IJMS
Vol 51, No 7, July 2026

Original Article

Development and Validation of Persian Speech Test Signal (PerSTS)

Sasan Etezadi[1], PhD candidate; Nariman Rahbar[1], PhD; Seyyed Jalal Sameni[1], PhD; Farhad Torabinezhad[2], PhD; Behrooz Mahmoodi-Bakhtiari[3], PhD

[1]Rehabilitation Research Center, Department of Audiology, School of Rehabilitation Sciences, Iran University of Medical Sciences, Tehran, Iran;
[2]Rehabilitation Research Center, Department of Speech and Language Pathology, School of Rehabilitation Sciences, Iran University of Medical Sciences, Tehran, Iran;
[3]Department of Performing Arts, College of Fine Arts, University of Tehran, Iran

Correspondence:
Nariman Rahbar, PhD;
School of Rehabilitation Sciences,
Iran University of Medical Sciences,
Madadkaran Street, Shah Nazari Street,
Madar Square, Mirdamad, Postal code:
15459-13487, Tehran, Iran
Tel: +98 21 22228051
Email: rahbar.n@iums.ac.ir
Received: 03 December 2025
Revised: 20 February 2026
Accepted: 13 March 2026

Abstract

Background: Digital hearing aids require precise fitting to enable realistic speech processing. The international speech test signal (ISTS) provides a standard reference, yet does not fully capture the phonetic and prosodic features of Persian. This study aimed to create and acoustically verify a Persian version of the ISTS—termed the Persian speech test signal (PerSTS)—to enhance verification accuracy for Persian-speaking users.

Methods: The North Wind and the Sun (NWS) passage was translated into Persian with careful phoneme distribution and natural syntax. Six native female speakers recorded the passage at 44.1 kHz and 24-bit resolution. These recordings were segmented, enhanced, and stitched into 10-15 second sequences that adhered to pitch and spectral criteria. Sequences that closely matched the Byrne female long-term average speech spectrum (LTASS; ±1 dB) and were rated as most natural by linguists were assembled into a 60-second PerSTS. The final audio was calibrated to 65 dB sound pressure level (SPL) and analyzed for its spectral, temporal, and amplitude characteristics in Tehran, Iran (2025).

Results: The PerSTS exhibited spectral peaks between 200–400 Hz and 3–4 kHz, characteristic of female Persian speech. Its LTASS closely matches the Byrne LTASS and ISTS, with minor deviations below 500 Hz and above 8 kHz. The crest factor (16.7 dB) and dynamic range (27.1 dB) were consistent with those of natural speech. Experts rated its naturalness as 4.35±0.28 on a five-point scale.

Conclusion: The PerSTS mimics the acoustic and temporal features of natural Persian speech within International Electrotechnical Commission (IEC) standards, providing a standardized tool for hearing aid verification and audiology.

Please cite this article as: Etezadi S, Rahbar N, Sameni SJ, Torabinezhad F, Mahmoodi-Bakhtiari B. Development and Validation of Persian Speech Test Signal (PerSTS). Iran J Med Sci. 2026;51(7):505-511. doi: 10.30476/ijms.2026.109949.4539.

Keywords ● Hearing aids ● Acoustics ● Speech ● Language

What's Known

- In 2010, the European Hearing Instrument Manufacturers Association introduced the international speech test signal as a standardized speech stimulus for verifying hearing aid performance. This signal has since been adapted into Polish, Brazilian Portuguese, and Korean.

What's New

- The Persian Speech Test Signal, as the Persian version of the International Speech Test Signal, simulates the acoustic and temporal features of natural Persian speech, providing a standardized and language-specific resource for hearing aid verification.

Introduction

Digital hearing aids are essential to modern hearing rehabilitation, offering advanced features, improved sound quality, and enhanced user satisfaction. Continued development in this field enhances accessibility, efficiency, and quality of life for individuals with hearing loss. Accurate hearing aid fitting and verification are critical to ensure these advantages, particularly when speech-based test signals are used to optimize performance.[1] Evidence-based hearing aid fitting protocols aim to improve speech audibility, with verification using speech or speech-like signals.[2] Speech spectral patterns, fundamental frequency (F_0), and temporal structure vary across speakers and languages.[3] Band importance functions tend to peak at higher frequencies for contextually weak, meaningless syllables, shifting toward lower frequencies as contextual weighting increases.[4] Languages with greater "phoneme distinctive content" in certain frequency regions, such as Arabic and Chinese, should receive increased gain in those regions to enhance speech clarity and understanding.[5] Most variability in speech spectra stems from individual differences and phonological distinctions between languages.[6]

Digital speech-like stimuli share some similarities with natural speech, yet lack its co-modulation features.[7] One such modulated signal is the International Collegium for Rehabilitative Audiology (ICRA) signal. Although the ICRA signal resembles natural speech in many ways, it has several notable limitations. For instance, it lacks speech-like periodic features such as F_0 and is frequently classified as noise rather than speech by hearing aid algorithms. Another significant limitation is that the ICRA signal does not allow audiologists to assess distortion levels in hearing aids. Its unnatural sound quality has also restricted its practical use.[3] Given the essential role of speech signals in hearing aid evaluation and fitting, the European Hearing Instrument Manufacturers Association introduced the international speech test signal (ISTS) in 2010 as a standardized stimulus for verifying hearing aid performance. The ISTS includes recordings from six female speakers representing six languages. However, it does not cover all languages and remains unintelligible in any single language.[3] Derived from "The North Wind and the Sun" (NWS) passage, the ISTS has limitations in fully representing the diversity of English dialects and human speech production.[8] The stimulus was developed using a limited set of languages and speakers,[3] and the distribution of speech pauses and their durations was derived from a specific speech corpus.[3] Consequently, the limited linguistic content of the ISTS, including both linguistic and supra-linguistic features, results from its multilingual composition. It is not phonetically balanced and lacks prosodic elements, which may compromise the success of hearing aid fittings for speakers of certain languages, such as Indo-Aryan, Japonic, Slavic, or Turkic.[9] Several language-specific adaptations of the ISTS have since been developed, including versions in Polish, Brazilian Portuguese, Korean, and Turkish, to better reflect the phonetic and prosodic characteristics of these languages and improve hearing aid verification outcomes in local populations.[2, 9-11] To the best of our knowledge, no existing stimulus accurately reproduces the electroacoustic features of the original ISTS while capturing the phonetic and linguistic characteristics of Persian. Therefore, this study aimed to develop and validate a PerSTS stimulus using the Persian version of the NWS passage.

Materials and Methods

The study protocol was approved by the Human Research Ethics Committee of Iran University of Medical Sciences, Tehran, Iran (code: IR.IUMS.REC.1402.988).

Speech Material

The NWS text was translated into Persian following a systematic approach to ensure linguistic and phonetic suitability. Initially, the original passage was analyzed to optimize phoneme distribution and balance. The translation process involved careful lexical selection and syntactic restructuring to align with Persian grammar, particularly by adopting the subject-object-verb word order. Less common Persian phonemes were intentionally incorporated to reflect the language's phonetic diversity while maintaining semantic clarity. Phonemes were transcribed using the International Phonetic Alphabet based on Martin's framework. The translated passage was evaluated by a panel of native Persian speakers, who assessed its clarity, naturalness, and overall linguistic appropriateness.[12]

Speech Recording

The stimulus construction method in this study was adopted from the protocol developed by Holube and colleagues (2010), whose work served as the primary reference for the development of the ISTS.[3] Six female Persian speakers who met the language inclusion criteria recorded themselves reading multiple versions of the translated tale, The NWS, at a natural speaking rate and with typical intonation. A unidirectional condenser microphone (Behringer, Germany) was positioned 20-30 cm from their mouths, at an angle of approximately 45 degrees. The text was placed approximately 50 cm in front of the speakers, slightly above eye level. Recordings were made with 24-bit resolution and a sampling rate of 44.1 kHz. All voice recordings were collected as part of a non-interventional acoustic study.

Segmentation

Annotation Pro software (Klessa, Poland) automatically divides recordings into monosyllabic speech units. It extracts the first 500-millisecond segments and calculates the energy in the final 400 milliseconds at 10-millisecond intervals. The point with the lowest absolute amplitude within the lowest-energy 10-millisecond interval is identified and marks the start of the next 500-millisecond segment. To prevent vowels and phonemes from being cut off or omitted, a researcher manually reviewed the computer-generated segmentation. The adjusted segments typically ranged from 100 to 600 milliseconds. Speech pauses longer than 100 milliseconds were included in the preceding speech segment to preserve natural timing. For longer segments (10 to 15 seconds), the individual segments were concatenated in a pseudo-random order.

Composition Rules

During concatenation, several constraints and modifications were applied to ensure a natural-sounding output. Each segment was tapered at both ends with a 1 ms Hanning window in Praat to prevent audible artifacts. No segment was reused within any 10- or 15-second sequence. To ensure natural pitch transitions, the F_0 difference between two voiced segments was restricted to 10 Hz. The F_0 was estimated over the first and last 30 ms of each segment through a modified autocorrelation method. Segment boundaries were classified as either "voiced" or "unvoiced" and joined accordingly.

Selection and Evaluation of Segments

Numerous sequences were generated and analyzed based on their long-term average speech spectrum (LTASS) and pause durations. Sequences that exhibited more than a 3 dB deviation in any one-third octave band from the Byrne female LTASS (200–5000 Hz), or that contained pauses longer than 650 ms, were eliminated. The remaining sequences were then assessed by a panel of five linguists—none of whom were involved in creating the stimuli—who rated their naturalness on a scale from 1 (least natural) to 5 (most natural) with 0.25-point increments. The highest-rated sequences were selected for the final stimulus compilation.

Filtering and Calibration

The final sequences were refined to align with the Byrne female LTASS curve within ±1 dB across all one-third octave bands. Root mean square (RMS) levels in the 200-5000 Hz range were matched across sequences. The overall level of the PerSTS stimulus was set to 65 dB sound pressure level (SPL) within this band, corresponding to typical conversational speech at 1 meter. Calibration was performed in accordance with Noah system standards.

Final Stimulus Construction

The final PerSTS was assembled from a carefully selected set of 10- and 15-second segments to create a continuous 60-second stimulus that met the acoustic and perceptual criteria described above.

Acoustic Analyses

The constructed PerSTS stimulus underwent multiple acoustic analyses to assess its similarity to natural speech and its appropriateness as a test signal. Percentile analysis (L30, L65, L99) was performed across one-third-octave bands from 125 Hz to 16 kHz, following the International Electrotechnical Commission (IEC) 60118-15 guidelines.[13] From these values, the spectral crest factor (L99−L65) and dynamic range (L99−L30) were calculated. The LTASS was determined using the 65th percentile levels. Additional analyses included pause duration, F_0 distribution, and segment naturalness ratings provided by an expert panel.

Results

Percentile Analysis

Figure 1 presents the percentile spectra (L30, L65, L99) of the PerSTS. The spectral shape shows clear peaks between 200 and 400 Hz, and again between 3 and 4 kHz, both of which are characteristic of typical female conversational speech.

Spectral Comparison

Figure 2 illustrates the deviation of the PerSTS spectrum from the Byrne female LTASS reference. When comparing the spectral features of PerSTS and ISTS, both signals exhibit broadly similar percentile patterns across the 125-8000 Hz range. Nevertheless, some differences are apparent. Larger discrepancies were observed above 8 kHz, which might stem from the microphone's recording characteristics limiting accurate capture beyond this frequency, as its frequency response above 8 kHz also appeared slightly flattened. Smaller variations were also observed in certain low- and mid-frequency ranges, such as 200–500 Hz and 3–5 kHz. These differences likely arose from recording conditions—including speaker articulation and the studio's acoustic environment—as well as minor effects introduced by segmentation and concatenation. Overall, despite these differences, the spectral characteristics of PerSTS remain consistent with ISTS, confirming its validity as a naturalistic speech test signal.

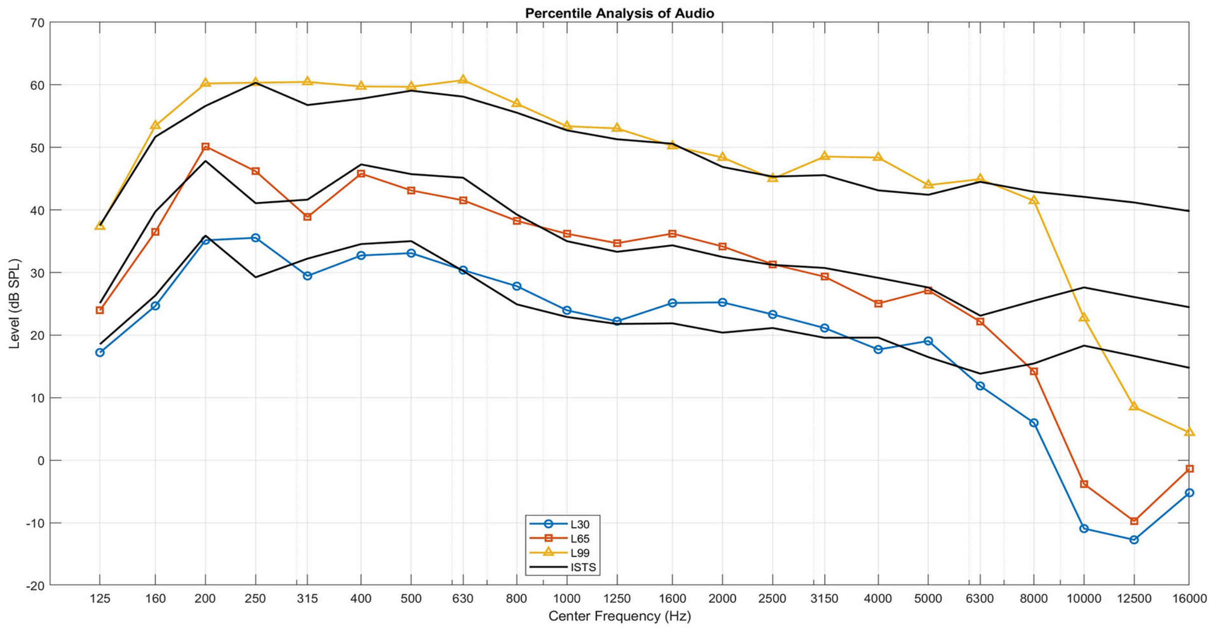

Figure 1: This graph represents the percentile spectra (L30, L65, L99) of the Persian speech test signal. The vertical axis represents sound intensity in dB SPL, and the horizontal axis shows center frequency from 125 to 16000 Hz. SPL: Sound pressure level; ISTS: International Speech Test Signal

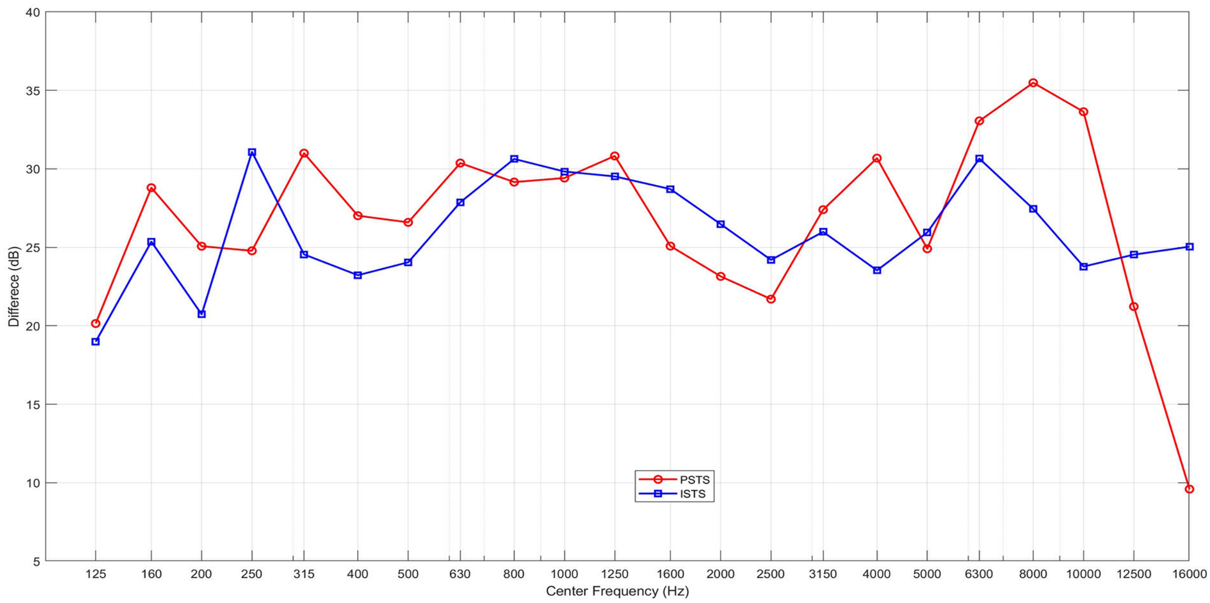

Figure 2: The red line represents the Persian speech test signal spectrum, and the blue line shows the Byrne female long-term average speech spectrum (LTASS). The vertical axis represents the difference of sound intensity in dB SPL, and the horizontal axis shows center frequency from 125 to 16000 Hz. SPL: Sound pressure level; ISTS: International speech test signal; PerSTS: Persian speech test signal

Crest Factor and Dynamic Range

Analysis of percentile levels revealed that the average spectral crest factor (L99−L65) across the frequency range of 200–5000 Hz was 16.7 dB. Meanwhile, the average dynamic range (L99−L30) was 27.1 dB, consistent with the amplitude variability typically observed in natural conversational speech.

Pause Duration and F_0 Continuity

All concatenated sequences adhered to the set thresholds: pauses did not exceed 650 ms, and F_0 discontinuities between voiced segments were kept at or below 10 Hz. This maintained smooth transitions without noticeable artifacts.

Expert Panel Evaluation

An independent panel of five linguists, none of whom were involved in creating the signals, assessed the naturalness of the PerSTS sequences. They rated each sequence on a five-point scale in 0.25 increments. The overall average score was 4.35±0.28, indicating that the stimuli were generally perceived as highly natural. The top-rated sequences were selected for the final 60-second PerSTS.

Discussion

In this study, we developed and validated a PerSTS stimulus using the Persian version of the NWS passage. The acoustic analysis demonstrated that the spectral and temporal features of the PerSTS closely matched those of the ISTS, confirming its suitability as a standardized speech-like stimulus for Persian speakers. The percentile analysis (L30, L65, L99) revealed a spectral profile similar to that of female conversational speech, with distinct peaks at 200–400 Hz and 3–4 kHz. These frequencies aligned with the main formant energy of Persian vowels and the high-frequency fricatives typical in Persian phonology. Although minor deviations were observed in the low (200–500 Hz) and mid-frequencies (3–5 kHz) ranges, the overall spectral balance of PerSTS remained within the tolerance limits specified by IEC 60118-15 for speech test signals.[13] The discrepancies observed above 8 kHz were primarily attributable to the recording microphone's limitations rather than to any linguistic factors.

The spectral crest factor (16.7 dB) and dynamic range (27.1 dB) of the PerSTS indicated that the signal maintained the amplitude modulation complexity characteristic of natural speech. PerSTS's dynamic range fell within the typical 20–30 dB range of speech.[3] Comparable crest factors were observed in original ISTS (9.7 dB), Polish speech test signal (13.7 dB), and the Brazilian Portuguese speech test signal (11 dB).[2, 3, 9] These comparisons demonstrated that PerSTS exhibited spectral peakedness and amplitude-modulation complexity similar to those of natural speech. The higher crest factor (peak-to-RMS ratio) suggested a broader range of instantaneous levels. Thus, the PerSTS effectively replicated the time-based energy pattern required for realistic speech-like stimulation in hearing aid testing.

Analysis of segment composition showed that the PerSTS maintained a suitable balance of voiced and voiceless sounds, aligning with Persian's phonetic structure. Persian features a relatively wide range of voiceless fricatives and affricates, which accounts for the higher proportion of unvoiced segments seen in the PerSTS.[14] Preserving this balance is crucial, as the ratio of voiceless fragments influences the spectral tilt and modulation depth of speech-like signals.[15] In contrast, the ISTS—derived from multiple languages—showed a more averaged spectral pattern that might not fully capture language-specific voicelessness features. By preserving these details, the PerSTS offers a spectral and temporal envelope closer to natural Persian speech, which could provide more accurate acoustic input for speech intelligibility testing.

Temporal analysis confirmed that the concatenation process effectively preserved natural rhythm and pause patterns. All speech segments satisfied the threshold criteria for pause duration (<650 ms) and F_0 continuity (≤10 Hz between voiced parts), resulting in seamless transitions without noticeable artifacts. As Holube and colleagues described for the ISTS, pauses naturally occurring at segment ends were limited (originally ≤600 ms) to avoid "dead air".[3] The shorter, more frequent speech segments observed in PerSTS mirror features of Persian prosody, such as rapid syllable shifts and brief inter-phrase pauses, as documented in previous acoustic research on Persian speech.[16, 17] In Sadat-Tehrani's analysis, an intonational phrase in Persian is typically followed by a brief pause and pitch reset.[18] Incorporating these prosodic features ensures that PerSTS captures both the linguistic and supra-linguistic aspects of Persian, which are essential for replicating the perceptual nuances of natural speech.

The perceptual evaluation conducted by expert linguists confirmed that the sequences sounded highly natural, with an average rating of 4.35±0.28. Their assessment also indicated that the concatenation and filtering methods effectively preserved important acoustic cues while minimizing artifacts. Maintaining this perceptual naturalness is crucial for ensuring realistic spectral-temporal characteristics.

Although this study did not include hearing aid fitting tests, the high acoustic similarity between the PerSTS and natural Persian speech suggested that PerSTS could serve as a valuable tool for clinical validation and electroacoustic analysis in Persian-speaking groups. Previous research demonstrated that using language-specific speech stimuli could enhance the precision of hearing aid gain settings and better represent speech audibility in everyday listening situations.[19] By providing a linguistically appropriate alternative to the ISTS, the PerSTS addressed an important gap in standardized testing materials for non-European languages. Its adoption could lead to more accurate assessments in Persian audiology and support the development of culturally and linguistically tailored rehabilitation protocols.

Several limitations of this study should be acknowledged. The recordings were made using only female voices, which, while consistent with the ISTS protocol, might not fully reflect the acoustic diversity of Persian speech, including male and child voices. Future research should incorporate speakers of different sexes, dialects, and speaking styles to enhance the signal's generalizability. Another limitation concerns the recording characteristics of the unidirectional condenser microphone used. Its frequency response above 8 kHz appeared slightly flattened, affecting its ability to accurately record sounds beyond this frequency range. Although the acoustic analyses verified the signal's validity, additional studies are required to assess its perceptual and electroacoustic equivalence to ISTS in hearing aid applications. Conducting real-ear verification and aided speech perception tests with Persian-speaking listeners would provide valuable insights into the clinical relevance and applicability of this approach.

Conclusion

In summary, this research successfully developed and validated the PerSTS as a linguistically appropriate and technically reliable indicator for hearing aid testing in Persian-speaking populations. Acoustic analysis demonstrated that the PerSTS closely matched the spectral and temporal features of natural Persian speech and the international ISTS standard. The stimulus preserves phonetically balanced voicing components and natural prosody, including appropriate pause durations and pitch continuity, resulting in high naturalness ratings from expert linguists. Its adoption could enhance the accuracy of electroacoustic assessments and speech clarity evaluations, addressing a critical gap in standardized audiological tools for Persian speakers and supporting the development of culturally appropriate rehabilitation practices.

Acknowledgment

This article was extracted from the PhD thesis of Sasan Etezadi in the Audiology major, which was submitted to Iran University of Medical Sciences, Tehran, Iran.

Authors' Contribution

S.E: Participated in designing the study, collecting data, analyzing results, and drafting the manuscript; N.R: Designed and oversaw the study; assisted with study design, data analysis, and critical manuscript revisions; S.J.S: Contributed to data analysis, developed methodologies, and revised the manuscript; F.T: Participated in designing the study, offered clinical and linguistic guidance, and thoroughly reviewed the manuscript; B.M.B: Offered theoretical guidance, assisted in interpreting results, and revised the manuscript to enhance its intellectual content. All authors have read and approved the final manuscript and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Declaration of AI

The authors declare that no AI tools were used in the preparation of this manuscript.

Conflict of Interest: None declared.

References

1 Picou EM. Hearing Aid Benefit and Satisfaction Results from the MarkeTrak 2022 Survey: Importance of Features and Hearing Care Professionals. Semin Hear. 2022;43:301-16. doi: 10.1055/s-0042-1758375. PubMed PMID: 36466566; PubMed Central PMCID: PMCPMC9715311.
2 Garolla LP, Scollie SD, Martinelli Iorio MC. Development of the speech test signal in Brazilian Portuguese for real-ear measurement. Int J Audiol. 2013;52:572-6. doi: 10.3109/14992027.2013.791031. PubMed PMID: 23713471.
3 Holube I, Fredelake S, Vlaming M, Kollmeier B. Development and analysis of an International Speech Test Signal (ISTS). Int J Audiol. 2010;49:891-903. doi: 10.3109/14992027.2010.506889. PubMed PMID: 21070124.
4 Ricketts T, Bentler RA, Mueller HG. Essentials of modern hearing aids: selection, fitting, and verification. San Diego: Plural Publishing; 2017.
5 Chasin M. Setting hearing aids differently for different languages. Semin Hear. 2011;32:182-8. doi: 10.1055/s-0031-1277240.
6 Byrne D, Dillon H, Tran K, Arlinger S, Wilbraham K, Cox R, et al. An international comparison of long-term average speech spectra. J Acoust Soc Am. 1994;96:2108-20. doi: 10.1121/1.410152.
7 Dreschler WA, Verschuure H, Ludvigsen C, Westermann S. ICRA noises: artificial noise signals with speech-like spectral

and temporal properties for hearing instrument assessment. International Collegium for Rehabilitative Audiology. Audiology. 2001;40:148-57. PubMed PMID: 11465297.
8 Deterding D. The North Wind versus a Wolf: short texts for the description and measurement of English pronunciation. J Int Phon Assoc. 2006;36:187-96. doi: 10.1075/eww.27.2.04det.
9 Habasińska D, Skrodzka E, Bogusz-Witczak E. Development of the Polish Speech Test Signal and its comparison with the International Speech Test Signal. Arch Acoust. 2023;48:321-35. doi: 10.24425/122373.
10 Lee K, Jin IK. A comparison of acoustical characteristics between International Speech Test Signal and Korean speech stimuli. Commun Sci Disord. 2018;23:169-79. doi: 10.12963/csd.17463.
11 Kara E, Akin Şenkal Ö, Çetin Kara H, Gulmez ZD, Akdogan S, Yener HM. An explanatory research on the development of the speech test signal in hearing-aid fitting. Arch Health Sci Res. 2022;9:198-202. doi: 10.12963/csd.17463
12 Etezadi S, Rahbar N, Sameni SJ, Torabinezhad F, Mahmoodi-Bakhtiari B. Translation and adaptation of the North Wind and the Sun passage for Persian: a phonetically balanced passage for the Persian language. Aud Vestib Res. 2025. Persian. doi: 10.18502/avr.v35i2.21198.
13 Dworsack-Dodge M, Switalski W. Current practices in modern probe microphone measurement. 2012.
14 Falahati R, Ghane Z. Handbook of phonetics and phonology of modern Iranian languages. Springer Handb Lang Linguist. 2025. doi: 10.1007/978-981-95-0383-4.
15 Lu Y, Cooke M. The contribution of changes in F0 and spectral tilt to increased intelligibility of speech produced in noise. Speech Commun. 2009;51:1253-62. doi: 10.1016/j.specom.2009.07.002.
16 Taghva N, Moloodi A, Abolhasani Zadeh V. Acoustic correlations of speech rhythms in Persian based on variability of between-speakers characteristics. J Res Linguist. 2020;12:27-50. Persian.
17 Nikravesh M, Aghajanzadep M. Speech prosody in Persian language. Zahedan J Res Med Sci. 2014;16:68-72.
18 Sadat-Tehrani N. The structure of Persian intonation. In: Proceedings of the 4th International Conference on Speech Prosody. Campinas: ISCA; 2008. p. 249-52. doi: 10.21437/SpeechProsody2008-54.
19 Matsumoto N, Suzuki N, Iwasaki S, Ishikawa K, Tsukiji H, Higashino Y, et al. Language-specific strategy for programming hearing aids - A double-blind randomized controlled crossover study. Auris Nasus Larynx. 2018;45:686-92. doi: 10.1016/j.anl.2017.11.007. PubMed PMID: 29157624.