DEVELOPMENT AND STANDARDIZATION OF PHONEMICALLY BALANCED WORD LISTS IN MALAYALAM LANGUAGE FOR ADULTS

SAFA M

Registration No: P01II24S023026

A Dissertation Submitted in Part of Fulfilment of the Degree of

Master of Science (Audiology)

University of Mysore

ALL INDIA INSTITUTE OF SPEECH AND HEARING

MANSAGANGOTHRI, MYSURU-570006

JUNE, 2026

CERTIFICATE

This is to certify that this dissertation entitled **"Development and Standardization of Phonemically Balanced Word Lists in Malayalam Language for Adults "** is a bonafide work submitted as part of fulfilment of the degree of Master of Science (Audiology) of the student with Registration Number **P01II24S023026**. This has been carried out under the guidance of a faculty of this institute and has not been submitted earlier to any other university for the award of any other Diploma or Degree.

M. Pushpavathi

Mysuru
8 June, 2026

Dr. M. Pushpavathi

Director
All India Institute of Speech and Hearing
Manasagangothri,
Mysuru-570006

CERTIFICATE

This is to certify that this dissertation entitled **"Development and Standardization of Phonemically Balanced Word Lists in Malayalam Language for Adults "**, has been prepared under my supervision and guidance. It is also being certified that this dissertation has not been submitted earlier to any other university for the award of any other Diploma or Degree.

Mysuru
June, 2026

Prajie
1/6/26

Mr. Prajeesh Thomas
Guide
Assistant Professor of Audiology
All India Institute of Speech and Hearing
Manasagangothri,
Mysuru-570 006

06/06/26

Dr. Brajesh Priyadarshi
Co-Guide
Associate Professor of Linguistics
All India Institute of Speech and Hearing
Manasagangothri,
Mysuru-570 006

DECLARATION

This is to certify that this dissertation entitled **"Development and Standardization of Phonemically Balanced Word Lists in Malayalam Language for Adults "** is the result of my own study under the guidance of Mr. Prajeesh Thomas, Assistant Professor of Audiology and co-guidance of Dr. Brajesh Priyadarshi, Associate Professor of Linguistics, All India Institute of Speech and Hearing, Mysuru and has not been submitted earlier to any other university for the award of any other Diploma or Degree.

Mysuru
June, 2026

Registration no. P01II24S023026

Acknowledgement

I express my sincere gratitude to my guide**, Mr. Prajeesh Thomas**, for his guidance, encouragement, and support throughout this dissertation. His mentorship made this research a meaningful and enriching experience.

I would also like to thank my co-guide, **Dr. Brajesh Priyadarshi**, for his valuable inputs and continued support during the course of this study. His expertise contributed significantly to the completion of this work.

I take this opportunity to thank **Dr. M. Pushpavathi**, Director, AIISH, and **Dr. Niraj Kumar Singh**, Head of the Department of Audiology, for providing the facilities and academic environment necessary for carrying out this study.

I am deeply grateful to **Dr. Vasanthalakshmi** and **Ms. Sahana** for assisting with statistics.

My sincere thanks to **Dr. Arivudai Nambi**, **Dr. Srikar Vijayasarathy**, and **Ms. Apoorva** for their valuable assistance and support during the course of my dissertation.

My special thanks to all the participants for taking part in the study.

Thank you, **Vardha** for being a wonderful dissertation partner and making this journey easier and enjoyable.

To my **Adil**, thank you for always being there for me, understanding me and making even the difficult days feel easier.

Dear **umma and uppa,** every step of this journey has been possible because of your love and support. Thank you.

Thank you, **Sahra, Vaishnavi, Karnika**, and my **unnimayam teams** for all the wonderful moments.

A special mention to **VINA FALUDA** for all the laughter and chayakada chats.

Finally, I would like to thank all my teachers, seniors, juniors, friends, and everyone who directly or indirectly contributed to the successful completion of this dissertation.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The present study aimed to develop and standardise multiple phonemically balanced (PB) word lists in Malayalam for adults to support speech identification testing in quiet and noise. The study was carried out due to the limited availability of standardised Malayalam PB word lists and the increasing clinical need for multiple equivalent word lists to obtain speech identification scores (SIS) for repeated assessment, hearing aid evaluation, and research purposes. A total of 1000 bisyllabic Malayalam words were initially collected from various linguistic sources. The collected words were subjected to familiarity ratings and content validation, after which highly familiar and socio-culturally appropriate words were organised into six phonemically balanced word lists, each consisting of 25 words. The developed word lists were recorded in an acoustically treated environment and analysed using spectral centroid estimation. Spectro-temporally matched noise was generated using spectral inversion techniques, after which pilot testing was conducted to determine the global Signal-to-Noise Ratio required to obtain 50% SIS (SNR-50). The finalised word lists were administered to 40 native Malayalam-speaking adults with normal hearing sensitivity in quiet and at different SNRs to assess list equivalency, homogeneity, psychometric characteristics, and Performance Intensity function for PB words (PI-PB) across different sensation levels. SIS were obtained across listening conditions and analysed using appropriate statistical methods. The findings revealed no significant difference in SIS obtained using the developed word lists, indicating good equivalency and homogeneity in listening difficulty. The psychometric functions demonstrated improved SIS with increasing sensation level and reduced performance under adverse noise conditions. Further, the developed word list exhibited good test-retest and inter-judge reliability, suggesting its potential usefulness in routine speech audiometry, hearing aid evaluation, and the evaluation of outcomes of amplification devices such as hearing aids, cochlear implants, and FM systems, as well as in research applications requiring multiple equivalent lists.

Keywords: SIS, word lists, Malayalam, phonemic balancing, Speech Perception, speech audiometry, psychometric

CHAPTER 1

INTRODUCTION

Speech perception plays a crucial role in daily life. For children, the ability to hear and understand speech is essential for language development, whereas for adults, it is important for effective communication in everyday activities. Hearing loss adversely affects speech perception, making it difficult to understand speech. Thus, the primary goal of the management and rehabilitation of an individual with hearing impairment is to improve speech perception. Speech perception testing plays an important role in evaluating an individual's listening (Noor & Arif, 2018).

Speech audiometry assesses sensitivity to speech and clarity when speech stimuli are presented (Katz et al., 2015). It is a fundamental tool in audiologic evaluation because it uses speech stimuli to measure hearing thresholds, assess speech intelligibility, evaluate hearing aid performance, and aid in the diagnosis of peripheral and central auditory disorders (Jerger et al., 1968). It also provides additional information to pure-tone testing by evaluating an individual's ability to recognise and discriminate speech sounds (Sargsyan & Rahne, 2021). The first speech audiometer was the Western Electric 4A, introduced in the 1920s, and was used for hearing screening (Hudgins et al., 1947).

Two basic speech tests include speech detection and speech identification /recognition. Speech detection threshold (SDT) is the lowest level of intensity at which an individual can detect the presence of speech stimuli. The speech recognition threshold (SRT) is the lowest level of intensity at which an individual can understand speech stimuli presented at least 50% of the time. In addition, the speech identification score (SIS) is a speech intelligibility test at the supra-threshold level. It is determined by the percentage of words correctly repeated from a list of phonemically balanced (PB) test words (Gelfand, 2009).

Different types of speech materials are used to assess speech recognition ability, which include sentences, nonsense syllables, and monosyllabic words (Ma et al., 2013). Studies have shown that among these, nonsense syllables are the most difficult for an individual to recognise, while sentences are the easiest (Martin & Clark, 2009). Recognition of monosyllabic words falls between nonsense syllables and sentences on the difficulty scale. Despite this, monosyllabic words remain the most commonly used material in clinical assessment. Due to a lack of contextual cues, nonsense syllables may yield better phonemic error identification, but they have poor content validity (Egan, 1948; Mendel & Danhauer, 1997). Even though sentences that represent everyday conversation have higher content validity, they are highly influenced by contextual cues, resulting in poor test reliability (Egan, 1948). McArdle and Wilson (2008) noted that using isolated words in assessment may yield a balanced representation of everyday communication and will reduce the influence of contextual and syntactic cues.

Speech audiometry uses standardised materials in a person's native language to assess the ability to perceive and understand linguistic stimuli (Vainutienė et al., 2024). Using controlled phonemic combinations, nonsense syllables or nonsense words, materials have been developed and are used in speech perception tests (Cheoy et al., 2021; Trimmis et al., 2012). In addition, monosyllabic words are utilised to evaluate word recognition, and these materials are constructed to be familiar to the listener and phonemically balanced (Han et al., 2009; Nissen et al., 2011). Dissyllabic words are also used in speech audiometry and are selected based on factors such as familiarity, phonemic representation, phonetic dissimilarity, and homogeneity of audibility (Harris et al., 2017; Noor & Arif, 2018; Wang et al., 2007). Sentence recognition is assessed using sentence materials, and the test items consist of groups of sentences with a fixed number of words and are phonetically balanced (Niemeyer, 1965).

Moreover, sentence materials have been developed and tested in various listening conditions, such as quiet and noise (Uslar et al., 2013).

1.1 Need for the Study

India's cultural and linguistic diversity make it impossible to use single assessment material for evaluating speech perception ability across all languages. Alusi et al. (1974) reported that the language used for testing is important when speech is used as a stimulus. Therefore, there is a need to develop material for individual languages.

Several attempts have been made to develop various speech audiometry materials in Indian languages, such as Indian English (Swarnalatha & Rathna, 1972), Hindi (De, 1973), Kannada (Yathiraj & Vijayalakshmi, 2005), Tamil (Dayalan, 1976), Telugu (Rathna Kumar & Mohanty, 2012), and Malayalam (Kapur, 1971). To overcome the drawbacks of earlier developed materials, some researchers have revised or developed new speech audiometry materials across various languages (Manjula et al., 2018; Dias et al., 2015; Sreedhar et al., 2005).

Lehiste and Peterson (1959) noted some of the limitations of already existing word lists, such as containing unfamiliar words, a lack of validity on clinical populations, and the use of test items in noise conditions, a smaller number of test items to assess various conditions, and the effect of dialect variations is neglected.

Malayalam is a Dravidian language, spoken by 38 million people in Kerala and the union territories of Pondicherry and Lakshadweep (Narne et al., 2024). Kacker and Basavaraj (1990) developed speech materials for speech recognition threshold (SRT) and speech identification score (SIS) in Malayalam, comprising two phonemically balanced word lists of 30 bisyllabic words. However, they were restricted to phonemic balancing, and psychometric functions to assess the equivalence of the lists were not developed (Narne et al., 2024). Carhart (1951) reported that to obtain precise test results, speech audiometry materials should be validated in an experimental setting with a test population. The development of the first

Malayalam word list dates back to the late 1970s, as reported by Kapur. This word list comprises a single phonetically balanced list of 25 bisyllabic Malayalam words. Hence, there is a need for developing phonemically balanced (PB) word lists that are linguistically and culturally appropriate for Malayalam speakers.

Monosyllabic words are commonly used to evaluate speech identification scores, but in South Indian languages, familiar monosyllabic words are significantly less common. Therefore, Kumar and Mohanty (2012) used bisyllables in their Telugu PB word lists. Also, when the number of syllables in a word increases, the likelihood of the word being heard correctly increases due to contextual cues in both normal and hearing-impaired individuals (Narne et al., 2024).

The now-available Malayalam word lists were developed by Kacker and Basavaraj (1990). Multiple lists are required in clinical settings and for research purposes. The clinical application of multiple SIS test lists is essential for tasks requiring repeated, reliable assessment without the confounding variable of practice effects. Its requirements in clinical settings include monitoring rehabilitation progress over time, testing different hearing aids during selection and fine-tuning, and evaluating performance-intensity functions for phonetically-balanced (PI-PB) words in cochlear and retrocochlear pathology cases. Similarly, for research purposes, it is important to have multiple equivalent word lists to meet methodological requirements, such as assessing the PI-PB function under various signal-processing conditions, administering multiple SIS assessments to evaluate the efficacy of advanced hearing aid features, and comparing with newly developed tools to assess speech perception.

Tillman and Carhart (1963) suggested that it is crucial that the same list not be used more than once for an individual, as memory and practice effects may interfere with scores. In addition, Dillon and Ching (1995) also suggested multiple lists because, if one uses equivalent lists, any item will be presented only once during the entire evaluation process.

The test's flexibility can be improved by increasing the number of equivalent lists. To overcome the practice effect, a good speech test should have a larger number of phonemically balanced word lists.

Hence, the present study aims to develop multiple word lists with familiar words used in everyday contexts. During speech identification score evaluation, at least two lists are required to assess performance in the right and left ears. Similarly, during hearing aid fitting, for the evaluation of unaided and aided performance, a minimum of two lists is required for each ear. Likewise, in a research setting, at least three lists are needed to assess PI-PB function in each ear. For these reasons, there is a need for multiple word lists. This study intends to develop PB word lists using bi-syllabic words. The developed word lists will be standardised for adults with normal hearing sensitivity.

1.2 Aim of the Study

The present study aims to develop and standardise multiple sets of phonemically balanced bi-syllabic word lists for evaluating speech identification score in adults with normal hearing sensitivity, and to assess the clinical utility of the developed word lists.

1.3 Objectives of the Study

1. To develop multiple Phonemically Balanced word lists in Malayalam for adults.
2. To standardise and assess the consistency of speech identification scores among different word lists in adults with normal hearing sensitivity.
3. To assess the homogeneity of audibility.

CHAPTER 2

REVIEW OF LITERATURE

Phonemically balanced (PB) word lists are an essential component of speech audiometry, which ensures that the distribution of phonemes in the test materials represents their natural occurrence in everyday speech. These lists enable accurate assessment of speech perception through representing the phonological structure of a language (Vainutienė et al., 2024). Careful consideration should be given to factors such as phoneme frequency, word familiarity, and phonemic composition when developing PB word lists. Words are generally taken from commonly used sources and further filtered through familiarity ratings, while equivalence and homogeneity across lists are preserved to allow different lists to be used interchangeably without affecting test results (Manjula et al., 2015; Chinnaraj et al., 2022). Additionally, facilitating homogeneity among test items through psychometric methods enhances the sensitivity, reliability, and clinical applicability of the speech materials (Ji et al., 2011).

Lehiste and Peterson (1959) stated that achieving perfect phonetic balance is not feasible because speech sounds vary in their acoustic characteristics across contexts. Therefore, phonemic balancing is used to represent the natural distribution of speech sounds. Nevertheless, achieving an exact balance is often difficult due to practical limitations, and an approximate approach is commonly employed. Regardless of these constraints, PB word lists provide a reliable measure of speech perception and accurately reflect natural communication (Ji et al., 2011).

Testing an individual in their native language is preferred; otherwise, it may yield an inaccurate result, as unfamiliar or meaningless words may lead to a low score. Hence, if test materials are available in the native language of an individual, speech tests should be administered in that language (Lehiste & Peterson, 1959). Carhart (1951) claimed that it is important to have standardized test material in every language in an experimental setting.

2.1 Various Test Materials for Speech Identification in the English Language

The early development of speech audiometry was largely influenced by studies focused on articulation testing and speech intelligibility measurement. Egan (1948) described articulation testing as a method to evaluate speech communication by measuring the percentage of correctly identified speech units such as syllables, words, or sentences, emphasising that intelligibility is influenced by multiple factors, including noise and transmission conditions. These articulation tests laid the foundation for quantifying speech perception abilities.

Subsequently, Hudgins et al. (1947) proposed important criteria for the development of speech test materials. They emphasised the importance of familiarity, phonetic dissimilarity, normal speech-sound sampling, and homogeneity to ensure accurate speech-perception evaluation. Their work also introduced structured recorded tests for assessing hearing loss using speech stimuli. Based on these, Hirsh et al. (1952) introduced standardised materials, such as spondee words and phonemically balanced (PB) word lists, for evaluating speech intelligibility. These studies contributed to the growth of speech audiometry by providing methods for intelligibility measurement and criteria for test construction, which led to the development of early phonemically balanced word lists used in clinical practice.

Speech audiometry is based on fundamental principles that emphasise the measurement of hearing using meaningful and standardised speech stimuli. Carhart (1951) described speech audiometry as a technique in which standardised samples of language are presented through a calibrated system to assess specific aspects of hearing ability, highlighting that proper standardization and calibration are essential for obtaining reliable clinical results. A key theoretical concept underlying speech audiometry is the articulation function, which describes the relationship between speech intensity and the percentage of correctly identified speech items, typically forming an S-shaped curve that reflects increasing intelligibility with increasing intensity. Furthermore, Carhart emphasised that speech tests must be developed separately for each language due to phonetic, intonational, and linguistic differences, making universal application of a single test inappropriate.

Expanding on the concept of intelligibility, Lehiste and Peterson (1959) defined intelligibility as a property of meaningful speech communication, stating that only meaningful linguistic units, such as words, carry intelligibility, whereas isolated sounds or nonsense syllables lack true communicative value. They also highlighted that intelligibility is influenced by factors such as dialect, listener familiarity, and phoneme distribution, and emphasised the importance of phonemically balanced materials that reflect natural language patterns. Together, these principles form the theoretical basis for speech audiometry and guide the development of reliable, clinically valid speech test materials.

The first monosyllabic word material was developed by Egan in 1948, known as the “PB 50 test”, which contains 25 lists of 50 words. Later, in 1952, Hirsh et al. modified the “PB 50-word list” into the CID W-22-word list, which consisted of four lists of 50 familiar words.

Lehiste and Peterson (1959) gave 10 lists of 50 phonemically balanced words. Later, Tillman and Carhart (1966) developed the Northwestern University Auditory Test No. 6 (NU-6), which controlled for phonemic equivalency and equal word familiarity. There were four equivalent lists, each containing 50 words. In 1929, sentence-level tests were developed by Fletcher and Steinberg at Bell Laboratories. Silverman and Hirsh (1955) developed the CID everyday sentences, consisting of 10 lists of 10 sentences each, with 50 keywords per list.

The speech perception in noise test, which contains eight lists of 50 sentences each, was developed by Kalikow, Stevens, and Elliot (1977). There were also high-quality recordings of SRT and SIS materials in a variety of languages other than English, including Spanish (Christensen, 1995), Italian (Greer, 1997), Japanese (Mangum, 2005), French (Nelson, 2004), and several Indian languages.

2.2 Speech Identification Tests Developed in India

India is a multilingual country, so it is crucial to develop and standardise language-specific material, as the language of the speaker and listener will influence the speech intelligibility. Various speech materials have been developed in Indian languages. Some of them are listed in Table 2.1

Table 2.1

Speech Identification Word Materials Developed in Various Indian Languages

Language	Author	Stimuli used	Number of wordlists
English (for Indians)	Swarnalatha (1972).	Bisyllabic words	2 lists
Hindi	De (1973).	Monosyllabic words	6 lists
Tamil	Samuel (1976).	Monosyllabic words	2 lists
	Menon and Thangaraj (2024)	Bisyllabic words	4 lists
	Chinnaraj et al. (2022)	Bisyllabic words.	25 lists
Manipuri	Devi (1985)	Multisyllabic words	4 lists
Kannada	Yathiraj and Vijayalakshmi (2005)	Bisyllabic words	8 lists
	Manjula et al. (2015)	Bisyllabic words	21 lists
Bengali	Ghosh (1988)	Multisyllabic words	3 lists
Telugu	Sreedhar et al. (2005)	Paired bisyllabic words	1 list
Marathi	Rathna Kumar et al. (2016)	Bisyllabic words	4 lists
Malayalam	Kapur (1971)	Bisyllabic words	1 list

2.3 Variables to be Considered for the Development of Speech Materials

Hudgins et al. (1947) described fundamental criteria for selecting while developing material for speech audiometry. These criteria include:

- Item Familiarity.
- Phonetic dissimilarity.
- Normal sampling of speech sounds.
- Homogeneity with respect to basic audibility.

2.3.1 Item Familiarity

Hudgins et al. (1947) emphasised that the items in the test material should be as simple and familiar as possible because the goal of the test is to measure how well speech can be understood, rather than a person's vocabulary or intelligence, while still meeting the other required criteria. Later studies on speech audiometry also highlighted how important lexical familiarity is. Egan (1948), in developing the PB-50 word lists, ensured that the selected words were commonly used in everyday English to help ensure consistent recognition by listeners.

Words are often selected based on their frequency of occurrence to ensure that they are commonly used and familiar to listeners, thereby improving the reliability of speech recognition assessment (Vainutienė et al., 2024). Owens (1961) reported that more familiar words tend to be chosen among competing responses than less familiar words. Familiar items in a list can be utilised in several ways. The test words may be more familiar to a person with high intelligence and verbal ability, resulting in a higher discrimination score than to a person with a lower intelligence and verbal ability. So, it will be difficult for someone with a lower level of education if the test items are unfamiliar. Several studies have included familiarity rating procedures, where native speakers from distinct educational and socio-economic backgrounds rated the selected words to ensure consistency in perception across listeners (Chinnaraj et al., 2022; Manjula et al., 2015; Durankaya et al., 2019).

Some of the word lists in foreign languages were developed by utilising words from a word corpus, according to their frequency of occurrence in those languages (Fu et al., 2011; Wang et al., 2007). The words can be taken from magazines, daily newspapers, novels, school textbooks, and dictionaries in that particular language. Individuals with different educational backgrounds and social statuses should rate the familiarity of the selected words to ensure consistent word perception across the test items.

2.3.2 Phonetic Dissimilarity

Hudgins et al. (1947) claimed that phonetic dissimilarity is essential because similar or rhyming words make the test harder needlessly. Similar words do not enhance the test's effectiveness, but they require finer discrimination. Using monosyllabic words in word lists can lead to phonetic similarity, which may create confusion for listeners. To overcome this, some researchers have developed bisyllabic word lists that reduce the likelihood of such similarities in test items (Chinnaraj et al., 2022). Hence, phonetic dissimilarity facilitates independent identification of words based on auditory perception, thereby improving the test's effectiveness.

2.3.3 Normal Sampling of Speech Sounds

Hudgins et al. (1947) also emphasised that speech materials should represent a normal distribution of speech sounds in the language, ensuring that they accurately reflect the phonetic structure of the language and typical everyday communication. To accomplish this, studies have employed approaches such as phoneme frequency analysis and phoneme allocation tables to establish the distribution of consonants, vowels, and, in some languages, tonal elements within each list (Ji et al., 2011).

However, achieving a perfectly normal distribution of speech sounds can be challenging because other criteria, such as familiarity and phonetic dissimilarity, restrict the selection of words. Despite these limitations, the authors emphasised that the selected items should provide the most representative possible sampling of speech sounds (Ji et al., 2011). Additionally, extensive analyses of spoken-language corpora have been employed to establish phoneme frequency and guide word selection (Trimmis et al., 2006).

While developing speech audiometry materials, the most frequently encountered issue is ensuring they are balanced phonetically or phonemically. Phonetic or phonemic balancing plays an important role in the speech discrimination score. A test has speech sounds that occur in everyday speech; it is said to be phonetically balanced (Egan, 1948). Boothroyd (1968) claimed that balancing can be achieved by using word lists that have the same proportion of phonemes in each list. Lehiste and Peterson (1959) introduced the concept of phonemic balancing. According to their view, phonemes are groups of speech sounds that native speakers of that language identify as similar.

2.3.4 Homogeneity

The fourth criterion proposed by Hudgins et al. (1947) is homogeneity of the test items. Homogeneity refers to the uniform level of difficulty across all words in the test list. Ideally, when spoken by the same speaker in a normal conversational tone, each word should have approximately the same level of intelligibility. He also claimed that to achieve homogeneity in a word list, two methods can be adopted. Either selecting words that, when spoken at a normal level, are heard at roughly the same threshold by the listener, or adjusting the recording levels such that all words are perceived at the same loudness. These approaches may also be combined to achieve greater consistency. Homogeneity is often evaluated using psychometric methods, for instance, performance–intensity (PI) functions, which assess the relationship between stimulus intensity and recognition performance (Ji et al., 2011).

Similarly, studies have demonstrated that homogeneous word lists establish equivalent performance throughout different intensity levels, reflecting consistent difficulty and reliability (Chinnaraj et al., 2022). Such equivalence across multiple lists allows for their interchangeable use (Hennig et al., 2018).

Hudgins et al. (1947) also reported that homogeneous test materials typically yield a steeper and more consistent articulation function, indicating that the test items are similar in difficulty and that changes in performance are mainly due to audibility rather than to material variability.

2.4 Speech Identification Tests Developed in Malayalam Language

The frequency of occurrence of phonemes has been reported by Sreedevi and Irfana (2013), who highlighted that speech materials should be based on the most frequently occurring phonemes in the language. Research has also shown that vowel combinations and phonemic balance significantly influence word recognition performance, particularly in disyllabic wordlists (Narne et al., 2024).

Various speech materials have been developed in Malayalam to evaluate speech identification score (SIS) and speech recognition score (SRT). These include the Malayalam Phonemically Balanced list (ISHA, 1990), the Malayalam spondee list (ISHA, 1990), which is currently available and the Picture test of speech perception in Malayalam (Mathew & Yathiraj, 1996). Konadath et al. (2013) developed sentence materials in the Malayalam language through collecting meaningful sentences that are commonly used, which have four to five words in each and are taken from textbooks, magazines and daily conversation. Equally difficult lists were formed by selecting sentences based on intelligibility in the presence of speech noise, and it was found to be equally difficult for individuals with normal hearing and those with hearing loss. Materials for speech in noise (SPIN) were developed by selecting a set of words and, after rating their familiarity, syntactically and semantically correct sentences were constructed. However, the material is no longer available in print or as a soft copy. These sentence lists were presented with multitalker babble at various SNR levels to evaluate speech perception in noise and to determine SNR-50 values (Prasad et al., 2017).

Low-frequency word lists in Malayalam were developed by Prabhu et al. (2017) using words with low-frequency energy and filtering them based on familiarity ratings. Thereafter, to obtain homogeneous words in terms of difficulty, psychometric functions were utilised. Nayana (2018) developed the Lexical Neighbourhood Test in Malayalam for children, by selecting words based on frequency of occurrence and lexical density, and then grouping them into easy and hard lexical items. However, existing Malayalam speech materials for speech identification scores are limited to phonemic balancing. The details of psychometric validation and equivalence across lists have not been established.

CHAPTER 3

METHODS

The present study involved the development and standardization of phonemically balanced word lists in Malayalam.

The study was conducted in 3 phases, which include:

1. Development of PB word lists in Malayalam.
2. Assessing the homogeneity in audibility
3. Standardization of the word lists

3.1 Phase 1: Development of Phonemically Balanced Word Lists in Malayalam

This phase was conducted in several steps: collecting words from various sources, rating the familiarity of collected words, conducting content validation, preparing word lists, and recording.

3.1.1 Collection of Words

A pool of 1000 words was collected from various sources, including novels, storybooks, dictionaries, and school textbooks up to the 8th standard. Recent magazines and newspapers published within the last six months were also considered during word collection. Words that were likely to cause ambiguity or perceptual confusion were excluded from the initial pool.

3.1.2 Familiarity Rating

The collected words were rated for familiarity on a five-point scale. Rating was carried out by 15 native Malayalam speakers from different levels of education and socio-economic status. Kuppuswamy's Socioeconomic Status Scale was used to assess socioeconomic status (SES) (Kumar et al., 2012). It includes five socio-economic classes: upper, upper middle, lower middle, upper lower, and lower. The familiarity rating scale in this study was defined as follows:

5- Most familiar (words which are well known and more frequently used in conversation)

4- Familiar (words well known but less in conversation)

3- Familiar but not used every day (known words but not used in conversation)

2-Not familiar (words heard but meaning not known)

1-Unknown (words never heard)

Only the words that received familiarity ratings of 3-5 from the raters were considered for inclusion in the final word lists.

3.1.3 Content Validation

The selected words were sent to 11 experts, including five from audiology and two each from speech-language pathology, linguistics, and psychology, for validation. The experts were asked to check whether the words met the criteria for familiarity, emotional and cultural appropriateness, relevance to speech audiometric evaluation and linguistic relevance. Particularly, the five audiologists and two speech language pathologists rated for familiarity, emotional and cultural appropriateness, relevance to speech audiometric evaluation and linguistic relevance. The two linguists rated same domains except for relevance to speech audiometric evaluation, while the two psychologists only rated familiarity and emotional and cultural appropriateness. Familiarity ratings were collected on a 5-point Likert scale, and the ratings were dichotomised into relevant (ratings of 3-5) and not relevant (ratings of 1 and 2). Cultural and emotional appropriateness and the relevance of speech audiometry were rated using a dichotomous (Yes/No) response format. Content validity was calculated using Lawshe's Content Validity Ratio (CVR) for each word and overall Content Validity Index (CVI) for each domain, following the method described by Gilbert and Prion (2016).

3.1.4 Preparation of Phonemically Balanced Words

Following familiarity rating and content validation, the retained words were organised into six phonemically balanced word lists, each consisting of 25 words. Phonetic balance was ensured by matching the frequency of occurrence of Malayalam phonemes reported by Sreedevi and Irfana (2013). For that, the number of individual phonemes in the selected words in each list was computed and compared with the frequency of occurrence in Malayalam. Adjustments were made to produce final lists with a balanced phoneme distribution. During list formation, the average spectral centroid frequency of the words was also considered to ensure that the lists were acoustically comparable while maintaining phonemic balance (as explained in the following section). Words that could not be appropriately accommodated within the lists were excluded from the final set.

After the preparation of word lists, the International Phonetic Alphabet (IPA) transcription of corresponding words was added with the help of an expert linguist.

3.1.5 Recording

All selected words were recorded in an acoustically treated room (ANSI 3.1, R2013). The speech material was recorded in Adobe Audition 25.2 using a high-fidelity microphone, held 10 cm from the speaker's mouth. Firstly, to find a speaker to record the final word list, a sample of 15 words was presented to five native Malayalam speakers (both male and female), who recorded the words. Then, the recorded samples were sent to five audiology experts and five native Malayalam speakers with diverse educational and socio-economic backgrounds. They were asked to rate the samples on naturalness, clarity, pronunciation, and pleasantness using a 5-point scale, where 1 indicates the poorest rating, and 5 indicates the best, with domain-specific descriptors provided for each. Afterwards, the responses were gathered, and a speaker with the maximum rating was selected for the final recording of all the word lists. The speaker was instructed to practice pronouncing the words naturally and clearly, with a neutral intonation and a constant vocal effort throughout the recording. The recorded waveforms were digitised with a 32-bit A/D converter at a sampling frequency of 44,100 Hz.

To generate a consistent set of Phonemically Balanced Malayalam words, recorded words were then edited using Adobe Audition version 25.2 software to trim extraneous intervals and to normalise the intensity.

3.1.6 Procedure for Spectral Centroid Estimation

The recorded bi-syllabic phonemically balanced Malayalam words were subjected to acoustic analysis to obtain their spectral centroid values using a custom MATLAB program developed with the assistance of an expert in signal processing. Each recorded word was first segmented and saved as an individual audio file in WAV format with standardised recording parameters. The audio files were imported into MATLAB, where preprocessing procedures, such as amplitude normalization and removal of leading and trailing silence, were applied to ensure uniformity across stimuli. The custom script then performed spectral analysis of each word using Fast Fourier Transform (FFT)-based processing. The spectral centroid, representing the centre of gravity of the frequency spectrum, was calculated by weighting each frequency component by its spectral magnitude over the signal duration. The obtained centroid values were used to examine and balance the spectral distribution of the words across the developed Malayalam PB word lists.

3.2 Phase 2: Assessing the Homogeneity in Audibility

3.2.1 Noise Generation

Each word recording was transformed to the frequency domain in order to generate Spectro-temporally matched noise for the test. Afterwards, intelligibility of the recording was removed through spectral inversion. This technique involved inverting the frequency components around the spectral centre using a custom MATLAB program developed with the assistance of an expert in signal processing. This ensures that the resulting noise preserves the temporal envelope and complexity of the original word recording. This method produced noise closely matching the spectral and temporal characteristics of the word stimuli.

3.2.2 Determination of Global SNR

A pilot study (N=5) was conducted to establish a global SNR at which word identification would be approximately 50% to reduce floor and ceiling effects during test administration. The recorded words were mixed with the spectrum-inverted noise at multiple SNR levels (0 dB, -2 dB, -4 dB, and -6 dB) and presented to a group of native Malayalam speakers, monaurally at a comfortable listening level. Participants were instructed to repeat the words.

The percentage of correctly identified words was scored at each SNR. The global SNR was defined as the level at which approximately 50% of words are correctly identified. The global SNR-50 ranged from -2 to -4 dB in the pilot study. This global SNR served as the reference point for constructing test lists and selecting additional SNR intervals: 0 dB, -2 dB, -4 dB and -6 dB for detailed speech perception assessment.

3.2.3 Mixing and Calibration

The stimuli and their corresponding spectrum-inverted noise were normalised to equivalent long-term RMS amplitude. Words and noise were digitally mixed at target SNRs based on the global SNR-50 and additional SNRs. The RMS levels for mixing were determined using precise amplitude measurements over a fixed analysis window (e.g., 30 ms), in accordance with established calibration practices.

3.2.4 Pilot Study

A pilot study was conducted with five native Malayalam speakers aged 18–50 years. Participants were instructed to repeat the presented recorded words verbatim. All word lists were administered under quiet as well as noise conditions at 0, −2, −4, and −6 dB SNR. Following test administration, a brief interview was conducted to evaluate the appropriateness of the recorded speech material when presented through transducers, the naturalness of the recorded words, the adequacy of the inter-stimulus interval, and the need for any practice items prior to testing.

Based on the pilot study and feedback, words demonstrating consistent identification performance across participants while maintaining the target phonemic balance were retained. The final word lists consisted of six phonemically balanced words with comparable acoustic properties and levels of listening difficulty.

Based on feedback from the pilot study, a practice list of five words not included in the original word lists was prepared. These practice items were presented initially along with the 1 kHz calibration tone prior to the commencement of testing. All participants were instructed on the test procedure and required to complete a practice trial before the administration of the actual speech identification test.

3.3 Phase 3: Standardization of Word Lists

3.3.1 Participants

After a routine audiological evaluation, the recorded word list was presented in a quiet room to 40 native Malayalam-speaking individuals aged 18 to 50 years with normal hearing sensitivity. Using a calibrated dual-channel audiometer, pure-tone thresholds, speech recognition thresholds using existing Malayalam spondee words, and speech identification scores using existing Malayalam PB words were estimated. To evaluate middle ear functional status, tympanograms and acoustic reflexes were obtained using a calibrated GSI Tympstar middle ear analyser.

Using the modified Hughson and Westlake method (Carhart & Jerger, 1959), pure-tone air-conduction thresholds were established at octave frequencies from 250 Hz to 8000 Hz, and bone-conduction thresholds from 250-4000 Hz for each participant. Tympanometry measurements were assessed using a probe tone at 226 Hz and 85 dB SPL, and acoustic reflexes were measured ipsilaterally and contralaterally with reflex-eliciting tones at 500, 1000, 2000, and 4000 Hz to track the acoustic reflex threshold.

Participants were selected based on these criteria: Pure tone threshold average within 15 dB HL, ‘A’/’As’ type (in case of ‘As’ type, a compliance of >0.30 with present acoustic reflexes is considered), tympanogram, and presence of acoustic reflexes in at least two frequencies. A change in admittance value greater than 0.03 ml was considered the criterion for the presence of reflexes (Katz et al., 2015). Participants with any history of otological, neurological, speech-language, or psychological disorders, or significant noise exposure (as determined through case history, interview, and observation) were excluded from the study.

3.3.2 Test Environment

All the testing was carried out in a sound-treated audiometric room having permissible ambient noise levels standards specified by the American National Standards Institute (ANSI S3.1-1999). This ensured a valid and reliable pure-tone and speech audiometric assessment.

3.3.3 Administration of Developed Word Lists

All 40 participants were presented with all the word lists in a quiet room at 40 dB SL (Ref: SRT). The word lists were given via a personal laptop and Sennheiser HDA 200 headphones, using a calibrated audiometer. Participants were instructed to repeat the words they heard verbatim, and their responses were recorded on a scoring sheet. In case of not understanding or not hearing properly, they were asked to guess the word (Katz et al., 2015). Two audiologists (one in the audiometric room and another in the tester room) scored participants' responses against a pre-prepared list of word lists in the order they occurred in the recorded material. The scorers have to check the corresponding columns for the correct (√) or incorrect (X) symbols. Correct responses were scored as 1 and incorrect responses as 0. In case of discrepancies in scores between two audiologists, the particular word list was re-administered after completion of the entire procedure and a short break. Further, the 6 lists were presented to 10 participants to assess PI-PB function at 0 dB SL, 10 dB SL, 20 dB SL, 30 dB SL, and 40 dB SL (Chinnaraj et al., 2022). To avoid order effects, the word lists were presented at different levels in a random order across participants. To minimise practice effect, the presentation order of the word lists and the intensity level (dB SL) were pseudo-randomised across participants. A predetermined presentation chart was used for this, which ensured that each participant heard each word list only once at a given presentation level, while all word lists were used equally across participants

3.4 Phase 4: Measuring the Reliability

3.4.1Test-Retest Reliability

The developed word lists were re-administered to five participants by the same audiologist one week later to assess test–retest reliability. The intraclass correlation coefficient (ICC) and Cronbach's alpha (α) were computed using a two-way mixed-effects model with absolute agreement.

3.4.2 Inter-Judge Reliability

To assess inter-clinician reliability, the developed word lists were administered independently by two audiologists to five participants under identical test conditions. The administrations were carried out on the same day, with an adequate interval between sessions to minimise fatigue. Percentage agreement was used to assess inter-judge reliability.

3.5 Analysis

The data obtained was tabulated and analysed using SPSS (Statistical Package for the Social Sciences) software version 26.0. The Shapiro-Wilk test was used to determine whether the data follow a normal distribution. Cronbach's alpha, mean, and standard deviation analyses were used for familiarity ratings. Centroid frequency across the word list was analysed using the Friedman test and repeated-measure ANOVA. Speech identification score obtained in quiet and across different SNR conditions was analysed using descriptive statistics and the Friedman test. As significance was observed at 0 dB and -4 dB SNR, the Wilcoxon signed-rank test with Bonferroni correction was used. Additionally, SNR 50 values were determined by linear interpolation and analysed using the Friedman test. Psychometric analysis was conducted for different sensation levels using the Friedman test and the PI-PB function.

CHAPTER 4

RESULTS

The present study aimed to develop and standardize phonemically balanced word lists in Malayalam language for adults to assess speech identification scores in adults with normal hearing sensitivity. In addition, the study evaluated the consistency of speech identification scores across different wordlists in adults with normal hearing sensitivity and examined the homogeneity of audibility among the developed word lists. The study included 40 native Malayalam speakers with normal hearing sensitivity, who underwent a detailed audiological evaluation before the administration of the developed Malayalam wordlists. The results of the present study are given under separate sections corresponding to the different stages involved in the development and evaluation of the recorded Malayalam PB word lists

4.1 Familiarity Ratings

A pool of 1000 bi-syllabic words in Malayalam was collected, and familiarity ratings of the words were obtained from 15 native Malayalam speakers using a five-point rating scale. The internal consistency of the ratings, as measured by Cronbach's alpha, was high ($\alpha = 0.79$), indicating good inter-rater agreement. Mean familiarity scores and standard deviation were computed for each word, and those words that had a mean score of ≥ 4 and a standard deviation ≤ 1 were retained for further development of lists. The Friedman test was performed across word lists and found no significant difference in mean familiarity scores (χ^2 (5) = 5.96, $p = .31$).

4.2 Content Validation

Following familiarity ratings, the selected words underwent content validation by 11 experts comprising five audiologists, two speech-language pathologists, and two experts each from linguistics and psychology. For each word, the Content Validity Ratio (CVR) was calculated using the following equation:

$$CVR = \frac{(n_e - N/2)}{N/2}$$

Where n_e is the is the number of experts rating the item as relevant (familiarity rating of 3—5 or "Yes" for appropriateness and relevance) and N is the total number of experts for that domain. Items meeting the critical CVR threshold for the respective panel size (Ayre & Scally, 2014) were retained for further development of Malayalam PB word lists. The overall CVI was computed as the mean of CVR across all the retained items in each domain. The domain-wise CVI were as follows: familiarity (CVI=0.95), cultural and emotional appropriateness (CVI=0.94), relevance to speech audiometry (CVI=1.00), and linguistic relevance (CVI= 0.94), indicating excellent content validity across all domains.

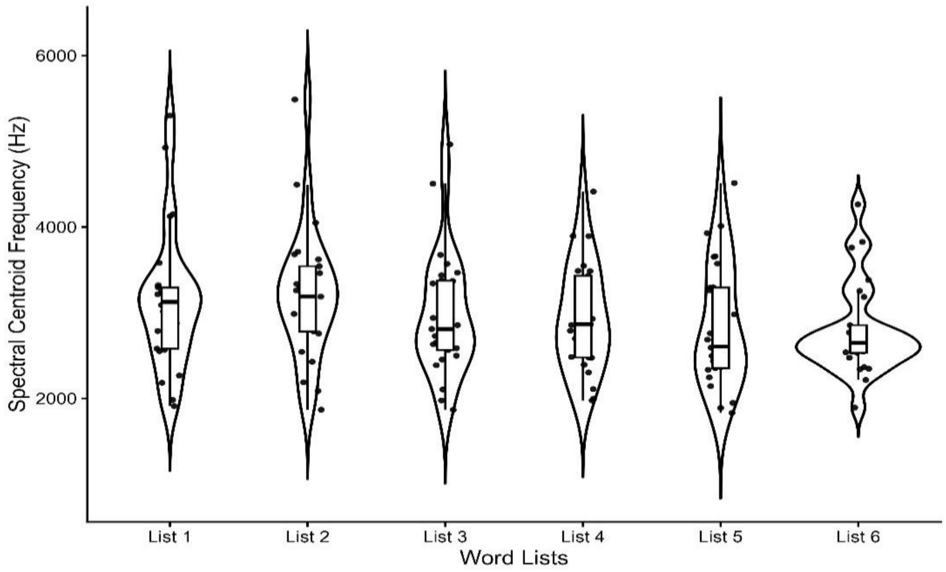
4.3 Centroid Frequency

The Shapiro–Wilk test revealed mixed normality findings across the six-word lists. Centroid values for List 1 and List 6 deviated significantly from the normal distribution ($p <$ .05), in contrast, the remaining lists showed normal distribution characteristics ($p > .05$). Therefore, a nonparametric Friedman test was primarily used for comparison of centroid frequencies across lists. The Friedman test revealed no significant difference in spectral centroid frequencies across the six Malayalam PB word lists ($\chi 2(5) = 6.371$, $p = 0.27$). The mean centroid values ranged from 2794.44 Hz to 3206.21 Hz, indicating comparable spectral energy distribution across the lists.

As some of the combinations showed normal distribution, a repeated-measures ANOVA was also performed, which again demonstrated no significant effect of word list on centroid frequency values (F (5, 120) =1.258, p=0.287), further supporting the acoustic equivalence of the developed lists.

Figure 4.1

Displays the Distribution of Spectral Centroid Frequencies Across the Six Word Lists



Note. Violin plot showing the spectral centroid frequency (Hz) across each list, where a filled circle represents an individual data point, the horizontal line represents the median, and the box indicates the interquartile range (25 -75 percentile).

4.4 Speech Identification Scores (SIS) in Quiet Across Word Lists

Speech Identification Scores (SIS) obtained in quiet were analysed across the six developed word lists. Descriptive statistics showed comparable mean scores across lists, with minimal variability.

Because the data were not normally distributed according to the Shapiro–Wilk test ($p < .05$), nonparametric analysis was performed using the Friedman test. The analysis revealed no statistically significant difference in SIS scores across the six-word lists, χ^2 (5) = 2.40, p = 0.79. The grand mean across the lists is 99.48% in quiet. These findings indicate that the developed word lists were equivalent in difficulty under quiet listening conditions.

Table 4.1

Mean, Median, Standard Deviation, Minimum, Maximum Scores, and Interquartile Range for the Six Word Lists Obtained in Quiet and at Different Signal-To-Noise Ratios (SNRs)

		Mean	**Median**	**SD**	**Minimum**	**Maximum**	**Interquartile Range**
List 1 quiet	Statistic	24.85	25.00	0.36	24.00	25.00	0.00
	Std. Error	0.05					
List 2 quiet	Statistic	24.90	25.00	0.30	24.00	25.00	0.00
	Std. Error	0.04					
List 3 quiet	Statistic	24.85	25.00	0.36	24.00	25.00	0.00
	Std. Error	0.05					
List 4 quiet	Statistic	24.92	25.00	0.26	24.00	25.00	0.00
	Std. Error	0.04					
List 5 quiet	Statistic	24.82	25.00	0.38	24.00	25.00	0.00
	Std. Error	0.06					
List 6 quiet	Statistic	24.87	25.00	0.33	24.00	25.00	0.00
	Std. Error	0.05					

		Mean	Median	SD	Minimum	Maximum	Interquartile Range
List 1 0 dB SNR	Statistic	21.10	21.00	0.37	20.00	22.00	0.00
	Std. Error	0.05					
List 2 0 dB SNR	Statistic	21.27	21.00	0.45	21.00	22.00	1.00
	Std. Error	0.07					
List 3 0 dB SNR	Statistic	21.10	21.00	0.30	21.00	22.00	0.00
	Std. Error	0.04					
List 4 0 dB SNR	Statistic	21.55	22.00	0.50	21.00	22.00	1.00
	Std. Error	0.07					
List 5 0 dB SNR	Statistic	21.37	21.00	0.49	21.00	22.00	1.00
	Std. Error	0.07					
List 6 0 dB SNR	Statistic	21.45	21.00	0.50	21.00	22.00	1.00
	Std. Error	0.07					
List 1 -2 dB SNR	Statistic	16.35	16.00	0.53	15.00	17.00	1.00
	Std. Error	0.08					

		Mean	Median	SD	Minimum	Maximum	Interquartile Range
List 2 -2 dB SNR	Statistic	16.35	16.00	0.48	16.00	17.00	1.00
	Std. Error	0.07					
List 3 -2 dB SNR	Statistic	16.32	16.00	0.52	16.00	18.00	1.00
	Std. Error	0.08					
List 4 -2 dB SNR	Statistic	16.57	17.00	0.71	15.00	19.00	1.00
	Std. Error	0.11					
List 5 -2 dB SNR	Statistic	16.37	16.00	0.49	16.00	17.00	1.00
	Std. Error	0.07					
List 6 -2 dB SNR	Statistic	16.55	17.00	0.63	15.00	18.00	1.00
	Std. Error	0.10					
List 1 -4 dB SNR	Statistic	12.80	13.00	0.40	12.00	13.00	0.00
	Std. Error	0.06					
List 2 -4 dB SNR	Statistic	12.77	13.00	0.42	12.00	13.00	0.00
	Std. Error	0.06					

		Mean	Median	SD	Minimum	Maximum	Interquartile Range
List 3 -4 dB SNR	Statistic	12.52	13.00	0.59	11.00	14.00	1.00
	Std. Error	0.09					
List 4 -4 dB SNR	Statistic	12.75	13.00	0.49	12.00	14.00	1.00
	Std. Error	0.07					
List 5 -4 dB SNR	Statistic	12.50	12.50	0.50	12.00	13.00	1.00
	Std. Error	0.08					
List 6 -4 dB SNR	Statistic	12.62	13.00	0.54	12.00	14.00	1.00
	Std. Error	0.08					
List 1 -6 dB SNR	Statistic	5.65	6.00	0.48	5.00	6.00	1.00
	Std. Error	0.07					
List 2 -6 dB SNR	Statistic	5.72	6.00	0.45	5.00	6.00	1.00
	Std. Error	0.07					
List 3 -6 dB SNR	Statistic	5.67	6.00	0.47	5.00	6.00	1.00
	Std. Error	0.07					

		Mean	Median	SD	Minimum	Maximum	Interquartile Range
List 4 -6 dB SNR	Statistic	5.75	6.00	0.43	5.00	6.00	0.75
	Std. Error	0.06					
List 5 -6 dB SNR	Statistic	5.55	6.00	0.50	5.00	6.00	1.00
	Std. Error	0.07					
List 6 -6 dB SNR	Statistic	5.87	6.00	0.64	5.00	8.00	0.00
	Std. Error	0.10					

Note. *SD-Standard Deviation; dB- Decibel; SNR-Signal-to-Noise ratio.

4.5 Effect of Signal-to-Noise Ratio (SNR) on Speech Identification Scores Across Word Lists

For the 0 dB SNR condition, the Friedman test revealed a statistically significant difference in SIS scores across the six-word lists, χ^2 (5) = 33.07, $p < .001$. Kendall's coefficient indicated low agreement among the pair differences ($W = .165$), suggesting a weak effect size (Peres, 2026). However, a post hoc Bonferroni pairwise comparison was conducted to identify the specific word lists that differed significantly. Post hoc analysis using the Wilcoxon signed-rank test revealed significant differences between several word-list pairs at 0 dB SNR. Significant differences were observed between List 1 and List 4, List 1 and List 5, List 1 and List 6, List 2 and List 4, List 2 and List 6, List 3 and List 4, List 3 and List 5, and List 3 and List 6. Kendall's coefficient of concordance indicated a weak effect size among all these pairs ($W < 0.10$). No significant differences were found between the remaining list pairs. The details of this analysis are presented in Table 4.2.

For the -2 dB SNR condition, the analysis revealed no statistically significant difference in SIS scores across the six-word lists, $\chi^2(5) = 8.81$, $p = .117$. This finding suggests highly comparable performance across the six lists.

For the -4 dB SNR condition, the Friedman test revealed a statistically significant difference in SIS scores across the six word lists, $\chi^2(5) = 13.01$, $p = .023$. Kendall's coefficient of concordance indicated a weak effect size ($W = .065$). Further, post hoc analysis using the Wilcoxon signed-rank test showed significant differences between List 1 and List 3, List 1 and List 5, List 2 and List 3, List 2 and List 5, and List 4 and List 5. Kendall's coefficient of concordance indicated a weak effect size among all these pairs ($W < 0.10$). No significant differences were found between the remaining list pairs. The details of this analysis are presented in Table 4.2.

For the -6 dB SNR condition, the analysis revealed no statistically significant difference in SIS scores across the six word lists, $\chi^2(5) = 7.61$, $p = .179$. This finding suggests highly comparable performance across the six lists.

Table 4.2

Pairwise Comparison with Bonferroni Correction Between List Pairs for 0 dB and -4 dB SNRs

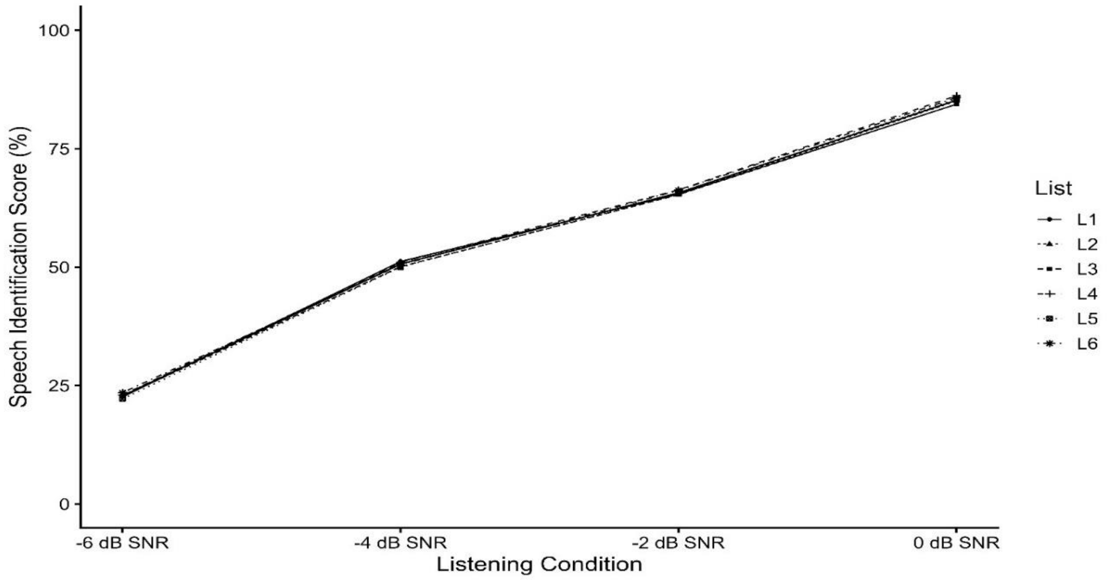
SNR	List Pairs	Z-value	p-value	p-adjusted
0 dB SNR	List 1_List 4	-3.53	0.00	0.00
	List 1_List 5	-2.67	0.008	0.12
	List 1_List 6	-2.99	0.003	0.05
	List 2_List 4	-3.32	0.001	0.02
	List 2_List 6	-2.33	0.020	0.30
	List 3_List 4	-3.84	0.00	0.00
	List 1_List 5	-2.52	0.012	0.18
	List 3_List 6	-3.13	0.002	0.03
-4 dB SNR	List 1_List 3	-2.67	0.008	0.12
	List 1_List 5	-2.35	0.019	0.29
	List 2_List 3	-2.36	0.018	0.27
	List 2_List 5	-2.67	0.008	0.12
	List 4_List 5	-2.24	0.025	0.38

Note. *SNR – Signal-to-Noise ratio.

Figure 4.2 presents the psychometric function, showing the speech identification score as a percentage for six word lists across different SNRs. The result indicates a clear improvement in SIS score as the SNR increases from -6 dB to 0 dB, and the psychometric functions for all six lists overlap, indicating that these lists are homogeneous in difficulty.

Figure 4.2

Psychometric Function Displaying the Speech Intelligibility Score (SIS In Percentage) of Six-Word Lists Across -6 dB, -4 dB, -2 dB and 0 dB SNRs



Note. SNR – Signal-to-Noise ratio, L1–List 1, L2–List 2, L3–List 3, L4–List 4, L5–List 5, L6–List 6.

Because the speech identification scores (SIS) obtained across the six word lists in quiet and at different SNR conditions showed either no significant difference or only a statistically significant difference with a small effect size (±1 word), the scores were averaged across the word lists for subsequent analysis of performance across the various SNR conditions.

4.6 SNR-50 Comparison Across Word Lists

The SNR-50 value was estimated from the psychometric function obtained across the tested SNR conditions (0, −2, −4, and −6 dB SNR). Since the exact 50% speech identification score was not directly measured, linear interpolation was used between the adjacent SNR conditions surrounding the 50% performance level to estimate the corresponding SNR-50 value. The formula used to obtain SNR-50 is

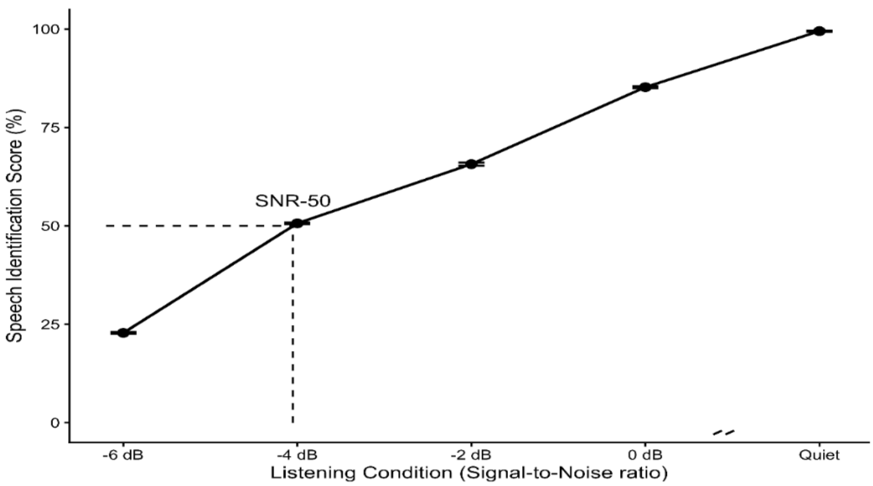
SNR-50 = SNR1 + (50-P1) (SNR2-SNR1) / (p2-P1) where P1, P2 are the Speech Identification Scores surrounding 50% and SNR1 and SNR2 are the corresponding SNR values.

The Friedman test revealed no statistically significant difference in SNR-50 across word lists, $\chi^2(5) = 10.42$, $p = .062$.

Figure 4.3 illustrates Psychometric functions for wordlists across different listening conditions. The SIS increased progressively from approximately 20% at -6 dB SNR to nearly 100% in the quiet condition. The dashed lines represent an SNR of 50, at which 50% speech identification was achieved, around -4 dB.

Figure 4.3

Psychometric Functions for Word lists Across Different Listening Conditions



Note. Psychometric function of wordlists displaying SNR-50.

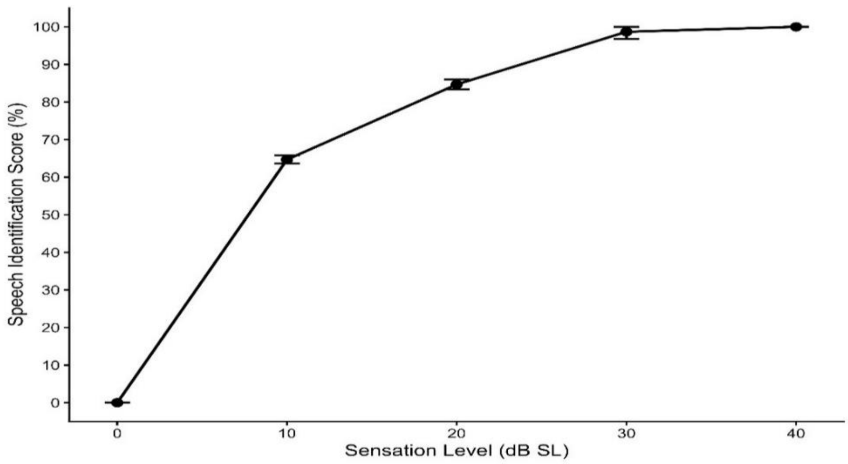
4.7 Psychometric Analysis of SIS at Different Sensation Levels (SLs)

A psychometric function was drawn using the Performance Intensity-Phonemically Balanced (PI-PB) function to assess whether all lists are homogeneous in difficulty across different intensities. Friedman's test revealed a statistically significant difference in SIS scores across the six word lists, χ^2 (4) = 40.00, $p < .001$. Kendall's coefficient of concordance indicated a large effect size ($W = 1.00$), suggesting high consistency in performance across the intensities.

The PI-PB test was conducted across five sensation levels (0, 10, 20, 30, 40 dB SL) with 10 participants, and the psychometric function was established for all six word lists. Figure 4.4 illustrates the psychometric function curve across five sensation levels (dB). From the figure, it is clear that SIS scores increased systematically with intensity level, with 0% at 0 dB SL and 100% above 30 dB SL, indicating that difficulty decreases as intensity increases.

Figure 4.4

Psychometric Function Curve Across Five Sensation Levels (dB)



Note. Psychometric function curve showing Speech Identification Score (SIS) across five sensation levels.

4.8 Reliability

4.8.1 Test–Retest Reliability

Test–retest reliability of the developed word lists was assessed using the Intraclass Correlation Coefficient (ICC) and Cronbach's alpha.

For several listening conditions, ICC estimation was not possible due to ceiling performance and zero variance in participant scores. In these conditions, all participants obtained identical scores (100% scores in both test and retest administrations), indicating perfect agreement across the two testing sessions.

Excellent test–retest reliability was observed for the analysed listening conditions, with ICC values of 1.00 and Cronbach's alpha of 1.00, indicating perfect agreement between test and retest administrations. These findings suggest that participant performance remained highly stable across repeated testing sessions under the analysed listening conditions.

For conditions in which sufficient variability in scores was present, Intraclass Correlation Coefficient (ICC) and Cronbach's alpha were computed to assess test–retest reliability. For instance, moderate test–retest reliability was observed in one of the listening conditions, with an ICC of 0.600 and Cronbach's alpha of 0.75, indicating acceptable agreement between the two testing sessions. Overall, the developed word lists demonstrated excellent temporal stability and consistency across repeated administrations.

4.8.2 Inter-Judge reliability

Inter-judge reliability was assessed using percentage agreement between the two judges across the developed Malayalam PB word lists. A high level of agreement was observed among the judges, indicating consistent scoring of speech identification responses. Cronbach's alpha and Intraclass Correlation Coefficient (ICC) could not be reliably estimated for several conditions due to the presence of constant scores and insufficient variance among observations, particularly under ceiling performance conditions.

Therefore, percentage agreement was considered a more appropriate measure of inter-judge reliability for the present data. The percentage agreement between the two judges ranged from 80% to 100%, indicating excellent inter-judge consistency across listening conditions and word lists. Table 4.3 shows the test-retest and inter-judge reliability of the SIS score obtained for six word lists across different listening conditions. Overall, the developed word lists show good test-retest and inter-judge reliability across different listening conditions.

Table 4.3

Test-Retest Reliability and Inter-Judge Reliability

Listening Condition	Test-Retest reliability		Inter-judge reliability
	ICC value	Cronbach's alpha (α)	Percentage agreement (%)
List 1 0 dB SNR	1.000	1.000	100
List 2 0 dB SNR	1.000	1.000	100
List 3 0 dB SNR	_	_	80
List 4 0 dB SNR	1.000	1.000	100
List 5 0 dB SNR	1.000	1.000	100
List 6 0 dB SNR	1.000	1.000	100
List 1 -2 dB SNR	1.000	1.000	100
List 2 -2 dB SNR	1.000	1.000	100
List 3 -2 dB SNR	0.600	0.750	100
List 4 -2 dB SNR	_	_	80
List 5 -2 dB SNR	1.000	1.000	100
List 6 -2 dB SNR	_	_	100
List 1 -4 dB SNR	1.000	1.000	100
List 2 -4 dB SNR	0.600	0.750	100
List 3 -4 dB SNR	0.667	0.800	100
List 4 -4 dB SNR	1.000	1.000	100
List 5 -4 dB SNR	1.000	1.000	100
List 6 -4 dB SNR	1.000	1.000	100
List 2 -6 dB SNR	0.600	0.750	80

Listening Condition	Test-Retest reliability		Inter-judge reliability
	ICC value	**Cronbach’s alpha (α)**	**Percentage agreement (%)**
List 3 -6 dB SNR	0.600	0.750	100
List 4 -6 dB SNR	1.000	1.000	100
List 5 -6 dB SNR	1.000	1.000	100
List 6 -6 dB SNR	1.000	1.000	80

Note. Intraclass Correlation Coefficient (ICC) value and Cronbach’s alpha (α) for measuring test-retest reliability, and Percentage agreement (%) for inter-judge reliability across different listening conditions (-6 dB, -4 dB, -2 dB and 0 dB) SNRs.

CHAPTER 5

DISCUSSION

The present study was undertaken to develop and evaluate recorded bisyllabic Malayalam phonemically balanced (PB) word lists for speech identification testing in quiet and noise conditions. The developed word lists were subjected to familiarity rating, content validation, acoustic analysis, psychometric evaluation, and reliability assessment to establish their clinical applicability. The findings of the study demonstrated that the developed lists were familiar, culturally appropriate, acoustically comparable, and homogeneous in difficulty across different listening conditions. Further, the developed lists showed good psychometric characteristics and high reliability across repeated administrations, supporting their potential utility in speech audiometry among native Malayalam speakers.

5.1 Familiarity Ratings and Content Validation

The collected 1000 bisyllabic words were subjected to familiarity rating, and the words that were highly familiar to the native Malayalam speakers were selected. A total of 700 words were shortlisted after the familiarity rating. This study followed the procedures used by Manjula et al. (2015) for developing PB word lists in Kannada and by Chinnaraj et al. (2022) for developing PB word lists in Tamil. Manjula et al. (2015) collected 1200 bisyllabic Kannada words and shortlisted 820 familiar words following familiarity rating by 15 native speakers using a 5-point scale. Whereas, Chinnaraj et al. (2022) collected 1015 bisyllabic Tamil words, and familiarity ratings were conducted with 20 native speakers. After familiarity ratings and content validation, 760 words were retained to develop 25 PB word lists.

Similar to these studies, the present study also used familiarity ratings on a 5-point scale, but words with a mean score$\geq$4 and a standard deviation $\leq$1 were retained for further standardization. However, compared with previous studies, the present study differs in the number of participants and the number of words retained. These differences may be due to variations in phonemic structure, word familiarity, dialectal influences and linguistic characteristics across different languages.

Owens (1961) noted that familiar words are easier to recognise than less familiar ones, and that discrimination ability is affected by word familiarity. According to Hudgins et al. (1947), the test materials used in speech audiometry should be familiar and simple so that the test assesses speech intelligibility rather than vocabulary or intelligence.

The high level of agreement among experts in the content validation, as reflected by domain-wise CVI values (familiarity=0.95; cultural and emotional appropriateness =0.94; relevance to speech audiometry=1.00; linguistic relevance=0.94), indicates that the selected words were considered appropriate and suitable for inclusion in the developed Malayalam PB word lists. These findings suggest that the developed test materials possessed good content validity and were linguistically, culturally, and clinically appropriate for speech audiometry assessment in Malayalam-speaking individuals.

The findings of the present study are similar to those reported by Chinnaraj et al. (2022), as both studies observed high expert agreement regarding the validity of the selected word. While the present study employed Lawshe's Content Validity Ratio (CVR) along with the Content Validity Index (CVI), Chinnaraj et al. (2022) assessed content validation using the Item-level Content Validation Index (I-CVI) and Scale-level Content Validation Index (S-CVI) measures. They reported I-CVI values ranging from 0.8 to 1, S-CVI/UA of 0.92, and S-CVI/Ave of 0.98, indicating excellent content validity. Despite methodological differences, both studies demonstrated strong expert agreement regarding the appropriateness and relevance of the selected words for speech audiometry assessment. In contrast, Manjula et al. (2015) mainly focused on familiarity ratings and equivalence testing rather than on formal content validation measures such as CVI or percentage agreement. Nevertheless, both studies emphasised the importance of selecting familiar, clinically appropriate words when developing phonemically balanced speech materials.

Content validation is an essential step in the development of speech audiometry materials because it helps ensure that the test items are familiar, appropriate, linguistically suitable, and clinically relevant to the target population (Hennig et al., 2018; Vaucher et al., 2022). Carhart (1951) highlighted that materials for speech audiometry should be independently developed and standardised for each language, as the linguistic and phonetic characteristics change across languages. Wang et al. (2007) reported that Mandarin disyllabic materials with high familiarity and phonological balance exhibited sufficient validity and reliability for clinical use.

A major strength of the present study was the inclusion of multidisciplinary professionals and non-expert native speakers as validators. The involvement of experts from different disciplines ensured the linguistic, clinical, and cultural appropriateness of the selected words. Additionally, inclusion of non-expert native Malayalam speakers helped ensure familiarity and real-world applicability of the test materials. Similar approaches involving expert and non-expert raters have been reported in studies by Chinnaraj et al. (2022), Hennig et al. (2018), and Vaucher et al. (2022), in which familiarity, appropriateness, and auditory recognition of speech materials were evaluated by both specialist and non-specialist judges.

After familiarity ratings and expert content validation, words that were familiar, culturally and emotionally appropriate, and relevant to speech audiometry were selected. The complete set of six Malayalam phonemically balanced (PB) word lists developed in the present study, together with their corresponding IPA transcriptions, is presented in Appendix A.

5.2 Centroid Frequency

The present study analysed the spectral centroid frequencies of developed Malayalam word lists to evaluate acoustic similarity among them. The spectral centroid frequency represents the distribution of spectral energy across frequencies and is the midpoint of the spectral energy distribution, reflecting the spectral characteristics of the sound signal (Kua et al., 2010; Sandell, 1995; Schubert & Wolfe, 2006). The results of the present study show no significant differences in centroid frequency across the six lists, indicating a comparable spectral distribution. Such acoustic similarity across the lists may reduce variability due to spectral differences. The distribution of acoustic energy across frequency is an important factor in the identification of speech sounds (Fry, 1964).

The wide range of centroid frequencies noted in each list indicated that significant acoustic energy was distributed across both low- and high-frequency regions of the speech spectrum. This finding may be related to the incorporation of different Malayalam speech sounds during the development of word lists. Studies related to broad phoneme classification and phoneme spotting have claimed that vowels, nasals, approximates, fricatives, and plosives possess distinct spectral and acoustic characteristics (Sangeetha et al., 2019). Frequent occurrence of fricatives and certain plosive sounds, which are characterised by greater high-frequency spectral energy, could contribute to relatively higher centroid frequencies in some lists (Deekshitha et al., 2018).

Previous studies on the development of PB words across languages have mainly focused on phonetic balancing and homogeneity in audibility rather than on acoustic spectral measures. So, a direct comparison of centroid frequencies across lists with previous studies on PB word list development is limited.

5.3 Speech Identification Scores (SIS) in Quiet Across Word Lists

The Speech identification scores showed comparable means across all developed word lists with no statistically significant differences between the lists in quiet listening conditions. The grand mean percentage correct across the lists is 99.48%. The result of the present study reveals that the developed six Malayalam PB word lists were equivalent in difficulty for individuals with normal hearing sensitivity.

These findings are similar to those of previous studies on the development of PB word lists across different languages. Chinnaraj et al. (2022) reported no significant difference in speech identification scores across 25 lists of the 25 developed PB word lists in Tamil, with a grand mean score of 99.77% in individuals with normal hearing. Similarly, Ullrich and Grimm (1976) noted that the mean score of 99.7% using the NU-6 word list at the most comfortable listening level, whereas Beattie et al. (1997), using the CID W22 monosyllabic word lists, reported a mean word recognition score of 95% in quiet for normal hearing listeners, which suggests excellent speech identification abilities under favourable listening conditions. According to Davis and Silverman (1960), individuals with normal hearing sensitivity generally achieve perfect speech identification performance at comfortable listening levels, results comparable to those of the present study.

According to Hirsh et al. (1952), speech audiometry materials should be standardised and homogeneous with respect to audibility. Lehiste and Peterson (1959) highlighted the importance of phonemic balancing and claimed that linguistic and phonemic characteristics affect speech recognition performance. The consistent SIS scores across the developed Malayalam PB word lists in the present study are linked to careful selection of familiar words, the validation procedure, and phonemic balancing followed during development.

Wilson and Margolis (1983) highlighted that to establish interchangeable speech materials, consistency between the lists is important. Therefore, the absence of significant differences across the developed Malayalam word lists in the Friedman analysis suggests homogeneity in difficulty level, indicating that the lists can be used interchangeably during speech audiometry evaluation. Hence, multiple equivalent lists will reduce the practice effect during repeated administration.

The ceiling effect observed in the present study may be due to administering the word lists at comfortable suprathreshold listening levels in quiet conditions among individuals with normal hearing sensitivity, where the highest Speech Identification Score is expected. Similar observations have been reported in previous PB word list development studies, in which normal-hearing individuals showed near-perfect speech identification scores at supra-threshold presentation levels in quiet (Chinnaraj et al., 2022; Manjula et al., 2015; Wang et al., 2007). Hence, variability in scores may become minimal under ideal listening conditions. Therefore, several studies assessed speech perception under more challenging listening conditions, such as varying sensation levels or the speech-in-noise paradigm, to better evaluate list equivalency and sensitivity (Nilsson et al., 1994; Wilson et al., 2008).

5.4 Effect of Signal-to-Noise Ratio (SNR) on Speech Identification Scores Across Word Lists

Speech Identification Scores obtained at different SNRs were analysed across the six developed word lists to determine the equivalency of the lists. As reported in the literature and as expected, the SIS scores reduced in the presence of noise compared to the quiet condition (Hornsby & Ricketts, 2001). The present study reveals that there was no significant difference across the lists at -2 dB and -6 dB SNR, whereas a significant difference was noted across lists at 0 dB and -4 dB SNR. Nevertheless, the Kendall's coefficient of concordance shows only small effect sizes across the significant list pairs, suggesting negligible variability across the lists. Detailed observation of the descriptive data revealed that the differences between the lists were within ±1 word, reflecting a small effect size. Hence, the difference of

±1 word performance across word lists is still considered comparable (Wilson et al., 2008). Similar findings were also noted by Wilson et al. (2008), who compared PB 50, CID W-22, and NU-6 word materials presented at -7, -2, +3, and +8 dB SNRs and found equivalent recognition performance across different word lists, with only minimal inter-list differences.

In the quiet-listening condition, most participants demonstrated ceiling performance, resulting in minimal variability in speech identification scores across individuals and word lists. Gifford et al. (2008) reported that such ceiling effects in speech recognition materials may influence performance across test materials and reduce the sensitivity of statistical comparisons. Therefore, multiple studies have demonstrated the importance of evaluating performance under degraded listening conditions with varying SNRs, as these conditions provide more realistic and sensitive measures of speech perception ability (Billings et al., 2016; Le Prell & Brungart, 2016; Nilsson et al., 1994). Le Prell and Brungart (2016) further reported that suprathreshold speech-in-noise measures may be more sensitive than traditional quiet-listening measures in identifying subtle auditory difficulties. Thus, in the present study, assessing performance across different SNR conditions revealed greater variability in speech perception and provided a more sensitive, statistically robust measure for evaluating the homogeneity and equivalence of the developed Malayalam PB word lists.

A study done by Manjula et al. (2015) on the development of PB word list in Kannada revealed that most word lists were equivalent in noise, whereas few showed significant differences and were excluded. Hagerman (1984) developed standardized speech materials in noise and reported that equivalent sentence lists with similar spectral characteristics and difficulty levels are essential for reliable speech reception threshold measurements in noise.

The minimal differences noted across list pairs at 0 dB and -4 dB SNR conditions in the present study could be attributed to phoneme confusions under moderately challenging conditions. Speech-in-noise studies demonstrated that speech recognition performance may vary depending on factors such as presentation level, speech material characteristics, and SNR condition (Plomp, 1986; Wilson et al., 2008).

5.5 SNR-50 Comparison Across Word Lists

The SNR-50 values across the six developed Malayalam PB word lists were compared to examine the equivalency of the lists in speech perception under noisy listening conditions. The present study demonstrated no statistically significant differences in SNR-50 values across the developed word lists, supporting their relative equivalency and interchangeability for speech-in-noise assessment. The derived SNR-50 was around -4 dB, indicating that approximately 50% speech identification was achieved at this listening condition.

Unlike the absolute speech identification score obtained at different sensation levels, SNR-50 is considered a more sensitive and statistically robust measure for evaluating the word list equivalency. Nilsson et al. (1994) demonstrated that percentage intelligibility measures obtained at fixed presentation levels are often affected by floor and ceiling effects and emphasised that SNR-50 reflects the signal-to-noise ratio required to achieve 50% speech identification performance and is derived from the psychometric function across multiple listening conditions. Thus, as SNR-50 captures performance across different listening conditions rather than at a single intensity level, it may provide a more sensitive measure for evaluating the equivalency and homogeneity of the developed word lists (Nielsen & Dau, 2009).

Reynard et al. (2022) claimed that speech-in-noise testing is clinically important, as speech understanding difficulties in background noise are common complaints reported by individuals during audiological evaluation.

The absence of significant differences in SNR-50 across the developed Malayalam PB word lists implies consistent difficulty with noise and reveals equivalency in speech-in-noise performance. Similar findings were reported in the study by Manjula et al. (2015), who noted that a 50% score was obtained at an SNR slightly above -3 dB and reported that most of the developed word lists showed equivalent performance, whereas only a few listed demonstrated significant differences. These observations are comparable to the present study findings, where approximately 50% was observed around – 4 dB SNR. Hence, these results revealed that developed Malayalam PB word lists can be used for speech audiometry for adults in noise conditions during routine evaluation.

The present findings are also comparable with other speech-in-noise studies. Leclercq et al. (2018) reported minimal variability in SNR-50 values across the three French VRB test (Rapid speech in noise) lists using sentence-based speech-in-noise materials, suggesting homogeneity and sensitivity of the speech stimuli.

Differences in exact SNR-50 values across studies should be interpreted cautiously because variations in speech materials, testing procedures, and noise conditions may influence speech perception performance. Manjula et al. (2015) reported that sentence materials contain contextual cues and linguistic redundancy, whereas word lists provide comparatively less redundancy. Such differences in speech material characteristics may influence speech identification performance in noise.

To conclude, the developed lists demonstrated consistent psychometric characteristics across SNR conditions. They can be used for clinical speech audiometry, hearing aid evaluation, auditory rehabilitation, and research applications among Malayalam-speaking individuals.

5.6 Psychometric Analysis of SIS at Different Sensation Levels (SLs)

A Performance-Intensity Phonemically Balanced (PI-PB) function was established across different sensation levels, and the present study revealed a gradual increase in the speech identification score with increasing intensity, with 0% at 0 dB SL and 100% above 30 dB SL. The obtained PI-PB functions showed similar psychometric properties across the developed word lists, indicating comparable audibility and difficulty levels.

The PI-PB function was established to assess the speech identification performance across different sensation levels (Boothroyd, 2008). Therefore, assessment across multiple intensity levels helps evaluate whether the lists exhibit the expected psychometric growth in speech identification performance with increasing intensity (Najem & Marie, 2021). Similar PI-PB functions across the developed word lists indicate homogeneity in difficulty and comparable auditory performance characteristics (Nissen et al., 2011). Furthermore, PI-PB analysis provides clinically relevant information on the optimal presentation level and overall performance of the recorded speech material in normal-hearing listeners.

Dirks et al. (1977) reported that PI-PB functions are clinically useful for evaluating speech recognition characteristics and changes in performance across intensity levels. The findings of the study are consistent with several previous studies in the development of PB word lists. Wang et al. (2007), during the development of Mandarin disyllabic speech audiometry materials, evaluated PI functions to establish the reliability and validity of the developed word lists and reported sufficient equivalency across lists. Chinnaraj et al. (2022) and Manjula et al. (2015) reported, during the development and standardisation of Tamil and Kannada PB word lists, that Speech Identification Scores increased progressively from 10 dB SL to 40 dB SL across all lists, indicating homogeneity in difficulty and audibility among the developed materials. Both studies demonstrated plateau performance above 30 dB SL. Comparable findings were reported by Najem and Marie (2021) in the development of Arabic PB wordlists, where no significant improvement was observed above 40 dB SL. In the present study, a plateau in SIS is observed above 30 dB SL, indicating that the developed word lists achieved maximum intelligibility at comfortable suprathreshold listening levels in individuals with normal hearing sensitivity.

The comparable PI-PB functions observed across the developed Malayalam PB word lists further support the homogeneity and equivalency of the developed materials. Similarly, Nissen et al. (2011) highlighted that psychometrically equivalent speech materials should possess comparable audibility characteristics and psychometric slopes across the lists.

5.7 Reliability

5.7.1 Test-Retest Reliability

Test-retest reliability was evaluated to assess the consistency and temporal stability of the developed Malayalam PB word list across repeated administrations. Reliability across sessions is important in speech audiometry, as it ensures Speech Identification Scores obtained are reproducible and are not influenced by random variations in listener performance or testing conditions (Ostergard, 1983).

In the present study, excellent test-retest reliability was observed across most listening conditions and lists, with perfect agreement between test and retest administrations. In a few listening conditions, moderate reliability was observed (ICC=0.60-0.67), indicating acceptable consistency across sessions. Similarly, Cheoy et al. (2021), during the development of the Mandarin Nonsense Word Identification test, reported an ICC score of 0.75-0.87 across repeated administrations, which indicates good test-retest reliability. Likewise, Noor and Arif (2018), in the development of the Urdu speech perception test, reported a significant test-retest value of 0.881, suggesting good consistency across sessions. Therefore, the findings of the present study revealed that developed Malayalam PB word lists demonstrated good temporal stability and consistency across repeated administration.

5.7.2 Inter-Judge Reliability

Inter-judge reliability was assessed to demonstrate the consistency of scoring across judges. Although the developed Malayalam PB word lists were presented using recorded speech materials, the speech identification scores were scored subjectively from listener responses. Hence, assessment of inter-judge reliability was important to confirm that the observed scores were consistent across examiners and were not influenced by scorer bias.

In the present study, the percentage agreement between two judges ranged from 80% to 100%, suggesting excellent inter-judge consistency across listening conditions and wordlists. Percentage agreement was considered, as Cronbach's alpha and ICC could not be reliably estimated for several conditions due to constant or insufficient score variability. Noor and Arif (2018) reported an inter-scorer reliability value of 0.598 during the development of the Urdu speech perception test, indicating acceptable agreement across scorers. Similarly, Hennig et al. (2018), while developing and validating disyllabic word lists, reported satisfactory evidence of reliability for the developed recorded speech materials. Therefore, the findings of the present study suggest that the developed Malayalam PB word lists provide consistent scoring across different examiners.

5.8 Clinical Implications

The findings of the present study reveal that the developed Malayalam PB wordlists can be used effectively for routine speech identification testing in both quiet and noise conditions among native Malayalam speakers. The developed word lists exhibited good equivalency, homogeneity, psychometric properties and reliability across listening conditions. Hence, it suggests their potential usefulness in routine speech audiometry testing, assessing the efficacy of different interventions and rehabilitation programs, evaluating the outcomes of amplification devices such as hearing aids, cochlear implants, and FM systems, and in research applications requiring multiple equivalent lists.

5.9 Limitations and Future Directions

The present study has some limitations that should be considered when interpreting the findings. The study included 40 individuals with normal hearing sensitivity, a relatively small sample. In addition, the developed word-lists were not validated in individuals with hearing impairment. Further, the test-retest reliability was evaluated on a relatively small sample. Moreover, the SNR-50 was estimated through interpolation between adjacent SNR conditions rather than measuring the exact 50% Speech Identification Score. Further studies can focus on the clinical validation of the developed Malayalam PB word lists in individuals with varying degrees of hearing loss, individuals with central auditory processing disorders, and older adults with and without hearing loss. Further research can also evaluate the usefulness of the developed lists for hearing aid and cochlear implant outcome measures.

CHAPTER 6

SUMMARY AND CONCLUSION

Speech audiometry plays an important role in assessing speech perception in individuals with hearing impairment (Noor & Arif, 2018). It is essential to use standardized speech material in the native language for reliable evaluation of speech identification abilities (Lehiste & Peterson, 1959; Vainutienė et al., 2024). Although several speech audiometry materials are available in different Indian languages, standardised, psychometrically validated Malayalam phonemically balanced word lists are limited. The existing Malayalam PB word lists are restricted in number and lack detailed psychometric validation (Narne et al., 2024). Therefore, there was a need to develop multiple bi-syllabic Malayalam PB word lists that are familiar, culturally appropriate, acoustically comparable and clinically applicable for speech audiometry evaluation.

The present study aimed to develop and standardize phonemically balanced bisyllabic Malayalam word lists for evaluating speech identification scores in adults with normal hearing sensitivity and to assess the homogeneity of audibility across the developed word lists. The study was carried out in different phases, including the development of phonemically balanced word lists in Malayalam, assessment of homogeneity in audibility, standardisation of the word lists, and measurement of reliability.

Initially, a pool of 1000 bisyllabic Malayalam words was collected from various sources. The collected words were rated for familiarity on a five-point scale by 15 native Malayalam speakers from diverse educational and socioeconomic backgrounds. Internal consistency analysis revealed good agreement among raters (Cronbach's α =0.79). Words with a mean familiarity score of ≥ 4 and a standard deviation of ≤ 1 were retained. Further, the Friedman test revealed no significant difference in familiarity rating across the developed lists (χ^2 (5) = 5.96, p = .31).

The selected words were then subjected to content validation by experts from different fields. The experts further rated the words for familiarity, cultural and emotional appropriateness, linguistic relevance and relevance to speech audiometric evaluation. The domain-wise CVI values were 0.95 for familiarity, 0.94 for cultural and emotional appropriateness, 1.00 for relevance to speech audiometry and 0.94 for linguistic relevance. These findings indicated a high level of expert agreement regarding the suitability and appropriateness of the selected words for speech audiometric assessment.

Following familiarity rating and content validation, six phonemically balanced word lists containing 25 words each were prepared. Phonemic balancing was done based on the frequency of occurrence of Malayalam phonemes reported by Sreedevi and Irfana (2013). The recorded words were then edited and normalised using Adobe Audition software. The word lists were then subjected to acoustic analysis using spectral centroid frequency estimation through a custom MATLAB program. The Friedman test revealed no significant difference in centroid frequencies across the six developed word lists ($\chi2(5) = 6.371$, $p = 0.27$), suggesting comparable spectral energy distribution across the lists.

The word lists were administered to 40 native Malayalam-speaking adults with normal hearing sensitivity to standardise the lists. Speech identification scores obtained in the quiet were analysed across the six developed Malayalam PB word lists. The result showed no statistically significant differences in scores across the six lists in quiet conditions (χ^2 (5) = 2.40, $p = 0.79$). The grand mean SIS across the lists was 99.48%, indicating equivalence in difficulty under quiet conditions.

Speech identification performance was further assessed under different signal-to-noise ratio conditions, including 0 dB, -2 dB, -4 dB, and -6 dB SNR. Significant differences were observed only at 0 dB SNR (χ^2 (5) = 33.07, $p < .001$, Kendall's W = .165) and -4 dB SNR (χ^2 (5) = 13.01, $p = .023$, Kendall's W = .065). However, the observed effect sizes were small, and differences between lists were negligible (approximately ±1 word). No significant differences were observed at -2 dB SNR (χ^2 (5) = 8.81, p = .117) and -6 dB SNR (χ^2 (5) = 7.61, p = .179), indicating overall homogeneity of performance across listening conditions.

SNR-50 values were estimated across the six developed Malayalam PB word lists to reduce floor and ceiling effects during test administration. The findings demonstrated no statistically significant difference in SNR-50 values across the developed lists (χ^2 (5) = 10.42, p = .062). Approximately 50% speech identification was observed at -4 dB SNR across the lists, suggesting comparable speech perception performance in noise.

Further, the six developed word lists were presented to 10 participants across different sensation levels (0 dB, 10 dB, 20 dB, 30 dB, and 40 dB SL) to establish PI-PB functions. Friedman test revealed a statistically significant effect of sensation level on speech identification scores (χ^2 (4) = 40.00, p < .001), with a large effect size (Kendall's W=1.00). The result revealed a gradual increase in speech identification performance with increasing sensation levels, ranging approximately 0% at 0 dB SL to 100% above 30 dB SL. Comparable PI-PB functions were observed across the developed word lists, indicating similar psychometric characteristics and audibility.

The reliability of the developed word lists was assessed using test-retest and inter-judge reliability measures. Test-retest reliability demonstrated excellent temporal stability with an intraclass correlation coefficient (ICC) value ranging from 0.60 to 1.00 and Cronbach's alpha (α) values ranging from 0.75 to 1.00 across listening conditions. Inter-judge reliability showed excellent agreement, with percentage agreement ranging from 80% to 100%

In conclusion, the present study highlighted that the developed Malayalam PB word lists were familiar, culturally appropriate, acoustically comparable, and homogenous in difficulty across the different conditions. Moreover, the developed lists demonstrated good psychometric characteristics and high reliability across repeated administrations. Hence, the developed Malayalam PB word lists can be used for speech identification testing in quiet and noise conditions among native Malayalam-speaking adults.

REFERENCES

Alusi, H. A., Hinchcliffe, R., Ingham, B., Knight, J. J., & North, C. (1974). Arabic speech audiometry. *Audiology*, *13*(3), 212–230. https://doi.org/10.3109/00206097409071679

Ayre, C., & Scally, A. J. (2014). Critical values for Lawshe's content validity ratio: Revisiting the original methods of calculation. *Measurement and Evaluation in Counseling and Development*, *47*(1), 79–86. https://doi.org/10.1177/0748175613513808

Beattie, R. C., Barr, T., & Roup, C. (1997). Normal and hearing-impaired word recognition scores for monosyllabic words in quiet and noise. *British Journal of Audiology*, *31*(3), 153–164. https://doi.org/10.3109/03005364000000018

Billings, C. J., Penman, T. M., Ellis, E. M., Baltzell, L. S., & McMillan, G. P. (2016). Phoneme and word scoring in speech-in-noise audiometry. *American Journal of Audiology*, *25*(1), 75–83. https://doi.org/10.1044/2016_AJA-15-0068

Boothroyd, A. (1968). Developments in speech audiometry. *British Journal of Audiology*, *2*(1), 3–10. https://doi.org/10.3109/00381796809075436

Boothroyd, A. (2008). The Performance/Intensity Function: An Underused Resource. *Ear & Hearing*, *29*(4), 479–491. https://doi.org/10.1097/AUD.0b013e318174f067

Carhart, R. (1951). Basic principles of speech audiometry. *Acta Oto-Laryngologica*, *40*(1–2), 62–71. https://doi.org/10.3109/00016485109138908

Carhart, R., & Jerger, J. F. (1959). Preferred method for clinical determination of pure-tone thresholds. *Journal of Speech and Hearing Disorders*, *24*(4), 330–345. https://doi.org/10.1044/jshd.2404.330

Cheoy, L. P., Chong, F. Y., Mazlan, R., & Lim, H. W. (2021). Development of the Mandarin nonsense word identification test. *International Journal of Audiology*, *60*(8), 578–587. https://doi.org/10.1080/14992027.2020.1864485

Chinnaraj, G., Neelamegarajan, D., & Ravirose, U. (2022). Development, standardization, and validation of bisyllabic phonemically balanced Tamil word test in quiet and noise. *Journal of Hearing Science*, *11*(4), 42–47. https://doi.org/10.17430/JHS.2021.11.4.5

Christensen, L. K. (1995). *performance intensity functions for digitally recorded Spanish speech audiometry* [Unpublished doctoral dissertation]. Brigham Young University.

Davis, H., & Silverman, S. R. (1960). *Hearing and deafness*. Holt, Rinehart and Winston**.**

Dayalan, S. J. (1976). *Development and Standardization of Phonetically Balanced Test Materials in Tamil Language. audiometry* [Unpublished independent project]. All India Institute of Speech and Hearing, Mysuru, 81–82.

De, N. S. (1973). Hindi PB lists for speech audiometry and discrimination test. *Indian journal of otolaryngology,* 25, 67-75. https://doi.org/10.1007/BF02993883

Deekshitha, G., Hathoon, K. I. H., & Mary, L. (2018). A novel two-stage system for spotting fricative and plosive regions from continuous speech. *2018 International Conference on Communication and Signal Processing (ICCSP)*, 0760–0764. https://doi.org/10.1109/ICCSP.2018.8524241

Devi, E. T. (1985). *Development and Standardization of Materials in Manipuri*. [Unpublished Master's dissertation]. University of Mysore, India.

Dias, M. A., Devadas, U., & Rajashekhar, B. (2015). Development of speech audiometry material in the Goan Konkani language.

Dillon, H., & Ching, T. (1995). What makes a good speech test? In: Plant, G., Spens, K.E. (Eds.), *Profound Deafness and Speech Communication* (pp. 308–309). Whurr Publishers.

Dirks, D. D., Kamm, C., Bower, D., & Betsworth, A. (1977). Use of performance-intensity functions for diagnosis. *Journal of Speech and Hearing Disorders*, *42*(3), 408–415. https://doi.org/10.1044/jshd.4203.408

Durankaya, M. S., Şerbetçioğlu, B., Dalkılıç, G., Gürkan, S., & Kırkım, G. (2019). Development of a Turkish monosyllabic word recognition test for adults. *The Journal of International Advanced Otology*, *10*(2), 172–180. https://doi.org/10.5152/iao.2014.118

Egan, J. P. (1948). Articulation testing methods. *The Laryngoscope*, *58*(9), 955-991.

Fry, D. B. (1964). Modifications to speech audiometry. *International Audiology*, *3*(2), 226–236. https://doi.org/10.3109/05384916409074090

Fu, Q. J., Zhu, M., & Wang, X. (2011). Development and validation of the Mandarin speech perception test. *The Journal of the Acoustical Society of America*, *129*(6), EL267–EL273. https://doi.org/10.1121/1.3590739

Gelfand, S. A. (2009). *Essentials of audiology* (3rd ed.). Thieme.

Ghosh, D. (1988*). Development and standardization of speech test materials in Bengali language.* [Unpublished Master's dissertation]. University of Mysore, India.

Gifford, R. H., Shallop, J. K., & Peterson, A. M. (2008). Speech recognition materials and ceiling effects: Considerations for Cochlear Implant Programs. *Audiology and Neurotology*, *13*(3), 193–205. https://doi.org/10.1159/000113510

Gilbert, G. E., & Prion, S. (2016). Making sense of methods and measurement: Lawshe's content validity index. *Clinical Simulation in Nursing*, *12*(12), 530–531. https://doi.org/10.1016/j.ecns.2016.08.002

Greer, L. F. (1997). *Performance intensity functions for digitally recorded Italian speech audiometry materials* [Doctoral dissertation]. Brigham Young University.

Hagerman, B. (1984). Clinical measurements of speech reception threshold in noise. *Scandinavian Audiology*. https://doi.org/10.3109/01050398409076258

Han, D., Wang, S., Zhang, H., Chen, J., Jiang, W., Mannell, R., Newall, P., & Zhang, L. (2009). Development of Mandarin monosyllabic speech test materials in China. *International Journal of Audiology, 48*(5), 300–311. https://doi.org/10.1080/14992020802607456

Harris, R. W., McPherson, D. L., Hanson, C. M., & Eggett, D. L. (2017). Psychometrically equivalent bisyllabic words for speech recognition threshold testing in Vietnamese.

International Journal of Audiology, *56*(8), 525–537. https://doi.org/10.1080/14992027.2017.1303202

Hennig, T. R., Vaucher, A. V. D. A., & Costa, M. J. (2018). Development and validation of lists of disyllabic words for speech audiometry testing. *Audiology - Communication Research*, *23*(0). https://doi.org/10.1590/2317-6431-2017-1915

Hirsh, I. J., Davis, H., Silverman, S.R., Reynolds, E.G., Eldert, E., & Benson, R.W. (1952). Development of materials for speech audiometry. *Journal of speech and hearing disorders*, 17, 321-337.

Hornsby, B. W. Y., & Ricketts, T. A. (2001). The effects of compression ratio, signal-to-noise ratio, and level on speech recognition in normal-hearing listeners. *The Journal of the Acoustical Society of America*, *109*(6), 2964–2973. https://doi.org/10.1121/1.1369105

Hudgins, C. V., Hawkins, J. E., Kaklin, J. E., & Stevens, S. S. (1947). The development of recorded auditory tests for measuring hearing loss for speech. *The Laryngoscope*, *57*(1), 57–89. https://doi.org/10.1288/00005537-194701000-00005

Jerger, J., Speaks, C., & Trammell, J. L. (1968). A new approach to speech audiometry. *Journal of Speech and Hearing Disorders*, *33*(4), 318–328. https://doi.org/10.1044/jshd.3304.318

Ji, F., Xi, X., Chen, A., Ying, J., Wang, Q., & Yang, S.-M. (2011). Development of a Mandarin monosyllable test material with homogenous items (I): Homogeneity selection. *Acta Oto-Laryngologica*, *131*(9), 962–969. https://doi.org/10.3109/00016489.2011.574646

Kacker, S.K.; Basavaraj, V (1990). Indian Speech, Language and Hearing Tests—The ISHA Battery. Indian Speech & Hearing Association, New Delhi, India.

Kalikow, D. N., Stevens, K.N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America*, 61(5), 1337-51.

Kapur, Y. (1971). Speech Audiometry in the Malayalam Language. *Speech Analysis in Indian Languages: Proceedings of a Seminar Held on February 11 and 12, 1971, Department of Health, Education, and Welfare, USA* (p. 66).

Katz, J., et al. (2015). *Handbook of Clinical Audiology* (7th ed.). Wolters Kluwer.

Konadath, S., Tanniru, K., Narne, V. K., Jain, C., Singh, N. K., & Sreenivas, K. J. R. (2013). *Sentence lists in Malayalam and Telugu*. [Unpublished ARF Project]. All India Institute of Speech and Hearing, Mysuru.

Kua, J. M. K., Thiruvaran, T., Nosratighods, M., Ambikairajah, E., & Epps, J. (2010). *Investigation of Spectral Centroid Magnitude and Frequency for Speaker Recognition*.

Kumar, N., Gupta, N., & Kishore, J. (2012). Kuppuswamy′s socioeconomic scale: Updating income ranges for the year 2012. *Indian Journal of Public Health*, *56*(1), 103. https://doi.org/10.4103/0019-557X.96988

Kumar, S. B. R., & Mohanty, P. (2012). Speech recognition performance of adults: A proposal for a battery for Telugu. *Theory and Practice in Language Studies, 2*(2), 193–204. https://doi.org/10.4304/tpls.2.2.193-204

Le Prell, C. G., & Brungart, D. S. (2016). Speech-in-noise tests and supra-threshold auditory evoked potentials as metrics for noise damage and clinical trial outcome measures. *Otology & Neurotology*, *37*(8), e295–e302. https://doi.org/10.1097/MAO.0000000000001069

Leclercq, F., Renard, C., & Vincent, C. (2018). Speech audiometry in noise: Development of the French-language VRB (vocale rapide dans le bruit) test. *European Annals of Otorhinolaryngology, Head and Neck Diseases*, *135*(5), 315–319. https://doi.org/10.1016/j.anorl.2018.07.002

Lehiste, I., & Peterson, G. E. (1959). Linguistic considerations in the study of speech intelligibility. *The Journal of the Acoustical Society of America*, *31*(3), 280–286. https://doi.org/10.1121/1.1907713

Ma, X., McPherson, B., & Ma, L. (2013). Chinese speech audiometry material: Past, present, future. *Hearing Balance and Communication*, *11*(2), 52–63. https://doi.org/10.3109/21695717.2013.794592

Mangum, T. C. (2005). *Performance intensity functions for digitally recorded Japanese speech audiometry materials*. Brigham Young University.

Manjula, P., Antony, J., Kumar, K., & Geetha, C. (2015). Development of phonemically balanced word lists for adults in the Kannada language. *Journal of Hearing Science*, *5*(1), 22–30. https://doi.org/10.17430/893515

Manjula, P., Antony, J., Kumar, K., & Geetha, C. (2018). Assessment of clinical utility of phonemically balanced word lists for adult speakers of Kannada. *Journal of Hearing Science*, *8*(3), 40–45. https://doi.org/10.17430/1002965

Martin, F.N., & Clark, J.G.(2009). *Speech audiometry* (10th ed.). Allyn and Bacon.

Mathew, P. (1996). *Picture test for speech perception in Malayalam* [unpublished Master's dissertation]. University of Mysore.

McArdle, R., & Wilson, R. H. (2008). Predicting Word-Recognition Performance in Noise by Young Listeners with Normal Hearing Using Acoustic, Phonetic, and Lexical Variables. *Journal of the American Academy of Audiology*, *19*(6), 507–518. https://doi.org/10.3766/jaaa.19.6.6

Mendel, L. L., & Danhauer, J. L. (1997). Audiologic evaluation and management and speech perception assessment. *San Diego: Singular Publishing Group, Inc.*

Menon, M. S., & Thangaraj, M. S. (2024). Development of phonemically balanced word list in Tamil for speech audiometry and evaluation of its effectiveness in adults. *Indian Journal of Otolaryngology and Head & Neck Surgery*, *76*(1), 545–551. https://doi.org/10.1007/s12070-023-04209-y

Najem, F. J., & Marie, B. (2021). Phonemically balanced Arabic monosyllabic word lists for speech audiometry testing in Jordan. *Journal of the American Academy of Audiology*, *32*(4), 246–253. https://doi.org/10.1055/s-0041-1722986

Narne, V. K., Mohan, D., Badariya, M., Avileri, S. D., Jain, S., Ravi, S. K., Krishna, Y., Hussain, R. O., & Almudhi, A. (2024). The influence of vowels on the identification of spoken disyllabic words in the Malayalam language for individuals with hearing loss. *Diagnostics*, *14*(23), 2707. https://doi.org/10.3390/diagnostics14232707

Nayana, P. V. (2018). *Lexical neighbourhood test for children in Malayalam (LNT-M)* [Unpublished master’s dissertation]. University of Mysore, Mysuru, India.

Nelson, A. (2004). *Performance intensity functions for digitally recorded French speech audiometry* [Unpublished master's thesis]. University of Colorado, Boulder.

Nielsen, J. B., & Dau, T. (2009). Development of a Danish speech intelligibility test. *International Journal of Audiology*, *48*(10), 729–741. https://doi.org/10.1080/14992020903019312

Niemeyer, W. (1965). Speech audiometry with phonetically balanced sentences. *International Audiology*, *4*(2), 97–101. https://doi.org/10.3109/05384916509074113

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise. *The Journal of the Acoustical Society of America*, *95*(2), 1085–1099. https://doi.org/10.1121/1.408469

Nissen, S. L., Harris, R. W., Channell, R. W., Conklin, B., Kim, M., & Wong, L. (2011). The development of psychometrically equivalent Cantonese speech audiometry materials.

International Journal of Audiology, *50*(3), 191–201. https://doi.org/10.3109/14992027.2010.542491

Noor, H., & Arif, M. (2018). Development and validation of phonetically balanced speech perception test in Urdu language. *Internet Journal of Allied Health Sciences and Practice*. https://doi.org/10.46743/1540-580X/2018.1746

Ostergard, C. (1983). Factors influencing the validity and reliability of speech audiometry. *Seminars in Hearing*, *4*(03), 221–239. https://doi.org/10.1055/s-0028-1091427

Owens, E. (1961). Intelligibility of words varying in familiarity. *Journal of Speech and Hearing Research*, *4*(2), 113–129. https://doi.org/10.1044/jshr.0402.113

Peres, F. F. (2026). Effect sizes for nonparametric tests. *Biochemia Medica*, *36*(1), 5–16. https://doi.org/10.11613/BM.2026.010101

Plomp, R. (1986). A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired. *Journal of Speech, Language, and Hearing Research*, *29*(2), 146–154. https://doi.org/10.1044/jshr.2902.146

Prabhu, P. (2017). *Development of low-frequency range word list in Malayalam*. [Unpublished AIISH Research Fund Project]. All India Institute of Speech and Hearing, Mysuru, India.

Prasad, S. M., Jain, S., & Vipin Ghosh, P. G. (2017). Development and standardization of sentences for speech in noise test in malayalam. *Journal of the All India Institute of Speech & Hearing, 36*.

Rathna Kumar, S. B., Mohanty, P., Ujawane, P. A., & Huzurbazar, Y. R. (2016). Conventional speech identification test in Marathi for adults. *International Journal of Otorhinolaryngology and Head and Neck Surgery*, *2*(4), 205. https://doi.org/10.18203/issn.2454-5929.ijohns20163467

Reynard, P., Lagacé, J., Joly, C.A., Dodelé, L., Veuillet, E., & Thai-Van, H. (2022). Speech-in-noise audiometry in adults: A review of the available tests for French speakers. *Audiology and Neurotology, 27*(3), 185–199. https://doi.org/10.1159/000518968

Sangeetha, S., Deekshitha, Anu, G., & Leena, M. (2019). Broad phoneme classification—a study. *2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*.

Samuel, J. D. (1976). *Development and Standardization of phonetically balanced material in Tamil* [Unpublished Master's dissertation]. University of Mysore, India.

Sandell, G. J. (1995). Roles for spectral centroid and other factors in determining "blended" instrument pairings in orchestration. *Music Perception*, *13*(2), 209–246. https://doi.org/10.2307/40285694

Sargsyan, S., & Rahne, T. (2021). Development of speech material for an armenian speech recognition threshold test. *Russian Open Medical Journal*, *10*(3), e0321. https://doi.org/10.15275/rusomj.2021.0321

Schubert, E., & Wolfe, J. (2006). *Does timbral brightness scale with frequency and spectral centroid? 92*.

Silverman, S. R., & Hirsh, I. J. (1955). Cx problems related to the use of speech in clinical audiometry. *Annals of Otology, Rhinology & Laryngology*, *64*(4), 1234–1244. https://doi.org/10.1177/000348945506400424

Sreedevi, N., & M., I. (2013). Frequency of occurrence of phonemes in Calicut and Eranakulam dialects of Malayalam. *Journal of the All India Institute of Speech & Hearing, 32.*

Sreedhar, J. S., Venkatesh, L., Nagaraja, M. N., & Srinivasan, P. (2005). Development and evaluation of paired words for testing of speech recognition threshold in Telugu: A preliminary report. J*ournal of Indian Speech Language & Hearing Association, 25*(2), 128– 136.

Swarnalatha, C. K., & Rathna, N. (1972). Development and standardization of speech material in English for Indians. *Student Research at AIISH, Mysore*, 65–66.

Tillman, T. W., & Carhart, R. (1963). Effect of competing speech on aided discrimination of monosyllabic words. *The Journal of the Acoustical Society of America*, *35*(11_Supplement), 1900–1900. https://doi.org/10.1121/1.2142749

Tillman, T. W., & Carhart, R. (1966). *An expanded test for speech discrimination utilizing CNC monosyllabic words: Northwestern University Auditory Test No. 6*. https://apps.dtic.mil/sti/html/tr/AD0639638/

Trimmis, N., Papadas, E., Papadas, T., Naxakis, S., Papathanasopoulos, P., & Goumas, P. (2006). Speech audiometry: The development of modern Greek word lists for suprathreshold word recognition testing. *The Journal of International Advanced Otology*, *2*(3), 117–126. https://advancedotology.org/index.php/pub/article/download/131/131

Trimmis, N., Vrettakos, G., Gouma, P., & Papadas, T. (2012). Speech audiometry: Nonsense monosyllabic lists in modern Greek. *Journal of Hearing Science*, *2*(3), 41–49. https://doi.org/10.17430/883512

Ullrich, K., & Grimm, D. (1976). Most comfortable listening level presentation versus maximum discrimination for word discrimination material. *International Journal of Audiology*, *15*(4), 338–347. https://doi.org/10.3109/00206097609071793

Uslar, V. N., Carroll, R., Hanke, M., Hamann, C., Ruigendijk, E., Brand, T., & Kollmeier, B. (2013). Development and evaluation of a linguistically and audiologically controlled sentence intelligibility test. *The Journal of the Acoustical Society of America*, *134*(4), 3039–3056. https://doi.org/10.1121/1.4818760

Vainutienė, V., Ivaška, J., Kardelis, V., Ivaškienė, T., & Lesinskas, E. (2024). Speech audiometry: The development of Lithuanian bisyllabic phonemically balanced word lists for evaluation of speech recognition. *Applied Sciences*, *14*(7), 2897. https://doi.org/10.3390/app14072897

Vaucher, A. V. D. A., Costa, L. D., Moraes, A. B. D., Menegotto, I. H., & Costa, M. J. (2022). Lists of monosyllables for logoaudiometric tests: elaboration, content validation and search for equivalence. *CoDAS*, *34*(3), e20210057. https://doi.org/10.1590/23171782/20212021057

Wang, S., Mannell, R., Newall, P., Zhang, H., & Han, D. (2007). Development and evaluation of Mandarin disyllabic materials for speech audiometry in China. *International Journal of Audiology*, *46*(12), 719–731. https://doi.org/10.1080/14992020701558511

Wilson, R. H., & Margolis, R. H. (1983). Measurements of auditory thresholds for speech stimuli. *Principles of speech audiometry*, 79-126.

Wilson, R. H., McArdle, R., & Roberts, H. (2008). A Comparison of recognition performances in speech-spectrum noise by listeners with normal hearing on PB-50, CID W-22, NU-6, W-1 spondaic words, and monosyllabic digits spoken by the same speaker. *Journal of the American Academy of Audiology*, *19*(6), 496–506. https://doi.org/10.3766/jaaa.19.6.5

Yathiraj, A. & Vijayalakshmi, C. S. (2005). *Phonemically Balanced Word List in Kannada.* Developed in the Department of Audiology, AIISH, Mysore.

APPENDIX: A

LIST 1	LIST 2	LIST 3

ഒപ്പം	/oppam/	നമ്മൾ	/nammaḷ/	വെട്ടം	/ʋeʈʈam/
ചെളി	/tʃeḷi/	കാറ്	/ka:rə/	കാറ്റ്	/ka:t̠t̠ə/
തള്ള്	/t̪aḷḷə/	തേടി	/t̪e:ḍi/	പൊട്ടി	/poʈʈi/
പാട്ട്	/pa:ʈʈə/	ചോറ്	/tʃo:rə/	ചെമ്മീൻ	/tʃemmi:n/
ശാസ്ത്രം	/ʃa:st̪ram/	പീലി	/pi:li/	കോട്ട	/ko:ʈʈa/
കരി	/kaɾi/	കാട്	/ka:ḍə/	ദൂരം	/d̪u:ram/
പടി	/paḍi/	യാത്ര	/ja:t̪ra/	നാളെ	/na:ḷe/
കഞ്ഞി	/kaɲɲi/	സർപ്പം	/sarppam/	സാമ്പാർ	/sa:mpa:r/
ഓട്	/o:ḍə/	നിങ്ങൾ	/niŋŋaḷ/	കുട്ടി	/kuʈʈi/
തൊലി	/t̪oli/	ഗുരു	/guɾu/	പുളി	/puḷi/
വാനം	/ʋa:nam/	തെക്ക്	/t̪ekkə/	വർഗം	/ʋargam/
തേച്ചു	/t̪e:t͡ʃt͡ʃu/	പുഴു	/puɻu/	തൊണ്ട	/t̪oɳḍa/
എന്നാൽ	/enna:l/	വേട്ട	/ʋe:ʈʈa/	കൂജ	/ku:dʒa/
നെല്ല്	/n̪ellə/	ഊഞ്ഞാൽ	/u:ɲɲa:l/	തന്നു	/t̪an̪n̪u/
കുട്ട	/kuʈʈa/	കഷ്ടം	/kəʂʈam/	പിന്നിൽ	/pin̪n̪il/
വാട്ടം	/ʋa:ʈʈam/	പക്ഷി	/pakʂi/	തോണി	/t̪o:ɳi/
കുപ്പി	/kuppi/	ദോശ	/d̪o:ʃa/	കല	/kala/
തീർന്നു	/t̪i:rn̪n̪u/	ഭാരം	/bʱa:ram/	പന്ത്	/pan̪t̪ə/
കറി	/kəri/	ചിട്ട	/tʃiʈʈa/	ചായ	/tʃa:ja/
എണ്ണ	/eɳɳa/	തൊട്ടു	/t̪oʈʈu/	കല്ല്	/kallə/
കായ	/ka:ja/	നെല്ലി	/nelli/	തയ്യൽ	/t̪ajjal/
നൂറ്	/n̪u:rə/	താക്കോൽ	/t̪a:kko:l/	ഏറെ	/e:re/
വഴി	/ʋaɻi/	വനം	/ʋənam/	വിഷു	/ʋiʂu/
കോട്ടം	/ko:ʈʈam/	കവി	/kaʋi/	നാട്	/n̪a:ḍə/
തേങ്ങ	/t̪e:ŋŋa/	നെയ്യ്	/nejjə/	തേക്ക്	/t̪e:kkə/

LIST 4		LIST 5		LIST 6	
നട്ടു	/naʈʈu/	നാണം	/na:ɳam/	നിറം	/niram/
ചോളം	/tʃo:ḷam/	കത്ത്	/kat̪t̪ə/	പൂജ്യം	/pu:dʒjam/
എട്ട്	/eʈʈə/	വാശി	/ʋa:ʃi/	എല്ല്	/ellə/
കമ്മൽ	/kammal/	കനൽ	/kanal/	പെട്ടി	/peʈʈi/
തൂക്കം	/t̪u:kkam/	പുക	/puka/	ഭാഷ	/bʱa:ʂa/
ചൊവ്വ	/tʃoʋʋa/	എപ്പോൾ	/eppo:ḷ/	ചിങ്ങം	/tʃiŋŋam/
എത്ര	/et̪ra/	രേഖ	/ɾe:kʰa/	താങ്കൾ	/t̪a:ŋkəḷ/
നാവ്	/na:ʋə/	കുരു	/kuɾu/	മഞ്ഞൾ	/maɲɲaḷ/
ഞായർ	/ɲa:jar/	ഞങ്ങൾ	/ɲaŋŋaḷ/	തട്ടം	/t̪aʈʈam/

എങ്കിൽ	/eŋkil/	ചൂല്	/tʃuːlə/	ശീലം	/ʃiːlam/
താടി	/t̪aːɖi/	തല്ല്	/t̪allə/	കുളിർ	/kuɭir/
വായു	/ʋaːju/	ഉയിർ	/ujir/	സൂര്യൻ	/suːɾjan/
പൊങ്ങി	/poŋŋi/	വെട്ടി	/ʋeʈʈi/	കെണി	/keɳi/
തെറ്റ്	/t̪et̠t̠ə/	കൊള്ള	/koɭɭa/	ഓട്ട	/oːʈʈa/
കണ്ണീർ	/kaɳɳiːr/	വട്ടം	/ʋaʈʈam/	കാവൽ	/kaːʋal/
സാധു	/saːd̪ʱu/	പാലം	/paːlam/	വേര്	/ʋeːɾə/
മുട്ട	/muʈʈa/	സൂചി	/suːtʃi/	തുക	/t̪uka/
പേര്	/peːɾə/	തോന്നി	/t̪oːnni/	നൂല്	/n̪uːlə/
മേഘം	/meːgʱam/	എത്തി	/et̪t̪i/	വാതിൽ	/ʋaːt̪il/
പൂട്ടി	/puːʈʈi/	പുട്ട്	/puʈʈə/	രീതി	/ɾiːt̪i/
നോക്കി	/noːkki/	നേട്ടം	/neːʈʈam/	കപ്പ്	/kappə/
പശു	/paʃu/	നാല്	/naːlə/	എന്റെ	/ent̠e/
താരൻ	/t̪aːɾan/	താഴെ	/t̪aːɻe/	കട്ടൻ	/kaʈʈan/
പാടി	/paːɖi/	കീരി	/kiːɾi/	പാറ	/paːra/
ജില്ല	/dʒilla/	ദാഹം	/d̪aːɦam/	ചുക്ക്	/tʃukkə/