ISSN: 2083-389X
eISSN: 2084-3127

J Hear Sci, 2026; 16(2): 35–47
DOI: 10.17430/jhs/224898
CC BY-NC-ND 3.0 PL, © The Authors

Received:
February 12, 2026
Revision accepted:
June 22, 2026
Published online:
August 12, 2026

Contributions:
A Study design/planning
B Data collection/entry
C Data analysis/statistics
D Data interpretation
E Preparation of manuscript
F Literature analysis/search
G Funds collection

MONOSYLLABIC SPEECH PERCEPTION UNDER MULTI-TALKER BABBLE FOR NORMAL ADULTS: A PHONOLOGICAL PERSPECTIVE

Mansi Pokiya*[1A-F], Anuj Kumar Neupane*[1A-F], Kristi Kaveri Dutta*[2A-F]

* These authors contributed equally.

[1] Department of Audiology, School of Audiology and Speech Language Pathology, Bharati Vidyapeeth (Deemed to be University) Pune, India
[2] School of Audiology and Speech Language Pathology, Bharati Vidyapeeth (Deemed to be University), Pune, India

Corresponding author: Anuj Kumar Neupane, Department of Audiology, Bharati Vidyapeeth (Deemed to be University) School of Audiology and Speech Language Pathology, Dhankawadi, 411043, Pune, India; email: anujkneupane@gmail.com

Abstract

Introduction: Speech perception in noisy environments is a crucial skill that often declines with age and varies across languages. Considering the need for a language-specific assessment tool, the present study aimed to develop the Speech-in-Noise Test–Hindi (SiNT-H) and to evaluate speech recognition performance and phonetic error patterns between younger and middle-aged adults at varying signal-to-noise ratios (SNRs).

Material and methods: The study included 27 young and 27 middle-aged native Hindi speakers with hearing thresholds ≤ 25 dB HL. The study was conducted in two phases. Phase I involved the development of the SiNT-H, its administration to young adults, analysis of performance across SNRs, and assessment of test–retest reliability. Phase II compared SiNT-H scores between young and middle-aged adults at different SNRs using four-talker babble noise. Additionally, a detailed phonetic error analysis was carried out across SNR conditions.

Results: The SiNT-H demonstrated no significant ear-wise or list-wise differences. Females outperformed males, with younger females showing better performance than middle-aged females, indicating a gender–age interaction. Performance significantly declined with decreasing SNRs, confirming the test's sensitivity to noise. Phonetic analysis revealed increased errors at lower SNRs, particularly in voicing, manner, place, and aspiration features. At 0 dB SNR, voicing, manner, and aspiration errors were more prevalent among middle-aged adults, reflecting declines in age-related processing.

Conclusions: The findings highlight the importance of incorporating language-specific acoustic-phonetic features in speech-in-noise assessment and underscore the influence of age on speech perception in noisy environments. The SiNT-H appears to be a sensitive tool for evaluating age-related changes in speech recognition in noise among Hindi speakers.

Keywords: Hindi • speech-in-noise test • monosyllabic words • speech babble • phonetic error analysis

ROZUMIENIE SŁÓW JEDNOSYLABOWYCH NA TLE GWARU PRZEZ DOROSŁYCH ZE SŁUCHEM W NORMIE: PERSPEKTYWA FONOLOGICZNA

Streszczenie

Wprowadzenie: Percepcja mowy w warunkach szumu stanowi podstawową zdolność komunikacyjną, która często ulega pogorszeniu wraz z wiekiem i wykazuje zróżnicowanie w zależności od języka. W rezultacie istnieje konieczność dostosowania narzędzi diagnostycznych do specyfiki danego języka. Celem niniejszego badania było opracowanie Hindi Speech-in-Noise Test (SiNT-H), dokonanie oceny wyników rozumienia mowy oraz zidentyfikowanie wzorców błędów fonetycznych u osób młodych i w średnim wieku przy różnych wartościach stosunku sygnału do szumu (SNR).

Materiał i metody: Badaniem objęto 27 młodych oraz 27 osób w średnim wieku, będących rodzimymi użytkownikami języka hindi, z progami słyszenia ≤ 25 dB HL. Badanie przeprowadzono w dwóch etapach. Etap I obejmował opracowanie testu SiNT-H, jego zastosowanie w grupie młodych dorosłych, analizę uzyskanych wyników dla różnych wartości SNR oraz ocenę rzetelności testu metodą test–retest. Etap II polegał na porównaniu wyników uzyskanych w teście SiNT-H przez osoby młode i osoby w wieku średnim dla różnych wartości SNR z wykorzystaniem gwaru. Ponadto przeprowadzono szczegółową analizę błędów fonetycznych w poszczególnych warunkach SNR.

Wyniki: W teście SiNT-H nie stwierdzono istotnych różnic związanych ani z badanym uchem, ani z zastosowaną listą materiału słownego. Kobiety uzyskiwały wyższe wyniki niż mężczyźni, przy czym młodsze kobiety osiągały wyższe wyniki niż kobiety w średnim wieku, co wskazuje

na korelację między płcią a wiekiem. Wyniki obniżały się istotnie wraz ze zmniejszaniem wartości SNR, co stanowi potwierdzenie czułości testu na obecność szumu. Analiza fonetyczna wykazała wzrost liczby błędów przy niższych wartościach SNR, zwłaszcza w zakresie cech takich jak: dźwięczność, sposób artykulacji, miejsce artykulacji oraz aspiracja. Przy SNR równym 0 dB błędy dotyczące dźwięczności, sposobu artykulacji i aspiracji występowały częściej u osób w średnim wieku, odzwierciedlając związane z wiekiem obniżenie sprawności przetwarzania mowy.

Wnioski: Uzyskane wyniki świadczą o tym, że w ocenie rozumienia mowy w warunkach szumu ważną rolę odgrywa uwzględnienie cech akustyczno-fonetycznych specyficznych dla danego języka, oraz wskazują na istotny wpływ wieku na percepcję mowy w hałaśliwym otoczeniu. Test SiNT-H wydaje się czułym narzędziem, właściwym do oceny zmian związanych z wiekiem w rozpoznawaniu mowy w warunkach szumu u użytkowników języka hindi.

Słowa kluczowe: hindi • test rozumienia mowy w szumie • słowa jednosylabowe • szum wielogłosowy • analiza błędów fonetycznych

Key to abbreviations	
AI	articulation index
F0	fundamental frequency
IEEE	Institute of Electrical and Electronics Engineers
IQR	interquartile range
MOC	medial olivocochlear
PTA	pure-tone average
RMS	root mean square
SiN	speech in noise
SiNT-H	Speech-in-Noise Test–Hindi
SNR	signal-to-noise ratio
SRT	speech reception threshold
WRS	word recognition score

Introduction

Effective speech communication depends critically on the listener's ability to perceive and comprehend spoken language at the word level. While pure-tone audiometry remains the gold standard for assessing auditory sensitivity, it provides limited insight into functional speech understanding. In contrast, speech perception testing offers a more ecologically valid perspective by evaluating how individuals process spoken language in real-world contexts [1].

Among the conventional tools, Speech Reception Threshold (SRT) and Word Recognition Score (WRS) are commonly employed to assess speech perception in quiet environments [2,3]. However, these measures inadequately represent the challenges listeners face in everyday noisy situations [4]. Understanding speech in noise (SiN) remains challenging even for individuals with normal audiometric thresholds [5–8]. Effective SiN perception requires the suppression of competing background noise and the selective attention to target speech – abilities that rely on an intricate interplay between auditory processing and higher-order cognitive functions [4,9]. Listener-related factors such as age, cognitive capacity, and peripheral hearing status further modulate performance in noisy environments [10].

Speech-in-noise tests provide critical information beyond what is available from traditional audiometric evaluations. These assessments capture complex auditory mechanisms, including temporal resolution, binaural integration, and the interface between linguistic and cognitive processing [1]. Various SiN measures differ in linguistic structure and cognitive demands. Sentence-based tests leverage contextual cues, engaging top-down processing mechanisms [11], whereas monosyllabic word tests emphasise bottom-up auditory perception by minimising linguistic redundancy [8]. This distinction holds clinical relevance as it aids in differentiating peripheral from central auditory impairments and is particularly valuable in evaluating the benefit of hearing aid features such as noise reduction algorithms [12,13]. Importantly, SiN testing is instrumental in identifying "hidden hearing loss", a phenomenon where individuals experience listening difficulties in noise despite normal pure-tone thresholds [14]. Over the decades, SiN testing has been extended to the diagnosis of auditory processing disorders, hearing aid candidacy, and post-intervention outcome monitoring [15,16]. However, a critical limitation remains: speech-in-noise perception is highly language-specific [17], necessitating the development of linguistically and culturally appropriate assessment tools [18].

In a multilingual country like India, Hindi is the most widely spoken language [19,20]. Yet, there is a conspicuous lack of standardised SiN tests in Hindi, especially those employing monosyllabic stimuli, which are essential for evaluating phoneme-level auditory deficits. This gap is particularly pertinent given the known age-related variations in SiN perception [21]. Research has shown that individuals across age groups employ distinct processing strategies and are differently influenced by both energetic and informational masking [21,22]. Additionally, performance in SiN tasks is known to vary at the phoneme level. Research has demonstrated that vowels are generally more robust against masking than consonants. For instance, Weber and Smits [23] reported vowel recognition accuracy at 78%, compared to 69% for consonants in multitalker babble. Similarly, Moreno-Torres et al. [24] observed enhanced vowel recognition in Spanish-speaking populations, with frontal consonants being particularly vulnerable to masking. Such differences are largely attributed to acoustic features such as frequency content, intensity, and temporal structure, where high-frequency consonants like /tʃ/, /s/, and /j/ tend to be recognised more reliably than mid- and low-frequency sounds like /r/, /m/, /n/, /p/, /t/, and /k/. The complexity of Hindi phonology further amplifies these challenges. Hindi comprises a rich array of phonemic distinctions – such as aspiration, retroflexion, and vowel length – that may influence susceptibility to noise-induced distortion [25]. Aspirated and retroflex

consonants, for instance, carry unique acoustic markers that may degrade more readily in noisy environments. Wang and Bilger [26] found that fricatives like /f/ and /h/ are particularly vulnerable to masking, while Cutler et al. [27] noted that consonant recognition improves with increasing signal-to-noise ratio (SNR), albeit with non-native listeners experiencing greater difficulty.

Given this context, the present study was undertaken to develop and evaluate the Speech-in-Noise Test in Hindi (SiNT-H) using monosyllabic word stimuli embedded in ecologically valid multitalker babble. By assessing both age-wise performance and phoneme-specific error patterns, this study aims to provide normative benchmarks and contribute to a deeper phonological understanding of speech perception in Hindi.

Material and methods

Participants

A total of 54 native Hindi-speaking adults participated in the study. The sample size was determined with reference to the findings of Vaidyanath and Yathiraj [28], who reported mean speech-in-noise scores of 22.74 ± 2.09 for young adults ($n = 34$) and 13.83 ± 3.93 for older adults ($n = 49$) with hearing thresholds ≤ 15 dB HL. Based on these values, an effect size (d) of 2.5 was considered. With a significance level (α) of 5% corresponding to a 95% confidence interval ($Z = 1.96$) and a test power of 80%, the required sample size was calculated to be 27 participants per group. Thus, the total sample size comprised 54 participants. Hence, the participants were equally divided into two age groups: young adults aged 20–39 years and middle-aged adults aged 40–59 years, with 27 individuals in each group. The young adult group consisted of 10 males and 17 females, whereas the middle-aged group included 16 males and 11 females.

All participants demonstrated normal hearing sensitivity, with pure-tone thresholds ≤ 25 dB HL across octave frequencies from 0.25 to 4 kHz in both ears. Middle ear function was confirmed using tympanometry, and individuals exhibiting any signs of middle ear pathology, otological complaints, tinnitus, speech-language or neurological disorders, or cognitive impairments were excluded. Ethical approval for the study was obtained from the Institutional Ethics Committee of Bharati Vidyapeeth (Deemed to be University) Medical College, Pune (Ref: BVDUMC/IEC/141). Written informed consent was obtained from all participants before initiating the procedure.

Study design and procedure

This cross-sectional study utilised purposive sampling and was carried out in two phases. In Phase I, the Speech-in-Noise Test in Hindi (SiNT-H) was developed. Phase II involved administering the SiNT-H to participants in two different age groups to assess their speech-in-noise performance.

Phase I: Development of the monosyllabic SiNT-H

Stimulus selection and familiarity ratings. A pool of 300 monosyllabic Hindi words was initially extracted from the standard word lists of De [29] and evenly distributed across six word lists. Five native Hindi-speaking adults independently rated each word on a 5-point familiarity scale (1 = "unknown" to 5 = "most familiar"), as adapted from Chinnaraj et al. [30]. Based on the mean familiarity scores, 177 words were identified as sufficiently familiar and were retained for subsequent steps. To minimise inter-item and inter-list variability, all test stimuli underwent a structured equalisation procedure. Familiar monosyllabic words with comparable phonological structure were selected as speech materials. Using Praat software, each recorded word token was individually amplitude-normalised to achieve a consistent RMS level. Noise samples were also normalised separately prior to mixing. Speech and noise were then combined at the required SNR conditions, after which the final mixed stimuli were re-equalised to ensure a uniform overall presentation level across all items and lists. In addition, spectral characteristics of the stimuli were reviewed to maintain broad acoustic comparability across lists.

Recording protocol. The total of 177 selected words were recorded in a sound-treated room by a 23-year-old native Hindi-speaking female. Recordings were made using Audacity (v3.6.3) on a Hewlett-Packard KTMP99C system, with a studio-quality microphone placed 15 cm from the speaker's mouth at 45°. Speech samples were digitised at a sampling rate of 44.1 kHz using a 16-bit analog-to-digital converter.

Recording speaker selection. From the pool of available speakers, one 23-year-old female speaker was selected based on perceptual evaluation of voice clarity, articulatory precision, consistency of vocal quality, and absence of noticeable accent or dialectal variation relative to the target language variety. This screening was conducted to ensure that the recorded stimuli would be linguistically neutral and easily intelligible for the intended participant population. During the recording process, multiple takes of each stimulus item were obtained. The recordings were then reviewed, and the most natural and consistent token for each item was selected based on clarity of pronunciation, stable intensity, appropriate speaking rate, and uniform prosodic characteristics (including stress pattern and intonation contour). Tokens containing hesitations, misarticulations, irregular pauses, or unintended prosodic emphasis were excluded. In addition, efforts were made to minimise acoustic variability across the final stimulus set. Recordings were monitored to maintain consistency in fundamental frequency (F0), overall speaking rate, and vocal intensity as far as practically feasible within natural speech production constraints. These procedures were undertaken to ensure that differences in participant performance were attributable to the experimental conditions rather than unwanted speaker-related acoustic variability.

Intelligibility ratings and refinement. To ensure perceptual clarity, five native Hindi speakers rated the intelligibility of the recorded words on a 5-point Likert scale (1 = "bad" to 5 = "excellent") based on IEEE (1969) guidelines. Words rated below a mean of 4 were re-recorded. This process led to the refinement of 16 items to enhance clarity and naturalness.

Table 1. Construction of two monosyllabic word lists based on the distribution of the acoustic energy across critical bandwidths in the selected words

Frequency band [Hz]	Total words in the band	Proportional distribution [%]	Number of words allocated to List 1	Number of words allocated to List 2
0–150	102	57.62	14	14
151–250	34	19.20	5	5
251–350	16	9.03	2	2
351–450	15	8.47	2	2
451–570	2	1.69	1	1
571–700	8	3.95	1	1
Total	177	100.00	25	25

Acoustic analysis and list construction. Acoustic analyses were conducted using PRAAT (v6.4.01). Formant frequencies (F0, F1, F2) and bandwidths (F1, F2, F3) were measured. Words were then categorised according to critical band centre frequencies as described by Zwicker [31], and articulation index (AI) values were calculated. Based on these parameters, two final test lists of 25 monosyllabic words each were created (**Table 1**). All items were RMS-normalised to −18.0 dB using Audacity to ensure uniform intensity across stimuli. Analysis of list performance revealed no statistically significant differences across word lists, suggesting functional equivalence and absence of measurable list bias under the present test conditions.

Background noise construction. To generate multitalker babble noise, four native Hindi speakers (two male, two female) recorded two standardised Hindi paragraphs titled /ˈmeːɾeː ˈd͡ʒiːʊən kɑː ˈləkʂjə/ and /ˈɦoːliː/ in cold running speech style, following the methodology of Gundmi et al. [32]. The individual recordings were merged to produce a 41-second, four-talker babble, which was RMS-matched to −18.0 dB to align with the target word stimuli.

Signal-to-noise ratios (SNRs). The final test material was created by mixing the normalised word lists with the multitalker babble at three SNRs (0, +5, and +10 dB, reflecting progressively easier speech-in-noise conditions. Additionally, five practice items were constructed at various SNR levels (+20 dB, +10 dB, +5 dB, and two at 0 dB) to familiarise participants with the testing format.

Phase II: Administration of the Speech-in-Noise Test in Hindi (SiNT-H)

In the second phase of the study, the finalised SiNT-H was administered to the 54 participants, comprising 27 young adults (20–39 years) and 27 middle-aged adults (40–59 years). All participants were native Hindi speakers with normal hearing sensitivity and no reported otological, neurological, or speech-language disorders.

Test environment and preliminary screening. The assessments were conducted in a sound-treated and acoustically calibrated room. Stimuli were presented binaurally at 60 dB SPL through Boat Bassheads 225 insert earphones connected to a Hewlett-Packard KTMP99C laptop. Prior to testing, a comprehensive case history was obtained, followed by otoscopic examination to screen for external or middle ear anomalies. Hearing screening was performed using the Hearing Test mobile application (e-audiologia.pl; version 2.7) [33] – a validated Android-based tool for estimating pure-tone hearing thresholds across standard audiometric frequencies. Testing was conducted in a quiet environment following standardised instructions. Pure-tone averages (PTAs) were calculated using thresholds at 0.5, 1, 2, and 4 kHz. The mean PTA across participants was 17.9 dB HL, indicating hearing sensitivity within normal limits. All participants met the study inclusion criteria based on this screening procedure.

Test procedure. Each participant was administered two word lists – List 1 and List 2 – each comprising 25 monosyllabic words (Supplement 1). The lists were presented under three signal-to-noise ratio (SNR) conditions: +10, +5, and 0 dB. The order of list presentation and ear dominance was randomised across participants to mitigate order and lateralisation effects. To ensure participants understood the test paradigm, practice trials were conducted prior to each SNR condition using five words presented at various SNRs. Participants were instructed to listen carefully to the target words embedded in multitalker babble and ignore the background noise. Responses were recorded manually by the examiner. A brief rest interval of 10 minutes was provided between the two lists to minimise listening fatigue and maintain performance consistency.

Scoring and reliability. Scoring for the SiNT-H was conducted at two levels: word-level accuracy and phoneme-level error analysis. For word-level scoring, each correctly identified monosyllabic word was awarded 4%, resulting in a maximum possible score of 100% per list. This scoring scheme allowed for straightforward quantification of speech recognition performance across different SNR conditions. In addition to overall accuracy, a detailed phoneme-level error analysis was performed to identify specific patterns of misperception and to gain deeper insight into the acoustic-phonetic challenges encountered by participants during speech-in-noise recognition.

To ensure the reliability of the SiNT-H, test–retest consistency was evaluated in a subset of participants. Approximately 10% of the total sample ($n = 12$) was re-assessed after an interval of 1 month using the same stimulus lists and identical testing procedures. This re-administration allowed for the examination of the stability and temporal reliability of the test scores, thereby contributing to the validation of SiNT-H as a dependable tool for assessing speech-in-noise perception in Hindi-speaking populations.

Statistical analysis

All statistical analyses were performed using IBM SPSS Statistics for Windows, Version 23.0. The normality of the SiNT-H scores across different SNR conditions was initially assessed using a Shapiro–Wilk test. The results indicated a significant deviation from a normal distribution ($p < 0.05$), necessitating the use of non-parametric statistical methods for all subsequent analyses. To examine ear-wise differences in performance, a Wilcoxon signed-rank test was applied separately to List 1 and List 2 for male and female participants. In order to assess potential list-related effects and gender differences within Group I (young adults), a Mann–Whitney *U*-test was employed using combined scores from both ears. The impact of varying SNR conditions (+10, +5, and 0 dB) on speech-in-noise performance was analysed using Friedman's test for repeated measures, allowing for within-group comparison of performance across noise levels.

Furthermore, between-group differences in SiNT-H performance were analysed using a Mann–Whitney *U*-test to compare the young adult group (Group I) with the middle-aged adult group (Group II). This comprehensive statistical framework enabled a detailed evaluation of test performance across demographic variables, test conditions, and linguistic complexity, contributing to the validation of SiNT-H as a reliable and clinically relevant assessment tool.

Results

The findings are presented in several parts: (1) Comparison of SiNT-H scores between ears; (2) Gender-wise performance analysis; (3) Assessment of list equivalency across genders; (4) Performance across different SNRs (0, +5, +10 dB) to determine the test's responsiveness to changes in listening difficulty; (5) Age-related performance differences across genders; (6) Evaluation of test–retest reliability; (7) Phonemic error analysis to identify perceptual confusions. Additionally, the performance was examined across different SNRs to determine the test's responsiveness to changes in listening difficulty.

1) Comparison of SiNT-H scores between ears

To compare the ear-specific performance in the SiNT-H test between ears, both male and female participants were compared (Supplement 2). The analysis was conducted separately for each list (List 1 and List 2) under three signal-to-noise ratio (SNR) conditions (0, +5, and +10 dB) and shown in **Figure 1**.

Figure 1 shows median and interquartile ranges for List 1 and List 2 scores between ears across genders at three different SNRs. Among males, List 1 showed marginally superior scores in the left ear compared to the right, whereas List 2 yielded nearly symmetrical distributions across ears for three conditions. Conversely, females exhibited a consistent pattern of better right ear performance, with higher median scores and broader IQRs noted in both lists at 0, +5, and +10 dB SNRs. Although certain patterns were observable, the Wilcoxon signed-rank test (Supplement 3) revealed no statistically significant differences across either list for male participants at any condition. Specifically, no significant differences were found for males at baseline for List 1 ($Z = 0.72$, $p = 0.473$) and List 2 ($Z = 0.69$, $p = 0.491$). Female participants also showed non-significant results at baseline (List 1: $Z = 2.09$, $p = 0.071$; List 2: $Z = 1.94$, $p = 0.052$). A similar trend was observed at +5 dB SNR, with male participants showing no significant differences (List 1: $Z = 0.21$, $p = 0.831$; List 2: $Z = 0.61$, $p = 0.542$). Female participants at this SNR also did not reach statistical significance (List 1: $Z = 1.92$, $p = 0.062$; List 2: $Z = 0.71$, $p = 0.473$). At +10 dB SNR as well, the results remained non-significant for males (List 1: $Z = 1.59$, $p = 0.112$; List 2: $Z = 0.82$, $p = 0.414$) and females (List 1: $Z = 1.11$, $p = 0.273$; List 2: $Z = 0.83$, $p = 0.414$). These findings support the ear-wise equivalence of the SiNT-H test. Consequently, data from both ears were pooled for subsequent analyses.

2) Gender-wise performance analysis

To assess the difference in performance between gender on the SiNT-H performance, a comparison of scores between male and female participants was conducted using the Mann–Whitney *U*-test (Supplement 4). Female participants were found to consistently achieve higher median scores than their male counterparts across both List 1 and List 2 under all SNR conditions. At 0 dB SNR, females performed significantly better than males on both lists [List 1 ($U = 506$, $p = 0.012$); List 2 ($U = 519.5$, $p = 0.011$)]. This gender-based performance difference remained statistically significant at +5 dB SNR [List 1 ($U = 450$, $p = 0.043$); List 2 ($U = 454$, $p = 0.042$)], and a similar trend was observed at +10 dB SNR [List 1 ($U = 497$, $p = 0.043$); List 2 ($U = 458$, $p = 0.026$)].

3) Assessment of list equivalency across genders

Further, the present study examined the psychometric equivalency of two monosyllabic word lists of the SiNT-H test at three SNR levels using a Wilcoxon Signed-Rank test. At 0 dB SNR, no significant differences were observed between the lists for males ($Z = 1.23$, $p = 0.212$) or females ($Z = 1.87$, $p = 0.074$). Similarly, at +5 dB SNR, comparisons revealed no significant difference for males ($Z = 0.95$, $p = 0.341$) or females ($Z = 1.40$, $p = 0.162$). At +10 dB SNR, performance improved further, but the lists remained equivalent with no significant differences for males ($Z = 0.65$, $p = 0.524$) or females ($Z = 1.10$, $p = 0.275$).

4) Performance across different SNRs

The performance on the SiNT-H test was compared across SNRs which revealed a decline in median test scores with decreasing SNRs as depicted in **Figure 1**.

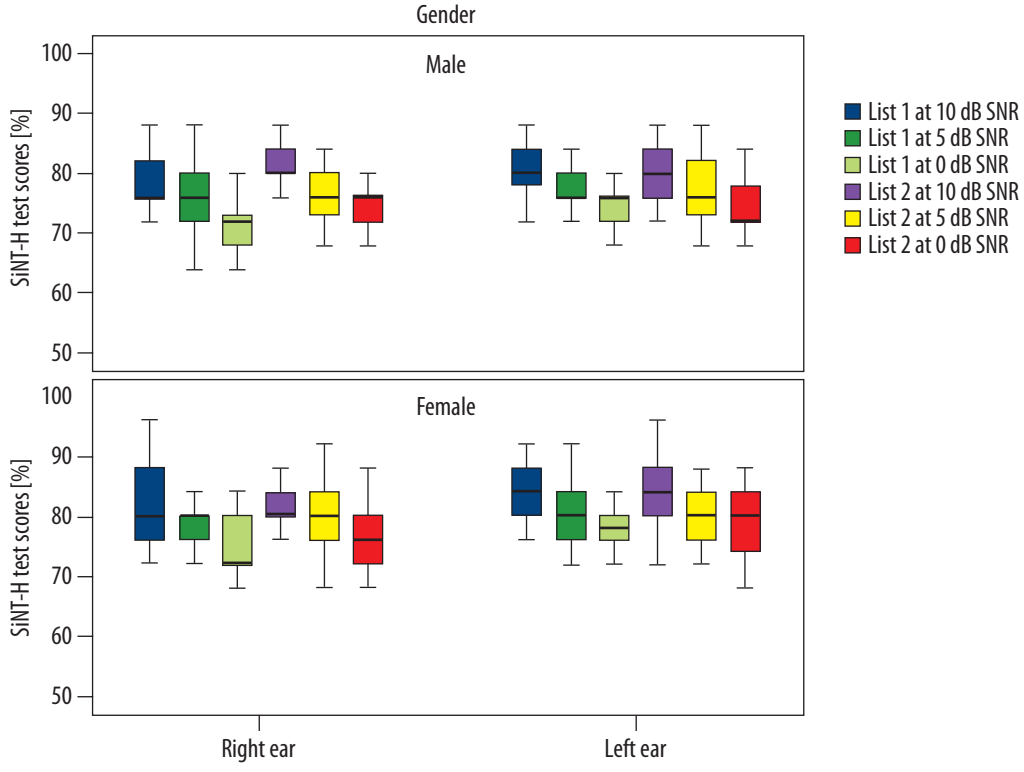

Figure 1. Scores in the SiNT-H test for both List 1 and List 2. The data is divided into that for males (top) and females (bottom); for right ears (left) and left ears (right); and for 0 dB, +5 dB, and +10 dB SNR (colour coding)

A Friedman test revealed a significant main effect of SNR on speech recognition scores, with performance declining notably at 0 dB SNR compared to +5 and +10 dB. Specifically, the test was applied to assess the impact of varying SNR levels on List 1 performance in the right ear. A statistically significant difference was observed across the three SNR conditions. To determine the specific contrasts, post hoc pairwise comparisons were conducted using the Bonferroni correction. For Group 1, the speech recognition scores at 0 dB SNR were significantly lower than those at +5 dB ($\chi^2 = 0.593$, $p < 0.001$) and +10 dB ($\chi^2 = 1.255$, $p < 0.001$). Additionally, the scores at +5 dB SNR were significantly lower than those at +10 dB ($\chi^2 = 0.662$, $p < 0.001$). Similarly, in Group 2, the performance at 0 dB SNR was significantly reduced in comparison to both +5 dB SNR ($\chi^2 = 0.620$, $p < 0.001$) and +10 dB SNR ($\chi^2 = 1.532$, $p < 0.001$), with +5 dB SNR scores also significantly lower than +10 dB SNR ($\chi^2 = 0.912$, $p < 0.001$).

5) Age-related performance differences across genders

An analysis of the performance in the SiNT-H test was done between young and middle-aged adults across genders (Supplement 5). It was found that the young female adults had significantly superior test scores than middle-aged females at 0 dB SNR ($U = 130$, $p = 0.025$) and +5 dB SNR ($U = 1347$, $p = 0.026$). However, the difference was not statistically significant at +10 dB SNR ($U = 1295$, $p = 0.107$). In contrast, male participants exhibited no significant age-related differences across any of the SNR levels [0 dB ($U = 1176$, $p = 0.373$); +5 dB ($U = 958$, $p = 0.404$); +10 dB ($U = 1005$, $p = 0.477$)]. These results indicate a statistically significant age-related decline in the SiNT-H performance among females, particularly under more challenging SNR conditions, whereas male participants showed stable performance across age groups.

6) Evaluation of test–retest reliability

To determine the consistency in the performance between tests and retest scores of SiNT-H, a test–retest reliability analysis was performed (Supplement 6). A subset of participants ($n = 12$; 6 males and 6 females) was reassessed after an interval of 15 days. The Wilcoxon signed-rank test was employed to compare paired performance scores between the initial and follow-up sessions. The results indicated no statistically significant differences across ears, lists, and SNRs. At 10 dB SNR, both List 1 and List 2 showed stable scores across ears – List 1: right ear ($Z = 0.45$, $p = 0.661$), left ear ($Z = 0.81$, $p = 0.412$); List 2: right ear ($Z = 1.00$, $p = 0.313$), left ear ($Z = 1.00$, $p = 0.317$). A similar stability was observed at +5 dB SNR – List 1: right ear ($Z = -0.707$, $p = 0.480$), left ear ($Z = 0.707$, $p = 0.480$); List 2: right ear ($Z = 2.449$, $p = 0.143$), left ear ($Z < 0.01$, $p = 0.999$). At 0 dB SNR, List 1 scores were: right ear ($Z = 1.933$, $p = 0.053$), left ear ($Z = 1.342$, $p = 0.180$), while List 2 scores were: right ear ($Z = 0.447$, $p = 0.655$), left ear ($Z = 0.378$, $p = 0.705$). Further, Spearman's rank-order correlation coefficient was calculated to examine the strength of association between test and retest scores across the different SNRs. At 0 dB SNR, strong positive correlations were

Table 2. Percentage of word recognition errors at different SNRs for List 1 and List 2 of the SiNT-H

No.	List 1 words	Total error [%]			List 2 words	Total error [%]		
		at 0 dB SNR	at 5 dB SNR	at 10 dB SNR		at 0 dB SNR	at 5 dB SNR	at 10 dB SNR
1	/tʃʰət/	10.60	10.61	5.30	/paːs/	31.82	31.82	15.91
2	/kʊtʃʰ/	2.43	3.03	0.76	/neːk/	11.36	11.36	3.03
3	/tiːn/	6.81	10.61	9.85	/pəl/	15.15	15.15	3.79
4	/pʰɪr/	42.42	53.79	45.45	/leːkʰ/	16.67	16.67	14.39
5	/gʊʈ/	71.97	58.33	28.79	/peːʃ/	57.58	57.58	31.82
6	/buːʈ/	38.64	33.33	23.48	/teːdʒ/	3.03	3.03	5.30
7	/ʈoːp/	21.21	25.00	22.73	/sɪkʰ/	24.24	24.24	26.52
8	/dʒ^{ɲ}iːl/	6.06	6.06	0.00	/poːʂ/	42.42	42.42	25.76
9	/bəl/	48.48	50.00	53.79	/sətʃ/	25.00	25.00	22.73
10	/bɛːl/	30.30	36.36	31.06	/dʒəj/	6.82	6.82	9.85
11	/eːk/	9.09	6.82	3.03	/mɛːl/	27.27	27.27	12.88
12	/ʈʰiːk/	0.00	0.76	3.79	/məʈʰ/	16.67	16.67	5.30
13	/tiːs/	0.76	3.03	7.58	/seːb/	58.33	58.33	49.24
14	/ɖeːʊ/	50.00	46.21	26.52	/rəs/	21.21	21.21	15.15
15	/aːm/	12.88	13.64	9.85	/aːg/	42.42	42.42	17.42
16	/kʰaːɖ/	40.91	37.12	27.27	/ɦəʈʰ/	7.58	7.58	6.06
17	/raːdʒ/	8.33	6.06	9.85	/tʃaːr/	43.94	43.94	28.79
18	/kəʈ/	16.67	12.12	6.06	/dʒaːn/	0	0	0
19	/nəl/	13.64	13.64	7.58	/loːbʰ/	43.18	43.18	17.42
20	/saːg/	93.18	92.42	86.36	/mən/	7.58	7.58	6.82
21	/baːʈ/	41.67	38.64	18.94	/dɪn/	22.73	22.73	3.79
22	/paːʈʰ/	30.30	26.52	13.64	/kaːm/	21.21	21.21	12.88
23	/kaːn/	4.55	6.82	2.27	/gaːʊ/	6.82	6.82	7.58
24	/gəm/	3.79	3.79	4.55	/tʃaːp/	42.42	42.42	31.82
25	/taːʃ/	0	0	0	/ɦoːʈʰ/	0.76	0.76	4.55

observed for the right ear in List 1 ($\rho = 0.84$, $p < 0.011$) and List 2 ($\rho = 0.86$, $p < 0.011$), as well as for left ear in List 2 ($\rho = 0.90$, $p < 0.012$), while a moderately strong yet significant correlation was found for List 1 ($\rho = 0.66$, $p = 0.022$). At +5 dB SNR, the correlations remained robust across all test conditions. For the right ear, strong associations were observed for List 1 ($\rho = 0.90$, $p < 0.012$) and List 2 ($\rho = 0.85$, $p < 0.013$). For the left ear, similarly strong correlations were noted for List 1 ($\rho = 0.88$, $p < 0.011$) and List 2 ($\rho = 0.87$, $p < 0.012$). At +10 dB SNR, test–retest associations continued to be strong across all conditions. For the right ear, high correlations were observed for List 1 ($\rho = 0.94$, $p < 0.013$) and List 2 ($\rho = 0.89$, $p < 0.014$). For the left ear, strong correlations were also noted for List 1 ($\rho = 0.80$, $p < 0.012$) and List 2 ($\rho = 0.72$, $p < 0.011$).

7) Phonemic error analysis

To elucidate the influence of background noise on speech perception, a comprehensive phonemic error analysis was conducted at three SNR conditions (0, +5, and +10 dB) (Supplement 7). This analysis encompassed both young and middle-aged adult participants. Therefore, each participant's response to the monosyllabic word stimuli was evaluated at the phoneme level to detect substitution patterns. Phonemic classification was guided by the articulatory taxonomy proposed by Shipra and Chandra [25], which organises phonemes based on place of articulation (e.g., bilabial, dental, retroflex, palatal), manner of articulation (e.g., stops, nasals, fricatives, affricates, liquids), as well as features such as voicing and aspiration. Word-wise

Table 3a. Phoneme substitution patterns in List 1 across three SNR conditions (0 dB, +5 dB, +10 dB)

Correct phoneme in List 1	Phoneme class	Incorrect phoneme in List 1	Phoneme class	0 dB SNR (no of ears)	+5 dB SNR (no of ears)	+10 dB SNR (no of ears)
/t̪/	voiceless unaspirated stop	/l/	dental liquids	9	9	3
/t̪/	voiceless unaspirated stop	/p/	bilabial voiceless unaspirated stop	2	5	3
/pʰ/	bilabial voiceless aspirated stop	/p/	bilabial voiceless unaspirated stop	5	5	9
/ɡ/	voiced unaspirated velar stop	/k/	voiceless unaspirated velar stop	6	1	4
/b/	bilabial voiced unaspirated stop	/uː/	back vowel	19	18	12
/ʈ/	voiceless unaspirated retroflex stop	/ɖ/	voiced unaspirated retroflex stop	3	3	0
/b/	bilabial voiced unaspirated stop	/p/	bilabial voiceless unaspirated stop	53	52	45
		/pʰ/	bilabial voiceless aspirated stop	3	2	9
/b/	bilabial voiced unaspirated stop	/m/	bilabial nasal	15	12	11
/e/	front mid-short vowel	/i/	front short vowel	3	3	5
/ʈʰ/	retroflex voiceless aspirated stop	/pʰ/	bilabial voiceless aspirated stop	1	3	3
/s/	dental fricative	/n/	dental nasal	3	4	3
/m/	bilabial nasal	/d͡ʒ/	palatal voiced unaspirated affricate	8	8	6
/d̪/	dental voiced unaspirated stop	/n/	dental nasal	12	10	5
/ɾ/	retroflex liquid	/l/	dental liquid	4	3	0
/s/	dental fricative	/p/	bilabial voiceless unaspirated 7 stop	43	41	22
/b/	bilabial voiced unaspirated stop	/p/	bilabial voiceless unaspirated stop	14	12	7
/ʈʰ/	retroflex voiceless aspirated stop	/t͡ʃ/	palatal voiceless unaspirated affricate	5	5	0

error percentages were computed for each stimulus in Lists 1 and 2 by dividing the number of incorrect responses by the total number of ears tested ($n = 116$) and multiplying by 100. This allowed for the identification of lexemes most vulnerable to noise-induced perceptual degradation across SNR conditions (**Table 2**). Further, the substitution errors were examined cumulatively across 116 ears for each word. Each erroneous phoneme was classified as per its articulatory features. The total number of phoneme-level errors was calculated separately for each SNR condition, providing a granular view of error patterns for both Lists 1 and 2 as given in **Tables 3a** and **3b**. The findings revealed that phoneme confusions were the most prevalent at 0 dB SNR and showed a progressive decline as the SNR improved to +5 and +10 dB. Notably, phonemes that shared articulatory properties – such as common place or manner of articulation, or similar voicing and aspiration characteristics, were seen to be more frequently misperceived. Additionally, middle-aged adults demonstrated a higher incidence of such confusions compared to their younger counterparts, suggesting an age-related decline in fine-grained phonological processing under acoustically challenging conditions.

Discussion

The present study developed the Speech-in-Noise Test in Hindi (SiNT-H) to assess monosyllabic word recognition using multi-talker babble among normal-hearing Hindi-speaking adults. By integrating linguistically relevant stimuli within a multi-talker babble context, the SiNT-H effectively mirrors real-world communication demands while maintaining diagnostic precision. This approach adheres to best practices recommended by Pichora-Fuller et al. [34], reinforcing the test's clinical and ecological validity.

Table 3b. Phoneme substitution patterns in List 2 across three SNR conditions (0 dB, +5 dB, +10 dB)

Correct phoneme in List 2	Phoneme class	Incorrect phoneme in List 2	Phoneme class	0 dB SNR (no of ears)	+5 dB SNR (no of ears)	+10 dB SNR (no of ears)
/s/	dental fricative	/n/	dental nasal palatal voiceless	14	11	9
		/t͡ʃ/	unaspirated affricate	6	5	4
/n/	dental nasal	/e/	front mid-short vowel	4	7	3
/l/	dental liquid	/e/	front mid-short vowel	2	10	5
/p/	bilabial voiceless unaspirated stop	/t̪/	voiceless unaspirated stop	8	11	3
		/d̪/	dental voiced unaspirated stop	6	7	7
		/ʊ/	bilabial glide	12	11	8
/s/	dental fricative	|/t̪/	voiceless unaspirated stop	10	9	3
		/p/	bilabial voiceless unaspirated stop	15	17	8
/p/	bilabial voiceless unaspirated stop	/ɦ/	glottal fricative	13	26	0
/s/	dental fricative	/t̪/	dental voiceless unaspirated stop	12	13	11
		/k/	velar voiceless unaspirated stop	5	4	2
/b/	bilabial voiced unaspirated stop	/m/	bilabial nasal	11	5	3
/m/	bilabial nasal	/ɦ/	glottal fricative	3	2	0
/ɾ/	retroflex liquid	/d̪/	dental voiced unaspirated stop	33	29	11
/d͡ʒ/	palatal voiced unaspirated affricate	/g/	voiced unaspirated velar stop	8	5	4
/bʰ/	bilabial voiced aspirated stop	/g/	voiced unaspirated velar stop	12	9	13
/m/	bilabial nasal	/n/	dental nasal	11	7	0
/p/	bilabial voiceless unaspirated stop	/ɾ/	retroflex liquid	22	23	13
		/j/	palatal glide	10	9	8

In the study, participants exhibited the highest speech recognition accuracy at +10 dB SNR, with performance declining progressively as the SNR decreased – demonstrating the expected SNR-dependent gradient in speech perception. This pattern reflects listeners' enhanced ability to extract phonetic and lexical cues under favorable listening conditions, with reduced performance in more challenging noise environments, particularly at 0 dB SNR. Such a decline is likely attributable to the increased impact of both energetic and informational masking, resulting in heightened cognitive listening demands. These findings align with the established literature which has consistently reported robust speech-in-noise performance among normal-hearing individuals under higher SNRs [35–39], as well as studies emphasising the cognitive load imposed by masking effects in degraded auditory environments [40,41]. Furthermore, the use of familiar, high-frequency monosyllabic words in SiNT-H likely facilitated word recognition by reducing lexical processing complexity and minimising intersubject variability – an approach supported by psycholinguistic evidence [42,43]. The exclusive use of monosyllables also minimised linguistic redundancy, thereby enabling a more focused evaluation of auditory processing.

In the study, no significant ear differences were observed in the test scores, reflecting bilateral symmetry in auditory processing among young, normal-hearing adults. These findings corroborate earlier studies [44–46] that attributed symmetrical performance to intact peripheral function and balanced cortical processing across hemispheres. Although some researchers have reported right-ear advantages associated with hemispheric language dominance and stronger medial olivocochlear (MOC) reflexes [47–50], such asymmetries are more likely to manifest in older adults or clinical populations. The present

findings affirm the suitability of the SiNT-H for bilateral testing in young adults without necessitating ear-specific norms. Additionally, the absence of ceiling or floor effects, and the comparable performance across Lists 1 and 2, provide empirical support for their equivalence and validate their interchangeable use.

A notable finding in the study was the presence of superior performance of female participants, particularly at lower SNRs. This observation aligns with prior research [38,46,51] and may be explained by a confluence of anatomical, hormonal, and cognitive factors. Females have been shown to possess shorter and more numerous outer hair cells [46], potentially contributing to enhanced cochlear amplifier function. Cortically, they may exhibit greater gray matter density in language-associated regions [52,53], along with more bilateral activation during linguistic tasks [54], which may confer advantages under degraded listening conditions. Given emerging evidence of earlier cognitive decline in women post-menopause, the reduced SiN performance in middle-aged females may reflect subtle, hormone-mediated auditory–cognitive interactions, which have yet to be fully understood [5]. Furthermore, the ability to engage top-down processing strategies more effectively [55] could enhance signal extraction from noise. Although some studies [56,57] have reported inconsistent gender effects, the present data support the inclusion of gender-specific normative values, particularly when testing at challenging SNRs.

The SiNT-H demonstrated high test–retest reliability, affirming its potential for repeated clinical use. Psychometric analysis confirmed that the two test lists were statistically equivalent across genders, supporting their interchangeability. This is critical for ensuring internal validity and avoiding vocabulary bias, as emphasised by Braza et al. [42] and Banks et al. [58]. Randomised test presentation further reduced the potential for order effects and preserved the cognitive consistency of the listening task.

Age-related comparisons revealed that younger females outperformed their middle-aged counterparts, whereas male participants did not exhibit significant age-related differences. These findings partly mirror the literature, where minimal decline is typically observed in individuals under 60 years [45,59]. However, the reduced performance in middle-aged females may indicate early signs of age-related auditory decline, even in the absence of audiometric hearing loss. Studies have linked age-related SiN deficits to reduced temporal resolution, decreased cortical inhibition, and weakened benefit from temporal dips in noise [10,60]. Neurocognitive factors such as slower processing speed and diminished working memory may contribute to these changes [5,6]. The absence of significant decline among males may reflect differential aging trajectories or compensatory neuroplastic mechanisms [61], but further research is needed to elucidate gender-specific auditory aging patterns.

Detailed phoneme-level error analysis revealed specific patterns consistent with the phonological characteristics of Hindi. Consonantal errors were markedly more frequent at lower SNRs, where masking effects were strongest. Voicing confusions (e.g., /b/ ↔ /p/, /k/ ↔ /g/, /ʈ/ ↔ /ɖ/) were prominent, underscoring the vulnerability of periodicity and voice-onset cues in noisy conditions, findings that align with Weber et al. [23] and Moreno-Torres et al. [24]. Errors in manner of articulation, such as /b/ ↔ /m/ and /d̪/ ↔ /n/, may reflect the degradation of nasal murmur and articulatory transitions.

The place of articulation errors, including retroflex–dental and palatal confusions (e.g., /ʈʰ/ ↔ /t͡ʃ/, /s/ ↔ /k/), suggest a loss of spectral localisation cues in multitalker babble. Aspiration, a phonemic feature in Hindi, was also significantly affected, with substitutions such as /pʰ/ ↔ /p/ and /bʱ/ ↔ /g/ pointing to the masking of low-intensity aspiration bursts. In contrast, vowel perception remained stable across SNRs, likely due to vowels' higher intensity, longer duration, and distinctive formant structure. This robustness is well-documented [24] and further supports the inclusion of consonant-heavy stimuli in SiN testing to maximise diagnostic sensitivity.

Conclusions

The present study developed the Speech-in-Noise Test in Hindi (SiNT-H) as a linguistically and culturally appropriate tool for assessing speech perception in noise among Hindi-speaking adults. Constructed using monosyllabic words and multitalker babble across various SNR levels, the SiNT-H demonstrated high test–retest reliability. The absence of significant ear-wise differences and the superior performance observed among females align with established gender-related auditory processing patterns. However, the decline in performance among middle-aged females may suggest early signs of auditory aging. It is important to note that the gender-related differences in the SiNT-H performance could also be influenced by factors such as sociolinguistic exposure, task familiarity, and cognitive engagement, which were not controlled in the present study. A detailed phonological error analysis revealed increased susceptibility to voicing, manner, place, and aspiration contrasts at lower SNRs (particularly among older participants), underscoring the test's sensitivity to language-specific phonetic demands. These findings collectively support the SiNT-H as a reliable and ecologically valid tool for clinical and research use in Hindi-speaking populations. Nonetheless, the inclusion of a broader sample of older adults could have further enhanced the study's value and should be considered in future investigations.

Limitations and future direction

The present findings should be interpreted in light of certain methodological considerations. Test–retest reliability was established on a relatively small subsample ($n = 12$), which, although adequate for preliminary estimation, may limit the precision and generalisability. Further, the interpretation of age- and gender-related effects warrants caution. Speech-in-noise performance is known to be sensitive to subtle auditory changes – such as cochlear synaptopathy, reduced temporal resolution, and suprathreshold processing deficits – that are not captured by conventional pure tone audiometry. In addition, the unequal gender distribution across age groups may have contributed to variability in group comparisons. Future investigations incorporating demographically balanced samples along with

extended audiological profiling would help delineate these effects more clearly.

Although acoustic normalisation and empirical list-equivalence procedures were implemented, formal psychometric equalisation at the item level was not undertaken. Consequently, a small degree of residual lexical or acoustic variability across test items cannot be entirely ruled out. Subsequent work employing item-level calibration using larger normative datasets would further enhance the standardisation, as well as the overall reliability and validity, of SiNT-H.

Acknowledgement

The authors acknowledge all the participants in the study.

Funding

The present study received no funds to report.

Other declarations

The authors declare no conflicts of interest. The data sets generated and/or analysed during the current study are available from the author on reasonable request.

Supplementary material

The supplementary materials are available at https://www.journalofhearingscience.com/

References

1. Middelweerd MJ, Festen JM, Plomp R. Difficulties with speech intelligibility in noise in spite of a normal pure-tone audiogram. Int J Audiol, 1990; 29(1): 1–7. https://doi.org/10.3109/00206099009081640
2. Lotto AJ, Holt LL. Speech perception: the view from the auditory system. In: Neurobiology of Language. Hickok G, Small SL, Editors. Elsevier eBooks; 2015, pp. 185–94. https://doi.org/10.1016/B978-0-12-407794-2.00016-X
3. Era P, Jokela J, Qvarnberg Y, Heikkinen E. Pure-tone thresholds, speech understanding, and their correlates in samples of men of different ages. Int J Audiol, 1986; 25(6): 338–52. https://doi.org/10.3109/00206098609078398
4. Beattie RC, Barr T, Roup C. Normal and hearing-impaired word recognition scores for monosyllabic words in quiet and noise. Br J Audiol, 1997; 31(3): 153–64. https://doi.org/10.3109/03005364000000018
5. Tremblay P, Brisson V, Deschamps I. Brain aging and speech perception: effects of background noise and talker variability. NeuroImage, 2020; 227: 117675. https://doi.org/10.1016/j.neuroimage.2020.117675
6. Dillard LK, Walsh MC, Merten N, Cruickshanks KJ, Schultz A. Prevalence of self-reported hearing loss and associated risk factors: findings from the survey of the health of Wisconsin. J Speech Lang Hear Res, 2022; 65(5): 2016–28. https://doi.org/10.1044/2022_jslhr-21-00580
7. Palva T. Studies of hearing for pure tones and speech in noise. Acta Oto-Laryngol, 1955; 45(3): 231–43. https://doi.org/10.3109/00016485509118154
8. Plomp R. A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired. J Speech Lang Hear Res, 1986; 29(2): 146–54. https://doi.org/10.1044/jshr.2902.146
9. Beattie RC. Word recognition functions for the CID W-22 test in multitalker noise for normally hearing and hearing-impaired subjects. J Speech Hear Disord, 1989; 54(1): 20–32. https://doi.org/10.1044/jshd.5401.20
10. Gordon-Salant S. Hearing loss and aging: new research findings and clinical implications. J Rehabil Res Dev, 2005; 42(4s): 9. https://doi.org/10.1682/jrrd.2005.01.0006
11. Miller GA, Heise GA, Lichten W. The intelligibility of speech as a function of the context of the test materials. J Exp Psychol, 1951; 41(5): 329–35. https://doi.org/10.1037/h0062491
12. Dubno JR, Dirks DD, Morgan DE. Effects of age and mild hearing loss on speech recognition in noise. J Acoust Soc Am, 1984; 76(1): 87–96. https://doi.org/10.1121/1.391011
13. Lutman ME, Brown EJ, Coles RRA. Self-reported disability and handicap in the population in relation to pure-tone threshold, age, sex and type of hearing loss. Br J Audiol, 1987; 21(1): 45–58. https://doi.org/10.3109/03005368709077774
14. Wilson RH, Strouse A. Word recognition in multitalker babble. American Speech Language-Hearing Association (ASHA) Convention. San Francisco, CA; 1999.
15. Carhart R, Tillman TW, Greetis ES. Perceptual masking in multiple sound backgrounds. J Acoust Soc Am, 1969; 45(3): 694–703. https://doi.org/10.1121/1.1911445
16. Davis H, Hudgins CV, Marquis RJ, Nichols RH, Peterson GE, Ross DA, Stevens SS. The selection of hearing aids Part II. Laryngoscope, 1946; 56(4): 135–63.
17. Killion MC, Niquette PA. What can the pure-tone audiogram tell us about a patient's SNR loss? Hear J, 2000; 53(3): 46–8. https://doi.org/10.1097/00025572-200003000-00006
18. Singh S, Black JW. Study of twenty-six intervocalic consonants as spoken and recognized by four language groups. J Acoust Soc Am, 1966; 39: 372–87.
19. Census of India. Sample Registration System, 2018. http://censusindia.gov.in
20. Mallikarjun B. Metamorphosis of 'Hindi' in Modern India: a study of Census of India. Language in India, 2019; 19(8).
21. Elliott LL. Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability. J Acoust Soc Am, 1979; 66(3): 651–53. https://doi.org/10.1121/1.383691
22. Schneider BA, Li L, Daneman M. How competing speech interferes with speech comprehension in everyday listening situations. J Am Acad Audiol, 2007; 18(7): 559–72. https://doi.org/10.3766/jaaa.18.7.4
23. Weber A, Smits R. Consonant and vowel confusion patterns by American English listeners. 2003. https://hdl.handle.net/11858/00-001M-0000-0013-2ED4-E
24. Moreno-Torres I, Otero P, Luna-Ramírez S, Heinze EG. Analysis of Spanish consonant recognition in 8-talker babble. J Acoust Soc Am, 2017; 141(5): 3079–90. https://doi.org/10.1121/1.4982251
25. Shipra, Chandra M. Hindi vowel classification using QCN-PNCC features. Indian J Sci Technol, 2016; 9(38). https://doi.org/10.17485/ijst/2016/v9i38/102972
26. Wang MD, Bilger RC. Consonant confusions in noise: a study of perceptual features. J Acoust Soc Am, 1973; 54(5): 1248–66. https://doi.org/10.1121/1.1914417

27. Cutler A, Weber A, Smits R, Cooper N. Patterns of English phoneme confusions by native and non-native listeners. J Acoust Soc Am, 2004; 116(6): 3668–78. https://doi.org/10.1121/1.1810292
28. Vaidyanath R, Yathiraj A. Influence of noise on the equivalence of word-lists in a phonemically balanced word test: comparison in young and older adults. Hear Bal Commun, 2019; 17(1): 42–50.
29. De NS. Hindi PB list for speech audiometry and discrimination test. Indian J Otolaryngol Head Neck Surg, 1973; 25(2): 64–75. https://doi.org/10.1007/bf02993883
30. Chinnaraj G, Neelamegarajan D, Ravirose U. Development, standardization, and validation of bisyllabic phonemically balanced Tamil word test in quiet and noise. J Hear Sci, 2022; 11(4): 42–47. https://doi.org/10.17430/jhs.2021.11.4.5
31. Zwicker E. Subdivision of the audible frequency range into critical bands (Frequenzgruppen). J Acoust Soc Am, 1961; 33(2): 248. https://doi.org/10.1121/1.190863
32. Gundmi A, Himaja P, Dhamani A. Effectiveness of multitalker babble over speech noise and its implications: a comparative study. Indian J Otol, 2018; 24(2): 88. https://doi.org/10.4103/indianjotol.indianjotol_24_18
33. Masalski M. The hearing test app for Android devices: distinctive features of pure-tone audiometry performed on mobile devices. Medical Devices: Evidence and Research, 2024: 151–63.
34. Pichora-Fuller MK, Schneider BA, Daneman M. How young and old adults listen to and remember speech in noise. J Acoust Soc Am, 1995; 97(1): 593–608. https://doi.org/10.1121/1.412282
35. Wilson RH, McArdle R. Intra- and inter-session test, retest reliability of the Words-in-Noise (WIN) test. J Am Acad Audiol, 2007; 18(10): 813–25. https://doi.org/10.3766/jaaa.18.10.2
36. Jain C, Narne VK, Singh NK, Kumar P, Mekhala M. The development of Hindi sentence test for speech recognition in noise. Int J Speech Lang Pathol Audiol, 2014; 2(2): 86–94. https://doi.org/10.12970/2311-1917.2014.02.02.5
37. Lee JY, Lee JT, Heo HJ, Choi C, Choi SH, Lee K. Speech recognition in real-life background noise by young and middle-aged adults with normal hearing. J Audiol Otol, 2015; 19(1): 39–44. https://doi.org/10.7874/jao.2015.19.1.39
38. Emami SF, Gohari N, Ramezani H, Borzouei M. Hearing performance in the follicular-luteal phase of the menstrual cycle. Int J Otolaryngol, 2018; 2018: 1–6. https://doi.org/10.1155/2018/7276359
39. Ghosh V, Devananda D, Harisanker SB, Kumar H. Speech perception in noise in Malayalam-speaking young adults with normal hearing. J Hear Sci, 2024; 14(2): 33–8. https://doi.org/10.17430/jhs/191377
40. Simpson SA, Cooke M. Consonant identification in N-talker babble is a nonmonotonic function of N. J Acoust Soc Am, 2005; 118(5): 2775–8. https://doi.org/10.1121/1.2062650
41. Kalikow DN, Stevens KN, Elliott LL. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. J Acoust Soc Am, 1977; 61(5): 1337–51. https://doi.org/10.1121/1.381436
42. Braza MD, Porter HL, Buss E, Calandruccio L, McCreery RW, Leibold LJ. Effects of word familiarity and receptive vocabulary size on speech-in-noise recognition among young adults with normal hearing. PLoS ONE, 2022; 17(3): e0264581. https://doi.org/10.1371/journal.pone.0264581
43. Puckett BJ, Ryherd EE, Bent T, Baese-Berk MM, Manley N. Medically related speech: impacts of age, familiarity, and noise on word recognition. J Acoust Soc Am, 2023; 153(Suppl 3): A167. https://doi.org/10.1121/10.0018538
44. Gu F, Kwong T, Wong L. Absence of ear advantage for segmental and tonal perception in noise. J Acoust Soc Am, 2017; 141(Suppl 5): 4032. https://doi.org/10.1121/1.4989294
45. Spyridakou C, Rosen S, Dritsakis G, Bamiou D. Adult normative data for the speech in babble (SiB) test. Int J Audiol, 2019; 59(1): 33–8. https://doi.org/10.1080/14992027.2019.1638526
46. Emami SF. The use of homotonic monosyllabic words in the Persian language for the Word-in-Noise Perception Test. Audit Vestib Res, 2023. https://doi.org/10.18502/avr.v33i1.14271
47. Philibert B, Veuillet E, Collet L. Functional asymmetries of crossed and uncrossed medial olivocochlear efferent pathways in humans. Neurosci Lett, 1998; 253(2): 99–102. https://doi.org/10.1016/s0304-3940(98)00615-6
48. Guinan JJ. Olivocochlear efferents: anatomy, physiology, function, and the measurement of efferent effects in humans. Ear Hear, 2006; 27(6): 589–607. https://doi.org/10.1097/01.aud.0000240507.83072.e7
49. Mertes IB, Wilbanks EC, Leek MR. Olivocochlear efferent activity is associated with the slope of the psychometric function of speech recognition in noise. Ear Hear, 2017; 39(3): 583–93. https://doi.org/10.1097/aud.0000000000000514
50. Mertes IB, Johnson KM, Dinger ZA. Olivocochlear efferent contributions to speech-in-noise recognition across signal-to-noise ratios. J Acoust Soc Am, 2019; 145(3): 1529–40. https://doi.org/10.1121/1.5094766
51. Yumba WK. Influences of listener gender and working memory capacity on speech recognition in noise for hearing aid users. Speech Lang Hear, 2020; 25(2): 112–24. https://doi.org/10.1080/2050571x.2020.1810491
52. Halpern DF. A cognitive-process taxonomy for sex differences in cognitive abilities. Curr Dir Psychol Sci, 2004; 13(4): 135–9. https://doi.org/10.1111/j.0963-7214.2004.00292.x
53. Luders E, Gaser C, Narr KL, Toga AW. Why sex matters: brain size independent differences in grey matter distributions between men and women. J Neurosci, 2009; 29(45): 14265–70. https://doi.org/10.1523/jneurosci.2261-09.2009
54. Jäncke L, Wüstenberg T, Scheich H, Heinze H. Phonetic perception and the temporal cortex. NeuroImage, 2002; 15(4): 733–46. https://doi.org/10.1006/nimg.2001.1027
55. Frye RE, Fisher JM, Witzel T, Ahlfors SP, Swank P, Liederman J, Halgren E. Objective phonological and subjective perceptual characteristics of syllables modulate spatiotemporal patterns of superior temporal gyrus activity. NeuroImage, 2008; 40(4): 1888–901. https://doi.org/10.1016/j.neuroimage.2008.01.048
56. McArdle R, Wilson RH. Predicting word-recognition performance in noise by young listeners with normal hearing using acoustic, phonetic, and lexical variables. J Am Acad Audiol, 2008; 19(6): 507–18. https://doi.org/10.3766/jaaa.19.6.6
57. Cruickshanks KJ, Tweed TS, Wiley TL, Klein BEK, Klein R, Chappell R, Nondahl DM, Dalton DS. The 5-year incidence and progression of hearing loss. Arch Otolaryngol Head Neck Surg, 2003; 129(10): 1041. https://doi.org/10.1001/archotol.129.10.1041
58. Banks B, Gowen E, Munro KJ, Adank P. Cognitive predictors of perceptual adaptation to accented speech. J Acoust Soc Am, 2015; 137(4): 2015–24. https://doi.org/10.1121/1.4916265
59. Schoof T, Rosen S. The role of age-related declines in subcortical auditory processing in speech perception in noise. J Assoc Res Otolaryngol, 2016; 17(5): 441–60. https://doi.org/10.1007/s10162-016-0564-x

60. Han JH, Kim JH, Park GK, Lee HJ. Preserved grey matter volume in the left superior temporal gyrus underpins speech-in-noise processing in middle-aged adults. J Int Adv Otol, 2024; 20(1): 62–8. https://doi.org/10.5152/iao.2024.231241

61. Pichora-Fuller MK. Cognitive aging and auditory information processing. Int J Audiol, 2003; 42(Suppl 2): 26–32. https://doi.org/10.3109/14992020309074641

Mansi Pokiya, email: mansipokiya1456@gmail.com • 0009-0009-0023-6930
Anuj Kumar Neupane, email: anujkneupane@gmail.com • 0000-0002-0320-4709
Kristi Kaveri Dutta, email: kristiaudiologist@gmail.com • 0000-0002-5040-587X