processes at the peripheral end of the auditory system (Wilson & McArdle 2005).

Another important consideration is set size and whether the answer set is open or closed. A closed set refers to a set of responses that is finite and often known by the patient. Alternatively, an open set test has an unlimited number of possible responses. For example, the CRM uses 32 specific color-number pairs, resulting in a closed set of answers that will be inherently easier than an open set task with words or sentences, all else being equal. That said, all else may *not* be equal; a closed-set test presented at a small/poor SNR, for example, might result in worse performance than an open-set task presented at a large/good SNR. It should be kept in mind that even a closed-set test requires cognitive processing by the patient to take advantage of the limited number of possible responses.

While speech-in-noise tests that use sentences are more reflective of real-world listening conditions and the use of context, their results are likely to be more affected by cognitive decline, for similar reasons as in the case of

closed sets discussed earlier (after all, knowledge about the set of possible answers is, like the surrounding words in a sentence, just another specific kind of contextual information). If an audiologist is concerned that a patient may be suffering from cognitive decline, it may be best to choose a test of words unless you hope to account for the cognitive decline as well (Wilson 2004; Wingfield 1996). In summary, with regard to signal type, clinicians should consider using sentence-level tests as a means to assess everyday communication in a functional way and use word or syllable tests to determine more bottom-up factors involving lower-level auditory function in the presence of noise.

### EFFECTS OF NOISE TYPE

Many different types of background noise are used in speech-in-noise testing. Noise types vary in content and complexity across frequency, level, and timing domains (i.e., in essentially every way that a sound can vary). To assist in considering the effects of background noise, the speech-in-noise literature has categorized masking in two general ways: energetic masking and informational masking. Energetic masking has been characterized as the overlap of the target and masker in time and frequency in the cochlea such that portions of the target are inaudible (Brungart 2001a b; Kidd et al. 2008). Informational masking, in contrast, cannot be explained by interactions in the auditory periphery and has its origins at higher levels in the auditory system. For informational masking, uncertainty (difference between what the listener actually hears and what the listener expects to hear) and similarity (the relationship between the target and the masker such that the listener is able to separate them from each other) cause increased understanding difficulties above and beyond what would be expected from the acoustics alone (Durlach et al. 2003). Unsurprisingly, higher-intensity noise provides more masking, all else being equal. For equal average intensity, the strongest speech maskers will be those whose other characteristics are most similar to speech (and, within that, most similar to the voice of the particular speaker they are masking). Thus, in the real world, speech-on-speech listening is typically the most difficult listening task.

Background noise progresses in speech-masking effectiveness as it becomes more like speech. Continuous broadband noise, such as white noise, pink noise, or speech spectrum noise, becomes more effective in masking as it becomes more similar spectrally and temporally to the target speech. It is important to keep in mind that continuous noise types typically do not contain significant amplitude or frequency changes over time (or, to be more precise, any instantaneous changes average out so as to be inconsequential on timescales relevant for human hearing); therefore, continuous maskers are less effective than maskers that vary over time similarly to target speech. Speech maskers with fewer than four talkers have fluctuations similar to speech targets and result in increased uncertainty for the patient and therefore increased difficulty. As the number of talkers increases beyond four, amplitude and frequency fluctuations begin to overlap across talkers, resulting in a more continuous-like noise and gradually shift from more informational masking to more energetic masking (Kidd et al. 2008). Most clinical speech-in-noise tests use speech as the background noise as a means to reflect real-world listening situations. Such a speech-on-speech listening condition reflects the more difficult speech-in-noise situations that listeners will experience in everyday life. When the background noise is speech, both energetic and informational masking are expected to occur, with the balance of masking tending toward more energetic masking as the talkers increase beyond four talkers. Audiologists should consider carefully what signal and noise type they are using and determine if that matches their purpose for performing speech-in-noise testing.

#### OTHER CONSIDERATIONS

Several other factors may influence the choice of which speech-in-noise test to use. Test time is a critical factor in most clinics as appointment lengths are limited. It may be that a speech-in-noise test could take the place of other more traditional tests that may not be needed. Wilson & McArdle (2005) have advocated for the possibility of replacing words-in-quiet testing with words-in-noise testing. It is noteworthy that most speech-in-noise tests take < 15

minutes to administer. High on the list of reasons to complete speech-in-noise testing would be the ability to counsel and instruct the patient using stimuli and conditions that have good face validity and match the difficulties that most patients face in their everyday lives. Test results may also be effective at helping patients set realistic expectations for intervention outcomes.

Speech-in-noise tests may help with the diagnostic process by differentiating, as per the Plomp (1978) model, between attenuation-based hearing losses and distortion-based hearing losses. Such a distinction could inform intervention strategies. For example, those that have a distortion-based hearing loss, or greater SNR loss, would benefit from use of tight-beam directionality, remote microphones, or accommodations like preferential seating (Etymotic Research 2006). Furthermore, counseling about the importance of lip reading and ensuring that visual cues are available would be helpful. However, those who have an attenuation-based hearing loss, or no SNR loss, may benefit substantially from a straightforward amplification approach to enhance frequencies that were previously inaudible.

#### Special Populations

Despite recommendations and support for speech-in-noise testing in the scientific literature, the prevalence of regular speech-in-noise testing by hearing professionals remains below 50% overall (ASHA 2019; Mueller 2010; Strom 2006; Clark et al. 2017). However, it is important to note that speech-in-noise testing is more frequently used in specific populations. We will highlight three of these populations: cochlear implant users, hearing aid users, and those who have listening difficulties beyond what would be expected based on audibility thresholds alone.

##### COCHLEAR-IMPLANT USERS

Speech-in-noise testing has more recently become an integral part of both cochlear-implant candidacy evaluations and post-implantation monitoring. Cochlear-implant candidacy guidelines are based on an individual's degree of hearing loss as well as aided speech-perception



scores. These results are necessary to document that an individual is not receiving sufficient benefit from acoustic amplification alone and, therefore, may obtain more benefit from a cochlear implant. The Food Drug Administration (FDA) approved candidacy indications specify the necessary aided speech-perception score on a test of open-set sentence recognition to be considered for cochlear implantation for each manufacturer. Early candidacy indications specified the use of the HINT; however, clinicians generally completed the evaluations in quiet rather than performing the test in the adaptive SNR format as originally designed. Therefore, testing was not representative of listening difficulties in real-world environments. Research later identified that the HINT was prone to ceiling effects when testing was completed only in quiet to assess pre-implant versus post-implant benefit. Updated recommendations based on the revised Minimum Speech Test Battery (MSTB 2011) suggest use of the AzBio sentence test in place of the HINT (Auditory Potential LLC, 2022; Gifford et al. 2008). Additionally, it was recommended that testing be completed both in quiet and in noise to better document an individual's performance in a more realistic range of listening scenarios (Auditory Potential 2022).

The revised MSTB (2011) now serves as a guideline for audiologists completing adult candidacy evaluations. Aided speech-perception testing generally includes CNC words, AzBio sentences in quiet, AzBio sentences in noise, and the BKB-SIN (the AzBio sentences provide a percent correct score at a fixed level, whereas BKB-SIN provides a SNR50 score). A presentation level of 60 dB SPL is recommended for all speech material; however, when administering the AzBio in noise, the clinic may choose their own SNR based on that clinic's candidacy evaluation protocol. Recent surveys by Prentiss et al. (2020) and Carlson et al. (2018) found that 68 or 89% (respectively) of respondents routinely perform speech-in-noise testing during cochlear-implant candidacy evaluations with either a +5 or +10 dB SNR. The majority of remaining respondents indicated that speech-in-noise testing was performed on a selective basis only if patient scores were considered

“borderline” after testing was completed in quiet. The same studies also inquired as to the routine use of the BKB-SIN during adult candidacy evaluations, and they found that only 32 or 9% (respectively) includes the BKB-SIN as an additional test in their protocol. Limited use of the BKB-SIN during adult candidacy evaluations may be related to time constraints and current FDA criteria, which do not base any requirements on SNR50 or any other SNR score.

Verification of benefit following implantation generally includes monitoring and comparing pre-implantation speech perception scores to post-implantation scores. Although research suggests that there is still variability in cochlear-implant users' performance, significant improvements in speech perception scores have been documented (Sladen et al. 2017). For new cochlear-implant users, the majority of aided speech-perception testing may be completed in quiet as this alone can be a difficult task for many patients. However, as the user adapts to their cochlear implant and their speech understanding improves, then speech-in-noise testing can be beneficial to verify benefit in a more realistic listening scenario.

#### HEARING-AID USERS

As with cochlear-implants users, hearing-aid users can benefit from speech-in-noise testing to demonstrate benefit. Three quarters of a century ago, Carhart (1946) and Davis et al. (1946) suggested that speech-in-noise testing could be very helpful as a pre-fitting measure to help determine candidacy for hearing aids. Mueller (2003) suggests several reasons for including a speech-in-noise measure as part of the pre-fitting testing: (1) help with selecting technology (e.g., more aggressive noise reduction for those with speech-in-noise difficulties), (2) determine frequencies to amplify (e.g., less low-frequency gain in noisy environments), and (3) use for counseling and setting realistic expectations. In fact, some of the currently available tests were designed specifically for pre-fitting hearing-aid testing. One of these is the Acceptable Noise Level (ANL) test, which seeks to determine the highest level of noise that is acceptable to the patient. Patients who can tolerate higher noise levels (i.e., ANLs

of  $< 7$  dB) are typically more successful with hearing aids as compared with those who cannot tolerate higher noise levels (i.e., ANLs  $\geq 8$  dB); those with ANLs  $> 13$  dB, on the other hand, may not be good candidates for hearing aids and perhaps should be encouraged to pursue hearing assistive technology systems (HATS) or may need additional counseling before and during hearing aid use (Nabelek et al. 2006). Recently, Davidson et al. (2021) completed a systematic review of the relationship between various pre-fitting measures (e.g., speech recognition in quiet, speech recognition in noise, subjective ratings, and dichotic speech tests) and hearing aid satisfaction, and they concluded that speech-in-noise tests had the highest association with hearing aid satisfaction.

Speech-in-noise testing can also be used in the post-fitting process to help patients understand more about their own performance in noise. The Performance-Perceptual Test (PPT) was designed as a post-fit measure to determine if and how patients misjudge their speech-in-noise performance. By comparing their actual performance with how they think they perform, a clinician is equipped to help the patient recalibrate and set expectations appropriately through counseling with the goal of improving hearing aid use among those reporting little benefit.

Ceiling effects in speech-in-noise testing are important to consider with hearing-aid users and other groups that might experience a wide range of performance variability. When fixed SNRs are used that are relatively easy for the patient, performance will bump up against the ceiling of 100%. As a result, it will be difficult to differentiate performance between conditions or tests. Therefore, a clinician may need to use lower SNRs or use an adaptive test, such as LiSN-S or HINT, or a progressive measure that tests a range of SNRs, such as WIN or QuickSIN.

#### INDIVIDUALS WITH LISTENING DIFFICULTIES

Speech-in-noise testing is especially important for individuals who have listening difficulties that are beyond what is suggested by their audiogram. Unexpected difficulties in complex listening environments could include challenges

with background noise or situations with rapid, reverberant, or otherwise degraded speech (Middelweerd et al. 1990; AAA 2010). Patients with normal-hearing thresholds and listening difficulties have been referenced in many ways in the literature including auditory inferiority complex (Byrne & Kerr 1987), auditory disability with normal hearing (Rendell & Stephens 1988), selective dysacusis (Narula & Mason 1988), obscure auditory dysfunction (Saunders & Haggard 1989), King-Kopetzky syndrome (Kopetzky 1948; King 1954; Hinchcliffe 1992), auditory dysacusis (Jayaram et al. 1992), and idiopathic discriminatory dysfunction (Rappaport et al. 1993). One common characteristic of these patients is greater-than-normal listening difficulties even with a normal audiogram. Reports from these patients show that a diagnosis of normal hearing combined with a lack of treatment recommendations may result in feelings of dismissal and confusion (Pryce & Wainwright 2008).

Currently, many audiologists may seek to diagnose patients who have listening difficulties in the presence of normal pure-tone hearing with what has been called APD: auditory processing disorder or deficit (ASHA 2005; AAA 2010). While APD's status as a distinct condition remains controversial and there are well-documented challenges in assessing the accuracy of such a diagnosis (see Vermiglio 2018a, 2018b, and Chermak et al. 2018, for a review), there is broad agreement that the listening difficulties reported by individuals who show up as normal on traditional audiologic assessments constitute a very real phenomenon. According to ASHA (2005) and AAA (2010), at least two domains of APD testing may include speech in the presence of background noise: monaural low redundancy and binaural interaction. Monaural low redundancy refers to signals presented to one ear that are degraded in some way to reduce the natural redundancy in speech, such as filtering a signal or adding background noise to a signal, and binaural interaction refers to using inputs to both ears (i.e., dichotic) to localize or lateralize sounds (AAA 2010; ASHA 2005). The tests presented in **Appendix A** may be especially useful with this population, that is, those with

normal or near-normal pure-tone thresholds and good speech understanding in quiet. For example, the LiSN-S (often categorized as a test of binaural interaction) was created to identify individuals with a spatial processing disorder. For these individuals, the LiSN-S has been used successfully to diagnose and monitor treatment (Brown et al. 2010; Cameron et al. 2012).

## METHODS

To illustrate differences in speech-in-noise performance across tests and groups of individuals, we tested individuals in a repeated-measures cross-sectional experiment. We used the QuickSIN (Killion et al. 2004), Words-in-Noise (WIN; Wilson & Burks 2005), Coordinate Response Measure (CRM; Bolia et al. 2000), and Listening in Spatialized Noise—Sentences (LiSN-S; Cameron & Dillon 2007) tests.

### Participants

Participants were 30 individuals separated into three groups based on age and hearing ability. The groups were 10 younger normal-hearing individuals (YNH) with a mean age of 27.1 years,  $SD = 7.2$ ; 10 older normal-hearing individuals (ONH) with a mean age of 67.2 years,  $SD = 5.1$ ; and 10 older hearing-impaired individuals (OHI) with a mean age of 68.8 years,  $SD = 5.9$ . Each of the three groups contained six female and four male participants. A two-sample  $t$ -test found that the difference in mean participant's age between the ONH and OHI groups was not statistically significant ( $t(18) = -0.645$ ;  $p = 0.527$ ). Pure-tone hearing thresholds of YNH and ONH participants were below 25 dB HL up to 4,000 Hz bilaterally. It should be noted that individuals in these "normal-hearing" groups may indeed have hearing deficits; we wish to generally categorize only the participant groups according to their pure-tone thresholds. OHI participants had approximately symmetrical mild-to-moderate sloping sensorineural hearing loss bilaterally. A PTA (pure-tone average of thresholds at 500, 1,000, and 2,000 Hz) was calculated for each participant, with mean PTA and standard

deviations calculated for each group (YNH:  $6.0 \pm 4.4$  dB; ONH:  $7.9 \pm 4.9$  dB; OHI:  $32.3 \pm 7.7$  dB); according to a two-sample  $t$ -test, the mean difference between normal-hearing groups was not statistically significant ( $t(18) = 0.95$ ;  $p = 0.355$ ). This research was completed with the approval of the local institutional review board and with the informed consent of all participants.

### Materials and Outcome Measures

The QuickSIN speech-in-noise test uses target sentences that originate from the IEEE corpus (IEEE 1969). Only five "key words" from each sentence are scored; the rest of the words do not affect the score. The target sentences were presented at 70 dB HL with four-talker babble in the background. The level of the target speech remains constant, while the background babble increases by 5 dB after each sentence. This produces SNRs that range from +25 to 0 dB, with one sentence per SNR.

The WIN uses target words from the NU-6 corpus (35 per trial), with each word being scored completely correct or incorrect. The target words were presented at a starting level of 84 dB HL with six-talker babble in the background at 60 dB HL. The signal level decreases by 4 dB after each five-word block, while the level of the background babble remains constant. This produces SNRs that range from +24 to 0 dB.

The CRM employs target sentences that follow a stereotyped, fill-in-the-blank format: "Ready [call sign], go to [color] [number] now" (e.g., "Ready Charlie, go to blue eight now"). The participant is asked to select the color/number combination (i.e., the "coordinates") associated with their assigned call sign (in our implementation, this was always "Charlie") by using a computer touchscreen with a  $4 \times 8$  grid of colors (red, blue, green, and white) and numbers (1–8). There is no single definitive version of this test, as even publicly released versions tend to be highly customizable by the user, and our version was run directly through custom MATLAB software. In our implementation of the CRM, the target signal was presented at 40 dB SL (re: spondee threshold); the signal level was held constant while the

background noise was gradually increased over the course of each run, producing SNRs ranging from +9 to -21 dB in 2-dB steps. We used three different types of background noise in separate runs to obtain a specific performance estimate in each noise type for each participant; these noise types were one-talker modulated (1TM), four-talker babble (4TB), and speech-shaped continuous (SSC). The SSC noise was created using the long-term average speech spectrum (LTASS) of the IEEE sentence corpus created previously (Billings et al. 2011) to spectrally shape continuous noise. The 1TM noise used this same LTASS but with the envelope modulated to mimic 10 concatenated IEEE sentences.

The LiSN-S target sentences were presented at 62 dB SPL, with each word scored as correct or incorrect. The distractor stimuli were children's stories initially presented at 55 dB SPL. The level of the target sentences remained constant while the level of the story changed according to an adaptive bracketing algorithm designed to pinpoint the SNR50 (see the following paragraph), and which was automatically performed by the software. The test was administered in four different conditions according to two parameters: (1) whether the target and distractor stimuli are presented by the same voice (SV) or two different voices (DV) and (2) whether the stimuli are presented with target and distractors both at 0-degree azimuth or with the distractors offset from the target by 90 degrees. The four different conditions (DV90, SV90, DV0, and SV0) are equivalent to the four different permutations resulting from this  $2 \times 2$  grid of test parameter values.

A single scalar metric, the SNR50 (the SNR threshold associated with a 50% correct rate), was used to quantify behavioral performance on each of the measures described earlier. To calculate each participant's SNR50 for the QuickSIN, CRM, and WIN, we first employed the Palamedes Toolbox (Prins & Kingdom 2018) in MATLAB to estimate each participant's psychometric function for each measure by fitting a four-parameter logistic curve (@PAL\_Logistic) to each participant's performance data using an iterative maximum-likelihood optimization algorithm (@PAL\_PFML\_Fit). The SNR50 was

then calculated as the point at which the fitted psychometric function intersected the 50% correct line. For the LiSN-S, on the other hand, we simply used the bracketed SNR50 reported by the software, since the LiSN-S is an adaptive SNR50-specific test that does not attempt to sample the rest of the psychometric function.

## Analyses

Three sets of repeated measures analyses of variance (RM-ANOVAs) were completed to test the following effects for statistical significance: (1) the effects of noise type using the CRM dataset, (2) the effects of group across all four tests, and (3) the effects of talker and azimuth as a function of group using the LiSN-S dataset.

## RESULTS

Table 2 presents means and standard deviations for all tests and conditions as a function of group. In general, and as expected, the YNH group demonstrated the best performance across tests and conditions followed by the ONH and then OHI groups. The four tests that were used provide the opportunity to explore the effects of noise type and, to some extent, signal type as a function of group.

### Effects of Noise Type and Group

As discussed in the introduction, noise type can have an important effect on performance. The CRM test used in this study provided a way to directly explore the effects of noise type. Fig. 3 shows results from the CRM test for the three groups and three noise types. The top portion of the figure shows the psychometric functions, demonstrating the effects of noise type across a range of SNRs. Participants had the most difficulty (highest SNR50s) with 4TB followed by SSC. Participants performed the best (had the lowest SNR50s) in 1TM noise. Such a pattern would be expected given the high degree of spectrotemporal similarity between the babble and the signal leading to poorer performance, and the gaps present in the modulated noise leading to better performance. The  $3 \times 3$  RM-ANOVA found the

Table 2 Mean SNR50 speech-in-noise test scores for the YNH, ONH, and OHI groups

Group	WIN84		Quick-SIN_SNR50		LiSN_DV90		LiSN_SV90		LiSN_DV0		LiSN_SV0		CRM_1TM		CRM_4TB		CRM_SSC	
	Mean (dB)	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
YNH	4.83	(1.31)	2.17	(1.05)	-15.53	(1.47)	-13.34	(2.95)	-11.0	(1.97)	-1.81	(0.46)	-20.66	(1.67)	-4.49	(2.25)	-8.51	(1.30)
ONH	8.34	(1.77)	3.61	(1.24)	-11.84	(1.74)	-9.72	(2.20)	-8.27	(2.24)	-0.53	(1.01)	-19.63	(2.61)	-2.50	(1.59)	-6.73	(2.06)
OHI	11.09	(2.44)	4.83	(1.69)	-3.65	(4.16)	-2.46	(3.29)	-1.81	(2.93)	2.08	(1.79)	-12.27	(3.13)	-1.43	(1.61)	-4.0	(2.13)

SD-standard deviation, dB-decibel, YNH-young normal hearing, ONH-old normal hearing, OHI-old hearing impaired

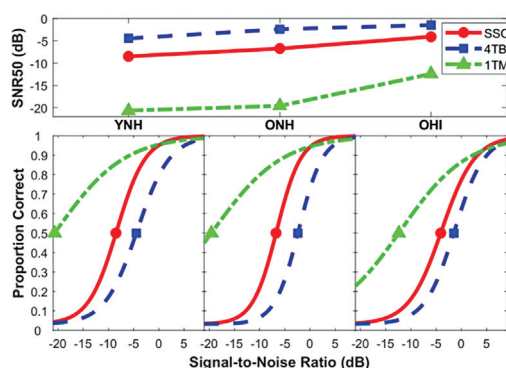
effect of Noise Type to be significant ( $F = 548.1$ ;  $df = 2, 27$ ;  $p < 0.0001$ ) as well as the between-subjects main effect of Group ( $F = 25.99$ ;  $df = 2, 27$ ;  $p < 0.0001$ ). As seen in Fig. 3 (bottom), the YNH group performed the best on average, followed by the ONH group, with the OHI group performing the worst. In addition to the aforementioned main effects, the Noise Type  $\times$  Group interaction was also found to be significant ( $F = 10.16$ ;  $df = 4, 54$ ;  $p < 0.0001$ ). Fig. 3 (top) reveals that this interaction was likely driven by OHI performance in 1TM noise; the OHI group showed a smaller improvement in 1TM, relative to the other noise types, than the YNH and ONH groups. In other words, the OHI individuals were not able to take advantage of the gaps in the 1TM noise to the same degree as the normal-hearing groups.

Fig. 4 shows the test results for the four tests that were completed as a function of group. For the tests with multiple conditions (CRM and LiSN-S), conditions were selected that were most comparable to the WIN and QuickSIN paradigms—namely, the CRM's 4TB condition (since WIN and QuickSIN both use multitalker babble as their maskers), and the LiSN-S's DV0 condition (since the other tests involved neither spatial separation nor a masker matching the target talker). One-way ANOVAs for each of the outcome measures displayed were completed to characterize the Group effect on each measure. In all cases, the effect of Group was found to be significant (CRM-4TB:  $F = 6.39$ ,  $df = 2, 27$ ,  $p = 0.0059$ ; WIN:  $F = 23.1$ ,  $df = 2, 27$ ,  $p < 0.0001$ ; QuickSIN:  $F = 8.29$ ,  $df = 2, 27$ ,  $p = 0.0018$ ; LiSN-DV0:  $F = 34.01$ ,  $df = 2, 27$ ,  $p < 0.0001$ ), demonstrating that Group was an important factor contributing to performance across the four tests. Generally, YNH individuals performed the best followed by ONH individuals and then the OHI individuals. It is important to note that there was overlap between groups for most of the tests.

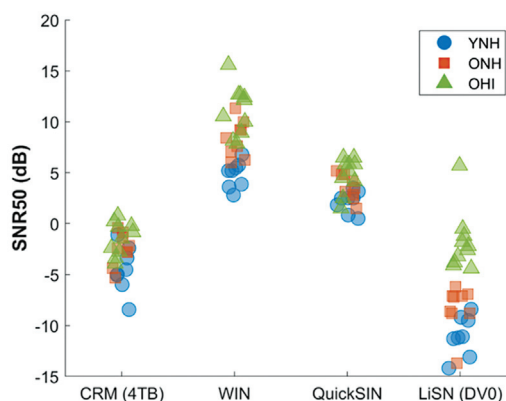
### Effect of Signal Type

The effect of signal type is somewhat apparent in Fig. 4, although only qualitatively, given that noise types and conditions are not equivalent





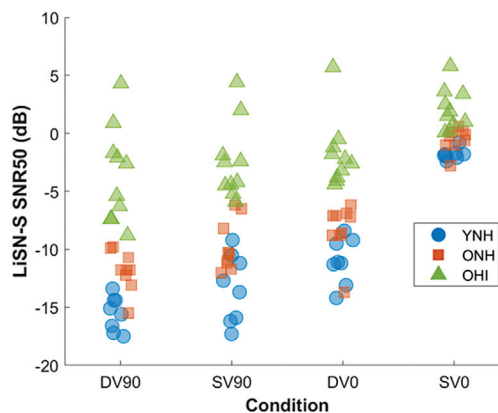
**Figure 3** Psychometric functions (bottom) and SNR50 values (top) for CRM testing as a function of Noise Type and Group. Error bars are present for SNR50s but are small compared with the symbol size (all standard errors were less than 1 dB). SSC, speech spectrum continuous noise; 4TB, four-talker babble noise; 1TM, one-talker-modulated noise; YNH, young normal hearing; ONH, older normal hearing; OHI, older hearing impaired.



**Figure 4** SNR50 values for the four different speech-in-noise tests as a function of participant group. The overall difficulty posed by each test is reflected by the “center of gravity” of each cluster, with WIN and QuickSIN resulting in the worst SNR50 values and the CRM and LiSN resulting in the best SNR values. Variability within a test is also demonstrated by the spread of SNR50 values within and across participant groups. The separation between groups was most pronounced for the LiSN.

across tests. However, it is noted that participants performed most poorly (i.e., had the highest SNR50s) on the WIN, an open-set word test with no syntactic or semantic context clues. Next in difficulty was the QuickSIN, also an open-set test. It is important to note that because the QuickSIN signals are sentences, the participants are able to use context cues (especially syntactic context clues) to help them understand and repeat words that may not otherwise have been recognized. Generally, the participants found the CRM to be an easier

test, which is not surprising given that the CRM is a closed-set test with a limited number of colors and numbers to choose from. Finally, the easiest test was the LiSN-S in the different voice, 0-degree condition. In this case, the signal was again sentences, like the QuickSIN; however, the LiSN-S sentences may have more context clues than IEEE sentences. Perhaps more importantly, the noise type was only a single talker, which may lead to being able to focus in on the target signal easier than in the multitalker babble of the QuickSIN.



**Figure 5** SNR50 values for LiSN-S test results plotted as a function of Group and Condition. Notice the larger spread of performance for the older hearing-impaired group. Effects of spatial separation and talker are also apparent.

### Effects of Spatial Processing and Voice

Fig. 5 shows the SNR50 results for the LiSN-S test, which varies talker voice (masker same vs. different from target) and azimuth (0 vs. 90 degrees of signal-masker separation). A  $2 \times 2 \times 3$  repeated-measures ANOVA was completed to determine the effects of Voice, Azimuth, and Group. Main effects of Voice ( $F = 314.2$ ;  $df = 1, 18$ ;  $p < 0.0001$ ), Azimuth ( $F = 317.5$ ;  $df = 1, 18$ ;  $p < 0.0001$ ), and Group ( $F = 43.96$ ;  $df = 2, 27$ ;  $p < 0.0001$ ) were found to be statistically significant. In addition, two-way interactions between Voice and Group ( $F = 15.22$ ;  $df = 2, 45$ ;  $p < 0.0001$ ), Azimuth and Group ( $F = 41.27$ ;  $df = 2, 45$ ;  $p < 0.0001$ ), and Voice and Azimuth ( $F = 32.60$ ;  $df = 2, 36$ ;  $p < 0.0001$ ) were also found to be statistically significant. The three-way interaction was not found to be significant ( $F = 1.061$ ;  $df = 2, 81$ ;  $p = 0.361$ ). From Fig. 5, it is likely that the same voice, 0-degree condition (SV0) played an important role in the two-way interactions, demonstrating poorer SNR50s than the other conditions and more overlap across groups.

### DISCUSSION AND CONCLUSION

Speech-in-noise testing has been a proposed part of the audiological test battery from the inception of the field after WWII (Carhart 1946; Davis et al. 1946). With the development and release of several commercially available tests in the last two decades, speech-in-noise

testing has increased (ASHA 2019). However, it could be argued that the proportion of use still lags behind the proportion of individuals who have special difficulties hearing in noise. Survey data obtained by the American Speech-Language-Hearing Association show that only about a third of audiologists report using speech-in-noise testing (ASHA 2019); however, more than 90% of patients report at least some difficulty hearing speech in noise (Kochkin 2010). The purpose of this article was to provide a resource for hearing health professionals seeking to learn more about speech-in-noise testing and to provide some data that illustrate some of the main considerations that are important when selecting tests to use.

### Benefits of Testing

There are many advantages to including speech-in-noise testing in hearing healthcare. Perhaps foremost among these advantages is the face validity of using a test that corresponds to one of the primary complaints of those seeking audiological care—that of understanding speech in background noise. The field of audiology has focused almost exclusively on performance near threshold or in quiet, likely due to the fact that treatment options have historically been mostly limited to increasing levels through basic amplification strategies. Certainly, the most important treatment for most individuals with hearing loss is restoring

audibility. However, more advanced noise reduction technologies and specialized HATS are providing opportunities for specialized auditory care. Therefore, to tailor specialized solutions to patients it will be important to include specialized testing such as speech-in-noise testing. For example, the QuickSIN manual currently recommends different treatments for different speech-in-noise test outcomes (e.g., directional microphones for moderate SNR loss versus FM systems for severe SNR loss).

A major challenge in hearing care is the wide range of variability that is seen between individuals, even individuals with very similar performance in quiet or pure-tone audiometric thresholds. Speech-in-noise testing gives the audiologist a direct measure of speech understanding in more complex environments that are more like the everyday environments in which patients experience difficulty. Assessing and treating speech-in-noise difficulties is currently among the most important challenges being addressed in the world of auditory research, with the potential to directly impact hearing health care by improving patient communication in daily life. At the very least, inclusion of speech-in-noise testing provides audiologists with a strong foundation for patient counseling and education. This is critical due to the variability of speech-in-noise test results among even those who have very similar pure-tone thresholds (Wilson 2004). Setting realistic expectations and helping patients to understand their difficulties is an important benefit of including speech-in-noise testing as part of routine clinical protocols. The possible treatments of those difficulties may need to include high-level noise reduction technology (be it noise reduction in hearing aids or assistive devices such as remote microphones), aural rehabilitation, focused communication strategies, family counseling, or any combination of solutions. Speech-in-noise testing along with case history can help deduce which strategies will be necessary for a particular patient. For example, sentence-level tests like QuickSIN and HINT can inform how patients communicate in a crowded restaurant, while a word test like the WIN can help reduce the effect of higher-order cognitive deficits on testing.

## Barriers to Testing

A discussion of this topic must also address the barriers that have so far prevented speech-in-noise testing from becoming the norm in most hearing clinics. Barriers like time and money are of great concern not only for those who rely on hearing-aid sales for income, but for all audiologists as populations with hearing loss continue to grow. The cost of training and learning new test procedures may also play a role in whether clinicians choose to perform speech-in-noise testing. Clinicians may feel uncomfortable using speech-in-noise testing because they do not have enough training or feel confident in selecting which test(s) to use. The inclusion of speech-in-noise testing as a topic of instruction in audiology programs is an important step toward gaining broader acceptance and use in audiology. For those who are beyond their degree programs, Table 1 and **Appendix A** were created as quick guides for education and selection of appropriate test materials with the goal of increasing use of speech-in-noise tests clinically.

Appointment length for hearing tests and hearing-aid fittings varies by setting; so, it is difficult to quantify the impact of adding an additional test to an audiologist's battery. That said, the thought of one more test which takes as few as 2 but as many as 15 minutes to conduct could be extremely difficult to implement into a schedule which is often already tight. However, speech-in-noise testing can also be completed quickly with some currently available tests (e.g., QuickSIN can take < 5 minutes), and the case has been made that speech-in-noise testing could replace a speech-in-quiet test (Taylor, 2003; Wilson 2011), which would result in no added test time to the appointment.

Cost can be a barrier in multiple ways. First, there is no specific billing code for speech-in-noise testing; it is typically either included in 92557 (comprehensive audiometry) or 92556 (speech audiometry threshold; with speech recognition) when assessing for hearing aid candidacy (ASHAc n.d.). Clinicians may find justification for using 92700, which is used for otorhinolaryngological procedures without a listed code (ASHAc n.d.). In the case of cochlear-implant candidacy and monitoring of

outcome performance, the clinician may utilize 92626 (evaluation of auditory function for surgically implanted devices); however, this is a time-based code and may not be widely covered by private insurers. This code covers the first 60 minutes of evaluation and can only be billed if at least 31 minutes is spent completing the evaluation of auditory function. Second, it can be costly (or otherwise difficult) to acquire test materials; for example, the ANL must be purchased along with Interacoustics equipment, and is not sold separately. Of course, clinicians could create their own assessments of speech understanding in noise; however, there is a benefit to using commercially available tests with norms to which patient performance can be compared.

Addressing these barriers has the potential to significantly increase the use of speech-in-noise tests. And, thankfully, the use of speech-in-noise tests is increasing; ASHA surveys of audiologists from 2014 to 2018 showed that the percentage of respondents who used speech-in-noise testing to validate hearing aids rose from 30 to 34% (ASHA 2019).

### Future Directions

More normative data are needed to improve the usefulness of speech-in-noise tests. Some of the tests that are presented in Table 1 and **Appendix A** have limited normative data. It will be important for additional testing and research to explore effects of aging, hearing impairment, and other conditions so that a patient's score can be compared with the general population or to subpopulations.

Speech-in-noise testing will likely be more useful in some situations and for some individuals than for others. Cochlear-implant candidates, hearing-aid users, and individuals with special listening difficulties are three such examples. Therefore, clinicians and researchers can benefit from carefully considering for which patients and under what conditions speech-in-noise testing would be beneficial. Another priority should be addressing the barrier of cost and lack of reimbursement for testing.

Speech-in-noise testing can be very helpful in audiology, with the potential to improve and augment auditory assessment and treatment in

some situations. It is important for the clinician to carefully consider the advantages and disadvantages of speech-in-noise testing in their own particular clinic and setting. In some populations, speech-in-noise testing is a vital component of the candidacy evaluation or the monitoring of auditory treatment (e.g., cochlear implantation, auditory processing disorders). Unfortunately, there is no agreed-upon standard for speech-in-noise testing; instead there are several tests to choose from with varying amounts of literature and data to support them. Nonetheless, given that understanding speech in noise is often one of the most difficult listening situations for patients, it is clear that audiologists who want to tailor treatment to the needs of individual patients will find speech-in-noise testing to be an important tool in providing top-quality clinical assessment and treatment.

### FUNDING/ACKNOWLEDGMENTS

This work was supported by the United States (U.S.) National Institutes of Health (NIDCD-R03DC10914) and by the U.S. Department of Veterans Affairs (RR&D-5I01RX003702). The contents do not represent the views of the U.S. Department of Veterans Affairs or the U.S. government. The authors would also like to acknowledge the contributions of Nashrah Maamor, Tina Penman, and Paul Pendergraft.

### CONFLICT OF INTEREST

None declared.

### REFERENCES

- American Academy of Audiology (AAA). (2010). American academy of audiology clinical practice guidelines: diagnosis, treatment and management of children and adults with central auditory processing disorder. Accessed May 31, 2023 at: [https://audiology-web.s3.amazonaws.com/migrated/CAPD%20Guidelines%208-2010.pdf\\_539952af956c79.73897613.pdf](https://audiology-web.s3.amazonaws.com/migrated/CAPD%20Guidelines%208-2010.pdf_539952af956c79.73897613.pdf)
- American Speech-Language-Hearing Association (ASHA). (2019). Audiology survey report: clinical focus patterns, 2014–2018. Accessed April 2, 2022 at: <https://www.asha.org/siteassets/surveys/2018-audiology-survey-clinical-focus-patterns-trends.pdf>

- American Speech-Language-Hearing Association (ASHA). (2005a). (Central) Auditory Processing Disorders. Accessed April 2, 2022 at: <http://www.asha.org/policy>
- American Speech-Language-Hearing Association (ASHA). (2005b). Billing and Coding for Audiology Services, N.D. Accessed April 2, 2022 at: <https://www.asha.org/practice/reimbursement/audiology-billing-and-coding-for-services-faqs/>
- American Speech-Language Hearing Association (ASHA). (2005c). Medicare CPT Coding Rules for Audiology Services. Accessed August 29, 2022 at: [https://www.asha.org/practice/reimbursement/medicare/aud\\_coding\\_rules/](https://www.asha.org/practice/reimbursement/medicare/aud_coding_rules/)
- Audiology Practice Standards Organization. (2021). Hearing Aid Fitting Standard for Adult and Geriatric Patients. Accessed April 2, 2022 at: <https://www.audiologystandards.org/standards/display.php?id=102>
- Auditec of St. Louis. (1971). Four-Talker Babble. St. Louis, MO, 63143–2105
- Auditory Potential. (2022). Distributor of AzBio sentence lists. Accessed April 2, 2022 at <https://www.audiorypotential.com/>
- Bench J, Kowal A, Bamford J. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *Br J Audiol*, 13(03):108–112
- Bilger RC, Nuetzel JM, Rabinowitz WM, Rzezczkowski C. (1984). Standardization of a test of speech perception in noise. *J Speech Hear Res*, 27(01):32–48
- Billings CJ, Bennett KO, Molis MR, Leek MR. (2011). Cortical encoding of signals in noise: effects of stimulus type and recording paradigm. *Ear Hear*, 32(01):53–60
- Billings CJ, Madsen BM. (2018). A perspective on brain-behavior relationships and effects of age and hearing using speech-in-noise stimuli. *Hear Res*, 369:90–102
- Billings CJ, Penman TM, Ellis EM, Baltzell LS, McMillan GP. (2016). Phoneme and Word Scoring in Speech-in-Noise Audiometry. *Am J Audiol*, 25(01):75–83
- Bolia RS, Nelson WT, Ericson MA, Simpson BD. (2000). A speech corpus for multitalker communications research. *J Acoust Soc Am*, 107(02):1065–1066
- Brown DK, Cameron S, Martin JS, Watson C, Dillon H. (2010). The North American Listening in Spatialized Noise-Sentences test (NA LiSN-S): normative data and test-retest reliability studies for adolescents and young adults. *J Am Acad Audiol*, 21(10):629–641
- Brungart DS. (2001a). Evaluation of speech intelligibility with the coordinate response measure. *J Acoust Soc Am*, 109(5, Pt 1):2276–2279
- Brungart DS. (2001b). Informational and energetic masking effects in the perception of two simultaneous talkers. *J Acoust Soc Am*, 109(03):1101–1109
- Brungart DS, Simpson BD, Ericson MA, Scott KR. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *J Acoust Soc Am*, 110(5, Pt 1):2527–2538
- Brungart DS, Walden B, Cord M, et al. (2017). Development and validation of the Speech Reception in Noise (SPRINT) Test. *Hear Res*, 349:90–97
- Byrne JET, Kerr AG. (1987). Deafness with normal pure tone audiometry. In: Kerr AG, Groves J, Booth JB, eds. *Scott Brown's Otolaryngology*. Great Britain, UK: Butterworth and Co.; 383–384
- 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. *J Am Acad Audiol*, 20(02):128–146
- Cameron S, Dillon H. (2007). Development of the listening in spatialized noise-sentences test (LISN-S). *Ear Hear*, 28(02):196–211
- Cameron S, Glyde H, Dillon H. (2012). Efficacy of the LiSN & Learn auditory training software: randomized blinded controlled study. *Audiology Res*, 2(01):e15
- Cameron S, Glyde H, Dillon H. (2011). Listening in Spatialized Noise-Sentences Test (LiSN-S): normative and retest reliability data for adolescents and adults up to 60 years of age. *J Am Acad Audiol*, 22(10):697–709
- Carhart R. (1946). Tests for selection of hearing aids. *Laryngoscope*, 56(12):780–794
- Carhart R. (1951). Basic principles of speech audiometry. *Acta Otolaryngol*, 40(1-2):62–71
- Carlson ML, Sladen DP, Gurgel RK, Tombers NM, Lohse CM, Driscoll CL. (2018). Survey of the American Neurotology Society on Cochlear Implantation: Part 1, candidacy assessment and expanding indications. *Otol Neurotol*, 39(01):e12–e19
- Chermak GD, Iliadou V, Bamiau D, Musiek FE. (2018). Letter to the editor: response to Vermiglio, 2018. Perspectives of the ASHA Special Interest Groups, SIG6, 3(02):77–82
- Cherry EC. Some experiments on the recognition of speech, with one and with two ears. *J Acoust Soc Am*, 25(05):975
- Clark JG, Huff C, Earl B. (2017). Clinical practice report card – Are we meeting best practice standards for adult hearing rehabilitation? *Audiol Today*, 29(06):15–25
- Cord MT, Walden BE, Attack RM. (1992). Speech Recognition in Noise Test (SPRINT) for H-3 Profile. Walter Reed Army Medical Center
- Cox RM, Alexander GC, Gilmore C. (1978). Development of the Connected Speech Test (CST). *Ear Hear*, 8(5, Suppl):119S–126S
- Cox RM, Alexander GC, Gilmore C, Pusakulich KM. (1988). Use of the Connected Speech Test (CST) with hearing-impaired listeners. *Ear Hear*, 9(04):198–207



- Davidson A, Marrone N, Wong B, Musiek F. (2021). Predicting hearing aid satisfaction in adults: a systematic review of speech-in-noise tests and other behavioral measures. *Ear Hear*, 42(06):1485–1498
- Davis H, Hudgins CV, Marquis RJ, et al. (1946). The selection of hearing aids. *Laryngoscope*, 56(03):85–115
- Durlach NI, Mason CR, Kidd G Jr, Arbogast TL, Colburn HS, Shinn-Cunningham BG. (2003). Note on informational masking. *J Acoust Soc Am*, 113(06):2984–2987
- Etymotic Research. (2005). Bamford-Kowal-Bench Speech-in-Noise Test (Version 1.03) [Audio CD and Test Manual]. Elk Grove Village, IL
- Etymotic Research. (2006). QuickSIN Speech-in-Noise Test (Version 1.3) [Audio CD and Test Manual]. Elk Grove Village, IL
- Gifford RH, Shalloo JK, Peterson AM. (2008). Speech recognition materials and ceiling effects: considerations for cochlear implant programs. *Audiol Neurotol*, 13(03):193–205
- Hardy WG. (1950). Hearing aids; procedures for testing and selection. *Postgrad Med*, 7(01):11–17
- Hearing Aid Research Lab (HARL). (n.d.). Connected Speech Test (CST) [Website]. Accessed June 10, 2022 at: <https://harlmemphis.org/connected-speech-test-cst/>
- Hinchcliffe R. (1992). King-Kopetzky syndrome: an auditory stress disorder. *J Audiol Med*, 1:89–98
- Hodgson M, Rempel R, Kennedy S. (1999). Measurement and prediction of typical speech and background-noise levels in university classrooms during lectures. *J Acoust Soc Am*, 105:226–233
- Holder JT, Levin LM, Gifford RH. (2018). Speech recognition in noise for adults with normal hearing: Age-normative performance for AzBio, BKB-SIN, and QuickSIN. *Otol Neurotol*, 39(10):e972
- Institute of Electrical and Electronic Engineers. (1969). IEEE Recommended Practice for Speech Quality Measures. New York: IEEE
- Interacoustics (n.d.). Acceptable Noise Level (ANL) test [Website]. Accessed January 20, 2022 at: <https://www.interacoustics.com/guides/test/audiometry-tests/acceptable-noise-level-anl-test>
- Jakien KM, Kampel SD, Stansell MM, Gallun FJ. (2017). Validating a rapid, automated test of spatial release from masking. *Am J Audiol*, 26(04):507–518
- Jayaram M, Baguley DM, Moffat DA. (1992). Speech in noise: a practical test procedure. *J Laryngol Otol*, 106(02):105–110
- Kalikow DN, Stevens KN, Elliott LL. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *J Acoust Soc Am*, 61(05):1337–1351
- Kidd G, Mason CR, Richards VM, et al. (2008). Informational masking. In: Yost WA, Popper AR, Fay RR, (eds.). *Auditory Perception of Sound Sources*. New York: Springer; 143–189
- Killion MC, Niquette PA, Gudmundsen GI, Revit LJ, 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, Pt 1):2395–2405
- King PF. (1954). Psychogenic deafness. *J Laryngol Otol*, 68(09):623–635
- Kochkin S. (2010). MarkeTrak VIII: consumer satisfaction with hearing aids is slowly increasing. *Hear J*, 63(01):19–32
- Kopetzky SJ. (1948). *Deafness, Tinnitus, and Vertigo*. New York, NY: Thomas Nelson & Sons
- Kuk F, Korhonen P. (2018). Using tracking of noise tolerance (TNT) as an outcome measure for hearing aids. *Hear Rev*, 25:16–23
- Markides A. (1986). Age at fitting of hearing aids and speech intelligibility. *Br J Audiol*, 20(02):165–167
- Mealings K, Yeend I, Valderrama JT, Valderrama JT, et al. (2020). Discovering the unmet needs of people with difficulties understanding speech in noise and a normal or near-normal audiogram. *Am J Audiol*, 29(03):329–355
- Middelweerd MJ, Festen JM, Plomp R. (1990). Difficulties with speech intelligibility in noise in spite of a normal pure-tone audiogram. *Audiology*, 29(01):1–7
- Miller GA. (1947). The masking of speech. *Psychol Bull*, 44(02):105–129
- Moore TJ. (1981). Voice communications jamming research. *AGARD Conference Proceedings*;311;2(1)–2(6)
- MSTB. New Minimum Speech Test Battery for Adult Cochlear Implant Users. 2011;1:1–15. Accessed July 18, 2022 at: <http://www.auditorypotential.com/MSTBfiles/MSTBManual2011-06-20%20.pdf>
- Mueller HG. (2003). Fitting test protocols are “more honored in the breach than the observance”. *Hear J*, 56(10):19–26
- Mueller HG. (2010). Three pre-tests: What they do and why experts say you should use them more. *Hear J*, 63(04):8–17
- Nabelek AK, Tucker FM, Letowski TR. (1991). Tolerant of background noises: relationship with patterns of hearing aid use by elderly persons. *J Speech Hear Res*, 34(03):679–685
- Nabelek AK, Freyaldenhoven MC, Tampas JW, Burchfiel SB, Muenchen RA. (2006). Acceptable noise level as a predictor of hearing aid use. *J Am Acad Audiol*, 17(09):626–639
- Narula AA, Mason SM. (1988). Selective dysacusis – a preliminary report. *J R Soc Med*, 81(06):338–340
- Nilsson M, Soli SD, Sullivan JA. (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(02):1085–1099
- Pearsons KS, Bennett RL, Fidell S. (1977). *Speech Levels in Various Noise Environments: Environmental Health Effects Research Series*. National Technical Information Service, Springfield

- Plomp R. (1977). Acoustical aspects of cocktail parties. *Acoustica*, 38:186–191
- Plomp R. (1978). Auditory handicap of hearing impairment and the limited benefit of hearing aids. *J Acoust Soc Am*, 63(02):533–549
- Prentiss S, Snapp H, Zwolan T. (2020). Audiology practices in the preoperative evaluation and management of adult cochlear implant candidates. *JAMA Otolaryngol Head Neck Surg*, 146(02):136–142
- Prins N, Kingdom FAA. (2018). Applying the model-comparison approach to test specific research hypotheses in psychophysical research using the Palamedes toolbox. *Front Psychol*, 9:1250
- Pryce H, Wainwright D. (2008). Help-seeking for medically unexplained hearing difficulties: a qualitative study. *Int J Ther Rehabil*, 15(08):343–349
- Rappaport JM, Phillips DP, Gulliver JM. (1993). Disturbed speech intelligibility in noise despite a normal audiogram: a defect in temporal resolution? *J Otolaryngol*, 22(06):447–453
- Rendell RJ, Stephens SDG. (1988). Auditory disability with normal hearing. *Br J Audiol*, 22:233–234
- Spahr AJ, Dorman MF, Litvak LM, et al. (2012). Development and validation of the AzBio sentence lists. *Ear Hear*, 33(01):112–117
- Saunders GH, Haggard MP. (1989). The clinical assessment of obscure auditory dysfunction—1. Auditory and psychological factors. *Ear Hear*, 10(03):200–208
- Saunders GH, Cienkowski KM. (2002). A test to measure subjective and objective speech intelligibility. *J Am Acad Audiol*, 13(01):38–49
- Schafer EC, Pogue J, Milrany T. (2012). List equivalency of the AzBio sentence test in noise for listeners with normal-hearing sensitivity or cochlear implants. *J Am Acad Audiol*, 23(07):501–509
- Saunders GH, Forsline A, Fausti SA. (2004). The performance-perceptual test and its relationship to unaided reported handicap. *Ear Hear*, 25(02):117–126
- Saunders GH, Forsline A. (2006). The Performance-Perceptual Test (PPT) and its relationship to aided reported handicap and hearing aid satisfaction. *Ear Hear*, 27(03):229–242
- Sladen DP, Gifford RH, Haynes D, et al. (2017). Evaluation of a revised indication for determining adult cochlear implant candidacy. *Laryngoscope*, 2017;127(10):2368–2374
- Smeds K, Wolters F, Rung M. (2015). Estimation of signal-to-noise ratios in realistic sound scenarios. *J Am Acad Audiol*, 26(02):183–196
- Stephens SDG. (1976). The input for a damaged cochlea—a brief review. *Br J Audiol*, 10:97–101
- Strom KE. (2006). The HR 2006 dispensing survey. *Hearing Review*, 13:16–39
- Studebaker GAA. A “rationalized” arcsine transform. *J Speech Hear Res*, 1985;28(03):455–462
- Taylor B. (2003). Speech-in-noise tests: How and why to include them in your basic test battery. *The Hearing Journal*, 56(01):40,42–46
- Teder H. (1990). Noise and speech levels in noisy environments. *Hearing Instruments*, 41(04):32–33
- U.S. Department of the Air Force. Manual 48–123. Medical Examinations and Standards, December 8, 2020. Accessed May 31, 2023 at: [https://static.e-publishing.af.mil/production/1/af\\_sg/publication/dafman48-123/dafman48-123.pdf](https://static.e-publishing.af.mil/production/1/af_sg/publication/dafman48-123/dafman48-123.pdf)
- U.S. Department of the Army. Army Regulation 40–501. Standards of Medical Fitness, 4 August 4, 2011. Accessed May 31, 2023 at: [https://www.qmo.amedd.army.mil/diabetes/AR40\\_5012011.pdf](https://www.qmo.amedd.army.mil/diabetes/AR40_5012011.pdf)
- U.S. Department of the Navy. NMCPHC TM-6260.51.99–3. Navy Medicine Hearing Conservation Program Technical Manual, July 2020. Accessed May 31, 2023 at: <https://www.med.navy.mil/Portals/62/Documents/NMFA/NMCPHC/root/Documents/health-promotion-wellness/general-tools-and-programs/blue-h-worksheet-semper-fit-Dec2020.xlsx>
- Vermiglio AJ. (2018a). The gold standard and auditory processing disorder. *Perspectives of the ASHA Special Interest Groups*, SIG6, 3(1):6–17
- Vermiglio AJ. (2018b). Response to the Letter to the Editor from Chermak. *Perspectives of the ASHA Special Interest Groups*, SIG6, 3(2):83–90
- Vermiglio AJ. (2008). The American English hearing in noise test. *Int J Audiol*, 47(06):386–387
- Vermiglio AJ, Leclerc L, Thornton M, Osborne H, Bonilla E, Fang X. (2021). Diagnostic accuracy of the AzBio Speech Recognition in Noise Test. *J Speech Lang Hear Res*, 64(08):3303–3316
- Wilson RH. (2003). Development of a speech-in-multitalker-babble paradigm to assess word-recognition performance. *J Am Acad Audiol*, 14(09):453–470
- Wilson RH. (2004). Adding speech-in-noise testing to your clinical protocol: Why and how. *Hear J*, 57(02):10–18
- Wilson RH, Burks CA. (2005). Use of 35 words for evaluation of hearing loss in signal-to-babble ratio: A clinic protocol. *J Rehabil Res Dev*, 42(06):839–852
- Wilson RH. (2011). Clinical experience with the words-in-noise test on 3430 veterans: comparisons with pure-tone thresholds and word recognition in quiet. *J Am Acad Audiol*, 22(07):405–423
- Wilson RH, McArdle R. (2005). Speech signals used to evaluate functional status of the auditory system. *J Rehabil Res Dev*, 42(4, Suppl 2):79–94
- Wingfield A. (1996). Cognitive factors in auditory performance: context, speed of processing, and constraints of memory. *J Am Acad Audiol*, 7(03):175–182

## APPENDIX A: SUMMARY OF AVAILABLE SPEECH-IN-NOISE TESTS

### Acceptable Noise Level (ANL)

**Purpose:** To measure a patient's tolerance of background noise and estimate their likelihood of successful hearing aid use (Nabelek et al. 1991). May also aid clinicians in determining the level of hearing aid technology a patient requires (Interacoustics n.d.).

**Materials:** Running speech presented in babble noise.

**Administration:** The test takes 5–10 minutes to administer. First, most comfortable listening level (MCL) is measured in quiet by playing a passage through a loudspeaker, during which the patient is directed to signal to the clinician whether to increase or decrease the volume until the MCL is achieved. Background noise is then added, and the patient is asked to indicate the “maximum level of noise that you would be willing to put up with for a long time while following the story.”

**Scoring:** The ANL value is calculated by subtracting the background noise level from the MCL (MCL background noise level = ANL). The smaller the ANL value, the better the predicted outcomes the patient has with hearing aids (Nabelek et al. 1991). (Note: for this calculation to be valid, both ANL and MCL must be quantified on a decibel scale relative to the same reference—e.g., dB HL for both or dB SPL for both, but not one of each.)

**Norms:** Scores of 7 and below are considered to predict good outcomes (Interacoustics n.d.; Nabelek et al. 2006). A higher value indicates the patient likely needs more counseling and/or noise reduction technology. Scores of 12.5 dB or higher are considered to predict poor outcomes. Using Fig. 2 in Nabelek et al. (2006), clinicians can estimate the likelihood of hearing aid success based on ANL score.

**Miscellaneous:** The official test from Interacoustics can be found as part of their AC440 audiometry module for multiple systems, but it is not sold separately. Clinicians may use their own clinic-available materials to administer a form of this test, with the caveat that norms may not be comparable. Note: The ANL has inspired the development of a similar test of noise tolerance, Tracking of Noise Tolerance Test, developed by Francis Kuk and colleagues (Kuk et al. 2018; the interested reader is also referred to <https://www.orca-us.info/en/research> for more information about this test).

### Arizona Biomedical Sentences Test (AzBio)

**Purpose:** To provide an objective estimate of listening ability in noise that lines up with how the patient perceives their own ability in typical environments. Has been adopted as a core component of the Minimum Speech Test Battery (MSTB 2011) for cochlear implant candidacy evaluations.

**Materials:** There are 15 AzBio sentence lists distributed for clinicians to utilize via CD and an additional 8 lists included in the MSTB. Each list contains 20 sentences in total, spoken by two males and two females, with 5 sentences recorded per speaker. The pace and volume were analyzed and found to be comparable to typical conversational speech (Spahr et al. 2012). The subject matter includes current social ideas and adult topics, while providing limited contextual cues to limit the listener's ability to “fill in” unintelligible words. List equivalency was demonstrated for the 15 lists (Spahr et al. 2012) and then reduced further to a subset of 10 lists (2, 3, 4, 5, 8, 9, 10, 11, 13, and 15) using a different statistical approach (Schafer et al. 2012).

**Administration:** The AzBio is a fixed-SNR test in which the clinician sets the SNR for their intended purpose, typically at +10, +5, 0, 5, or –10 dB. The sentences and 10-talker babble are often presented in sound field from a single loudspeaker at a 0-degree azimuth. However, the CD allows for the separation of the speech from noise to different channels if desired. Most research utilizes free-field loudspeakers; however, bilateral presentation under headphones is possible (Vermiglio 2021). Sentence presentation level should be set based on the objective of testing. For example, for testing conversational speech, a 60-dB SPL level could be used.

**Scoring:** Each word in the sentence is scored as entirely correct or incorrect. To obtain the overall score, the total number of correct responses is divided by the total number of words presented (which is noted at the end of each list). The resulting proportion value is multiplied by 100 to convert it to a percent-correct score.

**Norms:** The findings of Holder et al. (2018) suggest that participants without hearing difficulties are expected to score near ceiling at +10 and +5 dB SNRs regardless of age. Therefore, they argue, poorer scores at these SNRs are indicative of hearing difficulties and should not be attributed to age alone.

### **Bamford-Kowal-Bench Speech in Noise (BKB-SIN)**

**Purpose:** Created for administering to adults and children, as young as age 5, that may find the QuickSIN sentences too difficult (Bench et al. 1979). Recommended uses include demonstrating the benefits of amplification, predicting performance in noisy environments, testing cochlear-implant candidates or users, and screening children for auditory processing disorders (Etymotic Research 2005).

**Materials:** Paired lists of Bamford-Kowal-Bench (BKB) sentence lists; 18 lists are available. Difficulty equivalency was established at the level of the list pairs, not the individual sentences; therefore, in order for test results to be considered valid, both lists in a given pair must be administered, and the scores averaged. Each target sentence is preceded by the cue “ready” and presented in four-talker babble created by Auditec of St. Louis (1971). The BKB sentences use naturally connected speech that is meant to have high levels of redundancy and employ vocabulary common among children 8–15 years old, with the intention that test results be driven by ability to hear speech rather than linguistic ability (Bench et al. 1979). The BKB-SIN is progressive, with different ranges of steadily decreasing SNRs depending on the list. For List Pairs 1–8, there are 10 sentences in each list, with SNRs ranging from +21 to 6 dB in 3-dB steps. For most of the SNRs, the level of the target sentences remains constant at the nominal presentation level while the level of the babble progressively increases. For the negative SNRs (3 and 6 dB), however, to keep the overall presentation level similar to the other trials, the level of the babble is kept the same as it was for 0 dB SNR while the level of the target sentence is lowered. For List Pairs 9–18, there are 8 sentences with SNRs ranging from +21 to 0 dB, again in 3-dB steps. These shorter lists are not recommended for listeners with normal hearing because of the risk of floor effects on SNR50 (i.e., ceiling effects in terms of performance) (Etymotic Research 2005).

**Administration:** The test manual (Etymotic Research 2005) estimates administration and scoring to take 3 minutes for each pair of sentence lists. The presentation level and transducer should be chosen based on the purpose of the test and the patient being tested. Etymotic’s normative data were based on a presentation level of 65 dB SPL over loudspeaker for cochlear implant users and 70 dB HL over insert earphones for adults and children with normal hearing. However, the BKB test manual notes that in their laboratory’s experience, any level from 50 to 70 dB HL gives similar results for listeners with normal hearing. Normative data were collected from adults and children with normal hearing using binaural insert earphones at a presentation level of 70 dB HL. On adult cochlear implant users, normative data were obtained in sound field using 65 dB SPL. It is recommended to set the level above 70 dB HL if a hearing loss is more than mild.

**Scoring:** Scoring is based on the number of underlined key words from the sentence list that were correctly repeated by the patient. The first sentence in each list has four key words, and all subsequent sentences have three key words. The total number of correctly repeated key words is subtracted from 23.5 to obtain the patient’s SNR50 for that list. The SNR50 estimates obtained from each of the two lists in the pair are then averaged together; if more than one list pair was administered, then the results from each pair are further averaged together.

**Norms:** The test manual offers norms to which an adult or child BKB-SIN SNR50 can be compared with to discern a potential “SNR loss” (i.e., a higher-than-expected SNR50). These age-based norms, in terms of means and standard deviations are: for adults without hearing loss,  $2.5 \pm 0.8$  dB; for children 5–6 years old,  $3.5 \pm 2.0$  dB; for children 7–10 years old,  $0.8 \pm 1.2$  dB; for children 11–14 years old,  $0.9 \pm 1.1$  dB. SNR loss can be calculated by subtracting an individual’s score from their age group’s mean. SNR loss can be interpreted as the amount that SNR would need to be increased for the patient’s receptive speech-in-noise performance to be similar to that of an average person with normal hearing in their age group.

### Connected Speech Test (CST)

**Purpose:** Everyday speech intelligibility test which was originally designed for measuring hearing aid benefit. It can be adjusted to fit clinician needs (e.g., as pre- and post-fitting measure).

**Materials:** Speech includes passages recorded by a female speaker; subjects of the passages are familiar topics, and each contains 10 simple sentences of 7 to 10 words. The noise signal is a six-talker speech babble.

**Administration:** Requires less than 10 minutes if presenting two passages to each ear. There are no standard procedures available for this test; however, a summary of the original experimental procedures is provided here (HARL n.d.). In the study with normal hearing listeners, four CST passages were presented monaurally at 61 dB Leq with a 4 dB SNR. The passage was paused between each sentence to allow the patient to respond. In the study with hearing impaired listeners, the level of the passage was established as the level of conversational speech (61 dB Leq) added to half of the participant’s measured SRT. The SNR was then adjusted during the presentation of the practice passages with the aim of achieving a score between 50 and 80%; therefore, SNR was set individually (Cox et al. 1987).

**Scoring:** Each of the 25 key words per passage is recorded as correct or incorrect and the percent correct score for the passage is obtained. Scores can then be averaged to obtain an overall score. In the original experiments, the researchers used rationalized arcsine units (RAU) in lieu of percentage scores (Studebaker 1985).

**Norms:** No norms exist for this test at the time of writing this article. It has been validated with both normal-hearing and hearing-impaired listeners (Cox et al. 1988).

**Miscellaneous:** The audio for this test is available to download for free at the following link: <https://harlmemphis.org/connected-speech-test-cst/>

This link also offers administration and scoring software available for a \$50 purchase fee.

### Coordinate Response Measure (CRM)

**Purpose:** Highly customizable test of listening in a multi-talker situation based on experiments by Moore (1981). Originally created as an experiment to test spatial hearing in the presence of different types of noise.

**Materials:** Formulaic sentences following the pattern “Ready [CALL SIGN], go to [COLOR] [NUMBER] now” using every permutation of 8 call signs (from the N North Atlantic Treaty Organization (NATO) phonetic alphabet), 4 colors (red, green, blue, white), and numbers 1 through 8, with recordings of each taken from 8 different talkers (4 males, 4 females). In the most common paradigm, recordings from multiple talkers are presented at once, each saying a different call sign, color, and number; participants listen for their assigned call sign and select the color-number combination (“coordinates”) corresponding to that call sign.

**Administration:** The test can be adaptive or progressive according to a researcher’s needs.

**Scoring:** Scoring can also be adapted to the situation; for example, researchers can measure the number of correct call-sign identifications; obtain a 50% correct threshold for call sign and/or



color-number combinations (Brungart 2001a); reaction times (Bolia 2000); and binaural and monaural cues, including spatial release from masking (Brungart 2001b; Brungart et al. 2001).

**Norms:** Norms are not established for all test variations of the CRM; however, Jakien et al (2017) conducted a series of tests comparing spatial release from masking over headphones versus in an anechoic chamber. These tests showed that conducting spatial release from masking testing using headphones yielded similar results as those obtained in sound field.

**Miscellaneous:** The test is available for free by contacting Robert S. Bolia. It is also included as the Portable Automated Rapid Testing (PART) application (search “Portable Psychoacoustics” in the Apple App store) available for download.

### Hearing in Noise Test (HINT)

**Purpose:** Designed by Nilsson et al. (1994) for measuring the SNR threshold for sentence intelligibility in noise. It was created for American English speakers to measure functional hearing ability by avoiding the floor and ceiling effects of traditional fixed speech-in-noise measures, although its primary clinical function is to determine the benefit of directional microphones.

**Materials:** Modified versions of the BKB sentences, revised to be more reflective of American English and recorded by a male native speaker. The masking noise was designed to have the same long-term average speech spectrum as the recorded sentences.

**Administration:** Takes 5–10 minutes to administer. The level of the speech-spectrum noise remains constant at 72 dBA, while the level of the target sentences is adaptively varied based on patient’s performance: if the whole sentence is repeated correctly, the level goes down; otherwise, it goes up. This process continues until a 50% correct response rate is obtained.

**Scoring:** Whole sentences are scored based on correct repetition.

**Norms:** Norms for this test were created based on experiments with young, normal-hearing listeners. Measurements were completed in multiple noise conditions under headphones. To reference these norms, see Table 1 in Vermiglio (2008) which contains a percentile score in relation to the dB SNR threshold at which 50% of sentences are repeated correctly. Those scoring above the 50th percentile have good speech understanding in noise abilities, while those scoring below the 50th percentile struggle with speech in noise.

### Listening in Spatialized Noise—Sentences (LiSN-S)

**Purpose:** The LiSN-S is an adaptive sentence test developed by the National Acoustic Laboratories in Australia. The original test with Australian speakers (Cameron & Dillon 2007) was modified with North American English (Cameron et al. 2009). Originally developed for use in auditory processing disorder testing, the LiSN-S tests the auditory system’s ability to make use of voice and spatial cues to perform auditory stream segregation for the purpose of enhancing understanding of target sentences in the presence of competing speech.

**Materials:** The LiSN-S comes in the form of a software package, which plays the stimuli and is used to record the responses, and reports the results when finished. Target sentences are presented adaptively with a simultaneous competing children’s story read by a single talker in four different conditions: two spatial conditions (0 vs. +90 degrees azimuth separation between signal and noise) and two talker conditions (targets and distractors read by same talker vs. two different talkers).

**Administration:** After each target sentence is presented, the patient attempts to repeat it, and the examiner counts the number of words repeated correctly and enters this number into the software graphical user interface. The program then automatically adjusts the level of the target up or down accordingly (while holding the level of the distractor story constant) to try to bracket

the 50% correct point, and the process repeats. Once a certain number of reversals have occurred, the algorithm stops and considers the current SNR to be the SNR50 for that condition.

**Scoring:** Scores of each individual trial are entered by the examiner as described earlier. Overall scoring is performed automatically by the LiSN software, which reports five different outcome measures: two SNR50s (which the LiSN-S calls *low-cue SRT* and *high-cue SRT*) and three advantage scores (*talker advantage*, *spatial advantage*, and *total advantage*). The two additional SNR50s that are calculated by the program but not explicitly reported (for the conditions with only spatial cues or only talker cues) can be manually calculated by the examiner if desired by taking the relevant “advantage” score and adding it to the high-cue SRT (or subtracting it from the low-cue SRT).

**Norms:** Norms and test–retest information for the North American version of the LiSN-S were reported for children and young adults (Brown et al. 2010; Cameron et al. 2009) ranging in age from 6 years to 30 years. Additional normative data for individuals up to 60 years of age using the Australian version of the LiSN-S may also be useful (Cameron et al. 2011).

**Miscellaneous:** Additional versions of the LiSN besides the “S” are available; for example, LiSN-U (for “universal”) uses CVCV nonsense words in an attempt to be more generally applicable across languages (though differences in languages’ phonemic inventories may make true cross-language equivalency difficult).

### Perceptual Performance Test (PPT)

**Purpose:** Used to determine whether a patient underestimates or overestimates their performance when listening to speech in noise (Mueller 2010).

**Materials:** Sentences and maskers from the HINT. For each run in the *perceptual* condition, two 12-sentence lists are combined, with extended inter-sentence pauses removed, resulting in a 24-sentence passage emulating connected speech. For each run in the *performance* condition, one 12-sentence HINT list is used in unmodified form (Saunders & Cienkowski 2002).

**Administration:** The lists are presented via loudspeaker, with the speech held at a fixed level (e.g., patient MCL to HINT materials in quiet; Saunders & Cienkowski 2002). In the *perceptual* condition, the patient nods or shakes their head to indicate whether they can understand all the speech at the current SNR. Each head nod results in an SNR decrease and each head shake results in an SNR increase. The examiner keeps a log of the SNR used for each sentence. In the *performance* condition, the procedure is the same except that the patient is asked to repeat each sentence instead of nodding or shaking their head. A fully correct repetition results in an SNR decrease, and a repetition with one or more errors results in an SNR increase. In both conditions, the step sizes are 4 dB during the first four sentences of each run and 2 dB thereafter. Two runs are administered per condition.

**Scoring:** The *perceptual* threshold for a 24-sentence run is calculated as the average SNR across the final 10 sentences. The *performance* threshold for a 12-sentence run is calculated as the average SNR across the final 6 sentences. For both perception and performance, the thresholds from two different runs are averaged together to get the overall threshold measurement for the condition. The *perceptual–performance discrepancy* (PPDIS) is calculated by subtracting the performance threshold from the perceptual threshold. Therefore, a positive PPDIS suggests the patient tends to underestimate their ability to understand speech in noise, while a negative PPDIS suggests the patient tends to overestimate it, and a PPDIS close to zero suggests a relatively accurate self-estimation. The relationship between reported unaided and aided handicap may be of interest as an outcome measure (Saunders et al. 2004; Saunders & Forsline, 2006).

**Norms:** Not available.

### Quick speech in noise (QuickSIN)

**Purpose:** Originally developed to provide a 1- to 2-minute estimation of SNR loss to quantify a patient's ability to hear in noise (Killion et al. 2004). It can also help professionals in recommending amplification and accessory devices as well as counseling patients to have realistic expectations (Etymotic Research 2006).

**Materials:** There are 12 lists each containing six Massachusetts Institute of Technology recordings of Institute of Electrical and Electronics Engineers (IEEE 1969) sentences spoken by a female talker presented in four-talker babble noise created by Auditec. The IEEE sentences reportedly do not contain highly predictable words with respect to the surrounding context, such that the participant must rely on acoustic cues (Holder et al. 2018). A single list consists of six sentences, presented at SNRs descending from +25 dB to 0 dB, in 5-dB steps, with one sentence at each SNR. In addition to the 12 standalone lists on the QuickSIN CD, there are three list pairs (13/14, 15/16, and 17/18, whose within-pair averages are considered equivalent to the other lists even though the individual lists in the pairs are not) and three practice lists (A, B, and C; not equivalent to the numbered lists or list pairs).

**Administration:** The test manual (Etymotic Research 2006) recommends presenting at a level that the patient reports to be "loud, but okay"; however, the level may ultimately be decided by the test administrator, depending on the purpose and the patient. The presentation level is generally set at 70 dB HL for listeners who have a pure-tone average below 45 dB HL. The test may also be presented at 60 dB SPL, consistent with MSTB clinical recommendations for cochlear-implant users (Holder et al 2018). Performance differences between 60 and 70 dB SPL presentations are reportedly not statistically significant (SNR50 means of 1.8 dB at a 60-dB presentation level and 1.7 dB for 70-dB presentation level; Holder et al 2018). The transducer can be insert earphones or a free-field loudspeaker depending on if the listener is aided or unaided, or on what transducer is already being used for other testing.

**Scoring:** For each sentence, there are five underlined key words that are scored as correct or incorrect by the examiner; any additional words in the sentence do not contribute to the score. To obtain the patient's "SNR loss" in decibels (which is simply the SNR50 normalized to the average score of 2 dB SNR for a young, normal-hearing listener), the total number of correctly repeated words across all six sentences is subtracted from 25.5. To obtain the participant's raw SNR50 instead, the number of correct words would be subtracted from 27.5. Increasing the number of lists administered and averaging the results improves reliability. The rationale for using SNR loss is that it is easier to meaningfully compare this value across different tests and calibrations than it is for the raw SNR50.

**Norms:** If SNR loss is 3 dB, receptive speech-in-noise ability is considered approximately normal, and the listener is expected to hear at least as well as the typical normal-hearing listener in noise, if not better. SNR loss between 3 and 7 dB is labeled mild, with an expectation of hearing "almost as well as" normal listeners in noise. SNR loss between 7 and 15 dB is labeled "moderate," with an expectation that directional microphones will help. SNR loss > 15 dB is labeled "severe," with an expectation that maximum SNR improvement is needed and a recommendation to consider use of an FM system. If one list is administered, the 80 and 95% confidence levels are  $\pm 2.2$  dB and  $\pm 2.7$  dB, respectively. However, if more than one list is completed, the reliability improves; e.g., two lists averaged results in 80 and 90% confidence levels of  $\pm 1.6$  dB and  $\pm 1.9$  dB, respectively (Killion et al. 2004).

### Revised Speech Perception in Noise Test (R-SPIN)

**Purpose:** To evaluate the listener's ability to use context cues to understand speech in noise. The test was later refined by Bilger et al. (1984) using data from participants with hearing loss.

**Materials:** Sentences which are either *high-context*, meaning the words in the sentence are related to the final word of the sentence, or *low-context*, meaning the words are not related to the final word of the sentence. These are played against a backdrop of multitalker babble noise. The original study by Kalikow et al. (1977) developed 8 lists (or “forms”) of 50 sentences each. Bilger et al. (1984) recommended, however, that certain forms be discarded, or list pairs combined.

**Administration:** Kalikow et al. (1977) originally suggested multiple forms be presented at different SNRs to understand the benefit that a patient would receive from hearing aids. Bilger et al. (1984) presented their experiments based on a babble threshold which was calculated by transforming the critical bands for each octave present in the speech signal, then subtracting the average minimum audible pressure or minimum audible field measures (corresponding to headphone or sound field presentation) for each octave from the critical bands. For their hearing-impaired subjects, they replaced the minimum audible pressure and minimum audible field measures with the actual audiogram. High-context and low-context sentences are presented 50 dB above this babble threshold. Babble noise is then presented simultaneously with a fixed SNR of 8 dB. The patient repeats the last word of each sentence. Data for these calculations can be found in Appendix B of Bilger et al. (1984).

**Scoring:** Correct number of words are counted for each form.

**Norms:** No norms are established for this test.

**Miscellaneous:** Contact the University of Illinois Champaign - Urbana Department of Speech and Hearing Science (<http://www.shs.uiuc.edu/>) about how to obtain a copy of the R-SPIN.

### Speech Recognition in Noise Test (SPRINT)

**Purpose:** Established in 1992 (Cord et al. 1992) as a 200-word speech-in-noise test to assess fitness-for-duty of soldiers with hearing loss in the U.S. Army, motivated by inter-soldier variability in practical hearing-related tasks despite having similar audiograms. One of its goals is to prevent soldiers with hearing loss from being unnecessarily reclassified or separated from active duty. In 2013, the word list was shortened to 100 words, and the corresponding development and validation study was published in 2017 (Brungart et al. 2017).

Within the Department of Defense, branches of the military have different guidelines when evaluating auditory fitness for duty (US Department of the Air Force 2020; US Department of the Army 2011; US Department of the Navy 2020). The U.S. Army routinely utilizes the SPRINT to evaluate soldiers depending on their hearing profile classification: H1, H2, H3, or H4, which are defined by pure-tone threshold configuration. The H1 profile is meant to categorize normal hearing with a pure-tone average at 500, 1,000, and 2,000 Hz no greater than 25 dB with no individual level over 30 dB and 4,000 Hz not exceeding 45 dB. The H2 profile encompasses two possible hearing profiles. The first is a pure-tone average at 500, 1,000, and 2,000 Hz no greater than 30 dB with no individual level over 35 dB and 4,000 Hz not exceeding 55 dB. The second possible H2 profile focuses on the better ear with thresholds of 30 dB at 500 Hz, 25 dB at 1,000 and 2,000 Hz, and 35 dB at 4,000 Hz. With this second H2 profile, the poorer ear may be deaf. If audiometric thresholds exceed the H2 profile, they are considered an H3 profile which causes referral to a military occupational specialty medical retention board (MMRB) and medical evaluation board (MEB). A soldier with an H3 profile is administered the SPRINT to assist with deciding to retain, reclassify, or separate the soldier from active duty (US Department of the Army 2011).

**Materials:** The SPRINT test is composed of NU-6 words presented in six-talker babble. The CD utilized by the Department of Veterans Affairs contains eight 100-word lists. As of 2017, the 200-word version is still accepted; however, the 100-word SPRINT is now the standard within the Department of Defense.

**Administration:** This test is presented binaurally at 50 dB HL with a fixed SNR of +9 dB preset into the CD. If the service member guesses more than one word, the audiologist is encouraged to take the first word.

**Scoring and Norms:** A raw score of correct words places the service member in a performance category from 1 to 13, which corresponds to a percentile score. This performance category number is compared with the years of service on a graph provided by the SPRINT distributors to identify the recommendation represented by letters A–E.

The following recommendations are outlined by the U.S. Army Regulation 40–501 and a SPRINT development study by Brungart et al. (2017):

A: Retention in current assignment.

B: Retention in current assignment with restrictions.

C: Reassignment to (or retention in) nonhazardous Military Operational Specialty.

D: Discretionary.

E: Recommend Separation from Service.

### Words in Noise (WIN)

**Purpose:** Designed using words rather than sentences to minimize higher-order cognitive influences that may not be specific to the auditory system. Employs word lists that are already widely used for clinical testing of word recognition in quiet.

**Materials:** There are two lists of 35 monosyllabic NU6 words presented in six-talker babble (three males and three females) (Wilson 2003; Wilson & Burks 2005). Each list is randomized four times for a total of eight different tracks. The babble level is fixed, and the word level is varied to achieve seven different SNRs, ranging from 24 to 0 dB in 4-dB steps.

**Administration:** The six-talker babble is either presented at 80 dB SPL when the PTA is 40 dB HL, or at 90 dB SPL when the PTA is between 42 and 58 dB HL. The WIN has not been recommended when the PTA is 60 dB HL (Wilson 2011). Five words are presented at each SNR, in descending-SNR order. Administration is stopped early if none of the words at a given SNR are repeated correctly; all remaining words are then scored as incorrect.

**Scoring:** The total number of correctly repeated words is tallied and used to calculate an SNR50 score in decibels. The formula, derived from the Spearman-Kärber equation, is:  $26 - 0.8x$ , where  $x$  is the total number of words correct.

**Norms:** SNR50 norms for the WIN are provided for adults and are established from a 3,430 participant sample (Wilson 2011) of primarily Veteran participants (normal hearing, 6.0 dB; mild hearing loss, 6.8 to 10.0 dB; moderate hearing loss, 10.8 to 14.8 dB; severe hearing loss, 15.6 to 19.6 dB; profound hearing loss > 20 dB).