Research Article

Development and Validation of Azeri Version of Speech Recognitifon Threshold Test in Adult Azeri Population and Compiling Psychometrically Equivalent Lists

Asma Hammami Dizaj[1,2], Nariman Rahbar[1,2*], Behrooz Mahmoodi-Bakhtiari[3], Seyyed Jalal Sameni[1,2], Nayiere Mansouri[4]

[1.] Rehabilitation Research Center, Iran University of Medical Sciences, Tehran, Iran

[2.] Department of Audiology, School of Rehabilitation Sciences, Iran University of Medical Sciences, Tehran, Iran

[3.] Department of Performing Arts, University of Tehran, Tehran, Iran

[4.] Department of Audiology, School of Rehabilitation Sciences, Tabriz University of Medical Sciences, Tabriz, Iran

ORCID ID:
Asma Hammami: 0009-0009-4544-9032
Nariman Rahbar: 0000-0002-7724-632X
Behrooz Mahmoodi-Bakhtiari: 0000-0002-5671-4040
Seyed Jalal Sameni: 0000-0002-5520-3409
Nayiere Mansouri: 0000-0002-4230-7567

Citation: Hammami Dizaj A, Rahbar N, Mahmoodi-Bakhtiari B, Sameni SJ, Mansouri N. Development and Validation of Azeri Version of Speech Recognition Threshold Test in Adult Azeri Population and Compiling Psychometrically Equivalent Lists. Aud Vestib Res. 2025;34(1):?-?.

Article info:
Received: 30 Jan 2024
Revised: 02 Apr 2024
Accepted: 01 May 2024

* **Corresponding Author:** Department of Audiology, School of Rehabilitation Sciences, Iran University of Medical Sciences, Tehran, Iran. E-mail: rahbar.n@iums.ac.ir

Short running title: Development and Validation of Azeri Version…

Highlights:

- SRT is a widely employed method in speech audiometry examination
- This research presents the first SRT test conducted in the Azeri language
- The study generated five lists of ten disyllabic Azeri words

ABSTRACT

Background and Aim: One of the various measures used in speech audiometry is the speech recognition threshold. The aim of this study is to construct psychometrically equivalent lists containing Azeri two-syllable words to evaluate speech reception threshold in Azeri-speaking adults.

Methods: This study is a cross-sectional comparative study of test construction. Common two-syllable words were collected from Azeri books. Words with features of ease, familiarity, and relatedness as well as phonetic balance and phonetic dissimilarity were included in 5 lists of 10 words. The lists were evaluated for psychometric evaluations and validity and reliability on young Azeri population at 5 intensity levels from –5 to 15 dB HL. After two to four weeks, people were tested again under the same conditions.

Results: All 5 lists showed favorable face, content validity index and content validity ratio. The average speech thresholds obtained were 9.62, 9.77, 9.57, 9.67 and 9.57, respectively. Cronbach's alpha coefficient and halving were obtained above 0.8 in all 5 lists. Test-retest reliability was also significant ($p<0.001$).
Conclusion: The review and analysis performed on the data show that all 5 lists have psychometric homogeneity and alignment and show a high degree of validity and reliability; which makes them suitable for evaluating speech reception threshold in hearing centers of Azari regions.
Keywords: Psychometric function; speech reception threshold test; Azeri; language; validity; reliability

Introduction

The Speech Reception Threshold (SRT) is a widely employed method for assessing an individual's speech comprehension ability during a speech audiometry examination. The SRT is determined as the minimum sound level at which an individual can consistently comprehend 50% disyllabic words [1, 2]. The speech reception threshold has a plethora of clinical applications, including establishing a correlation with pure tone thresholds, determining appropriate levels for conducting speech recognition tests beyond threshold levels, evaluating the efficacy of hearing aids, deciding the need for aural rehabilitation, and evaluating the hearing sensitivity of infants and young children [3]. The selection and documentation of word materials should be based on familiarity, homogeneity, phonetic balance and phonetic dissimilarity, word materials should be chosen and documented [4, 5]. In order to ensure that a hearing evaluation truly measures hearing acuity and not receptive vocabulary, it is imperative to use SRT materials that make use of words that are commonly used and well-known in a language [6, 7].

Homogeneity is a key psychometric property that should be considered when designing testing instruments. Dillon emphasized the importance of matching the difficulty and understandability of each test item or question set as a function of intensity to achieve homogeneity [7]. Homogeneity of audibility is also crucial, as noted by Ramkissoon, to ensure that all test stimuli are equally audible, thereby enhancing the reliability of the measure and ensuring a fair assessment [5]. It is important to note that the presence of similar or rhyming words in the list imposes an auditory discrimination that is useless in threshold tests. Therefore, phonetic dissimilarity between items is crucial for accurate testing results [4]. The criterion of phonetic balance is less important among these criteria [8, 9].

Azeri, a member of the Turkic language family, is the native language of nearly 20 million individuals. The Azeris speak Azeri Turkish, a distinct variant from Istanbul Turkish and other Turkish dialects spoken in Iran, including Qashqai and Turkman [10]. In audiology, standardized speech audiometry material is crucial to assess hearing ability. However, such materials were not readily available in Azeri Turkish, highlighting the need for their development to improve audiological care for this language community.

Methods

The present study was designed to develop a new assessment tool for Azeri speakers using the cross-sectional method. Prior to conducting the study, the medical ethics committee of Iran University of Medical Sciences approved the research protocol. Participants were informed of their right to withdraw from the study at any time for any reason. The study consisted of two main stages. The first stage involved the development of test lists and the examination of their content validity, while the second stage focused on determining the equivalency and reliability of the lists. Overall, this study represents a rigorous and comprehensive approach to the development and validation of a new assessment tool. All statistical analyses were performed with the SPSS version 16 software package.

Materials

The process of selecting speech materials is crucial in designing effective speech audiometry tests. In this regard, disyllabic words have been considered as one of the most suitable units for testing SRT. The current study focused on selecting disyllabic words from a famous Azeri literary work, Heydar Baba Salam, authored by Mohammad Hossein Behjat Tabrizi, also known as Shahriar. Being a prominent Iranian Azeri poet who wrote in both Azeri and Persian, his literary works are known for their rich cultural and linguistic values. The selection process involved choosing 1004 disyllabic words from diverse sources of Azeri literature. The chosen words were carefully scrutinized to eliminate those with unsuitable cultural and semantic loads. The final selection consisted of disyllabic nouns with the most frequent syllable structures of Azeri, including Consonant Vowel Consonant Vowel Consonant (CVCVC), Consonant Vowel Consonant Consonant Vowel Consonant (CVCCVC), Consonant Vowel Consonant Vowel (CVCV), and Consonant Vowel Consonant Consonant Vowel (CVCCV). Proper nouns such as names of people or cities, plural words, pseudowords, and numbers were excluded from the final list.

Validity of materials

In the realm of psychometric testing, assessing the validity of materials for word lists is a critical process. Validity can be evaluated through various means, including face validity, content validity, and construct validity In this context, the content validity of disyllabic words was assessed by obtaining the opinions of eight audiologists and speech pathologists. They were asked to provide feedback on a questionnaire containing a set of selected words to determine whether the chosen words were suitable for achieving the intended test goals. These experts hailed from different Azari-speaking provinces of Iran, ensuring that the word list contained common and familiar words from diverse Azeri regions of the country. The suitability of each word was evaluated to ensure that the test items were measuring the intended purposes. Words that were deemed unsuitable or unnecessary were eliminated, and necessary changes were made based on the experts' suggestions, such as corrections and additions. Subsequently, the Content Validity Ratio (CVR) was calculated for each item and, to calculate the CVI, the average CVR scores of the items that were retained was used [11, 12]. To ensure face validity, participants were asked to provide their opinions on the familiarity and clarity of words. This rigorous process of validity assessment ensured the selection of appropriate words for the intended test goals. Also, construct validity through factor analysis was used to check and determine the number of factors from the interpreted variance test. Logistic regression model was used to plot third degree polynomial psychometric curves.

Participants

In order to perform psychometric evaluations in young population, 40 Iranian adult Azeri (20 males and 20 female) individuals aged between 18–25 (mean: 21.57) who were native speakers of Azeri participated. All participants had no history of ear disease or surgery, or neurological disorder. The outer ear of the participants was carefully examined with a Heine mini 300 LED otoscope (Germany) to ensure the ear canal was free of cerumen and the tympanic membrane was normal with all landmarks visible. Tympanometry was obtained using an Inventis Clarinet Tympanometer (Italy), in which each participant had static acoustic admittance between 0.3 and 1.4 mmhos, with peak pressure between –100 and +50 dapa [13, 14]. Moreover, pure-tone audiometry was conducted to confirm normal air conduction thresholds for each participant, using an Inventis Piano clinical audiometer (Italy), in which each participant had pure-tone air conduction thresholds less than 15 dB HL in all octave frequencies ranging from 250 to 8000 Hz. The mean Pure-Tone Average (PTA) for all participants was 5.53 dB HL.

Developing lists

The present study involved the creation of five distinct lists, each comprising ten words, through the systematic observation of phonetic balance and the avoidance of phonetic similarity between words. To minimize variability in list presentation, a male Azeri speaker was recruited to record the lists in a controlled acoustic environment. The recorded lists were subsequently edited using Adobe Audition (v:3) software, and the intensity of the words was standardized to match the calibration tone of 1000 Hz [15]. This rigorous approach to list creation and standardization is critical to ensure reliable and valid results in subsequent experimental tasks that utilize these lists.

Presenting lists

The lists were presented to the participants in 5 intensity levels from –5 to 15 dB HL in ascending order. The participants returned after two to four weeks for retesting. The results were recorded.

Reliability

To check the reliability of the test, Cronbach's alpha and Split Half coefficients and test-retest reliability were conducted.

Results

Face and content validity

The number of 220 words had the minimum acceptable amount of CVR according to the number of experts. CVI values for 5 lists were obtained as 0.975, 0.975, 1, 1, and 0.975 respectively. The average face validity percentage calculated for each list using Cooper's equation was 97%, 96%, 97.5%, 98%, and 97% respectively.

Construct validity

In order to measure the sampling adequacy, the Kaiser-Meyer-Olkin (KMO) was employed. The resulting values for each list were found to be greater than 0.6, indicating that factor analysis could be performed without any

necessary corrections to the lists (Table 1). The one-dimensional variance resulting from the factor analysis of the words contained in all five lists was found to be in excess of 50%. Moreover, the addition of a second factor resulted in an increase in variance of less than 10%, indicating the one-dimensional validity of the lists (Figure 1 and Table 1).

Reliability

Internal consistency

The investigations carried out to check the reliability and internal consistency showed that the value of Cronbach's alpha for each complete list and each half of the list is above 0.7, thus the reliability of the test was confirmed (Table 2).

Test-retest reliability

The values of the correlation coefficient for each of the lists in all intensity levels are more than 0.6 and a strong correlation between the first test and its repetition is seen in all intensity levels (Table 3).

Comparison of average pure tone thresholds of people and obtained speech thresholds

Considering that the speech thresholds confirm the obtained pure tone thresholds, the obtained values were compared, and there was a difference of about 5 dB between the average PTA obtained for each list and the pure tone results.

Psychometric functions and homogeneity

Psychometric functions were constructed using the third-degree polynomial method for each of the five lists under examination. The PTAVE of the participants was 5.53. The analysis of the average slope of the psychometric functions reveals a consistent rate of change of about 7% in decibels between the 20% to 80% range and at the point of 50% (threshold) of word reception across all lists. The regression slope and constant value in these five lists were found to be highly similar, indicating a high degree of homogeneity among them in terms of hearing threshold indicators, 50% point-slope, regression coefficient, and 20–80% slope (Table 4). The graphical representations of these results further confirm the homogeneity among these lists. In conclusion, the evidence supports the notion that these five lists are highly similar to each other in terms of the aforementioned indicators, and the homogeneity between them has been confirmed (Figure 2).

Discussion

The primary objective of the present study was to develop a comprehensive set of Azeri disyllabic words that exhibit homogeneity and psychometric consistency and can be utilized for measuring SRT in adult individuals whose native language is Azeri. To achieve this goal, a total of five lists of disyllabic words were compiled by selecting appropriate and familiar words, and each list consisted of ten disyllabic words. The scoring process was conducted by different experts, based on content validity, to ensure the reliability of the results. The current section of the study delves into the discussion of the outcomes obtained through this rigorous process. The present study employed a CVR for all words by eight experts, and a CVI for each list, based on the calculation formula and CVR results. For the selection of the words of each list, attention has been paid to phonetic balance. Then, the psychometric characteristics of the lists have been examined. Furthermore, a face validity check was conducted, which revealed that all words were deemed familiar and possessed the necessary characteristics from the perspective of the sample population.

It is imperative for any assessment tool to accurately measure the intended constructs. In the present study, five sets of ten disyllabic words were devised and administered with the aim of assessing the accuracy of word recognition, while controlling for extraneous variables. The construct validity of the test was scrutinized through factor analysis. The findings revealed that one factor accounted for over 50% of the variance, while the remaining factors had less than 10%, thereby indicating unidimensionality. The similar slopes of the psychometric functions demonstrate the homogeneity and alignment of the lists, which is crucial in minimizing inter-word and inter-listener variables. This study found that the psychometric functions of the lists possess similar characteristics.

In 1977, Beattie et al. [16] reported the average slope of the psychometric function for English words to be 12, in 2003 Harris et al. [17] obtained a slope of 9 for male talker and 8 for female talker in Korean language. Also in another study conducted by Harris et al. [18], in Vietnamese language a slope of 9 for male talker and 8 for female talker was obtained. In current study, the average slope of the lists was 7. The observed variation in the properties of psychometric functions across studies can be attributed to various factors such as speech calibration, levels of presentation, and speaker gender [19].

Data analysis showed that the Azeri version of the SRT test has high validity and reliability. Therefore, this test has good psychometric properties that make it suitable for local use. Also, the results showed that there is no statistically significant difference between the average performance of the sample population in the test based on gender ($p>0.005$). The psychometric results of this study and other studies are shown in Table 5.

Conclusion

This research presents the first speech threshold test conducted in the Azeri language. The study generated five lists of ten disyllabic words, which were assessed based on familiarity, phonetic balance, dissimilarity, and psychometric homogeneity. The test can be used to evaluate speech thresholds among Azeri speakers and as a reference for above-threshold speech tests.

Ethical Considerations

Compliance with ethical guidelines

All participants have signed an informed consent form to participate in this study, which is in accordance with the Declaration of Helsinki Ethics and approved by the Research Ethics Committee of Iran University of Medical Sciences with the ethics code IR.IUMS.REC.1402.228.

Funding

This research did not receive any grant from funding agencies in the public, commercial, or non-profit sectors.

Authors' contributions

AHD: Study design, acquisition of data, and drafting the manuscript; NR: Study design and supervision, interpretation of the results, and critical revision of the manuscript; BMB: Study design and supervision; SJS: Designing a statistical analysis and Interpretation of the results; NM: Assistance in sampling.

Conflict of interest

The authors declare that there is no conflict of interest to be reported.

Acknowledgments

The authors would like to thank all participants for them cooperation.

References

1. Guidelines for determining threshold level for speech. ASHA. 1988;30(3):85-9.
2. Harris RW, Nissen SL, Pola MG, McPherson DL, Tavartkiladze GA, Eggett DL. Psychometrically equivalent Russian speech audiometry materials by male and female talkers. Int J Audiol. 2007;46(1):47-66. [DOI:10.1080/14992020601058117]
3. Gelfand SA, Calandruccio L. Essentials of Audiology. 5th ed. New York: Thieme; 2022.
4. Hudgins CV, Hawkins JE, Kaklin JE, Stevens SS. et al. The development of recorded auditory tests for measuring hearing loss for speech. Laryngoscope. 1947;57(1):57-89. [DOI:10.1288/00005537-194701000-00005]
5. Ramkissoon I. Speech Recognition Thresholds for Multilingual Populations. Commun Disord Q. 2001;22(3):158-62. [DOI:10.1177/152574010102200305]
6. Nissen SL, Harris RW, Jennings LJ, Eggett DL, Buck H. Psychometrically equivalent trisyllabic words for speech reception threshold testing in Mandarin. Int J Audiol. 2005;44(7):391-9. [DOI:10.1080/14992020500147672]
7. Dillon H. The effect of test difficulty on the sensitivity of speech discrimination tests. J Acoust Soc Am. 1983;73(1):336-44. [DOI:10.1121/1.388815]
8. Martin FN, Champlin CA, Perez DD. The question of phonetic balance in word recognition testing. J Am Acad Audiol. 2000;11(9):489-93; quiz 522. [DOI:10.1055/s-0042-1748141]
9. Mahdavi ME, Rabiei A. Psychometric function characteristics of Persian consonantvowel-consonant words. Aud Vestib Res. 2021;30(1):50-5. [DOI:10.18502/avr.v30i1.5311]
10. Erfani P. Azeri Morphosyntax: The Influence of Persian on a Turkic Language. [Thesis]. Vancouver: Simon Fraser University; 2012.
11. Mohsen SM, Jabri F, Al Maidani S, Ammar A, Alkhatib R. Construction and Standardization of the Syrian Version of the Word Rec-ognition Test for Children with Hearing Impairment. Aud Vestib Res. 2022;31(4):264-74. [DOI:10.18502/avr.v31i4.10730]
12. Yusof MSB. ABC of Content Validation and Content Validity Index Calculation. Educ Med J. 2019;11(2):49-54. [DOI:10.21315/eimj2019.11.2.6]
13. Roup CM, Wiley TL, Safady SH, Stoppenbach DT. Tympanometric Screening Norms for Adults. Am J Audiol. 1998;7(2):55-60. [DOI:10.1044/1059-0889(1998/014)]
14. Guidelines for screening for hearing impairment and middle-ear disorders. Working Group on Acoustic Immittance Measurements and the Committee on Audiologic Evaluation. American Speech-Language-Hearing Association. ASHA Suppl. 1990;(2):17-24.
15. Wang S, Mannell R, Newall P, Zhang H, Han D. Development and evaluation of Mandarin disyllabic materials for speech audiometry in China. Int J Audiol. 2007;46(12):719-31. [DOI:10.1080/14992020701558511]
16. Beattie RC, Forrester PW, Ruby BK. Reliability of the Tillman-Olsen procedure for determination of spondee threshold using recorded and live voice presentations. J Am Audiol Soc. 1977;2(4):159-62.
17. Harris RW, Kim E, Eggett DL. Psychometrically equivalent Korean bisyllabic words spoken by male and female talker. Commun Sci Disord. 2003;8(1):244-70.

18. Harris RW, McPherson DL, Hanson CM, Eggett DL. Psychometrically equivalent bisyllabic words for speech recognition threshold testing in Vietnamese. Int J Audiol. 2017;56(8):525-37. [DOI:10.1080/14992027.2017.1303202]
19. Wilson RH, Carter AS. Relation between slopes of word recognition psychometric functions and homogeneity of the stimulus materials. J Am Acad Audiol. 2001;12(1):7-14. [DOI:10.1055/s-0041-1741115]

Table 1: Kaiser-meyer-olkin value and variance of the first factor

List number	Kaiser-mayer-olkin	Variance of the first factor
List 1	0.755	64.67
List 2	0.702	60.00
List 3	0.688	55.60
List 4	0.776	62.47
List 5	0.774	59.61

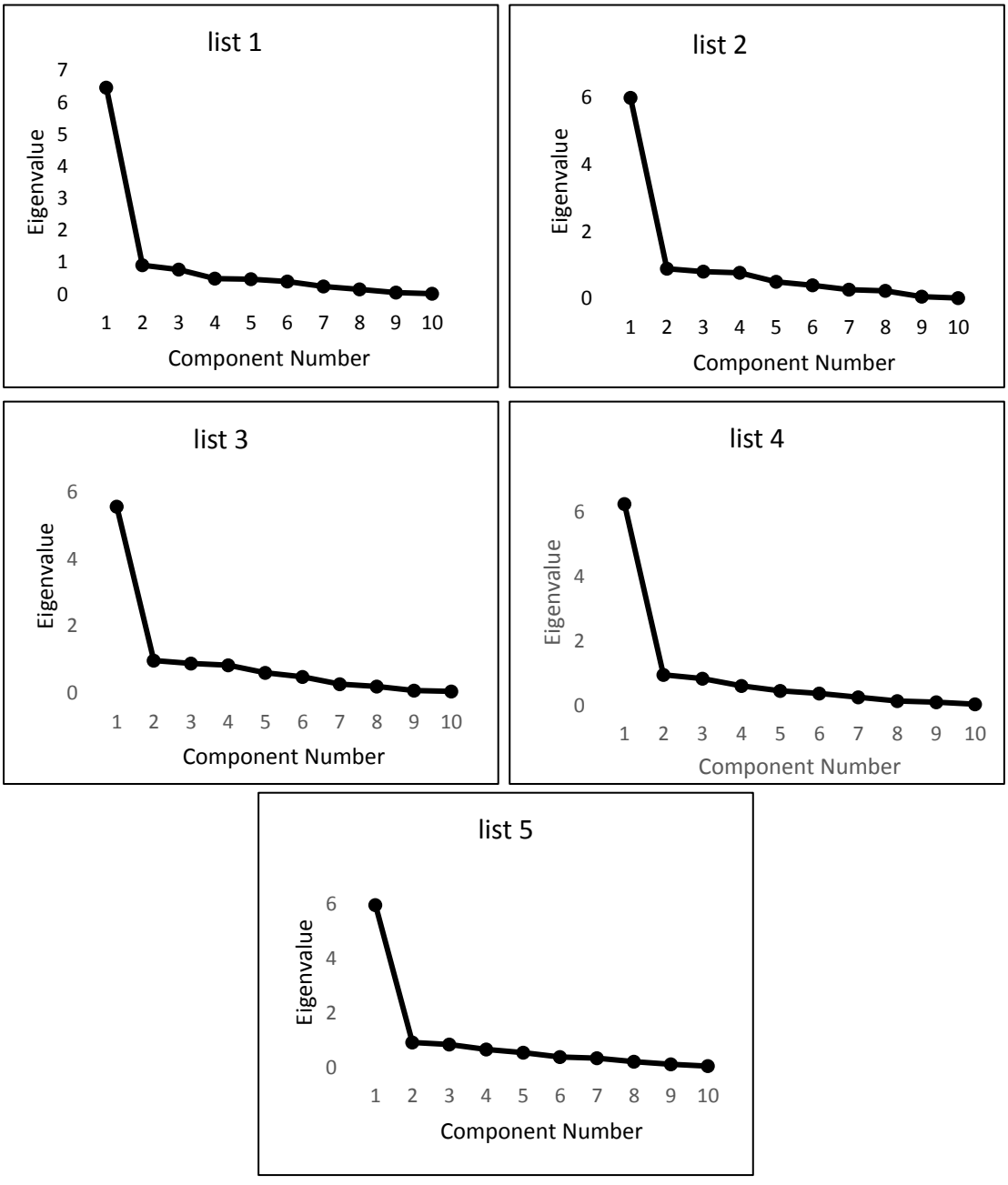

Figure 1. The factorial analysis graph for five lists

Table 2. Reliability using Cronbach's alpha method and split-half method

Number list	Number of words	Cronbach's alpha of list	Cronbach's alpha of the first half	Cronbach's alpha of the second half	Correlation between two halves
First	10	0.933	0.869	0.871	0.890
Second	10	0.921	0.825	0.871	0.880
Third	10	0.908	0.835	0.821	0.845
4th	10	0.928	0.869	0.855	0.883
5th	10	0.919	0.839	0.859	0.852

Table 3. Value of test-retest reliability coefficients

Correlation coefficient	5 dB HL	10 dB HL	15 dB HL
List 1	0.821	0.790	0.607
List 2	0.885	0.834	0.763
List 3	0.819	0.822	0.717
List 4	0.848	0.810	0.833
List 5	0.875	0.940	0.823

Table 4. Psychometric performance characteristics of lists

List number	Constant	SD	Regression coefficient	Thresholds (dB HL)	Slope 50%	Slop 20-80 %
First	–2.2	1.60	1.80	9.60	7.70	7.60
Second	–2.2	1.60	1.80	9.80	7.70	7.40
Third	–2.2	1.70	1.80	9.60	7.80	7.50
4th	–2.3	1.70	1.80	9.60	7.70	7.40
5th	–2.2	1.70	1.80	9.60	7.70	7.40

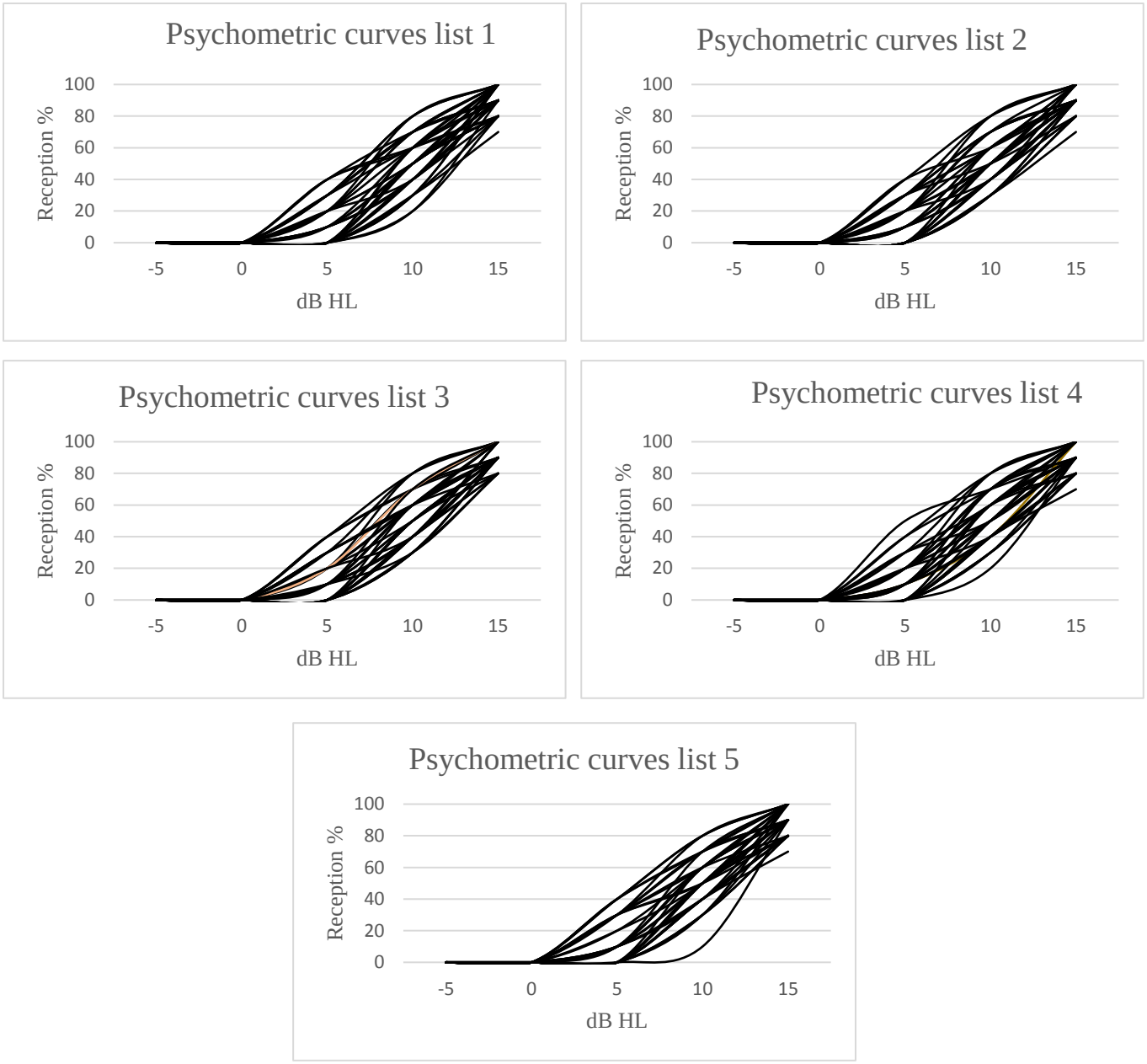

Figure 2. Psychometric curves of the lists

Table 5. Psychometric performance characteristics of bisyllabic words are reported for the present study, and other languages

Study	Year	Language	Slope (50%) %/dB	Slope (20–80%) %/dB
Beattie et al [16]	1977	English	-	12
Harris et al [17]	2003	Korean male	10.5	9.1
		Female talker	9.3	8.0
Harris et al [18]	2017	Vietnamese male	11.0	9.6
		Female talker	9.4	8.2
This study	2024	Azeri male talker		
		List 1	7.72	7.63
		List 2	7.71	7.45
		List 3	7.78	7.51
		List 4	7.67	7.41
		List 5	7.68	7.42