Research Article

Developing and Validating Monosyllabic Speech Materials with Psychometric Homogeneity for Young Adult Turkish-Azeri Speakers in Iran and Compiling the Equivalent Word Lists

Amirreza Karami[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*

Use your device to scan and read the article online

Citation: Karami A, Rahbar N, Mahmoodi-Bakhtiari B, Sameni SJ, Mansouri N. Developing and Validating Monosyllabic Speech Materials with Psychometric Homogeneity for Young Adult Turkish-Azeri Speakers in Iran and Compiling the Equivalent Word Lists. Aud Vestib Res. 2024;33(4):349-60.

https://doi.org/10.18502/avr.v33i4.16653

Highlights

- This study developed word recognition test materials in the Turkish-Azeri language
- Four valid and reliable lists of monosyllabic words were prepared
- The prepared word lists can be useful in evaluating hearing problems of adults

Article info:
Received: 04 Jan 2024
Revised: 03 Mar 2024
Accepted: 11 Mar 2024

ABSTRACT

Background and Aim: For the Word Recognition Score (WRS) test, homogenous lists of words with similar difficulty levels are needed. This study aimed to develop and validate Turkish-Azeri monosyllabic words with phonetic balance and psychometric homogeneity for the WRS test in young adult Turkish-Azeri speakers in Iran and to compile four 25-item word lists.

Methods: In this cross-sectional/comparative study, four lists of 25 monosyllabic words with phonetic balance were created by extracting common words from Turkish-Azeri dialects and assessing them in terms of ease of use, familiarity, and relevance. Then, the lists were tested on 40 young adult Azari speakers aged 18–25 years to determine the validity and reliability.

Results: All four lists showed adequate face and content validity. Cronbach's alpha and split-half values for all four lists were above 0.9, indicating acceptable internal consistency and reliability. Construct validity was confirmed the factor analysis with one-dimensional variance of 77.9%, 80.0%, 79.9%, and 88% for the word lists 1 to 4, respectively, and reported the single-factor solution of the index in all four lists. Also, test-retest reliability with a two-week interval with Pearson correlation coefficients of the lists were 0.94, 0.97, 0.97, and 0.96, respectively.

Conclusion: The four developed Turkish-Azeri word lists have phonetic balance and psychometric homogeneity with a high level of validity and reliability, which makes them suitable for evaluating the recognition of monosyllabic words by young adults in hearing centers of Azeri cities in Iran.

Keywords: Speech audiometry; speech discrimination test; Azeri; monosyllabic; validity and reliability; psychometric function

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

Introduction

An audiological examination typically includes pure-tone and speech audiometries. One of the measures used in speech audiometry is the Word Recognition Score (WRS) [1], which is used for diagnosing peripheral and central hearing disorders, prescribing and adjusting hearing aids, and assessing rehabilitation needs and progress [2]. It is crucial to make sure that the word recognition tests are reliable and have effective diagnostic abilities. Carhart's primary guidelines for constructing word lists indicate the necessity for words to be recognizable to minimize instructional variability. Moreover, they emphasized the phonetic balance of word lists for equivalency in results across different lists [3]. Egan's criteria for word list selection emphasize monosyllabic structure, uniform difficulty between and within lists, and maintaining phonetic balance between the lists and the commonly used words [4], since frequently employed words tend to show enhanced recognition compared to less commonly used words [5].

Homogeneity or equivalence can be related to either individual test items or the entire list of items. To achieve homogeneity, it is necessary to match the difficulty and intelligibility of the items as a function of sound intensity [6]. Homogeneity can be established by plotting the psychometric performance-intensity functions for each word [7]. The psychometric performance has two key aspects: threshold and slope. The 50% threshold refers to the intensity level at which 50% correct recognition achieves, while the slope represents the rate of changes in correct recognition relative to changes in intensity level. The slope of the psychometric performance denotes the consistency of difficulty across speech materials [8].

Iran is a country with cultural and linguistic diversity, having a significant number of native speakers proficient in four predominant languages of Persian, Azeri, Kurdish, and Arabic. Despite the absence of a precise census on the Azeri population in Iran, it has been estimated that Iranian Azeris make up about 24% of the total population of Iran [9]. Although most of them reside in the northwestern region, they can also be found in various provinces of Iran. The Azeri language in Iran is mainly spoken in Tabriz, Urmia, Ardebil, and Zanjan, where the Tabriz dialect is considered as a standard Azeri language in Iran. Turkish-Azeri language has 33 phonemes, including 24 consonants and 9 vowels [9, 10].

The purpose of the present study was to develop Turkish-Azeri speech audiometry materials by producing standardized Turkish-Azeri word lists for measuring the WRS to help audiologists in Iran for testing individuals whose native language is Azeri which can provide accurate information about hearing problems in the Azeri population.

Methods

This is a cross-sectional study for a test development that was approved by the medical ethics committee of Iran University of Medical Sciences.

Test materials

First, common monosyllabic words were extracted using the basic rules for designing speech materials from the books in Azeri language. The syllabic structures of the selected words were Consonant-Vowel-Consonant (CVC) and Consonant-Vowel-Consonant-Consonant (CVCC). The name of digits, conjunctions and prepositions, having an unusual and non-cultural meaning and improper name were excluded. Finally, 1152 most frequent monosyllabic words remained. The words were given to four experts from four Azeri dialects to choose the words that are more common and familiar in that dialect, leading to the selection of common words from four Azeri dialects. To ensure the validity of the test, face validity, content validity, and construct validity were determined. For determining content validity, 523 words from the word bank were assessed by eight experts including audiologists and speech-language therapists. The experts were informed about the purpose of the test and were asked to comment whether the selected words could fulfill the purpose. They analyzed the words in terms of familiarity, clarity, and appropriateness using a Likert scale. Finally, based on the Content Validity Ratio (CVR) and Content Valid Index (CVI) scores, 289 words remained.

We also measured the prevalence of sounds in daily Azeri speech. For this purpose, sound samples from various Azeri TV and radio programs, Azeri movies, and Azeri contents found on social media were randomly selected and recorded. The percentage of the occurrence of each sound was calculated according to the samples collected from the daily speech in the Azeri language (Table 1). Then, for 289 words, 4 lists were developed based on their phonetic balance.

Table 1. Frequency of phonemic occurrence in four lists and common speech

Number	Speech sound	First list	Second list	Third list	Fourth list	Azeri speech
1	a	11.59	11.76	10.95	10.95	12.42
2	Æ	5.79	5.88	5.47	5.47	7.78
3	I	4.34	4.41	4.10	4.10	5.84
4	D	7.24	5.88	5.47	5.47	5.65
5	ɾ	4.34	4.41	5.47	5.47	5.04
6	ɒ	4.34	2.94	4.10	4.10	4.74
7	J	5.79	5.88	4.10	4.10	4.14
8	B	4.34	4.41	5.47	5.47	4.03
9	S	2.89	4.41	2.73	2.73	3.46
10	U	2.89	2.94	2.73	2.73	3.38
11	Z	4.34	4.41	5.47	5.47	3.29
12	T	4.34	4.41	4.10	4.10	3.10
13	M	1.44	2.94	2.73	2.73	2.94
14	G	2.89	2.94	4.10	4.10	2.90
15	O	2.89	2.94	2.73	2.73	2.88
16	E	2.89	2.94	2.73	2.73	2.80
17	ʃ	2.89	2.94	4.10	4.10	2.70
18	L	4.34	2.94	4.10	4.10	2.69
19	Ts	2.89	1.47	1.36	1.36	2.05
20	ɣ	1.44	2.94	2.73	2.73	2.13
21	Y	2.89	2.94	2.73	2.73	2.09
22	ɟ	1.44	1.47	1.36	1.36	1.93
23	K	1.44	2.94	2.73	2.73	1.82
24	ɯ	1.44	1.47	1.36	1.36	1.68
25	P	1.44	1.47	1.36	1.36	1.54
26	Ø	1.44	1.47	1.36	1.36	1.30
27	X	1.44	1.47	0.00	0.00	1.24
28	Dz	1.44	0.00	1.36	1.36	1.12
29	H	1.44	0.00	0.00	0.00	0.91
30	V	0.00	0.00	1.36	1.36	0.89
31	F	1.44	1.47	0.00	0.00	0.80
32	ʒ	0.00	.	0.00	0.00	0.28
33	ʔ	0.00	0.00	0.00	0.00	0.09
Total	-	100.00	100.00	100.00	100.00	100.00

Participants

Forty normal-hearing healthy young adults (mean age: 21.62±1.35 years, ranged 18–25) including 20 males and 20 females participated in this study. All participants were native speakers of Azeri language and had no history of ear disease or surgery. They had pure-tone audiometry thresholds of less than 20 dB HL at all octave frequencies of 250–8000 Hz, static acoustic admittance of 0.3–1.4 mmhos, and peak pressure between –100 and +50 daPa [11].

Presentation of word lists

To prevent changes in speech materials during the presentation, they were recorded in a double-walled, soundproof acoustic chamber at the Faculty of Rehabilitation, Tabriz University of Medical Sciences, using a cardioid condenser microphone (Audio-Technica AT2020, Audio-Technica Co., UK) having windscreen and a Steinberg UR12 power supply, Adobe Audition 2022 software, and a sampling rate of 44100 with 16-bit render. The speaker was at a distance of 20 cm from the microphone and 10 cm from the windscreen at 0° azimuth. The lists were recorded by one male and one female native Azeri speaker. During the recording, the speaker was asked to pronounce each monosyllabic word four times. Two audiologists judged the recorded lists. The recording of the word list involved selecting the speaker with the highest score based on the voice quality, standard accent, and pronunciation. After a rigorous selection process, one of the recorded versions of the male speaker was recognized as being suitable and better for the test. This version was designated as the final sound file. All words in the recorded file were digitally edited using Adobe Audition 2022 software and a 1000-Hz calibration tone, to ensure that they all have an average intensity same to the calibration tone [12].

During the evaluations, the word lists were presented at different intensity levels from 0 to 40 dB HL in 10 dB steps. After a two-week interval, each participant's responses were recorded for each intensity level, and psychometric functions were obtained for each person and word using the third-degree polynomial method. The parameters of the psychometric function including the threshold 50%, the slope at the threshold level or slope 50%, and slope 20–80% for each word and participant were measured using logistic regression analysis to determine psychometric homogeneity [13].

Validity and reliability of word lists

To ensure the accuracy and consistency of prepared speech recognition word lists, we assessed face validity, content validity, and construct validity (using factor analysis [14]). To assess reliability and internal consistency, we employed split-half method, Cronbach's alpha [15], difficulty coefficient [16], and test-retest method [17]. All statistical analyses were performed in SPSS v.16 software, considering $p<0.05$ as statistically significant.

Results

In each word list, there were 23 words with CVC structure and two words with CVCC structure. The number of each vowel in each list was equal. The lists were tested on 40 individuals (20 females). The results of the non-parametric Mann-Whitney U test showed no significant difference in the scores between females and males. Therefore, the gender did not affect the recognition rate of the lists.

Validity assessment of word lists

Content validity

After evaluation based on the opinions of eight experts in terms of familiarity, ease of use, and relevance, of 289 words, 216 had a CVR score of 1 and 73 had a score of 0.75 [18]. By applying the mean CVR formula for the lists 1 to 4, the CVI scores were obtained 0.96, 0.99, 0.98, and 1, respectively.

Face validity

To determine the appropriateness of each item for participants, the agreement between participants were calculated using Cooper's equation, expressed as $Pa=[Ag/(Ag+Dg)]$ [14], where Pa represents the percentage of agreement, Ag is the number of individuals who agreed, and Dg represents the number of individuals who disagreed. Based on the sample size of 40, the mean face validity for the lists 1 to 4 was estimated at 95%, 100%, 92.5%, and 95%, respectively.

Construct validity

Based on the results obtained from factor analysis, it can be said that the four lists demonstrated acceptable levels of construct validity. the variance determined by a single factor in lists was 77.9%, 80.0%, 79.9%, 88.0% respectively; This is evident from the fact that the words used in each list (n=25) explained more than 50% of the one-dimensional variance. After introducing the second factor to this list, the overall increase in variance was less than 10%, which further confirms the validity of the words used in this list (Figure 1).

Item difficulty

The difficulty coefficient of an item refers to the percentage of test-takers who answer an item correctly. Higher difficulty coefficient indicates that the item was more easy. In the present study, all word difficulty coefficients were in the accepted range (moderate difficulty). Therefore, the selected words were considered suitable for further analysis of their reliability (Table 2).

Internal Consistency

Cronbach's alpha

The Cronbach's alpha coefficient, which is a measure of internal consistency, was calculated for the four lists. The findings indicated a high degree of reliability, with coefficients of 0.97, 0.97, 0.97, and 0.98 for the lists 1 to 4, respectively (Table 3).

Split-half reliability

The results indicated that the correlation coefficient between the two halves of the test for each list was very high, ranging from 0.92 to 0.98. These findings suggested that the word lists had high split-half reliability (Table 3).

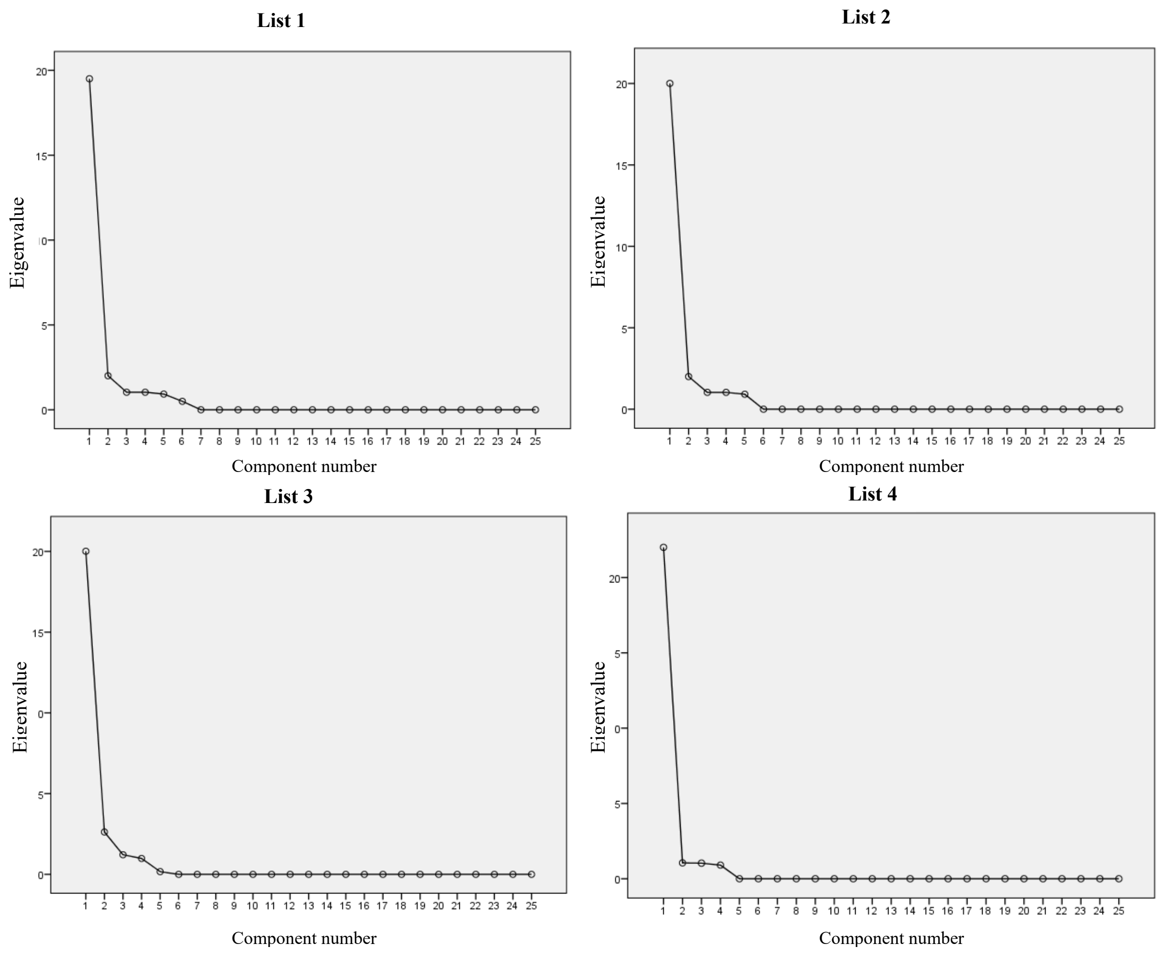

Figure 1. The factorial analysis graph for each list

Table 2. Difficulty index and p-value for each word in the lists

Item	First list		Second list		Third list		Fourth list	
	DI	P	DI	p	DI	p	DI	p
1	64	0.64	62.5	0.62	60.5	0.60	64.5	0.64
2	60.5	0.60	62	0.62	57	0.57	58	0.58
3	62	0.62	64.5	0.64	59.5	0.59	61	0.61
4	62	0.62	57	0.57	63.5	0.63	65.5	0.65
5	59	0.59	62.5	0.62	65	0.65	53	0.53
6	60.5	0.60	52	0.52	62.5	0.62	56.5	0.56
7	60	0.60	56.5	0.56	64	0.64	59	0.59
8	67	0.67	62	0.62	58.5	0.58	56.5	0.56
9	60	0.60	60	0.60	55	0.55	59.5	0.59
10	60.5	0.60	58	0.58	63	0.63	62	0.62
11	58	0.58	66	0.66	59.5	0.59	62	0.62
12	61.5	0.61	52.5	0.52	58.5	0.58	63	0.63
13	56.5	0.56	55.5	0.55	57.5	0.57	62	0.62
14	63	0.63	60.5	0.60	60	0.60	56.5	0.56
15	58.5	0.58	58	0.58	52.5	0.52	55	0.55
16	60	0.60	63	0.63	58	0.58	56.5	0.56
17	55	0.55	68	0.68	55.5	0.55	62.5	0.62
18	58	0.58	60.5	0.60	64	0.64	64.5	0.64
19	59.5	0.59	66	0.66	65.5	0.65	67	0.67
20	60	0.60	58.5	0.58	57	0.57	64	0.64
21	54.5	0.54	59.5	0.59	58.5	0.58	58	0.58
22	60.5	0.60	60	0.60	57.5	0.57	59.5	0.59
23	56.5	0.56	59	0.59	60.5	0.60	54	0.54
24	55.5	0.55	57	0.57	60	0.60	55.5	0.55
25	56.5	0.56	55.5	0.55	54	0.54	59	0.59

DI; Difficulty Index

Test-retest reliability

There was strong correlation between the percentage of word recognition at baseline and two weeks after (when the test was repeated). The observed correlation value was statistically significant ($p<0.05$). The Pearson correlation coefficients of the lists were 0.94, 0.97, 0.97, and 0.96, respectively.

Psychometric homogeneity

The findings from the parameters of the psychometric functions for words and participants (Figure 2) indicated that the hearing thresholds across all four lists were comparable, with 50% correct word recognition occurring at an intensity of about 15 dB. Moreover, the results demonstrate that the rate of change in the slope at 20–80% correct word recognition was 4%/dB for all four lists, and there was no considerable difference between these four lists in this slope. Furthermore, the regression slope and constant value in these four lists were highly similar. Therefore, it can be concluded that the prepared four lists had close similarities in terms of hearing threshold, slope 50%, regression coefficient, and

Table 3. Reliability using Cronbach's alpha and split-half methods to check internal consistency

List	Number of words	Cronbach's alpha of whole words	Gottman coefficient by halving method (12 and 13)	Cronbach's alpha of the first half	Cronbach's alpha of the second half	Correlation between two halves
First	25	0.974	0.980	0.959	0.931	0.979
Second	25	0.976	0.972	0.967	0.935	0.961
Third	25	0.972	0.931	0.979	0.894	0.925
Fourth	25	0.986	0.983	0.986	0.951	0.982

Table 4. Hearing threshold values, and point and interval slopes for each list based on participants

	List 1	List 2	List 3	List 4
Threshold 50% (dB HL)	15.26	15.04	15.26	15.16
Slop 20–80% (%/dB)	4.05	4.09	4.13	4.12
Slop 50% (%/dB)	4.25	4.35	4.35	4.25

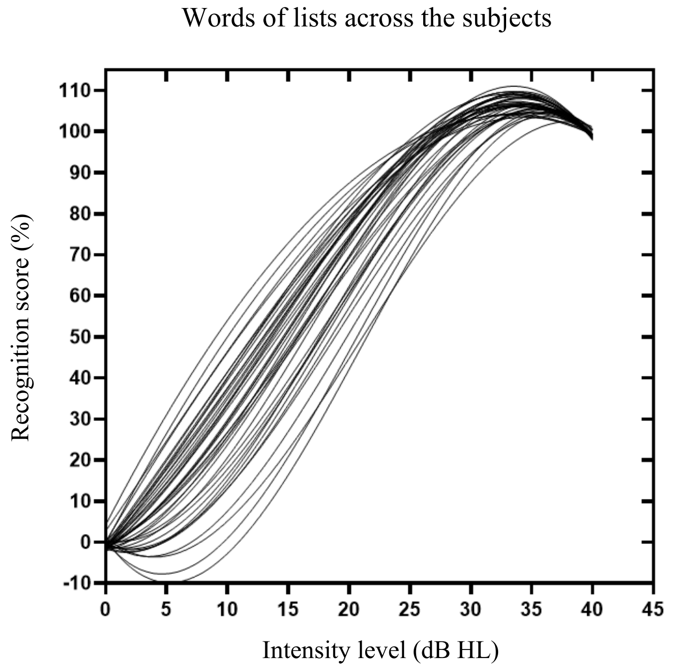

Figure 2. Psychometric functions of word recognition scores of participants

slope 20–80%. The homogeneity within and between the lists was confirmed based on the parameters of the psychometric function (Table 4). The graphical representations of the data further support these findings (Figure 3).

Discussion

The main purpose of this study was to develop a set of Turkish-Azeri monosyllabic words with phonetic balance and homogeneity to be utilized for measuring the WRS in people in Iran who speak the Azeri language, regardless of their regional dialect. To achieve this, four phonetically balanced and equivalent lists of 25 monosyllabic words were developed. Data analysis indicated that the Azeri version of the word recognition test had high validity and reliability. Hence, the Azeri version of the WRS test has acceptable psychometric

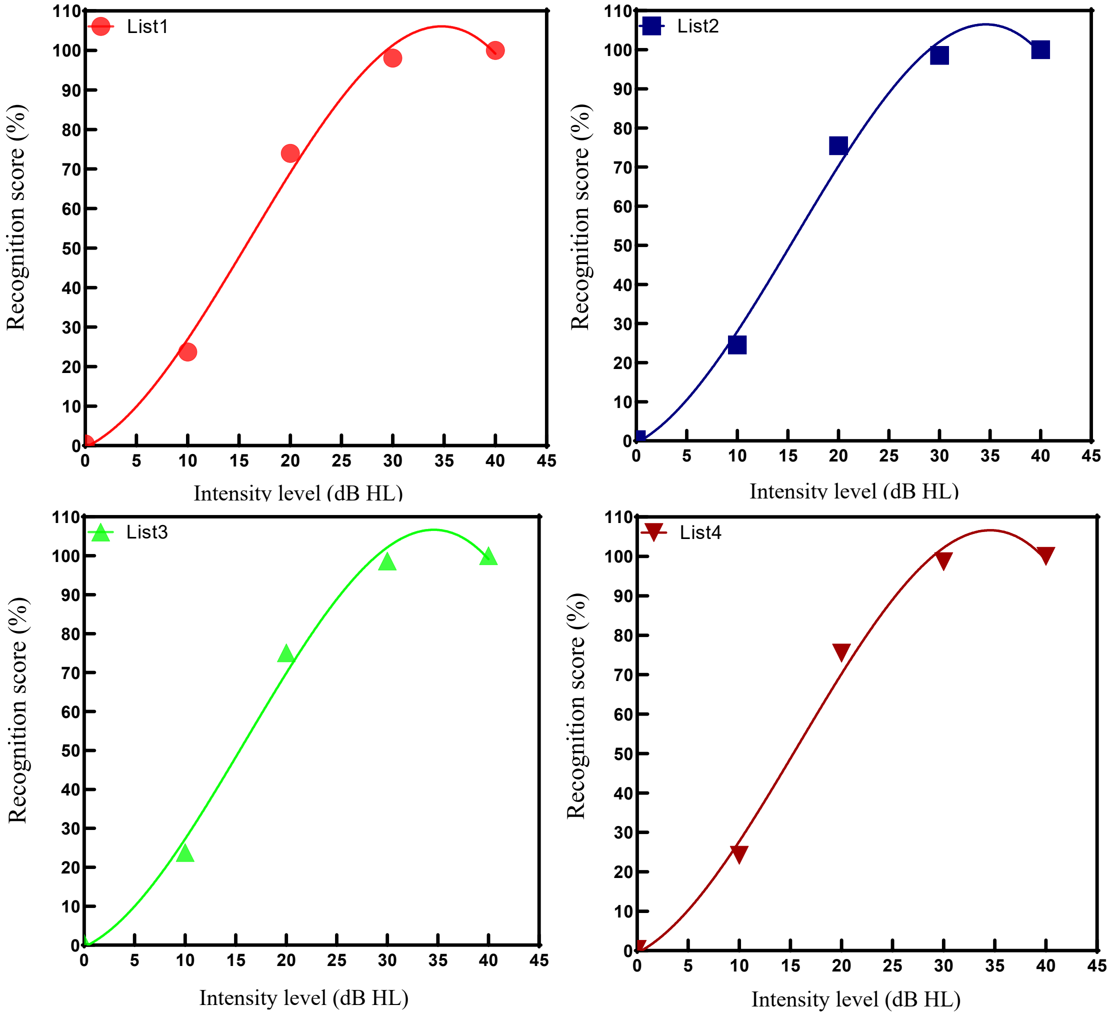

Figure 3. Psychometric functions of the lists based on recognition scores of participants

properties, making it suitable for local use in Iran. There was no statistically significant difference in the WRS between participants based on gender.

Achieving homogeneity for test materials in phonetic balance has been a topic of interest among researchers. However, there is a lack of consensus on the effectiveness of this approach [2, 19-21]. In this study, we adopted a phonetic balance approach with a particular focus on high-frequency sounds such as vowels that are commonly used in everyday speech. In contrast, Mohsen et al. [14] categorized sounds into three frequency-based groups (phonetic rate percentage) as very common (>5.1%), common (2.1–5%), and rare (0–2%), and created phonetic balance within the list based on these categories. While most studies on phonetic/phonemic balance are based on individual phonemes/phonetics, this phonetic/phonemic balance approach attempts to match the frequency of different phonemes/phonetics in each word list to the speech of the language and other word lists [22-24]. Further study is needed to fully understand the implications of phonetic balance on the homogeneity of test materials.

In the field of homogeneity, there is ongoing debate about the use of different monosyllabic structures. Some studies only used the most frequent structure, i.e. CVC [1, 8, 22, 25], while others used a more balanced approach that reflects the frequency of structures in everyday speech [2, 19]. In the present study, the frequency of occurrence was obtained. As a result, we opted to use 23 CVC and two CVCC structures in each word list. This approach was used to provide a more accurate depiction of the frequency and distribution of monosyllabic structures in spoken language.

Table 5. Psychometric performance characteristics of monosyllabic words for the present study and various languages/ dialects

Study	Year	Language/dialect	Threshold 50% (dB HL)	Slope 50% (%/dB)	Slope 20–80% (%/dB)
Harris et al [29]	2003	Korean (male voice)	11.4	5.0	4.4
		Korean (female voice)	10.7	5.1	4.4
Harris et al. [22]	2007	Russian (male voice)	11.5	5.8	5.0
		Russian (female voice)	11.6	5.6	4.9
Tsai et al. [25]	2009	Mandarin	11.1	4.5	4.1
Durankaya et al. [23]	2014	Turkish	13.5	6.2	5.4
Mahdavi and Rabiei [8]	2020	Persian	8.7	7.6	6.6
Garadat et al. [24]	2021	Arabic list 1	13.8	6.0	5.7
		List 2	13.3	6.1	5.6
		List 3	13.6	5.8	5.3
		List 4	14.1	5.8	5.3
Current study	2024	Azeri list 1	15.42	4.1	4.09
		List 2	15.14	4.31	4.1
		List 3	15.23	4.34	4.13
		List 4	15.20	4.33	4.1

In accordance with Mohsen et al.'s methodology [14], we determined face validity, content validity, and construct validity of the lists to confirm their sufficient validity. Additionally, the reliability and internal consistency of the lists were assessed using Cronbach's alpha, split-half and test-retest methods using Pearson's correlation test, indicating that the lists had stable psychometric results over time and did not undergo significant changes. To assess the construct validity of these lists, we used the factor analysis method. The findings revealed that single- factor solution accounted for a significant amount of the variance, while the two-factor solution contributed to the variance less than 10%, indicating that the lists were one-dimensional and successfully measured word recognition ability. According to the factor analysis, a factor is considered valid if it can explain more than 40% of the variance, where 80% or higher variance explained is highly preferable to establish a general factor. The results of the analysis could be affected by the small sample size (n=40), which can be a study limitation; at least 100 participants are required for factor analysis [13, 14]. In addition, the factor analysis results can be based on both statistical and theoretical principles. We did not anticipate any other factors that can affect the accuracy of the analysis, and the first factor could exceed the minimum acceptable level of variance. Therefore, it can be said that this factor represents the objective of the analysis, i.e., speech discrimination [14].

The results of psychometric functions showed that the Azeri word lists were homogeneous and equivalent, which is important for reducing inter-word and inter-listener variables. Interestingly, word recognition was not affected by intensity, indicating cross-list equivalency for these words. Furthermore, there were no changes in recognition of words within each list as a function of intensity, indicating that the materials also had inter-list equivalency. These findings have crucial implications for the development and testing of speech recognition models. The threshold 50% and slope values of psychometric functions for monosyllabic words in all four lists ranged 15.42–15.14 dB HL and 4.34–4.09%/dB, respectively. These values are comparable with those reported in other languages. Interestingly, the slopes of the psychometric functions for the constructed word lists can vary from even within the same language. Previous studies using Northwestern University Auditory Test

No.6 (NU-6) and W-22 monosyllabic words reported a slope range of 3.6-5.6%/Db [25, 26]. Wilson and Oyler [26] observed that the psychometric functions of W-22 and NU-6 words recorded by the same carrier and speakers varied, with the 50% threshold being 15.6 dB HL for W-22 and 13.4 dB HL for NU-6 word lists. Moreover, the list of NU-6 words had a slightly steeper slope than the W-22 words. Similarly, Heckendorf et al. [27] reported a slope of 4.1%/dB for W-22 and 1.9%/dB for psychoacoustic laboratory-phonetic balance 50 words (PAL-PB50).

The results of our study regarding the characteristics of psychometric functions for the Azeri word lists indicated that the threshold, the slope 50%, and the slope 20–80% were comparable across all four lists, indicating their equivalency. Therefore, the four word lists can be used interchangeably. The comparison of the results of psychometric functions with those of recent studies is shown in Table 5. The results indicate that psychometric functions have different characteristics in different languages. This claim is supported by several studies, including those conducted by Harris et al. in Korean and Russian languages [21, 28]. For the Mandarin language, Tsai et al. [24] reported a 50% threshold of 11.1 dB, a 50% slope of 4.5%/dB, and a 20–80% slope of 4.1%/dB. For Turkish language, Durankaya et al. [22] found that the 50% threshold was 13.5 dB, the 50% slope was 6.2%/dB, and the 20–80% slope was 5.4%/dB. Mahdavi and Rabiei [8] used the words in Persian language and reported a 50% threshold of 8.7 dB, a 50% slope of 7.6%/dB, and a 20–80% slope of 6.6%/dB. Garadat et al. [23] created four monosyllabic lists in Arabic and found that the characteristics of their psychometric functions were similar to those reported by Durankaya et al., [28], with a 50% threshold of around 13.5 dB, a 50% slope of 6.0%/dB, and a 20–80% slope of 5.4%/dB. The discrepancy in the characteristics of psychometric functions can be attributed to a multitude factors such as the gender of the speaker, the intensity levels of the presented words, and the calibration of the spoken content [29]. Additionally, the employed statistical models as well as the syllable (CVC vs. CVCC) of monosyllabic words and the average pure tone (500–2000 Hz) of participants can contribute to the observed discrepancies. It is important to note that differences may exist between tests, even within the same language, due to differences in participant characteristics and test protocols. However, the lack of change in the slope between words and lists in our study is evidence of homogeneity within and between lists.

This is a novel study on the development of word recognition test materials for Turkish-Azeri speakers in Iran. However, there is a need for further research in this area and in the field of speech audiometry to increase the knowledge in this field. To ensure the accuracy and consistency of the developed word lists in this study for individuals with hearing impairment, it is recommended to conduct further research with a larger sample size to examine their validity, reliability, homogeneity, equivalency, and phonetic balance.

Conclusion

The four developed Turkish-Azeri word lists with psychometric homogeneity and phonetic balance can be useful for evaluating the hearing problems and central auditory processing of young adult Azeri speakers in Iran. They can also be used as suitable and practical resources for designing auditory processing tests with monosyllabic word materials in Turkish-Azeri language.

Ethical Considerations

Compliance with ethical guidelines

All parents 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.1402.149.

Funding

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

Authors' contributions

AK: 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.

Acknowledgments

The authors would like to thank the audiologist Asma Hamami and the staff and students of the Faculty of Rehabilitation of Tabriz and Iran University of Medical Sciences for their high cooperation in the hearing evaluations of this project.

References

1. Tillman TW, Carhart R. An expanded test for speech discrimination utilizing CNC monosyllabic words. Northwestern University Auditory Test No. 6. SAM-TR-66-55. Tech Rep SAM-TR. 1966:1-12. [DOI:10.21236/ad0639638]

2. Jarolahi F, Delphi M, Tahaie SA, Modarresi Y, Kamali M, Jafari M. [Selection of preeminent list in word recognition score test for adult with normal hearing]. J Res Rehabil Sci. 2012;8(2):212-8. Persian. [DOI:10.22122/JRRS.V8I2.343]

3. Carhart R. Basic principles of speech audiometry. Acta Otolaryngol. 1951;40(1-2):62-71. [DOI:10.3109/00016485109138908]

4. Egan JP. Articulation testing methods. Laryngoscope. 1948;58(9):955-91. [DOI:10.1288/00005537-194809000-00002]

5. Han D, Wang S, Zhang H, Chen J, Jiang W, Mannell R, et al. Development of Mandarin monosyllabic speech test materials in China. Int J Audiol. 2009;48(5):300-11. [DOI:10.1080/14992020802607456]

6. 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]

7. Young LL Jr, Dudley B, Gunter MB. Thresholds and psychometric functions of the individual spondaic words. J Speech Hear Res. 1982;25(4):586-93. [DOI:10.1044/jshr.2504.586]

8. Mahdavi ME, Rabiei A. Psychometric function characteristics of Persian consonant-vowel-consonant words. Aud Vestib Res. 2021;30(1):50-5. [DOI:10.18502/avr.v30i1.5311]

9. Rezaei S, Latifi A, Nematzadeh A. Attitude towards Azeri language in Iran: a large-scale survey research. J Multiling Multicult Dev. 2017;38(10):931-41. [DOI:10.1080/01434632.2017.1342652]

10. Ghaffarvand Mokari P, Werner S. Azerbaijani. J Int Phon Assoc. 2017;47(2):207-12. [DOI:10.1017/S0025100317000184]

11. 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.

12. Wilson RH, Strouse A. Psychometrically equivalent spondaic words spoken by a female speaker. J Speech Lang Hear Res. 1999;42(6):1336-46. [DOI:10.1044/jslhr.4206.1336]

13. MacPherson A, Akeroyd MA. Variations in the slope of the psychometric functions for speech intelligibility: a systematic survey. Trends Hear. 2014;18:2331216514537722. [DOI:10.1177/2331216514537722]

14. Mohsen SM, Jabri F, Al Maidani S, Ammar A, Alkhatib R. Construction and Standardization of the Syrian Version of the Word Recognition Test for Children with Hearing Impairment. Aud Vestib Res. 2022;31(4):264-74. [DOI:10.18502/avr.v31i4.10730]

15. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951;16(3):297-334. [DOI:10.1007/BF02310555]

16. Green SB, Yang Y, Alt M, Brinkley S, Gray S, Hogan T, et al. Use of internal consistency coefficients for estimating reliability of experimental task scores. Psychon Bull Rev. 2016;23(3):750-63. [DOI:10.3758/s13423-015-0968-3]

17. Crocker LM, Algina J. Introduction to classical and modern test theory. 1st ed. New York, NY: Holt, Rinehart and Winston; 1986.

18. Lawshe CH. A quantitative approach to content validity. Pers Psychol. 1975;28(4):563-75. [DOI:10.1111/j.1744-6570.1975.tb01393.x]

19. Egan JJ. Basic aspects of speech audiometry. Ear Nose Throat J. 1979;58(5):190-3.

20. Martin FN, Champlin CA, Perez DD. The question of phonetic balance in word recognition testing. J Am Acad Audiol. 2000 Oct;11(9):489-93; quiz 522. [DOI:10.1055/s-0042-1748141]

21. 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]

22. Durankaya SM, Şerbetçioğlu B, Dalkılıç G, Gürkan S, Kırkım G. Development of a Turkish Monosyllabic Word Recognition Test for Adults. Int Adv Otol. 2014;10(2):172-80. [DOI:10.5152/iao.2014.118]

23. Garadat SN, Alkharabsheh A, Almasri NA, Hagr A. Pediatric Arabic Closed-Set Word-Recognition Test: Development and Evaluation of Psychometric Characteristics. J Am Acad Audiol. 2021;32(8):547-54. [DOI:10.1055/s-0041-1732440]

24. Tsai KS, Tseng LH, Wu CJ, Young ST. Development of a mandarin monosyllable recognition test. Ear Hear. 2009;30(1):90-9. [DOI:10.1097/AUD.0b013e31818f28a6]

25. Beattie RC, Edgerton BJ, Svihovec DV. A comparison of the Auditec of St. Louis cassette recordings of NU-6 and CID W-22 on a normal-hearing population. J Speech Hear Disord. 1977;42(1):60-4. [DOI:10.1044/jshd.4201.60]

26. Wilson RH, Oyler AL. Psychometric functions for the CID W-22 and NU Auditory Test No. 6. Materials spoken by the same speaker. Ear Hear. 1997;18(5):430-3. [DOI:10.1097/00003446-199710000-00008]

27. Heckendorf AL, Wiley TL, Wilson RH. Performance norms for the VA compact disc versions of CID W-22 (Hirsh) and PB-50 (Rush Hughes) word lists. J Am Acad Audiol. 1997;8(3):163-72.

28. Harris RW, Kim E, Eggett DL. Psychometrically Equivalent Korean Monosyllabic Speech Discrimination Materials Spoken by Male and Female Talkers. Commun Sci Disord. 2003;8(1): 217-43.

29. 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]