DEVELOPMENT AND STANDARDIZATION OF SPEECH PERCEPTION TESTS IN ORIYA FOR ADULTS

S M E E T A . B
Register No: 02SH0019

A Dissertation submitted in part fulfillment of Final year M.Sc. (Speech and Hearing), University of Mysore, Mysore

ALL INDIA INSTITUTE OF SPEECH AND HEARING
NAIMISHAM CAMPUS, MANASAGANGOTRI
MYSORE-570 006

MAY - 2004

Certificate

This is to certify that this dissertation entitled **"Development and Standardization of Speech Perception Tests in Oriya for Adults"** is bonafide work in part fulfillment for the degree of **Master of Science (Speech and Hearing)** of the student **(Register No. 02SH0019).**

Prof. M. Jayaram
Director
All India Institute of Speech and Hearing
Mysore

Mysore,
May, 2004

Certificate

This is to certify that this dissertation entitled **"Development and Standardization of Speech Perception Tests in Oriya for Adults"** has been prepared under my supervision and guidance. It is also certified that this dissertation has not been submitted earlier in any other University for the award of any Diploma or Degree.

Guide

Dr. Asha Yathiraj
Reader and HOD,
Department of Audiology
All India Institute of Speech and Hearing
Mysore-570006

Mysore,
May, 2004

DECLARATION

This is to certify that this dissertation entitled **"Development and Standardization of Speech Perception Tests in Oriya for Adults"** is the result of my own study under the guidance of **Dr. Asha Yathiraj,** Reader and Head, Department of Audiology, All India Institute of Speech and Hearing, Mysore, and has not been submitted earlier in any other University for the award of any Diploma or Degree.

Mysore

May, 2004 Register No. 02SH0019

ॐ भूर्भुवः स्वः तत्सवितुर्वरेण्यं भर्गो देवस्य
धीमहि धियो यो नः प्रचोदयात् ।

Dedicated

to

Mummy & Papa

ACKNOWLEDGEMENTS

I would like to express my heartfelt thanks to **Dr. Asha Yathiraj**, Reader and HOD, Department of Audiology, All India Institute of Speech and Hearing, Mysore, for her constant guidance throughout the completion of this project. Ma'am, your perfectionism is admirable!

I thank **Dr. M. Jayaram**, Director, All India Institute of Speech and Hearing for giving me an opportunity to carry out this project.

I thank **Mr. R. K. Mishra**, whose noble guidance paved me in the initial part of the study.

I also thank **Dr. Lancy D'souza** for helping me carry out the statistics.

I immensely thank all my subjects, without whose cooperation my study wouldn't have been possible.

I thank **Mr. Satyanarayan Mohapatra** for the timely help.

Mummy & Papa – words cannot express my feelings for u. Ur affection and support has always helped me overcome the hurdles.

Mini – U've given me the strength to face the challenges in life.

Jitu – Thanx 4 being a dear brother.

Tanu, Richa & Gauri – I thank God 4 our friendship. Thanx 4 being there when I needed u guyz the most.

Ramya – U've been a very dear junior. Be the same always!

Ashish – U mean a lot to me.

Sapna – U've been a constant source of inspiration 4 me.

I would like to thank the **library staff**.

*Thanx to the **Graphics computers** for giving the final touches.*

I would also like to thank all those who have directly or indirectly helped me in completing this project.

*I'm grateful to **God** for giving me the strength and courage in finishing this project.*

TABLE OF CONTENTS

LIST OF TABLES

INTRODUCTION

Speech perception is defined as the process of decoding a message from the stream of sounds coming from the speaker (Borden & Harris, 1980). The study of speech perception is concerned with the listener's ability to perceive the acoustic waveform produced by a speaker as a string of meaningful words and ideas (Goldinger, Pisoni & Luce, 1990). In the past, different terms have been used to refer to measures of speech perception like intelligibility, articulation, discrimination and recognition/identification. The detennination of which term to use depends on the type and purpose of the test being presented. These terms may refer to phoneme, syllable, word, or sentence perception in particular whereas all of these terms refer globally to speech perception. When using each of these terms, it is beneficial to determine whether the task being performed is described accurately.

Assessment of speech perception is basic to almost every aspect of audiology, including its research and theoretical foundations, the fundamental understanding of how the ear functions, and the clinical administration of diagnostic and rehabilitative services for patients with hearing problems. It is impossible to avoid the influence of speech perception in the practice of audiology. There are a variety of purposes for which speech perception test materials have been used in many clinical and research contexts (Mendel and Danhauer, 1997). They are as follows:

> to provide a measure of how well listeners understand speech,

> to reflect the degree of communication handicap created by the hearing loss,

> to provide information for planning and managing auditory (re)habilitation,

> to monitor listeners performance throughout the therapeutic process,

> to assess the success of different types of medical and surgical treatments,

> to monitor subjects' performance in research paradigms,

> to classify the degree and type of hearing loss,

> to be used as a baseline measure for other test procedures, and

> to be used in various forms of research.

A variety of speech perception assessment materials have been developed for a wide array of purposes. Consideration should be made when selecting tests to use as to whether the tests chosen are capable of performing the desired assessments. The speech perception testing has become an integral part of the routine audiological procedure especially the speech audiometry. Two tests, which have gained a lot of importance in the recent years with respect to speech audiometry, are the speech recognition threshold (SRT) and the speech identification (SI) testing.

The SRT serves many clinical purposes. The basic purpose is to quantify the listeners hearing level for speech. It serves as a validity check for the pure tone audiograra SI testing on the other hand makes it possible to evaluate the functional integrity of the auditory system. The poorer the speech identification scores, greater is the involvement of the sensori-neural mechanism. Speech identification scores (SIS) can be used to differentiate cochlear from retrocochlear pathology in addition with other test results (Goetzinger, 1972).

Various kinds of speech stimuli have been used to determine the SRT. They are sentences, connected discourse, spondaic words, spoken digits etc. the kind of stimuli used for speech discrimination testing are monosyllables, nonsense syllables, synthetic speech etc. In case of SRT tests, spondaic words are most widely used test stimuli and monosyllabic words in case of speech identification tests (Carhart, 1971).

NEED FOR THE STUDY:

An individual's perception of speech is influenced by his mother tongue (Singh & Black, 1966). De (1973) found that people consistently produced better and optimum discrimination score in their mother tongue as compared to other languages. Hence, administering the test in a subject's native language is considered to be ideal. A few of the speech recognition threshold tests developed in India were by Rajashekhar (1976), Tanuza (1984), Ghosh (1986) and Mallikarjuna (1990). Among the speech identification tests developed for the adults were by Abrol (1971, cited in Nagaraja, 1990), Kapur (1971, cited in Nagaraja, 1973), Swarnalatha (1972), De (1973), Nagaraja (1973), Mayadevi (1974), Samuel (1976), Tanuza (1984), Ghosh (1986) and Mallikarjuna (1990).

No speech test material for evaluating the speech recognition threshold and speech identification ability is available in Oriya, the language spoken in Orissa. Hence, there is a need to develop and standardize speech material in Oriya for assessing the hearing abilities of subjects who know only Oriya.

AIM OF THE STUDY:

The aim of the study was to carry out the following:

- Construction of a bisyllabic speech recognition test in Oriya, containing two lists
  - Check the equality of the two SRT lists in terms of establishing threshold and their intelligibility
  - Check the effect of presentation level on the intelligibility of the SRT lists

- Construction of a monosyllabic speech identification test in Oriya, containing two lists
  - Check the equality of the two speech identification lists
  - Check the equality of the half lists
  - Check the effect of presentation level on the intelligibity of the lists

REVIEW

Speech is one of the most important vehicles of human communication system. To be able to hear and comprehend normal speech, requires normal auditory integrity. For the purpose of evaluating the auditory integrity, speech audiometry is essential.

The sounds of speech have come to occupy an important place among the auditory stimuli that are used in clinical audiometry. By measuring a client's ability to use his hearing in ways that are close to everyday auditory experience, speech audiometry has not only added a kind of validity to pure-tone audiometry, but also certain speech tests have appeared to have diagnostic and prognostic value as well (Hirsh et al., 1952).

Speech stimuli aid in detecting disturbances which may go unnoticed if pure tones alone are used. Pathologies in the retrocochlear region and higher auditory pathways may not manifest itself in pure tone hearing loss despite significant difficulty in speech discrimination (Goetzinger, 1972). Speech is the preferred test material for assessing higher cortical functions. It is also helpful in assessing success in otological surgery (Kasden, 1970). They are used for hearing aid evaluation. Speech materials also contribute to the assessment of communication ability (Berger, Keating &Rose, 1971).

Assessing the speech perception abilities of an individual constitute an important part of the hearing evaluation. Two such tests, which seem to very sensitive in this regard, are the speech recognition threshold (SRT) and the speech identification (SI) tests.

According to Elkins (1984), sensitive tests of speech perception should have a clearly established purpose; designation of the population of individuals for whom the test in designed; validity (predictive, content, construct, and face); reliability including measurement of the standard error; typical subject variance for the population of interest; and reliable and equivalent alternative test forms and procedure for test administration, scoring, and interpretation.

Since, the early 1900's several tests have been developed to measure the speech recognition threshold (SRT) and speech identification (SI) abilities. These tests vary in the kind of material used and their complexity.

SPEECH RECOGNITION THRESHOLD TESTS

In general most SRT tests utilize bisyllabic or spondaic word stimuli (Mendel & Danhauer, 1997). The spondaic words are presented to determine the lowest level at which subjects can correctly repeat or understand the stimuli (i.e. the threshold of speech). Apart from spondaic words, sentences could be used to establish SRT (Hagerman, 1982). In the past these tests had been termed as speech reception threshold tests. Houghson and Thompson coined the term speech reception threshold (SRT). However ASHA (1988, cited in Martin, 1991) recommended that they should be called as speech recognition threshold tests as a listener is asked to recognize rather than receive the words used in the test. Therefore, the conventional terms speech reception threshold has been replaced by speech recognition threshold (as cited in Martin, 1991). There are several SRT tests that have been constructed over the years. The material and the method of construction varies from test to test. The following review describes the method used for the construction of various SRT tests. The ones developed in India are reported separately.

1. Harvard or Psycho-Acoustic Laboratory (PAL) Auditory Test No. 9

This test was originally developed by Hudgins, Hawkins, Karlin & Stevens (1947, cited in O'Neill & Oyer, 1966). The purpose of the test was to measure the threshold of intelligibility of speech. The criteria followed for constructing the list was

a. the words were familiar to the listeners,

b. the test items were dissimilar in phonetic construction,

c. they had a normal representation of English speech sounds and

d. they had similar audibility values.

It was an open set test using spondaic words as stimulus and consisted of two lists, each with forty-two disyllabic words with equal stress placed upon both syllables in each of the words. Six scrambled versions were made of each of two tests resulting in a total of twelve equivalent lists. The stimuli were recorded on phonograph discs. The carrier phrase used was "Number one, Number two....". In the recordings of the lists the words were attenuated through a range of 24 dB with an attenuation of 4 dB being provided for each group of six words and an interstimulus interval of 6 seconds. A 1000 Hz calibrating tone was recorded on the disc. This test was found to be quite applicable to the testing of hearing in most situations (Hirsh, 1947, cited in O' Neill & Oyer, 1966).

2. Psycho-Acoustic Laboratories Auditory Test No. 14

This test was also developed by Hudgins et al. (1947, cited in O'Neill & Oyer, 1966), which was similar to Auditory Test No. 9. The only difference occurred in the manner of the recording of the test materials. In this test all of the spondee words were recorded at one level. No attenuation of the level of the words occurred.

3. CID Auditory Test W-l

In 1952, Hirsh et al. modified the Psycho-Acoustic Laboratories (PAL) lists without changing the basic concept to improve the clinical applicability of the materials. They developed the Central Institute for the Deaf (CID) Auditory Test W-l, which is an open set test to measure the threshold of intelligibility for speech. From a group of eighty-four spondee words in PAL Test No. 9, only thirty-six were retained. This was accomplished by first eliminating those words judged to be either too easy or two difficult. For the remaining thirty-six spondees, those considered still too difficult were emphasized by adding 2 dB to the signal. Those considered still too easy were reduced in intensity by 2 dB. These corrections permitted all of the spondees to fall within $\pm$ 2 dB of an average intensity reading, adding greater homogeneity of audibility. The thirty-six spondees were scrambled into six lists and identified by alphabet letters A through F. This test consisted of disc recordings of six lists that used different orderings of thirty-six spondaic words. The words were recorded at a constant level but with a carrier phrase 'say the word', preceding the test word. A 1000 Hz tone was provided for calibration purposes and was recorded at the same level as the carrier phrase. The procedure for obtaining SRT was as follows: Before starting the test, the listeners were familiarized with the thirty-six spondees. The input selector was set to tape or disk. The 1000 Hz calibration tone was adjusted. The instructions given were "You are now going to hear some words that contain two syllables. The words will gradually get softer. Before each word you will hear the phrase "say the word", which will be followed by the word you are to repeat. Repeat each of the two-syllable words to me even though you have to guess some of them. Are there any questions?"

The speech reception test was presented at a level 10 dB above subject's average loss for 500, 1000 and 2000 Hz, for the ear being tested. The subject was given a "run" of six correct responses at this level. The intensity level was then dropped 2 dB for every two correct responses out of three presentation.

When the subject reached a point where he gave only one correct response for each three words, the intensity level was increased in 1 dB steps until a threshold point was reached, i.e. three correct responses for six presentations or two correct responses for four presentations. When the actual threshold was reached the judgement was based on four or six words. Subject responses were verbal; they orally repeated the word they thought they had heard. The examiner followed the subject responses on an answer sheet. A similar procedure was followed for the opposite ear using different lists. If there was a difference of 30 dB between the average loss for each of the ears, white-noise masking was introduced into the better ear. The masking level was set at 20 dB above the better-ear average loss. The absolute threshold was obtained at 14.3 dBSL for normal hearing individuals.

4. CID Auditory Test W-2

This test also developed by Hirsh et al. (1952) was designed for rapid estimation of the intelligibility threshold. The recorded lists were the same as those used in Auditory Test W-l. However, the intensity of the words had been attenuated in such a fashion that there was a drop in intensity of 3 dB for every three words. The initial trio of words was recorded at a level that corresponded to the level of a 1000 Hz calibration tone recorded on the disc. The carrier phrase for the W-2 recording was presented at the same level as the first group of three words and remained constant until after the ninth word. Thereafter, the carrier phrase was attenuated at the rate of 3 dB with each new group of three words. Instead of presenting a whole list or a portion of a list at a fixed intensity or several intensities, this test sweeped through an intensity range of 33 dB by attenuating the level of the test words 3 dB every three words. The test continued until the listener had missed five consecutive words. A count was made of the words correctly identified and this number was subtracted from the reading on the hearing level dial for the SRT in decibels. The mean absolute threshold by W-2 list was obtained at 17.7 dBSL.

Along with CID W-l and CID W-2 as being the SRT tests a speech identification test CID W-22 was also developed by the Harvard Psycho Acoustic Laboratory (PAL), which will be discussed later.

5. Speech reception threshold testing using sentence stimuli

To measure the SRT for sentences as accurately as possible, a special test was developed (Plomp & Mimpen, 1979, cited in Plomp, 1986). This open-set test consisted of ten carefully selected lists of thirteen simple meaningful sentences of eight or nine syllables, each which was recorded by a trained female speaker. In the selection procedure only sentences with approximately equal chances of correct recognition in noise were retained. These were divided into ten lists with equal numbers of the various phonemes. The standard, adaptive, procedure for measuring the SRT with these equivalent lists consisted of the following steps:

a) The first sentence of the list was presented repeatedly by increasing the sound level (step size 2 dB) until the listener could reproduce the sentences correctly.

b) The level was decreased by 2 dB and the second sentence was presented.

c) If the listener was able to reproduce the sentence correctly, then the level was decreased by 2 dB and the next sentence was presented.

d) The previous step was repeated for all remaining sentences of the list.

e) Then the average level adjusted after sentences 4 through 13 was taken as the SRT for that particular condition.

This was the simple up-down procedure originally given by Levitt and Rabiner (1967, cited in Plomp, 1986). With a group of 10 normal hearing listeners it was verified that the test-retest reliability with this test procedure was quite good. Including differences among the lists, the standard deviation for SRT in noise was about 1 dB (Plomp & Mimpen, 1979, cited in Plomp, 1986).

6. Speech intelligibility in Noise

The main purpose of this test was to develop a speech material suitable for clinical measurement of speech intelligibility in noise. The test has also been used to establish threshold. Hagerman (1982) developed this sentence test in Swedish in J>r which each word was scored. Ten sentences of five words each were chosen to constitute the original list, which was phonetically balanced. The original list was recorded in a non-reverberant room with a Nagra tape recorder. It was read by a female speaker. This material was recorded digitally and computer edited to obtain new lists with exactly the same content of sound, but with new sentences. A noise was synthesized from the speech material by the computer to produce exactly the same spectrum of speech and noise. The noise was also amplitude modulated by a low frequency noise filtered around 2.1 Hz to make it sound more natural. Thirteen lists of sentences and thirteen replicas of the noise were recorded on each channel of a stereo tape recorder. The pause between the sentences was 7 sec and the total time of each list was 110 sec.

Listening tests were conducted on normal-hearing subjects to study homogeneity of the lists, reliability, learning effects, intelligibility curves, and influence of the speech level with constant S/N ratio.

The material was tested monaurally on twenty normally hearing subjects. The subjects were asked to repeat verbally the perceived words of the sentences. Before doing the testing, the following instructions were given by the experimenter "We want to investigate how difficult it is to perceive speech in background noise. You will now listen to sentences consisting of five words. Depending on the background noise it is sometimes easy and sometimes difficult to perceive the words. Please repeat clearly the words you have perceived. You may guess, but please do not hesitate long with your answer as to miss the word beginning in the next sentence. Don't say anything about the words you cannot perceive".

The experiment was divided into two sessions. The first session was divided into two parts, denoted A and B and the second session was divided in to four parts, namely C, D, E and F. The purpose of the thresholds measured in parts A and F was to estimate the reliability and the learning effect. The thresholds were all measured in a descending way (i.e. 3 dB steps for list A and F and 2 dB steps for list C) and the 50 % level interpolated from two points at each side of this level. Part B was administered to assess the homogeneity of the lists. In parts B and E the S/N ratio was chosen separately for each subject in order to obtain discrimination scores around 50 %. The intelligibility curves in part C were all measured in an ascending way. The corresponding curves without noise, measured in part D, however, were achieved with levels more randomly ordered.

The results showed that the lists were homogenous. Only a small part of the learning effect was due to learning of the word material. Intelligibility curves in the presence of noise and without noise showed maximum steepness of 25 and 10% per dB respectively. At constant (S/N) ratio the best performance was achieved at a speech level of 53 dB.

This material was intended to be used especially for hearing aid evaluation, but also to measure the patient's speech discrimination in noise with earphones during hearing aid fitting. This may be a way to predict the possible benefit of a hearing aid.

7. Hearing in Noise Test (HINT)

A recorded test for measuring sentence intelligibility in quiet and in noise was developed by Nilsson, Soli and Sullivan (1994) called the Hearing in Noise Test (HINT). It was composed of twenty five phonemically balanced lists of ten sentences that had been normed form naturalness, difficulty and reliability. These sentences had been taken from Bamford-Kowal-Bench (BKB) sentences (Bench & Bamford, 1979). The 336 BKB sentences were revised to equate their lengths. Then ten native speakers of American English evaluated the sentences for naturalness on a seven-point scale (7 = natural, 1 = artificial). All revised sentences were rated again by another set of six subjects. The revised materials were recorded by a male professional voice actor at a signal level of 70 dBSPL. The sentences were sampled and later edited. Only nine sentences were scaled up. After equating the waveform levels, the average long-term spectrum of the sentences was computed. This served as the target for generation of a specially matched masker to be used in measurements of sentence SRTs (sSRTs) in noise.

Equating of sentence difficulty was done by computing the mean-squared (MS) amplitudes for each sentence waveform. The MS level of the sentences was increased if intelligibility was below average, decreased if intelligibility was above average, or left unchanged if intelligibility was approximately at the average. Seventy-eight native English speaking male and female subjects in the age range from 17 to 45 years and having normal hearing were chosen for the study. Groups of six to eight listeners were tested at a fixed signal-to-noise (S/N) ratio. All sentence tests were presented diotically under TDH-50 headphones. Noise was mixed with speech using a Grason Stadler GSI-16 audiometer. The subjects were instructed to listen and repeat aloud whatever was heard or understood. Scoring was done on a word-by-word basis. Average percent correct scores for each sentence was calculated from the number of words repeated correctly.

After matching the sentences for intelligibility, lists of sentences were formed for use in the measurement of sSRTs. The sentences were rewritten in the International Phonetic Alphabet (IPA). Twenty-one lists of twelve sentences which matched the phonemic distribution of the entire sentence set were formed. For each phoneme in each sentence list, the difference between the target. Phoneme count and the obtained phoneme count were tabulated. Finally only 10 sentence lists were formed.

The next step in the development of the SRT test was determination of inter-list reliability. Eighteen male and female native speakers of English having normal hearing and in the age range of 18 to 43 years were selected. A PC-based, digital processor (Ariel DSP-16) with two dual-cascaded programmable low-pass filters and two programmable attenuators presented the materials. The two channels of output were filtered attenuated, and then mixed. The sSRTs were measured in four alternating blocks with and without noise, with each block containing five threshold measurements (i.e. five lists). The sentence order was randomized by the computer. The procedure involved was as follows: The listeners were asked to listen to each sentence and repeat aloud whatever was heard or understood. An adaptive up-down strategy determined the sentence presentation levels; the first sentence was presented below threshold and was increased in level by 2 dB steps until it was repeated correctly. The subsequent sentences were presented once each, with the presentation level dependent upon the accuracy of the preceding response. Presentation levels were attenuated by 2 dB after a correct response and increased by 2 dB after an incorrect response. A comparison was done by the experimenter to judge accuracy. The results indicated that the average sSRTs in quiet were 23.91 dB(A). Average sSRTs in 72 dB(A) noise were 69.08 dB(A).

8. Sentence SRT Test in Dutch

A method for the development of speech materials for efficient measurement of the speech reception threshold was given by Versfeld, Daalder, Festen and Houtgast (2000). The speech material used was sentences. The sentence texts were selected from large databases (automated selection and manual selection), whereafter recordings of these materials were made, using two male and two female speakers. Next, a subset of the new materials was sampled and evaluated in a listening experiment, where the speech reception threshold in noise was measured for a subset of sentences with twelve normal-hearing subjects in the age range from 20 to 53 years. For this purposes a homogenous set of sentences that were equally intelligible in a variety of listening conditions was created. Sixteen lists and one practice list were created by random sampling of sentences. Each list comprised by twenty-five sentences. Within each list, the sentence order was kept fixed. The stimuli was converted to analog signals and low pass filtered at 20 kHz. Noise with the long term average spectrum of speech (LTASS) of the respective speaker served as the master. The noise level was kept fixed at 70 dBA, and the S/N ratio was varied by changing the level of the sentences.

The subjects' best ear was subjected to testing. The subject was informed about his task by a set of written instructions. The testing was carried out in a sound insulated room. The subjects received the sentences monaurally to the test ear over headphones and were instructed to repeat it as accurately as possible. In a list, the S/N ratio was varied according to the adapted procedure described by Plomp and Mimpen (1979). The first sentence of each list was presented at a S/N ratio of 8 dB and was repeatedly presented, each time at a 4 dB higher S/N ratio, until it was correctly reproduced. A response was considered only if the entire sentence was reproduced without mistakes. The subsequent sentences of the list were presented only once. Based on the subject's response, the S/N ratio of the successive sentence was either increased or decreased by 2 dB.

The SRTs for the four speakers ranged room -2.6 to 1.1 dB. SRTs of speakers differed significantly, and the aim of producing equivalent lists across speaker was not met. So a new method was described to form subsets for efficient SRT measurements. Here, no adaptive procedure was used and S/N ratio was kept fixed at either -1 or -4 dB. The new set was evaluated by determining the intelligibility of speech in a background of LTASS noise. Forty-eight subjects with normal hearing were evaluated on all 1272 sentences. These sentences were uttered both by a female speaker and male speaker. The speech was masked by noise with the LTASS of the speaker. The level of the noise was kept fixed at 70 dBA.

The testing was carried out in a sound proof booth. A set of written instructions were given and the subjects' task was to listen carefully to the stimuli and to type it on a computer keyboard. A sentence was presented only once. Subjects could edit the typed text as often as desired, and when they were satisfied with it, they pressed "Enter". If the response was correct, the next sentence was presented automatically after a delay of a few seconds. If the response was incorrect, the correct sentence was displayed, and the next sentence was presented only after the subject pressed the

"Enter" key again. Prior to the actual session, a practice list was given consisting of ten sentences which was uttered by each speaker.

The results indicated that at a S/N ratio of -1 dB, 76% and 71% of the responses were correct for the first and second speaker respectively. At -4 dB, the values were 43% and 37% respectively. Hence, for the first speaker, the SRT was -3.4 dB with a slope of 11.7% dB whereas for the second speaker this was -2.9 dB with a slope of 11.8% dB. In comparison with the previous experiments, SRTs were more than 0.5 dB better, but slopes were slightly less sleep. The difference may be due to the experimental method. The SRTs for the individual subjects ranged from - 1.3 to - 5.1 dB with an SD of 0.8 dB.

Two sets each comprising of thirty-nine lists of thirteen sentences were created. A sentence occurring in one set did not occur in the other set. Thus, in total 1014 new sentences were generated that were equally intelligible when presented in speech shaped noise at an equal rms level.

This section of the review was focussed on the tests which measure the SRT. Though the tests differ in terms of the speech stimuli and the method to establish SRT, the overall applicability remains the same. While most tests make use of spondaic words, the Speech Intelligibility in Noise Test, HINT and Sentence SRT Test used sentences.

SPEECH RECOGNITION THRESHOLD TESTS DEVELOPED IN INDIA

1. SRT Test material in English for Indians

Swarnalatha (1972) attempted to standardize speech material in English for application on the Indian population. In the development of the SRT test for adults, eighty-four words from PAL Auditory Test No. 9 and No. 14 were used. Fifty-seven words from the children's spondee list were used, for children. Two hundred adults in the age range of 16 to 25 years and 200 children aged 7 to 15 years rated the spondaic words. The subjects were asked to rate the words as 'familiar', 'not so familiar' and 'not familiar'. The final spondee lists for adults and children had fifty and twenty-five words respectively. The lists were recorded in a sound treated room. All the test items were recorded preceded by a carrier phase 'say the word'. A time interval of 5 sec was allowed for the subject to respond. A calibrated diagnostic audiometer (Madsen model OB 70) and type recorder was used for the study.

For standardization of SRT test the same subjects were used. The instruction given was as follows: "You will hear a list of words through your ear phones. These will consist of two syllable words such as foot-ball, playground. You must repeat into the microphone any word just as you hear it. If you are not sure, guess. Do you have any questions?"

The above instructions were common to adults and children. For testing the adults, each subject was tested only at one intensity, once in the right ear using the list I and then in the left ear using the list II, at a different intensity. The intensities at which the lists were presented varied from 0 to 35 dB at intervals of 5 dB (i.e., 0, 5, 10, 15, 20, 25, 30 and 35 dB) above the pure-tone average of each subject.

In case of children, as there was only one twenty-five item spondee word list, the test was administered to the right ear only. Otherwise the procedure was same as that for adults.

The subjects were instructed to respond orally. The tester bias in scoring had not been controlled. Test-retest reliability was determined by administering the list to randomly selected subjects from the same group. Adults and children obtained an SRT of 9 dB (i.e., PTA 10dB). A speech discrimination test was also developed which is discussed later.

2. SRT Test for adults and children in Kannada

This speech test was given by Rajashekhar (1976). It dealt with two aspects

a) Procedure for obtaining test materials and

b) Procedure for obtaining SRT

The test materials were obtained from the Central Institute of Indian Languages (CIIL). About 104, bisyllabic or polysyllabic words were selected which could be picturable. A familiarity check was done using a three-point rating scale i.e., most familiar, familiar and not familiar. The subjects chosen for the familiarity test were 100 graduate and postgraduate students in the age range of 17-25 years and twenty-five children in the age range of 3-5 years. The mother tongue of all the subjects was Kannada.

Out of 104 words, sixty-four words were subjected for intelligibility. These formed four lists each of sixteen words. The lists were presented at -6, -4, -2, 0, +2, +4 and +6 dB. Words intelligible at lower levels and those that were difficult at higher levels were discarded and the equally intelligible words (at -2 to +2 dBHL) were selected for formation of final test materials. Finally forty words were randomly sorted into two lists.

Pictures were used to test familiarity with children. Words which were not familiar to any of the children were discarded and only those that were familiar were selected. These were subjected to intelligibility tests and the equally intelligible words were chosen. The appropriate pictures were drawn and the ambiguous ones were redrawn.

The recordings were done on a Madsen tape recorder at 0 VU meter deflection for the last word of standard carrier phrase. It was recorded by a nineteen-year-old male whose mother tongue was Kannada. The interstimulus interval was 5 sec. The recordings were done in a sound treated two-room situation.

For standardization of the speech tests with adults, fifty-five subjects in the age range of 16-25 years were selected. Thirty school going children in the age range of 5-10 years were chosen for standardization in children. The instruction used was "You will hear a list of words through your earphone. These will be preceded by the carrier phrase 'idannu heli' (adults) or 'idannu torisu' (children). You will have to repeat the words or pick up the appropriate picture. If you are not sure, guess. Do you have any questions?"

Two lists of twenty words were chosen for the adults and one list of fifteen words for the children. In case of children, fifteen pictures were randomly divided on individual sheets along with two more pictures that were not included in the test.

Hence, each page contained five pictures (three stimulus words and two others) the presentation level varied from 0 to 25 dBHL at intervals of 2 dB.

The number of correct responses were noted down and converted into percentage of correct response at each intensity level. The results showed that for adults the mean SRT was 19 dBSPL for list I and 19.5 dBSPL for list II. The mean SRT for children was 21 dBSPL.

3. SRT Test in Manipuri

Tanuza (1984) aimed at construction and standardization of SRT test material in Manipuri language. To construct the test, polysyllabic words were collected from phonetic books, magazines, books and normal conversational speech. A familiarity test was done using a three-point rating scale (not familiar, familiar and most familiar). Words rated as most familiar were collected and from them eighty polysyllabic words were chosen for the construction. Four lists were developed. The polysyllabic word lists contained twenty items each.

All the test materials were tape recorded through a Philips cassette tape recorder which was done by an adult female talker whose mother tongue was Manipuri. The items were spoken with an interstimulus interval of 5 sec. English spondees of Swarnalatha (1972) were also recorded in a similar way by the same talker. A carrier phrase 'Say the word' was used for the spondees.

Five Manipuri speaking graduate students who had normal hearing (3 males and 2 females) served as subjects. The instrument used was a 2 channel diagnostic audiometer Beltone 200 C and a cassette deck. The recorded words were played by the tape recorder and fed to the tape input of the audiometer, which in turn was fed to the earphone (TDH 39) coupled with MX-41/AR ear cushion. The testing was carried out in a sound treated two-room situation. Pure tone thresholds at 500 Hz, 1 kHz and 2 kHz for each subject was found using an up 5 down 10 method of threshold measurement.

For the standardization of the SRT test material, four lists of twenty-five items each were used. The presentation level was (0, 5, 10, 15, 20 and 25) dBHL w.r.t. 0 dB SRT. Each list was randomized into six lists and each randomized list was presented at only one intensity level. A time gap of 5 sec was given to respond. The lowest level at which the subjects repeated correctly 50% of the test items was considered as the SRT level. The results indicated that the mean SRT was obtained at 13 dB ref.O dBHL. A speech identification test was also developed by Tanuza (1984), which is described later.

4. SRT Test in Bengali

This test was given by Ghosh (1986). The objective of the study was to construct and standardize speech materials in Bengali language to facilitate the speech audiometry procedure. The polysyllabic words were collected from phonetic books, periodicals, journals and spontaneous speech. A familiarity test was done using a three-point rating scale (highly familiar, familiar and unfamiliar). Sixty polysyllabic words were chosen for assessing the SRT. The recording was done by using a Philips deck cassette tape recorder in a sound treated room. An adult male talker whose mother tongue was Bengali did the recording. All the test items were recorded preceded by a carrier phrase. The interstimulus interval was 5 seconds. A two channel diagnostic audiometer GSI-16 was used. A cassette deck (Philips) was used to feed the speech material. The recorded words were played by tape recorder and fed to the tape input of the audiometer, which in turn was fed to earphones (TDH-39) coupled with MX-41/AR ear cushion.

Six subjects whose mother tongue was Bengali were chosen. The subjects were in the age range of 18-25 years. All the subjects were subjected to routine audio logical testing, which included pure tone testing at 500, 1000 and 2000 Hz using a 10 down and 5 up method of threshold measurement.

For the standardization of the SRT test material, the polysyllables were divided into three lists (consisting of twenty words) and each list was randomized into six lists to overcome practice effect. The level of presentation for the lists were (0, 5, 10, 15, 20 and 25 dBHL ref. 0 dBHL. The subjects were asked to repeat the spondees and the responses were converted into percentages. The level at which the subject repeated correctly 50% of the test items was taken as the SRT level. The mean SRT was obtained at 12 dBHL ref. OdBHL. Average PTA was attained at 10.16 dB. The difference between the PTA and SRT for polysyllabic word list was 1.84 dB, which showed that all the three lists yielded almost equivalent scores at different hearing levels. Apart from the SRT testing, a SI test was also developed using monosyllabic words, which is discussed later.

5. ***SRT Test in Gujrati***

Mallikarjuna (1990) developed a list of sixty spondee words in Gujrati language, for which the articulation as a function of intensity was determined. A positive correlation of 0.73 and significant at 0.01 level was established with spondee list in English for Indian population (as standardized by Swarnalatha, 1972). The list was found to be valid test and could be used to establish the SRT in Gujrati language.

The studies done in India, pertaining to the SRT, do not differ much from the western studies but these tests have been modified with reference to the Indian context. Except for the test developed by Swarnalatha (1972) and Rajashekhar (1976), the number of subjects on whom the tests were presented was way too small. Hence, they cannot be considered as standardized tests.

SPEECH IDENTIFICATION TESTS

In the past, many authors have used the term "discrimination" for "identification". However, discrimination refers to differentiation among stimuli as "same" or "different". Hence, speech discrimination is considered an inappropriate term (Olsen & Matkin, 1979). Instead the term identification is currently used because identification/recognition refers to the ability to repeat the stimulus item (Mendel & Danhauer, 1997). Different speech materials are used to assess the identification ability. This includes monosyllabic words, phonemes and nonsense words. The following section gives some of the speech identification tests, reported in literature. The ones developed in India are given separately.

I. Fletcher and Steinberg Nonsense Syllable Tests

Fletcher and Steinberg (1929, cited in Mendel & Danhauer, 1997) gave one of the earliest tests to measure speech identification. It determined the most appropriate ways for ascertaining the recognizability of speech sounds and to assess the transmission characteristics of communication systems. Constant-vowel-consonant (CVC) monosyllabic nonsense syllables were chosen out of which ten lists of ninety syllables were framed. Twenty-two introductory sentences were used to make the test more like connected speech. This open-set format test was presented through monitored live voice and twenty-two different introductory phrases were used as carrier phrases.

2. Word tests for Deaf People

Fry and Kerridge (1939, cited in Olsen & Matkin, 1979) described five lists of twenty-five CVC words each, called Word Tests for Deaf People. Each list contained all but four sounds of English spoken language. The list of words were to be spoken by a friend or relative in a quiet room at an agreed upon distance. The listener closed his/her eyes and responded to each word. Four points were given for each word, one per sound or phoneme, and one point was allowed if a response was made but was in error for all the sounds of the spoken item. Fry and Kerridge advised that any person whose score was less than 35% should be tested with a list from their sentence test to allow the listener the advantage of context.

3. Sentence Tests for Deaf People

Fry and Kerridge (1939, cited in Olsen & Matkin, 1979) prepared sentence tests for patients who obtained scores of 35% on less on their word lists. These

"Sentences Tests for Deaf People" consisted of five lists of "short common place" sentences. There were twenty-five sentences per list with four to seven words per sentence. A correct sentence received a score of four points; one point was deducted for each error, but inaccuracies for "the" and "a" were ignored.

4. PAL Auditory Test No. 8

This test was developed by Hudgins et al. (1947, cited in Mendel & Danhauer, 1997) to determine listener's ability to hear simple sentences in the presence of interfering noise. It used a multiple-choice format. The stimulus consisted of 100 sentences with multiple-choice responses, which were recorded on a cassette tape.

5. *PAL PB - 50 word lists*

This speech test was originally designed to assess the efficiency of communication systems. It was for this purpose that, during the Second World War, Egan (1948, cited in Goetzinger, 1972) developed the PB-50 monosyllables. The Harvard phonetically balanced word tests were developed from a pool of 1200 monosyllabic words. The general criteria for the selection of the words were that they should be familiar but not easy. The phonetic balance of the word lists was such that each list contained the elements of the English language approximately in the same proportion as they occurred in the English language. Twenty lists of fifty words each were eventually compiled with reference to the following criteria:

- equal average difficulty,
- equal range of difficulty,
- equal phonetic composition,
- representative of English speech and
- words in common usage.

Later, eight of the lists (5, 6, 7, 8, 9, 10, 11 and 12) were recorded at Central Institute for the Deaf by Rush Hughes, a professional announcer in the St. Louis area. The articulation function was derived from research using the Rush Hughes recordings of the PAL PB tests. Eldert and Davis (1951, cited in Goetzinger, 1972) noted as a result of their investigation that the eight-recorded Rush Hughes tests were not equal in difficulty. They found tests 5, 6, 7, 9 and 10 to be equivalent. Test 8 was the most difficult and tests 11 and 12 the easiest. On the basis of their research it was recommended that 2% be added to the PB max score for test 8 and that 2% be subtracted from tests 11 and 12 when used with conductive and mixed hearing loss cases. Davis (1948, cited in Goetzinger, 1972) also recommended that the tests be administered routinely at 110 dBSPL for hearing losses of 55 dB or less. He indicated that 120 dBSPL should be used for hearing losses greater than 60 dB, unless this level caused discomfort. Carhart (1951-1952, cited in Goetzinger, 1972) recommended that the Rush Hughes test be administered at 35 dB level.

6. CID W-22 Word Lists

Hirsh et al. (1952) constructed PB word lists for use with adults. These lists were constructed with the aim of increasing average familiarity of words, in comparison with the Harvard PB lists. This was done with the aim of making the test suitable even for subjects with minimum education.

The words for the CID W-22 lists were chosen from Thorndike's tabulation of 20,000 familiar words. The words chosen were monosyllabic, like those in the Harvard PB-50 lists. The words so chosen were grouped into four lists of fifty words each. All the lists were phonetically balanced. The word order in the lists was randomized to get six scramblings to each list. Ira Hirsh was the talker.

The test was standardized on a small group of fifteen subjects. It appeared that the speech material was presented directly through an amplifier and a loudspeaker. The subjects were divided into three groups of five each. The first group of subjects listened to all the twenty-four lists at 100 dB (re 0.0002 microbar). Then they heard each of these lists at levels 10 dB apart ranging from 20 to 70 dB. Lists and levels were randomized. The only constraint applied while randomizing was that, no list would be heard at the same level by any subject. This did not rule out the possibility that the subject might have heard the same list at different levels consecutive which could have led to practice effect.

The second group of subjects listened to the lists under the same condition as did group I. In addition, they heard the lists at one more level viz., 15 dB. The third group heard each word order at 50, 40, 30, 20 and 15 dB (re 0.0002 microbar). 100% correct responses were obtained at 50 dB itself and therefore the levels 60 and 70 dB were not used.

Hirsh et al. (1952) had not reported the mean scores obtained at the different levels for the four lists. The scores obtained could be derived only from the articulation curve reported by them. To demonstrate the significance of difference among lists and among levels, no statistical procedure was employed. The conclusions were based only on the articulation function. In addition, the equivalence of the scramblings of each list had not been statistically validated.

L

The CID W-22 lists were found to be easier than the Harvard PB lists. The former gave high scores at a sensation level of 25 dB. To obtain the same scores, the Harvard lists had to be presented at about 40 dBSL (Carhart, 1965). This difference had been attributed to the greater familiarity of the words and speaker intelligibility (Owens, 1961, and Goetzinger, 1972).

However, the utility of W-22 lists has been questioned. In fact Hirsh et al. (1952) themselves pointed out that the preliminary experiments with the lists indicated, " .... W-22 did not satisfactorily separate patients with mixed deafness from patients with pure conductive deafness. The older recordings of Egan lists were effective in this respect" (pg. 335).

Berger (1978) opined that the W-22 lists had very few words, which were of sufficient difficulty for any listener, except to those with very poor discrimination. Similarly Goetzinger (1972) stated, "the W-22 words although highly familiar were too easy for fine differential diagnosis" (pg 167).

The dissatisfaction with the CID W-22 lists, stemmed from the fact that they were inadequate as a diagnostic tool in the cases of mild losses and progressive losses. Linden (1965, cited in Geffner & Danovan, 1974) reported that in a case with slowly progressive hearing loss, with normal hearing at frequencies below 2000 Hz, discrimination score obtained using W-22 were not affected. In subjects with a loss of about 60 and 80 dBHL at 2000 Hz and above, the intelligibility functions appeared similar to those in ears with mild loss. In ears with normal healing or mild hearing loss, one would obtain maximum scores at 16 dB (re SRT) using W-22 lists (Geffner & Danovan, 1974). W-22 lists were thus inadequate in uncovering the discrimination problems, if they were severe. Thus, Geffner & Danovan (1974) concluded that there was a need for a more sensitive discrimination test.

While the CID W-22 lists were not useful in cases of mild hearing loss (Geffner & Danovan, 1974), they were also found to be inefficient in uncovering the discrimination difficulties presented by cases with retrocochlear lesion (Johnson, 1966). Twenty-five out of 163 patients were observed to have good discrimination despite a retrocochlear lesion. Ten out of these twenty-five subjects could score 90% and above on the W-22 lists.

Based on the above reports, it was concluded that there was a need for a more difficult test than the CID W-22 test.

7. CID Everyday Sentences

The sentences most frequently used to assess speech intelligibility were commonly referred to in the USA as the 'CID everyday sentences' or the 'CHABA sentences' (Silverman & Hirsh, 1955, cited in Kruger & Mazor, 1987). They constructed 100 sentences of two to twelve words in length to represent 'everyday American speech". Interrogative, declarative and imperative sentences were used. Ten sentences were incorporated in each list of the ten tests with fifty key words being considered as the test items in each test. The materials were not recorded originally but later recorded by Auditec of St. Louis by a male talker with general American English Dialect. They were particularly useful for evaluation of the patient with severe recognition problems, such as the geriatric patient. The test materials has been used for auditory, visual and auditory-visual speech perception testing. Hinkle and Binnie (1979, cited in Martin, 1990) in a study of test equivalence of ten lists of the CID Sentences, found that although 'several groups of sentence tests were not statistically different on the basis of mean data .... standard deviations and correlations were not equal (and) equivalency was rejected'.

8. Multiple- Choice Intelligibility Test

This test was developed by Black (1957). The purpose of developing this test was to determine a speakers ability to be heard correctly (speaker rating was determined by the number of times the words spoken were recorded correctly by listeners); to determine a speakers intelligibility under quiet and noise conditions; to measure listener's efficiency as a reflection of listener's performance in communication situations. This closed set multiple-choice format test consisted of twenty-four multiple-choice lists each consisting of twenty-four test items. They were forms A and B (Black & Haagen, 1963) and forms C and D (Black, 1957). They were constructed by:

> selecting materials from a master population of words;

> collecting the error responses for the words through written down tests;

> assembling of trial form of multiple choice list and answer form and

> assembling of test forms C and D and A and B.

All words with Thorndike ratings one to ten were screened to remove proper nouns, homonyms and homographs. The remaining words were typed on cards to form certain lists, which was recorded by ten speakers on a Presto disk recording system. These disks were played back to panels of eight to twelve listeners each to determine the preliminary scores in quiet and noise conditions. 750 words of one or two syllables each were selected for trial forms of the multiple-choice test (twenty-five lists of thirty words). The criteria for the selection of the items was:

> a word must have fallen within the range 15-85 percent intelligible in both quiet and noise on the basis of the screening process, and

> a word must have been responded to erroneously with at least three incorrect words (the same ones in quiet and noise) a minimum of ten times in noise and five times in quiet.

The twenty-five trial lists were equated with respect to the mean level of difficulty of the items, administered in noise to twenty-four panels and subjected to item analysis. Continuing Haagen's forms A and B (forms of twelve lists each), twenty-four new lists of twenty-seven items each were constructed on the basis of the multiple-choice values to 15-85 percent intelligible, matched in mean score and item variance within each list, and assigned to one of the two forms, designated C and D.

The speaker list was comprised of sequences of three items to be read, 'Number 1 (word word word); Number 2 (word word word) ... Number 9 (word word word), with not pause between the carrier phrase and the items or within a group of three items. A group of three test items, exclusive of the carrier phrase, usually contained four or five syllables.

The results showed that the mean intelligibility scores for the different forms were: Form C, 68.0 % in noise; Form D, 71.5%; Form A, 75.8%; and phonetically balanced (PB) 61.8%. The mean for form C, in 'quiet' was 72.7%. No sequential effect was noticed. No significant difference in the intelligibility of the speakers was seen under test-retest situation. The characteristics of the tests were as listed below:

> Scoring error was less than one percent by an experienced scorer,

> Specified responses were possible, which reduced the importance of linguistic sophistication among listeners in testing the intelligibility. It also made

possible the study of confusion characteristics among fixed population of word,

> It changed the relative intelligibility value of words that it had a write down item,

> It limited the range of scores and the useful range was 25-100 percent,

> The time spent in administering was shorter and the scoring was easier and accurate.

The limitation of this test lay in the rigidity of the answer forms. The item could not be scrambled from one experimental session to another and presumably could not be used repeatedly with the same panel of listeners.

9. Rhyme test

This test was developed by Fairbanks (1958, cited in Olsen & Matkin, 1979) on the basis of the multiple-choice test. It was designed to emphasize the auditory phonemic factors and to minimize the linguistic factors. Five lists of rhyming monosyllables with fifty items per list were used in the Fairbanks rhyme test. Only the initial consonant differed in a set of five rhyming words, but the initial consonants were not given on the response sheet. Hence, the response task of the listener was to write the initial consonant for the stem provided in the answer sheets. This test was not a classic closed-response test. However, it was not an open-set either, because the listener was required to recognize the initial consonant of each test item based on perception of the initial consonant and vowel transition. Thus, the Fairbanks rhyme test was more correctly labeled as a phoneme recognition, actually a consonant recognition test.

10. Constant-Nucleus-Consonant (CNC) word lists

Lehiste and Peterson (1959, cited in Goetzmger, 1972), cognizant of the inadequacies in the phonetic balance of the Harvard PB lists, developed a new monosyllabic word test. In the construction of the test 1263 monosyllabic words of the CNC type were selected from the Thorndike list of one million words in which these words had occurred at least once. The frequency of occurrences of each initial consonant, vowel nucleus and final consonant was determined, and incorporated into the construction of 10 lists of 50 words each. The advantages of their lists were:

> Only CNC words were used,

> Each list matched the phonetic balance of the parent list of 1263 words rather than English generally and

> The lists consisted of familiar words.

Carhart in 1965 noted that the Lehiste-Peterson lists and the Havard PB lists gave comparable results although they differed in terms of the criterion used for achieving phonetic balance and in terms of the composition of words. In addition, Harvard PB lists consisted of constant vowel (CV) as well as vowel conconant (VC) type of combinations, while the Lehiste and Peterson lists contained only CNC combinations.

11. PAL S-I Sentence test

The PALS-I sentence test developed by Davis and Silverman (1960, cited in Brandy, 2002) was designed as a supra threshold measure of speech intelligibility (recognition). There were two equivalent lists of twenty sentences, and in each sentence there were five key words (four monosyllabic and one bisyllabic). For this open-response test, the listener wrote down or repeated each sentence. Scoring was based on recognition of the five key words in each sentence. Clinical experience suggested that scores on this test were somewhat higher than scores on the W-22 and NU-6 tests with talker and presentation level held constant. This was probably due to increased redundancy for sentence material.

12. Revised CNC lists

The original CNC lists were revised by Peterson and Lehiste (1962), to give more uniform distribution of word familiarity by eliminating literary and rare words and proper names while maintaining phonemic balance. The revised lists contained 192 items which occurred less than five times in one million words. 262 of the total items were the same as those found in the original PB-50 lists developed by Egan, 1948. Elkins (1970, cited in Causey, Hood, Hermanson and Bowling, 1984) analyzed the phonetic balance and word familiarity of these lists. The results indicated generally good phonetic balance to the English language and uniformity of word familiarity.

13. CAL-PBM Lists

The Commonwealth Acoustic Laboratories (CAL) adopted the use of twenty-five word lists as an alternative to the PAL PB lists to estimate PB max. These lists were based on the British lists and were called the Phonetically Balanced Monosyllable (PBM) lists to be used in Australia, which was developed by Macrae, Woodroffe and Farrant (1962, cited in Dermody & Mackie, 1987). To compensate for the reduced reliability of the twenty-five word lists, it was recommended that the lists be administered using three estimates of the Performance-Intensity (P-I) curve. The initial presentation level was at 10 dB above the most comfortable listening level for speech. Another list was then presented a the most comfortable listening level while a third list was presented at 10 dB below the most comfortable level or at about the 50% recognition point.

In 1969, the CAL audiology manual changed the recommended adult speech test protocol, abandoning the PI function and recommended the use of average three frequency hearing loss for predicting the presentation levels to obtain a PB max score. The recommendations included a presentation level of 90, 100, 110, 120 and 130 dB SPL for average hearing levels of 0-40, 41-60, 61-85, 86-100 and greater than 100 dBHL respectively.

14. NU Auditory Test No. 4

Tillman, Carhart and Wilber (1963, cited in Goetzinger, 1972), utilizing the Lehiste and Peterson lists developed the Northwestern University (NU) Auditory Test No. 4. They selected ninety-five of the 500 words comprising the ten 50 word lists of Lehiste and Peterson and added five to construct lists I and II of the NU test No. 4 Research with the two lists has shown them to be interchangeable.

15. Modified Rhyme test

House, William, Hecker and Kryter (1965,cited in Martin, 1994) developed the Modified Rhyme Test (MRT) in which the patients were supplied with six rhyming words from which they had to select the one they thought they had heard. Fifty sets of items were presented to the patient, along with a noise in the test ear. Half of the word sets varied only on the initial phoneme and the other half differed in the final phoneme. The MRT had the advantage of assessing both initial and final consonants. The MRT might be called a phoneme or consonant identification test.

Kruel, Nixon, Kryter, Bell and Lang (1968) adapted the MRT to make it more clinically useful. As in the original MRT, there were three forms (1-3), each of which had six lists (A-F) of fifty, 6-word ensembles. Each ensemble was in a different random order. To minimize order effects, the test word was chosen at random from a 6-member ensemble without replacement of each MRT list. The result was eighteen, 50-word tests, i.e., three test forms with different test-word orders for each of the six lists. The answer sheet for each test form had fifty, 6 word ensembles of five foil words and one test word. Each ensemble was arranged in two columns within a box. The word form was either CVC, CV, VC. In all cases only a single initial or final consonant was varied, the remainder of the word was consistent with its foils. This was an attempt to maintain phonemic balance by including representatives of each of the major speech sound categories. Each word was uttered within the carrier phase, "Number -, You will mark the word please". The carrier phrase was chosen so that the test word would be preceded by a neutral vowel and followed by a voiceless stop plosive. This was to reduce the coarticulation variation of the second formant that typically accompanied vowels displaced toward the extremes of the vowel triangle (Ohman, 1966a, 1966b, cited in Kruel et al., 1968).

The recording was done by three speakers, two males and one female who recorded one of the test forms. They were chosen as representative speakers of General American English. The speakers monitored the level of their speech with a VU meter. They were instructed to peak the word 'mark' in the carrier phrase at a constant VU reading at the centre of the meter. All recordings resulted in master types with S/N ratio in excess of 40 dB. A 10 second sample of 1000 Hz pure tone and a 10 second sample of masking noise were recorded as calibration signals at the beginning of each tape. The recorded sample was presented to 38 young male and female college students with normal hearing through an Ampex 351 tape recorder which was connected to Kron-Hite active fitter adjusted to band pass 200-6000 Hz with a rejection rate of 20 dB/ octave. A pilot study was conducted in which each of the eighteen MRT lists were presented in four levels of noise to examine:

> equivalence of the six test lists for any given speaker,

> differences between speakers, and

> learning resulting from repeated listening to the same tests.

The S/N values with reference to the calibration tone or noise, were -5, 0, +10, and +30 dB. The differences between lists were statistically different for each of the two male talkers but only at + 10 S/N. From the pilot study it was found that + 30 S/N resulted in near perfect word identification by normal listeners for all three speakers.

16. Synthetic Sentence Identification (SSI)

Speaks and Jerger (1965) devised a method of assessing speech understanding using synthetic sentences. The sentences were synthetic in that words were selected at a random from the 1,000 most common words in the Thorndike and Lorge (1944, cited in Speaks & Jerger, 1965) list to form first order imitation of sentences which were formed by selecting one word at random, reporting it to an individual, then asking that individual to select a second word, reporting that second word to another person and asking him or her to select a third word, and so on. The word selected had to follow reasonably from the preceding word in the sentence and be from the 1,000 most common words of the Thorndike and Lorge list. Third-order synthetic sentences were generated by giving selected word pairs to an individual for her or his selection of a third word, then asking another individual for a fourth word after hearing only the second and third words of the sentence, and so on. Eg "Forward March, said the boy a" is an example of such a synthetic sentence. Twenty-four lists of ten sentences each were developed, nine first-order sets, nine second-order sets, and six third order sets. The third order sets were used more than the other sets (Jerger, Speaks & Trammel, 1968). In administering these test materials, the ten sentences in a set were presented on a panel or answer sheet and the task of the listener was to identify the sentence presented. Speaks and Jerger (1965) named this test the Synthetic Sentence Identification (SSI).

In order to make the test more difficult, the same talker reading a story was recorded as competition on the second channel of the tape recording. This competition could be mixed at various levels with the sentences to vary the message-to-competition ratio (MCR). Martin and Mussel (1979) noted that pauses in the competition allowed perception of one or a few words in the sentences and, thereby, identification of the target sentence. Therefore, they added speech spectrum noise to the competition at a level 6 dB below the overall level of the continuous prose. The continuous noise filled in the pauses and made it more difficult for the subjects to identify the sentences on the basis of the one or a few words, yet allowed perception of the story. Relatively large practice or learning effects have been observed for the SSI.

17. NU Auditory Test No. 6

In view of the limited applicability of the NU No. 4 test, consisting of only two lists, Tillman and Carhart (1966, cited in Goetzinger, 1972) developed another test which they labeled NU Test No. 6. This test consisted of four lists, each with fifty words. They selected 185 words from the parent list of 1263 words of Lehiste and Peterson and fifteen from outside sources. The four lists were randomized four times. Research with normal hearing and sensori-neural hearing loss subjects indicated the inter-list reliability to be high. Like the W-22s, these were recorded on tape by a single male talker. Before each test word, the talker recorded the carrier phrase "say the word' at 0 VU and let the test word complete the sentence with a natural, downward inflection. These lists were subsequently recorded on tape by a single female speaker.

Detailed studies using NU Auditory Test No. 6 were conducted by Rintelmann, Schumaier and Jetty (1974). One of the experiments, conducted on normal hearing subjects showed that all four lists of form A were equivalent. The study also demonstrated that the inter subject variability was higher at lower sensation levels. It was also noted that the original recordings of the lists were easier than that used by Rintelmann et al. (1974) although the talkers were reported to have a comparable dialect.

Rintelmann, Schumaier and Burchfield (1974) investigated if the four scrambles of the lists were equivalent, when used with young normal hearing adults. The general trend of results was similar to that obtained using form A alone (Rintelmann et al., 1974), indicating that all the four forms of the test were essentially equivalent. List I was the most difficult and List IV was the easiest. List II, III and IV were essentially equivalent.

On comparing the discrimination scores obtained by normal hearing and hearing impaired subjects using half lists (both first and second halves) and full lists, Schumaier and Rintelmann (1974) found that the difference was very small.

Presbycusic subjects, however, were likely to give larger difference. Based on the good equivalency between the half and the full-lists, the authors suggested that equating the tests based on familiarity could prove to be better than equating them based on phonemic balance.

Contradictory to the findings of Schumaier and Rintelmann (1974) were the results obtained by Jirsa, Hodgson and Goetzinger (1975) who found a poor correlation between the discrimination scores obtained using half list (first half) and the full list. They found that the half list reliability of CID W-22 list was better than that of the NU Auditory Test No. 6 (List IA). These contradictory results need to be resolved by further studies.

Schumaier and Rintelmann (1974) compared the performance of two groups of normal hearing subjects who represented two different dialects of English. Both the groups performed similarly on the NU 6. The order of difficulty of the lists from easy to difficult was list IV, list II, list III and List I.

18. Rhyming Minimal Contrasts

Griffiths (1967) refined the Modified Rhyme Test further in 1967. In Griffith's test each word within a given set differed from another word in the same set in only one of the distinctive features characterizing speech sounds i.e., manner of articulation, place of articulation or voicing. Griffith's basis for developing word lists of this type was that earlier work had demonstrated that listener confusions in manner of articulation, place of articulation, and voicing were essentially independent, therefore, word sets of minimal contrasts ("rhyming minimal contrasts" as labelled by Griffiths) would enhance the diagnostic utility of the test. The rhyming minimal contrast test was made up of 250 monosyllables (150 of the words from the modified rhyme test and 100 new times), arranged in five 50-word lists. The initial consonant was the variable in twenty-five sets and the final consonant was the variable in the other twenty-five. No attempt was made to achieve phonetic balance similar to that of English. From column to column, the five lists were balanced, however no variable element appeared twice in one column without appearing in all other columns. Thirteen initial and thirteen final consonants appeared in all lists. The order of the words were subjected to three randomizations and recorded by two talkers making thirty lists. The talkers attempted to speak the words at a constant vocal effort and had a VU meter available for monitoring. No carrier phrase was used. The words were recorded on an Ampex 601A tape recorder at 3-sec intervals while the speaker was seated in an anechoic chamber.

Each listener was provided with a response form, which consisted of fifty response sets: one set of five words (also randomized) to a line, ten lines to a page. The listener was instructed to line out the word that was heard in each set and to guess if not sure of the proper response. The test was evaluated in a manner similar to House et al. (1965) to determine whether their results and conclusions also applied to this modification. The word lists were presented to twelve college undergraduate women at seven signal-to-noise (S/N) ratios, +8, +4, 0, -4, -8 and —12 dB in addition to being presented without noise. The speech was presented monaurally to the listener through TDH-39 earphones in a quiet room. The results indicated that the perception of final consonants was degraded more rapidly than that of initial consonants as the S/N ratio was reduced. The different test lists were similar in terms of responses to the two halves of each list.

19. Isophonemic Word Lists

Boothroyd (1968, cited in Olsen & Matkin, 1979) described so-called isophonemic word lists with ten words per list and each of fifteen lists were phonemically balanced. All of the test items were of the CVC type. The same twenty consonants and ten vowels occurred once in each list. The phonemes chosen were those occurring most frequently in the CVC words. The lists were recorded by a male talker with Northern English accent. Boothroyd also recommended scoring each phoneme. Hence, thirty phonemes were scored in each list of this test rather than just the ten words. According to Boothroyd, one of the advantages of scoring phonemes rather than words is that the phoneme scores gave a more valid estimate of the subjects ability to recognize the acoustic features of speech sounds. In addition to allowing for phoneme error analysis, these short word lists had the advantage of being faster to administer. It took less than six minutes to plot a P-I function of at least three levels on the slope and three on the plateau.

20. Kent State University (KSU) Word Identification Test

The Kent State University Word Identification Test was a closed set sentence format test devised by Berger (1969). The development of this test began with collection of groups of five phonetically similar words, which were put within sentences. A total of 150 sentences containing a total of 750 key words were thus developed. The criteria for key word selection was that it never appeared as the initial words in the sentence, but may be in any other location, except that the key word was not the second word in a sentence where the first word was an aiticle. Neither proper nouns nor contractions were avoided, but the final sentences contained no key words that were contractions. Key words of both one syllable and two syllables were included but each group of five key words had the same number of syllables and, for the two-syllable words, the same stress pattern. Word difficulty and word frequency varied and no efforts were made to equate these to the several lists of word familiarity. The criteria for selecting the test sentences was: The test sentences were variable in length but never shorter than four words or longer than nine words, including the key word. Contractions were not avoided, as they were common in conversational speech. No efforts were made to achieve phonetic balance in either the test sentences or the key words.

To evaluate the relative discrimination difficulty of the 750 key words and sentence combinations they were presented to three groups (Conditions I, II, III) of normally hearing subjects. The testing was carried out in a sound-treated two-room suite. Sentences were presented through a Grason Stadler Model 162 speech audiometer. Recorded speech was played through an Ampex model 601 tape recorder.

The identification ability was evaluated in three different types of conditions. These were:

Condition I: Fifty subjects aged 18-22 years heard the sentences at 30 dB above SRT with a continuous white noise presented through a BC vibrator fastened to the centre of the subjects forehead by velcro strips. The competing noise was at a constant level for all subjects. Sentences were administered by a tape recorder prepared by a male speaker through monitored live voice (MLV) testing. The mean error score was 18.9%.

Condition II: Fifteen subjects aged 20-32 years heard the sentences at 10 dB above SRT in quiet. The sentences were spoken with MLV by a male speaker. The mean error score was 16.1%.

Condition III: Twelve subjects heard the sentences through a 1140 Hz low pass filter at 30 dB above SRT. Sentences were spoken with MLV by a female speaker. The mean error score was 17.0%. The key words, which produced similar, mean error scores under each of the three conditions were accepted for actual test use.

An evaluation of the logical sentence was also made by presenting the 150 sentences to an additional 156 subjects. They were asked to read the sentences silently and to indicate which key word in each group of five was the most logical in that sentence and which key word was the least logical. For use in the final forms the keyword had to be chosen as most logical in atleast 20% and to be least logical in not more than 30% of the responses. Only 134 key words and their accompanying sentences were acceptable from the original pool of 750. The acceptable key words and their sentences were ordered in difficulty from those answered 100% correct under all the three conditions. Eight lists of equal difficulty and thirteen sentences in each were constructed. The following instructions were given:

"You will hear some sentences read to you. However within each sentence you will hear only one of the five key words (which are in capital letters on your answer sheets). You may hear any one of the five key words. As your hear each sentence, draw a line through the key word you hear. If you are not sure of a key word, please guess. Are there any questions?"

The examiner monitored the VU meter at "zero" while saying the key word; the remainder of the sentence varied around or just below the "zero" level. The examiner uttered the carrier phrase "Number _ " before each sentence, paused for 1 sec, then read the sentence. The test was presented at SRT plus 25 dB (or SRT +26 dB when using audiometers with -2 dB attenuator increments).

A unique scoring system was employed: the number of the sentence in which the error occurred was subtracted from 100. For example, if, in a given list, a listener made errors in sentence 8 and 13, their sum, 21, was subtracted from 100 for a score of 79%.

Berger, Keating and Rose (1971) observed that the KSU test was less sensitive to hearing impairment, when compared to W-22 lists. However, this test was better than W-22, in predicting how efficiently one could use his hearing for communication purposes.

21. Multiple Choice Discrimination Test (MCDT)

The development of a closed-response format for the CID W-22 word lists was reported by Schultz and Schubert (1969) for each of the fifty words in each of the four CID W-22 lists, the test word and four alternative choices were available on the answer sheets provided to the listeners. The response task was to mark the item presented each time. Various sets in the Multiple-Choice Discrimination Test (MCDT) of Schultz and Schubert included foils allowing for confusion of either the initial consonant or final consonant. Three general guidelines were followed in the construction of response foils:

1. It was desired that response foils be composed of words of the same level of familiarity as the stimulus word.
2. In the absence of estimates of the discriminability power of individual consonants, it was desirable that all possible substitutions had an equal probability of occurrence.
3. Each of the four response foils per stimulus item differed from the stimulus word by a single consonant phoneme substitution. The recording of the stimulus items was the same as in CID W-22 test. The carrier phrase used was "You will say..." This was a short test suitable for a group presentation which was convenient to administer. It used error analyses to differentiate among populations.

22. The Gardner High Frequency word Lists

Gardner (1971) compiled a list of words to meet the need for testing individuals having **a** high frequency hearing loss. According to him accurate measurements of the effects of modifications in tubing diameter, earmold design or acoustic filter placement, which results in the critical enhancement of high frequency information are essential. With improvement in technology and availability of instruments whose specification suggests their suitability of cases, there are few clinical methods of demonstrating the benefits of amplification. In order to test the subtle perceptual changes that the acoustical or electro acoustical modification brought about, he designed the high frequency word list.

It contained seven voiceless consonants / p, t, k, s, f, o, h / used in conjunction with the vowel *lil.* These consonants have been known to result in confusion when identification is attempted by persons with high frequency hearing loss. The fifty words were arranged in random order and assigned alternatively to two 25-word lists. The Gardner high frequency word lists were recommended for use in live voice presentation or with a tape recording of a female (or high pitched) talker. It was also recommended that the lists of stimuli provided be randomized for different list presentations, especially when performing hearing aid evaluations. Thought the test was specially designed for application in hearing aid selection it may be used for auditory training as well.

The drawbacks of the Gardner test was that no standardization information was reported for the different talkers presentation modes or randomized lists, therefore the sensitivity of this test is doubted. Also the stimulus must be tested under specific conditions to determine if they are sensitive enough to provide the kind of information desired from the test.

23. University of Oklahoma Closed Response Speech Test (UOCRT)

The UOCRT developed by Pederson and Studebaker (1972) was a closed response set, **CVC** monosyllabic word speech discrimination test consisting of three independent subtests. The subtests were an initial consonant subtest, a final consonant subtest, and a medial vowel subtest (referring to the position of the test phoneme that was varied across the items in each set). Each subtest had five, randomized, test item presentation orders (scramblings) and four response sheets counterbalanced for the printed positions of the words in each response set.

The following criteria were applied to the selection of the test stimuli:

1. Meaningful words were selected as test items whenever possible.
2. All test items were CVC words with the exception of those with the absent phoneme in the test positions.
3. The phonemes in the nontest positions within each set were identical and
4. Test items were chosen for familiarity.

Two additional requirements that were applied in the case of the consonant subtests were:

1. The variable phonemes in each set were selected to vary only in the place of articulation with the following exception. It was necessary to deviate from this criterion for one word within each consonant test set. In this instance, the variable phoneme position was left vacant.
2. Identical test phoneme sets were used in both the initial and the final positions consonant subtests.

Three phonemes with identical manners of production were selected to be evaluated within each test set. The vowel subtest consisted of one, 8-item, closed-response set. Each of the eight words was used as a test item eight times during the subtest producing a sixty-four item subtest.

Each test item was recorded as a natural speaking continuation of the carrier phrase "The word is..."All the test items were spoken by a single male talker with General American dialect. Five copies of each subtest were recorded. Four of the recordings for each sub test were cut and spliced in other sentences to develop the five different scramblings of each sub test. A final re-recording of all the scramblings was done.

Each OUCRT subtest was presented with ipsilateral white noise at five different S/N ratios to develop intelligibility functions for the subtests with normal hearing subjects. The subjects were ten male and ten female students in the age range of 18-25 years with normal hearing. The subjects' task was to mark the item, which he felt, had been presented over the earphone from among the selections in the appropriate closed-response set on a printed response sheet. All subtests, test scramblings, noise conditions and response sheets were counter balanced with respect to temporal order and coincidence across all subjects.

Results for the final consonant subtest indicated that the two subtest halves and various subtest scramblings were not statistically different. As the S/N ratio increased the percentage of correct response increased. The initial consonant subtest resulted in a correct response percentage difference of 2.7% between males and females. For the vowel subtest, the two sexes performed differently again favoring the male group. The slopes of the articulation-signal/noise function obtained were 2.3 and 2.4 percent per dB for the 2 consonant subtests and 6.4 percent per dB for the vowel subtest.

24. The Pascoe high frequency test

Pascoe (1975) developed this test to assess the speech perception abilities of individuals who are hearing impaired, using words with difficult phonemes. The list included fifty monosyllables words that emphasize phonemes that are difficult for hard-of-hearing subjects. Only three vocalic nuclei were used /i/, */ai/,* and /ou/ in order to increase the weight of the consonants in the correct identification of words.

Voiceless fricatives and plosives formed 63% of the number of consonants. The rest were nasals, laterals and voiced plosives. The words were recorded by a male and female talker.

The experiment by Pascoe consisted of two parts, one in which eight hearing impaired subjects were tested with a binaural master hearing aid with five different frequency responses and the second part in which the Pascoe high frequency test was compared with a phonetically balanced word list in quiet and in noise. The results indicated a high condition between the subjects adjusted hearing levels, using a high frequency band and the identification scores in a non-PB list (i.e., the high frequency list). The Pascoe high frequency test was advantageous in that it provided standardization information on a male and female talker version of the test.

However, it is limited in that it uses two types of word lists, which differ in several dimensions. The high frequency lists were presented as a closed or known set. They were phonetically unbalanced in favour of high frequency phonemes and every list was identical to the others, except for differences in word order.

25. California Consonant Test

Owens and Schubert (1977) developed a four alternative forced-choice response test called the California constant test (CCT). The purpose was to develop a test of consonant identification for listeners with hearing impairment that will have application in rehabilitation and hearing aid evaluations.

The items selected for inclusion on the test were based on data gathered previously which assessed impaired subjects (Owens & Schubert, 1968; Owens et al., 1971; Sher & Owens, 1974). Of the 125 items initially selected, twenty-five were eliminated on the basis of extremely low item scores. The remaining 100 of the original recorded items were retained to make two equivalent forms of the test; items were rearranged so that the first and last fifty were equated as closely as possible in phonetic composition, item difficulty and point biserial correlations. Each of test item consisted of four words, the stimulus word and three foils, which differed only on either the initial or final consonant presentation. There were thirty-six initial consonant and sixty-four final consonant test words per list. The test words were recorded by a male speaker of General American Dialect, experienced in recording procedures. For each item the speaker first said "repeat item (number)" and allowed five seconds each for the subject to read the four response foils aloud. Then the speaker presented the VU monitored carrier phrase "check the word", followed by the test word. Items were presented via an audiometer (Allison 22) from a tape deck to subjects wearing TDH-39 earphones in a sound treated room. The presentation was to one ear of each subject at a sensation level (SL) of 40 dB. Subjects were patients with sensori-neural hearing loss and reduced speech discrimination scores on W-22 lists, except for a few patients with high tone loss. Results showed that the two lists were not equivalent clinically and it was time consuming. Hence, another experiment was carried out to reduce the time of administration. All speaker utterances were removed except the carrier phrase "check the word" followed by the test word. The subject was instructed to mark his choice for one item and then immediately say the words of the next item without prompting. This way the test item was reduced by 3 minutes for a total of 12 minutes. Owens and Schubert (1977) indicated that a potential use of the results of measurement of speech perceptual skills with the CCT was in identifying consonant confusions to assist with the planning for aural rehabilitation.

Work by Schwartz and SUIT (1979) indicated that the CCT was more sensitive to high frequency hearing losses than was NU No. 6. Givens and Jacobs-Condit (1981) and Townsend and Schwartz (1981) observed some consistency in errors made by hearing impaired persons on the CCT in terms of more errors for place and manner of articulation than for voicing and nasality. Errors were more frequently for final consonants (Owens & Schubert, 1977 and Givens & Jacobs-Condit, 1981).

26. The Speech Perception in Noise Test

The Speech Perception in Noise (SPIN) test was designed by Kalikow, Stevens and Elliot (1977) to assess a listener's ability to use context uses and linguistic information to perceive auditory information. These eight lists of the open response SPIN test were each composed of fifty sentences, each being five to eight words in length. The last word of each sentence was the test item. Twenty-five of the sentences contained test items that was classified as having "high predictability", which meant that the word was readily predictable given the context of the sentence. Conversely, twenty-five sentences had test items with "low predictability". The predictability was based on contextual, syntactic, and prosodic cues.

The key words were monosyllabic nouns, which were selected from items in the Thorndike-Lorge list with frequency counts in the range 5 to 150 per million words. Tape recordings of these sentences were prepared with the sentences on one channel and the babble of twelve voices reading continuous test on the second channel. Thus the user could mix the test sentences and babble at various signal-to-channel. Thus the user could mix the test sentences and babble at various signal-to-babble ratios. The task of the listener was to repeat (or write) the last word of each sentence (Olsen & Matkin, 1979). Scoring was based on correct identification of the key words. An overall score was derived from the difference between the scores on high and low productivity items. The results of the eight forms indicated a shift of 1 dB in S/N ratio for high predictable sentences and 2 dB for low predictable sentences. Comparison of a listener's performance on these two types of test items provided insight regarding whether the listener took advantage of linguistic or contextual cues or not. This information could be helpful to clinicians planning rural rehabilitation services (DeBonis & Donohue, 2004).

Morgan, Kamm, and Velde (1981) and Bilger, Neutzel, Rabinowitz, and Rzeczowski (1984) found that all forms of the SPIN test were not equal in difficulty. Morgan et al. (1981) suggested discarding some of the lists. Bilger et al. (1984) suggested that the best items from all forms should be selected to produce new forms that were equivalent in difficulty. Using this strategy, Bilger (1984) reported eight forms for the revised SPIN test. He noted that with this revised version, the reliability of Form 2 was slightly lower than the reliability of the other seven forms. He suggested that, when possible, Form 2 should be used and its cognate, Form 1, could be avoided.

27. Tonisson Sentence List

The only set of sentence lists designed for use in Australia and available for use were those developed by Tonisson (1977, cited in Dermody & Mackie, 1987) although other sentence lists were being standardized. Tonisson's lists reflect the

Central Institute of the Deaf Everyday sentences. They consisted of nine lists, each containing ten sentences or common phrases and fifty key words. The phraseology used and the length of the sentences made these sentences appropriate for older children and adults, rather than the younger hearing-impaired child.

28. Closed Nonsense Syllable Test

Levitt and Resnick (1978, cited in Bess & Humes, 1995) at the City University of New York (CUNY) developed a closed-set nonsense syllable test (NST) The test was composed of CV and VC syllables categorized into eleven subtests of seven to nine syllables each. The syllables in these subtests differed in the following:

a) consonant voicing (voiced or voiceless)

b) syllable position of the consonant (initial or final position), and

c) vowel context (/a, u, i/).

Voiced consonants with vowel /a/ were incorporated into three lists, unvoiced consonants followed vowels /a/, /u/ or /i/ in three other lists, and unvoiced consonants followed by /a/ made up another list of the nonsense syllable test. Because the vowels were constant within a given subtest or list, the task of the listener was one of phoneme on consonant identifications. Items included in each list were those most frequently confused with one another by normal hearing and hearing-impaired individuals. With this test, it was possible to obtain a somewhat detailed picture of the type of errors made by the listener and not just an indication of the total number of errors made. More detailed knowledge of the types of perceptual errors made by the client would lead to more effective remediation of their speech understanding difficulties.

In this investigations using NST material Dubno and Dirks (1982) and Dubno, Dirks and Langhofer (1982) found that responses from hearing impaired subjects were highly reliable. They noted that voiced versus voiceless consonant confusions were rare, that place errors were more common than errors in manner of articulation, and that fricatives were more likely to be confused with plosives. Overall, place errors were more common than errors in manner of articulation, and that fricatives were more likely to be confused with plosives. Overall, place errors constituted about 49% of the errors, manner errors accounted for about 22% of the confusions, and about 29% of the misidentifications were errors of both place and manner of articulation. More errors occurred for final consonants.

29. Bamford-Kowal-Bench (BKB) Sentences

This open set sentence test was originally developed for children by Bench and Bamford (1979) but could be used for adults also. For constructing the sentences lists, utterances of over 240 children aged eight to fifteen years were tape-recorded. The children were asked to describe a set of coloured drawings, which depicted scenes and events from common place, play or family environments". The recordings were transcribed and subjected to grammatical analysis using "The Language Acquisition Remediation and Screening Procedure"(LARSP) profile of Crystal et al. (1976, a parsing of each word, and a vocabulary count. The basic approach was to use vocabulary and grammar which were familiar to children. Thus, for grammar, subject-verb-complement /object, subject-verb-adverb, and subject-verb sentence were allowed at the clause level; the ten phase structure most frequently elicited were used at the phrase level; and the seven most frequent morphological structures were allowed at the word level. The words which had a high frequency of occurrence were chosen. The minimum requirement was that any word used in the sentence list had to appear at least twice in the total collected vocabulary. Further, the sentence length was restricted within seven syllables. The grammar and first word of the sentences were balanced across the lists of sentences. So, twenty-one lists each containing sixteen sentences and fifty 'key' words which made up the 'Standard BKB Sentences Lists for children', and the eleven lists, each of which had sixteen sentences and fifty key words, which made up the 'Picture-related BKB Sentence Lists for Children' were tape-recorded by a female speaker and standardized on a large sample of partially hearing children. Each sentence was preceded on the tape by a low-pass filtered white noise warning signal, recorded at the same level as the speech peaks.

Sixteen partially hearing children aged eleven to thirteen years, with pure tone hearing losses of between 30 dB and 80 dB and an average WISC performance IQ of 90-120 points were assessed by speech audiometry. Five BKB sentence lists were presented to the child, one sentence at a time from a tape-recorder through head phones at selected relative hearing levels, and the corresponding five randomized key word lists were administered one key word at a time at the some relative hearing levels as the sentence lists from which they were derived. In both cases the method of scoring was to measure the percentage of correctly repeated key words for each lists. Results indicated that the sentences were significantly better repeated than the randomized key words.

The BKB Standard Sentence Test was appropriate for assessments, which were related to natural listening conditions. The test using the randomized key words represented 'natural' vocabulary better than did tests of phoneme-balanced words, not only in the sense that they were well known to the children, but also because they were not of uniform length, which was an unnatural feature of PB word lists.

30. Distinctive Feature Discrimination Test [DFDT]

The purpose for developing this test by McPherson and Pang-Ching (1979) was to provide more diagnostic information regarding the type (s) of discrimination difficulties a patient experienced as opposed to a correct score. The DFDT lists were constructed primarily from the corpus of words found in the Modified Rhyme Test. Fifty monosyllabic CVC stimulus words and three rhyming error responses for each stimulus word were selected. Of the fifty stimulus words in each list, twenty-five had the initial consonant as the variable phoneme while the other twenty-five had the final consonant as the variable phoneme. The vowel nuclei for the words in each list reflected the frequency of occurrence of vowels in the English language. The error responses in each of the fifty sets of four words in each of the lists included words in which the variable phoneme was first, second and third distinctive feature removed from the corresponding phoneme in the stimulus word. The distinctive features were those proposed by Miller and Nicely (1955, cited in McPherson & Pang-Ching, 1979) i.e., voicing, nasality, affrication, duration and place of articulation.

The stimulus words for each of the four lists were recorded on magnetic tape by a 39-year old male speaker of General American Dialect. A 1 kHz calibration tone was first recorded on the master tape and peaked at VU=0. Each of the stimulus words were recorded three times using the carrier phases "Please mark the............" Twelve forms of the test were made. The subjects' chosen were six normal hearing women aged 20-24 years. SRT's using CID W-l lists were established for each ear using a 2-dB ascending approach. All SRT's fell between 18 and 22 dBSPL. Each word list was then presented to each ear under each of the three conditions [unfiltered (actually bandpassed at .2-6 kHz), bandpassed at .2-1.2 kHz, and bandpassed at .2-6 kHz] at each of five sound pressure levels (20, 25, 30, 35 and 40 dB). The .2-1.2 and the .2-6 kHz bandpass conditions were rated by subjects in a pilot study as being of moderate difficulty and extreme difficulty while listening to connected discourse.

The various forms of each list were counterbalanced to control for learning and order effects. Each condition required about 8 minutes. Subjects were given a 2 minute rest period between conditions. The responses were scored simply as correct or incorrect or weighted with one point for a one-feature error, two points for two-feature error and three points for a three-feature error. P-I functions were presented for unweighted scores and for scores where error responses were weighted in terms of distinctive feature differences from stimulus items. The four DFDT lists generated, and their three randomizations each appeared to have a reasonable degree of inter-list reliability and it was felt that the twelve resultant lists might be used interchangeably.

31. Nonsense Syllable Test

Edgerton and Danhauer (1979), described trisyllabic test material using constant-vowel-consonant-vowel (CVCV) nonsense syllables. Twenty consonants and ten vowels were used. The test list consisted of twenty-five CVCV items each, randomized six times. This open set test was recorded by a male talker and the carriers phrase used was "Say......". At 55-dB sensation level above SRT a group of normal hearing subjects attained mean scores on the order of 95%, 90%, 86% and 66% for vowels, phonemes, consonants, and whole bisyllable recognition, respectively. Mean performance value for a group of patients with sensori-neural hearing losses were approximately 80%, 72%, 62% and 24% for vowels, phonemes, consonants, and bisyllables, respectively. Edgerton and Danhauer (1979) concluded that their data did not support use of bisyllabic scoring and "that while the data supported the use of phoneme and consonant scoring for nonsense stimuli, phoneme scoring was clearly superior". They also indicated that "Nonsense syllables are optimal stimuli for discrimination testing because they can be easily combined into equivalent lists, are low in information redundancy, and are relatively insensitive to familiarity effects.

32. Picture Identification Task

A picture identification task for adult patients unable to respond orally or in writing was developed by Wilson and Antablin (1980). The word lists were compiled to approximate traditional word-recognition test materials. Consonant-vowel nucleus - consonant (CNC) words and three-rhyming alternatives were chosen from the Teacher's Word Book of 30,000 words (Thorndike & Lorge, 1947, cited in Wilson & Antablin, 1980). From this pool of words, four lists of fifty words each were developed. The 200 test words consisted of 175 nouns and twenty-five action verbs distributed among the four lists. Of the test words chosen, 183 had alternative words that rhymed in the final position where as the remaining seventeen alternative words that rhymed in the initial position. The test words were phonemically balanced and similar to Lehiste-Peterson's list. However, in the vowel position some discrepancies occurred among the four Picture Identification Task lists.

The test words and rhyming alternative words were illustrated in color by a commercial artist. The individual drawings were then grouped into response plates comprising of the test word and the three alternative words randomly positioned in the quadrants. Two randomizations (form A and form B) of each of the four lists were recorded by Auditec of St.Louis. The carrier phrase, "Show me," preceded each test word. A 1000 Hz calibration tone corresponding to the peaks of the carrier phrase preceded each randomization.

Two experiments were carried out. In both experiments, the Picture Identification Test was compared with NU Test No. 6. In experiment 1, the relationship between the performance-intensity functions obtained in quiet and in noise with the Picture Identification Task and the NU No. 6 was examined. An open-set response paradigm, in which the subjects wrote their responses, was used with both materials. The noise conditions were included to study the performance with these materials at high stimulus presentation levels. The CNC words in the quiet conditions were presented in 4-dB increments from 2 to 26 dB re: the speech reception thresholds (SRT). When the noise conditions were administered, the CNC words were presented at eight signal-to-noise (S/N) ratios from -12 dB to 16 dB. The level of the broad band noise was held constant at 70 dB sound-pressure level (SPL, re: 20 µPa).

In experiment 2, the Picture Identification Task in a closed-set response paradigm was compared with the NU No. 6 both in an open-set and a closed-set response paradigm. The subjects responded to the Picture Identification Task by pointing to a picture in one of the quadrants. In contrast, the subjects responded to the NU No. 6 open-set by writing the stimulus word on an answer sheet, and to the NU No. 6 closed-set by marking the appropriate word on a multiple-choice answer sheet. Each test word in the closed-set modification of NU No. 6 was grouped with three rhyming words. The three materials were presented in 4 dB increments from -2 to 26 dB (re: the speech reception threshold).

The speech materials were reproduced on a tape recorder (Sony, Model TC 377), passed through a 1 dB step attenuator and fed into an amplifier-attenuator complex that also generated the broadband noise. The speech was then delivered monaurally to a TDH-39 earphone encased in an MX-41/AR cushion. The non-test ear was covered with a dummy earphone. Testing was conducted in a sound treated room.

Before presenting the CNC lists in the quiet conditions of both experiments, two SRTs were established according to the procedure described by Tillman and Olsen (1973) and Wilson, Morgan, and Dirks (1973). The lower of the two trials was used as the threshold of speech.

All the subjects had normal hearing for pure tones presented in octave intervals from 250 Hz to 8000 Hz. The sixteen subjects in Experiment I participated in four one-hour sessions. The twenty-four subjects in Experiments 2 participated in three one-hour sessions.

The results from the first experiment demonstrated no significant differences between the picture identification task and the NU No. 6 in an open-set paradigm. The results from the second experiment revealed significant differences among the three response tasks. The easiest task was a closed-set response to words, the next was a closed-set response to pictures, and the most difficult task was an open-set response. At high stimulus presentation levels, however, the three tasks produced similar results. Picture Identification Task has been used clinically and is found to provide good estimate of word recognition performance.

33. Clark PB word lists

Clark (1981, cited in Bench, 1987) developed word lists, which were derived from the Northwestern University Auditory test No. 6 but were specially designed for Australian English. Clark's criteria for the design were:

i. Monosyllabic CVC word structure,

ii. Exact inter-list phonological balance,

iii. Minimal intra-list phonotactic redundancy,

iv. High lexical familiarity, and

v. Phonological distribution generally compatible with that for monosyllabic words in Australian English.

34. The Minimal Auditory Capabilities (MAC) Battery

Fourteen tests comprising the Minimal Auditory Capabilities (MAC) battery were described by Owens, Kessler, Telleen and Schubert (1981). The MAC Battery included thirteen auditory tests and one lip reading test. Of the thirteen auditory tests, twelve were composed primarily of spoken materials. The first four tests evaluated some basic abilities underlying the understanding of speech.

A second group of tests required identification of phonemes. In a test of vowel recognition each item contained four alternative choices, one of which was the correct (test) word. The response formats were the same as that of the vowels. The consonant items were designed to test whether nasality, voicing and glide features of speech may be detected.

Multiple-choice items employing spondee words contributed a relatively easy test of word recognition to the battery. All of the above tests employed a multiple choice response format. In the scoring process, an allowance was made for a guessing factor inherent in multiple-choice items.

The Spondee Recognition Test presented twenty-five two-syllable words for the patient to repeat. Only a few patients with cochlear implants had succeeded in recognizing any spondee words presented in an open response mode. Thus, success on this test as well as on the Everyday Sentence Test, which seemed comparable in difficulty, appeared to be a reasonable objective for cochlear implant patients.

The two most difficult tests on the battery were the "NU Auditory No. 6" which was a conventional speech discrimination test and the "words in context". The other groups of tests dealt with recognition of environmental sounds and visual enhancement.

The total administration time was approximately two hours. The battery was administered in preferably an audiometric sound-treated test suite. The materials were presented through a tape deck. All responses by the patients were verbal with the exception of three tests in which the patient was required to mark his selection in a multiple-choice format.

Neither the practice or sample items nor any of the test items can be utilized as test materials because MAC serves primarily as an evaluative pre- and post- training / surgical measure of auditory responses. Currently this battery is used for evaluation of cochlear implantee patients. Another possible use of MAC is in the determination of the aural rehabilitative need of patients with profound loss, whether amplification be from an implant or from a hearing aid, and in evaluating the results of an aural rehabilitative program.

35. Distinctive Feature Difference (DFD) Test

A closed-response nonsense syllable test, the Distinctive Feature Difference (DFD) test, was described by Feeney and Franks (1982). The consonants having an error probability of 0.3 or greater from the data of Owens and Schubert (1968) were selected. The thirteen consonants / p, b, t, d, k, f, v, θ, s, ʃ , tʃ, dʒ/ were inserted in nonsense syllables having the same intervocalic context, for example, [ˆbil], [ˆpil], and so forth. These thirteen consonants were tape-recorded in the same intervocalic context by a general American English speaker. Each utterance was preceded by the carrier phrase, "You will mark ". Recordings were made in a quiet recording studio by using a microphone with a reel-to-reel tape recorder. The vowels of both carrier phase and stimulus utterance were monitored at 0 dB on the VU meter. Eleven additional scramblings were dubbed from the original recording by using a tape recorder and a mixer. A 1000 Hz calibration tone was recorded on the test tape.

A list of the thirteen phonemes was provided for each stimulus phoneme and the subjects were simply asked to select and circle the consonant heard. The thirteen phonemes were represented on the answer sheet.

The DFD test was scored both on the basis of distinctive feature differences and on the basis of the identification of the correct phoneme (phonemic scores). An error consisted of failure to perceive one of the features, which distinguished the correct phoneme from the other twelve in the test. Seven features for each phoneme (high, back, anterior, coronal, continuant, voice, and strident) were scored. A correct response was given 7 points, but a response of /d/ for /t/ (voicing error) received 6 points, /b/ for /t/ (place and voicing errors) received a score of 5 points, /v/ for /t/ (place, voicing and manner errors) earned 3 points, and so on.

A number of advantages were claimed for the DFD test. One advantage was that the DFD test, dealt with the actual perceptual features of the code rather than phonemes only. This contributed to stronger content validity and increased the number of elements contributing to the discrimination score, which permitted a shorter test while maintaining stability of scores on repeated tests. Moreover, the DFD test was a closed response test with an unambiguous responses mode and scoring (Feeney & Franks, 1982).

36. The Maryland CNC Test

Causey et al. (1984) described the Maryland CNC test, named after the University of Maryland where it was developed. These investigators reemphasized that phonemes preceding and following the middle nucleus sound affected the production, and hence the acoustics, of that nucleus sound. Furthermore, they pointed out that the acoustic parameters of consonant sounds were influenced by the transitional shifts, which occurred where a vowel and a consonant joined. This phenomenon was referred to as "coarticulation" and the authors believed this was a factor that should be controlled when obtaining word recognition scores.

Causey et al. (1984) took 500 of the original NU CNC words and imbedded each of them in the phrase "Say the (test word) again" because they believed the same "acoustic surround" should be used for the best control of the coarticulation factor. They noted that the schwa sound, which occurred at the final position of the word "the" and at the initial position of the word "again", was a phoneme that provided minimal influence on adjacent stimuli because of its lack of stress and its neutral articulatory position. By surrounding the test word in this way, the authors believed that the effect of co-articulation would be minimized.

Ten PB lists, each with fifty different words, were recorded on tape by a single trained male talker. Vocal output was monitored at 0 VU during the "say the" part of the phrase, each test word was allowed to occur naturally with a downward inflection, and the final syllable of "again" was monitored at 0 VU. Performance intensity functions were completed for a group of sixty normal hearing, college-aged subjects and a group of forty hearing-impaired patients ranging in the age from thirty to forty years. There was a 2.1 % increase per decibel in word recognition in the linear portion of the performance-intensity function for normal listeners and a 1.3% increase per decibel for hearing-impaired listeners. Out of ten, six lists were judged to be equivalent and there was good test-retest reliability. Hearing-impaired subjects demonstrated a wide range of scores on the Maryland CNC test, suggesting that this measure could distinguish among varying degrees of word recognition ability.

37. Connected Speech Test (CST)

This test of intelligibility of everyday speech had been developed primarily for use as a criterion measure in investigations of hearing aid benefit (Cox, Alexander & Gilmore, 1987). The test consisted of 48 passages of conversationally produced connected speech. The initial pool of test items included seventy-two passages about familiar topics such as common plants, animals and household objects. Each passage contained ten syntactically simple sentences, seven to ten words in length. To control word familiarity, the basic vocabulary was derived from a children's educational reading source. The passages were spoken by a female and this was recorded audiovisually at ± ldB.

Each passage contained forty-five to fifty-five potential key words that could be used for scoring purposes. No sentence contained more than seven potential key words. These words were selected by testing only the audio portion. The video signal was not available to the subjects. This scoring applied to the audio test only.

Thirty-college undergraduate with a normal hearing audited the test passages. Five subjects served in each of six signal-to-babble (S/B) ratio conditions. The CST passages and competing babble were replayed by the optical disk player and routed to attenuators to allow independent adjustment of S/B ratio. The two outputs were then mixed, amplified, and delivered to an insert earphone that was coupled to the listener's ear using a compressible foam earplug. Calibration of the playback system was achieved with the output of the insert earphone delivered to a Zwislocki-type ear simulator coupled to a precision sound level meter.

The test passage was presented at the level of normal conversational speech i.e., at 55 dB Leq outside the listener's ear. The competing babble was presented at S/B ratios of -3, -4, -5, -6, -7, and -8 dB. Each subject heard all seventy-two passages at a constant S/B ratio.

Subjects were seated in a sound treated room viewing a video monitor that briefly displayed the passage topic before the presentation. They listened monaurally to each of the passages, which were presented one sentence at a time. After each sentence, both speech and babble were halted while the subject repeated the sentence or as much of it as he/she had heard.

The potential key words for the sentence were displayed on the video monitor. The examiner scored the words correctly identified by entering the corresponding number on a keypad. Words containing additions, substitutions, or omissions were scored as incorrect.

Two passages were used for the practice session. Out of seventy-two, fifty-seven CST passages were equated in terms of average intelligibility and range of key words intelligibility across subjects. Finally forty-eight passages having the best correlations with true scores were selected as the CST test passages. In the typical passage, about half the key words were monosyllables and one-third were two-syllable words.

A validity check was done to find the difficulty and equivalence of the final CST passages. There were no systematic differences in difficulty across the forty-eight test passages. The performance-intensity (P-I) function for the CST had a slope of 12 rau (rationalised arcsine unit) per dB S/B ratio. The CST appeared to meet many of the criteria for reliability and validity not found in other sentence tests and thus holds promise as a diagnostic tool.

38. The Four Alternative Auditory Feature test (FAAF)

The Four Alternative Auditory Feature (FAAF) test was a speech identification test developed by Foster and Haggard (1987), to meet a particular set of objectives:

- Ease of response task and lack of response practice effects,
- Face validity,
- Trans-dialect applicability within the United Kingdom,
- Applicability to modest literacy levels,
- Discriminatory power for good channels where intelligibility ceilings were approached with conventional speech materials, and
- Maximum applicability of directional error information to diagnostic assessment.

The four-alternative forced choice tests comprised of twenty sets of four binarily and minimally paired words, giving an eighty-item test. Nine sets varied in the initial constant and eleven in the final consonant. To remove the effect of sequence learning, the recordings were prepared with five different pseudo-random orderings of the stimulus items for the complete test, with corresponding response sheets. The four 20-item pages were roughly equivalent in terms of the type of material included. The test items were spoken in the carrier phrase, "Can you hear. . . . . .clearly?" by a male speaker. Each test was preceded by five practice items for familiarization.

The FAAF's format offered the listener three foils (opportunities for error), one on each of the two features, and one on their combination. The features had different intrinsic difficulty, and single-feature contrasts were by probability theory more difficult than double-feature ones. The items were presented against a steady noise background with the same long-term frequency spectrum as the speech. For obtaining normative data, different S/N ratios were used i.e., -12.5, -10.0, -7.5, -5.0, -2.5, 0, 2.5 and 5 dB. The scoring was done on a response sheet by noting down the total number of correct scores or by error types.

No significant differences were found between the scores for the two ears within each S/N ratio condition or over all the S/N ratios. There was no right ear advantage. This test also indicated word frequency effects upon phonemes in initial position and effects of imageabilty upon phonemes in final position. However these effects were not large enough to undermine the use of FAAF as an acoustical phonetically structured material reflecting the analysis of auditory information.

39. *German Sentence Test*

This test was developed by Kollmeier and Wesselkamp, 1997 to assess speech intelligibility. The discrimination function was measured for each of the 324 sentence i.e., the number of correctly repeated words divided by the total number of words in the sentence as a function of the S/N ratio. The results from forty normal hearing listeners aged between 19 to 31 years were averaged for each S/N ratio and each sentence. Signals were presented monaurally with headphones in a sound insulated booth. Their task was to orally repeat each sentence presented over headphone. A handheld computer was used with a LCD-touch screen on which the target sentence was displayed. Each correctly identified word for each subject was stored by the computer. Along with the digitized sentences, a speech-simulating continuous noise was stored digitally. The noise was generated by superimposing all words of a monosyllabic rhyme test produced by the same speaker. The long-term spectrum of the sentence material was similar to the noise spectrum. The noise was presented at a level of 65 dBSPL.

In a pilot experiment with four listeners, the S/N ratio was determined which yielded an average total intelligibility of the speech material of 50%. In the first experiment, sentence intelligibility was measured with twelve subjects with all 324 sentences at this S/N ratio. All the sentences were grouped into six sets. A second experiment was performed with a different group of twelve subjects for all 324 sentences. In a third experiment, another group of twelve subjects measured the intelligibility of all 324 sentences. The SNR for the third experiment was determined such for that each set of sentences the three measurement points sampled the region between 20% and 80% sentence intelligibility for all the experiments. Each effective sentence was presented once to each subject. A sentence-specific discrimination function was obtained. Performance-intensity curves were measured for each individual sentence. The twenty test lists, were highly equivalent with respect to their performance-intensity curves, the number of words within each test list, the number of phonemes within each test list, and approximately the frequently distribution of the German language. The discrimination function obtained with each list were shown to be very close to the "ideal" discrimination function.

This test could be used to assess the effective impairment in noise for hearing impaired clients and the possible use of a hearing aid. It could be used to assess the training progress of children wearing cochlear implants. Another application would be assessment of binaural interaction in normal and hearing impaired subjects which can be used for assessing the benefit of bilaterally worn hearing aids.

The review of literature indicates that there are a host of tests available to establish the speech identification ability of a person. Several of these tests are modifications of tests developed earlier. The modifications were incorporated to overcome the drawbacks of existing tests.

To increase the complexity of the tests, either more complex speech material such as synthetic sentences have been used, or the introduction of a competing signal has been suggested. Most of the tests have several lists, which have been evaluated for their evaluation. While most tests have been evaluated on normal hearing subjects, a few such as CID W-22, NU Auditory No. 6, CUNY NST, Nonsense Syllable test and Maryland CNC Test have been evaluated on both normal as well as hearing-impaired subjects.

SPEECH IDENTIFICATION TESTS DEVELOPED IN INDIA

1. Phonetically Balanced (PB) words list in Hindi

Development of spondee and phonetically balanced word lists in Hindi was done by Abrol (1971, cited in Nagaraja, 1990). He analyzed 800 commonly used words of Hindi for syllabic constructions. The majority of the words were found to have a CVC structure. The frequency of the initial and final consonants were rated on the basis of frequency counts for all consonants as done by Ghatage (1964, cited in Nagaraja, 1973). Similarly the frequency of the vowels was also computed. Finally the familiarity of words was rated, according to their frequency of occurrence. Two lists of fifty words each were prepared based upon the fr'equency counts and familiarity of the words. No word was common to both the lists and also two lists of thirty-eight spondee words each were prepared from most commonly used words. Thirty normal subjects with SRT ranging from 10 to 30 dB were studied by presenting the material with a carrier phrase "Say the word" at 10 dB above the presentation level of the test word. Two seconds were given for the subject to respond. At 10 dB above SRT slightly more than half of the population repeated 90% of words and at 30 dB above SRT all the subjects repeated all the words presented to them (i.e. 100% scores were achieved). The optimum for Hindi PB was tentatively kept at 20 dB above SRT.

Some of the limitations of the Abrol's study are as follows:

- Practice effect was not maintained
- SRT levels were not maintained
- Articulation curves were not given.

2. Speech Perception Test in Tamil, Telugu and Malayalam

Development of hearing and speech test materials based on Indian languages (Tamil, Telugu and Malayalam) was done by Kapur (1971, cited in Nagaraja, 1973). Except for the nature of the material used in the construction of these tests and methods for the selection, methodology for tests in all these three languages were similar. In Tamil language though he succeeded in collecting the familiar monosyllables, the list failed to represent all the sounds, which do occur in Tamil language and are used as distinctive feature in the perception of speech in today's Tamil. Speech audiometric tests in Malayalam was done by selecting words which were very common, for both SRT and PB word lists as very few monosyllabic words were available. Two hundred disyllabic words were found to be most familiar. Six subjects with normal hearing and otological findings were selected for the study and were given the 200 disyllabic words. The recorded words were presented at threshold +4, +2, 0, -4 and -6 dB relative to the pure tone average thresholds of the subject. Thirty-five familiar spondee words were selected after rating for familiarity. For PB the maximum score of 97 percent was obtained at 45 dB.

3. Speech Discrimination Test material in English for Indians

Swarnalatha (1972) also developed and standardized speech discrimination test material in English for Indians apart from developing an SRT test. For the speech discrimination tests for adults, 200 monosyllabic words by PAL and another 200 monosyllabic words of CID Auditory Test W-22 were combined and used after eliminating the common words in the two tests. For developing the children's speech discrimination test, 150 monosyllabic words from the kindergarten PB word lists developed by Haskins (1949) were used. Two-hundred adults and 200 children rated the monosyllabic words. The procedure for familiarity was the same as the SRT test.

No data on frequency of occurrence of phonemes in Indian English was available. Hence, the relative frequency of occurrence of phonemes obtained by analyzing telephone conversation (Fletcher, 1965) was used as a basis for preparing the monosyllabic words. Finally, there were four monosyllabic word lists, which were phonetically balanced, two lists of equal familiarity for adults and two lists of equal familiarity for children. The recording procedure and test environment was same as discussed earlier in the SRT procedure. Fifty-six adults and children were selected for standardization. The following instructions were given: "You are going to hear a list of twenty-five one syllabic words with which everybody is familiar. You are to repeat each word into the microphone as best as you can. If you are not sure, guess. Do you have any questions?

Each subject was tested only at one intensity, one in the right ear using the list I and then in the left ear using the list II at a different intensity. The intensities at which the two lists were administered were (0, 10, 20, 30 35, 40, 45 and 50) dB above the subjects PTA. The responses were then noted down. Similar procedure was used for children. The subjects were instructed to respond orally. The tester bias in scoring had not been controlled. Since each list consisted of twenty-five words, each word was given a weightage of 4%. Therefore, each error was penalized twice as much as it would have been, if the list had contained fifty words. 100% discrimination score was obtained at 42 dB (re: PTA 10dB) and 45 dB (re: PTA 13 dB) for adults and children respectively.

The problem with the test was that it had only two lists with one form each (i.e., no scramblings were available). When one was interested in determining the P-I functions, the same lists had to be used repeatedly, which could bring in practice effect. In addition, list equivalency had not been statistically established. Yet another practical difficulty with the use of the test was the non-availability of the recorded version of the lists.

4. Hindi PB list for Speech Audiometry and Discrimination Test

De (1973) developed spondee and PB word list in Hindi and claimed that it could be used all over India. The sources for obtaining the test material were taken from Hindi primers, Hindi daily newspapers, day-to-day conversational, Hindi in the Armed Forces and commonly occurring Hindi words from Hindi dictionary. The material was transcribed phonetically by using a method of broad transcription.

The preparation of the lists began with collection of all the CVC, VC and CV types of monosyllabic words and they were compiled into a list. From this list, words that were unfamiliar or difficult were discarded. The remaining words were used to prepare eighteen lists of fifty words each. All the lists were phonetically balanced. The lists were tested on normal Hindi speaking subjects. The number of subjects used in the study was not mentioned. The words that were constantly missed were replaced by suitable easier words. After initial trials a final set of six 50-word PB lists of Hindi words were accepted for clinical trial and use. Standardization of the test material was obtained by carrying out clinical tests with these tests. A few minor alterations were done as a result of these trials so as to obtain a fair degree of uniformity amongst the lists.

The audiometer used was Allison's 4C type audiometer. The tape deck was used to record and reproduce the word lists. Recordings were done by an experienced male talker who used the carrier phrase 'boliye' before pronouncing each word of the lists. The testing was done in a sound treated room. Individuals were tested through earphones. They were asked to repeat the words, as they perceive them. The score was kept by the operator and the same expressed as percentage of words heard correctly or percent discrimination score. The results were as follows:

> All normal persons whose mother tongue was Hindi irrespective of education and service background produced a discrimination score of 98-100% at the most comfortable level (MCL).

> All normal persons whose mother tongue was a language other than Hindi but who had some knowledge of Hindi language produced an average score of 92-96% at MCL.

> Discrimination score in pure conductive deaf patients were the same as normal persons under similar circumstances. The MCL in these cases were considerably higher.

> Early otosclerotics discriminated like normal persons. Advanced otosclerotics showed a tendency to produce discrimination loss depending on the degree of deafness.

It was felt that in a multilingual country like India, Hindi speech material could be accepted for routine speech audiometry for all persons who have even minimum knowledge of Hindi (De, 1973).

5. *Development of Synthetic Sentence Identification Test in Kannada*

Nagaraja (1973) constructed synthetic sentences using most commonly used words in Kannada language. Ten first-order sentences and ten second-order sentences were constructed. To make the task more difficult these sentences were recorded with a continuous competing speech message. First, three lists of 1^{st} order and three lists of 2^{nd} order sentences were presented to thirty normal hearing subjects for getting performance pattern and to find the presentation level and message-to-competition ratio (MCR) level at which performance was maximum.

As this tape was very lengthy and time consuming and also as fatigue might have affected the performance of the subjects, a second tape was prepared. It consisted of one list of first order sentence, another list of second order sentence on one track and competing speech on the other track. This tape-2 was presented to thirty normal hearing subjects for computing normative data. Next it was administered to eight conductive hearing loss cases, four mixed hearing loss cases, ten sensori-neural cases and two high fi-equency hearing loss cases to study their performance on this test. PB word test was administered to all those cases and normals who knew English and Kannada language. The data obtained on this was compared with that of SSI test.

To compare the performance of four subjects of different audiogram pattern in terms of area, PI function curves were drawn for PB and SSI scores obtained at different intensity levels. The reliability of the test was computed by analyzing the data obtained on two different days for the same test on ten randomly selected subjects. The results of this test were as follows:

- The performance on the SSI test varied directly with the level of presentation and inversely with MCR.
- Normals obtained maximum performance scores on SSI test at 40 dBSL and at 0 dB MCR level.

- > There was no significant different between normals and conductive loss groups on the SSI second order sentences when compared to the SSI first order sentences.
- > The sensori-neural hearing loss group showed significant difference in performance when compared to normals and other clinical group except the conductive loss group on the SSI second order.
- > The mixed hearing loss group showed significant difference in performance when compared to the normals and other clinical group on the SSI 2^{nd} order sentences except for gradually sloping curve.
- > The high frequency hearing loss group differed significantly in performance from clinical group except in gradual sloping curve from normal group.
- > The comparison of performance on the PB word lists and the SSI test did not yield any valid conclusions, as a comparison of an SSI test in Kannada with an English PB test may not be appropriate.
- > The SSI first order test seemed not be to as valid as the SSI second order test as the former showed meaning discrepancies in the results.
- > The SSI second order sentences were a valid clinical tool and may be used as a test for speech discrimination especially with subjects who know only Kannada.

6. Common Speech Discrimination Test for Indians

Mayadevi (1974) attempted to construct a speech discrimination test that could be used with the speakers of all Indian languages. She chose the common monosyllables of CV combination found common in Indian languages. These monosyllables were selected from phonetic readers of Indian languages written by linguists. The list of monosyllables was not phonetically balanced, as all the sounds were in different proportions in different languages. An intelligibility test was carried out by presenting six lists at different levels to ten normal subjects. The materials were tape recorded in a sound treated room spoken by a twenty-one year old female whose mother tongue was Kannada. She was however proficient in other languages which included Hindi, English, Telugu, Tamil, Sanskrit, Marathi and Konkani. Each monosyllable was recorded using a carrier phrase 'i:ga idannu he:li'. The interstimulus interval was 10 sec. Similarly six English PB list standardized by Swarnalatha (1972) were recorded by the same speaker, for use in listing concurrent validity.

The following instructions were used: "Now you are going to hear in you right or left ear some speech sounds like ba, ma etc. They are preceded by a Kannada phrase 'i:ga idannu he:li'. You need not repeat the phrase again, but you have to repeat the syllables which you hear in the end". The instructions were translated to different languages depending on the subjects. The level of presentation was kept constant i.e., at definite sensation level above the individuals' pure tone average (PTA) level. The test procedure was first standardized by presenting the test list of twenty monosyllables on thirty normal ears and comparing their verbal and written responses. With clinical population, three responses were elicited for the same sound. As the testing was done in a one-room situation and no talkback system was used, oral responses were chosen as the chief criteria. The validity and reliability was checked on normals and clinical groups (conductive loss, mixed loss, sensori-neural loss and high frequency hearing loss).

The results of this study was follows:

> Normals obtained optimum scores ranging from 90 to 100 on this test at 40 dBSL (ref. PTA)

> The performance of normal speaking different Indian languages followed the same pattern.

> There was no difference in the scores of verbal and written responses of the subjects.

> Sex difference in terms of performance was found insignificant.

> There was no difference in the performance of the right and left ear on this test.

^ The performance of the SN loss cases was different from that of normal, conductive and high frequency loss cases. But their performance was not different from the mixed loss cases.

> The conductive loss group resembled normal and high frequency loss casesin their performance.

> Performance of mixed loss cases differed from that of high frequency loss cases.

> The high frequency loss cases performed like normals.

As discrimination testing is an important test battery for differential diagnosis, this test could be used as a common speech discrimination test in all the clinics.

7. *PB Test in Tamil*

Samuel (1976) developed meaningful, familiar monosyllabic word lists, which were phonetically balanced. The familiar words were collected from Rajaram's (1972) study of "Recall vocabulary in Tamil" and the words were administered on ten people to ensure familiarity. Later, four lists of twenty-five words each were constructed, using the functional load of sounds in Tamil as reported by Meenakshisundaram (cited in Samuel, 1976). All the word lists were recorded on a magnetic tape using KBR 71 Deck Type tape recorder with an appropriate carrier phrase in Tamil. It was then administered on ten subjects to ensure the equivalency of all the four lists. After verifying the equivalency of all the lists, thirty subjects selected from the student population of the Mysore University who knew Tamil, were tested to plot the articulation gain function of the lists. It was found to be highly reliable on test retest situations. The concurrent validity of this test was checked by presenting English PB list to normals having knowledge of English and analyzing the scores. Normals obtained optimum scores at 35 dB SL (re SRT). All the four lists were found to be essentially equivalent and can be used interchangeably.

8. Speech Identification in Manipuri

This study was given by Tanuza (1984) who aimed at constructing and standardizing a speech identification (SI) test material in Manipuri. To construct the list, monosyllabic words were collected from phonetic books, magazines, books and normal conversational speech. A familiarity test was done which has been previously described. Words rated as most familiar were collected and from them 100 monosyllabic words were chosen for the construction. Four monosyllabic word lists were developed which contained twenty-five items. Five scramblings were made of the list to avoid practice effect. The monosyllabic words were not phonetically balanced as studies were not available. The recording procedure, instrumentation and test environment was the same as discussed earlier in her SRT procedure. The lists were presented at 5, 10, 20, 30 and 40 dB above the subjects' established SRT. Five subjects had to repeat the words and the number of correct responses were noted down. It was then converted into percentage at every intensity levels. The maximums score was attained at 40 dBSL. The test cannot be considered to be standardized due to the small number of subjects it has been evaluated on.

9. Speech Identification test in Bengali

Ghosh (1986) besides constructing an SRT using polysyllabic words also constructed monosyllabic word lists for measuring speech discrimination. Seventy-five monosyllabic words were collected for assessing speech discrimination ability. These words were rated to be most familiar. The recording procedure, subjects' selection, instrumentation and test environment was same as previously discussed in his SRT procedure. For the standardization, the monosyllables were divided into three lists and each of the lists was further randomized into five lists. Each of these randomized lists was presented at (5, 10, 20, 30, 40) dBSL (ref SRT). The subjects were asked to repeat the monosyllabic words heard and the number of correct responses were converted into percentages at every intensity level. The maximum speech discrimination was obtained at 30 dBSL (ref SRT) for list A & C. For the list B, the maximum level was attained at 40 dBSL (ref SRT). As done with his SRT test, the evaluation of the test was established on a very small population.

10. Speech Identification Test in Gujrati

Mallikarjuna (1990) developed phonetically balanced (PB) monosyllabic words in Gujrati language. Based on Egan's (1948) three point criterion, 150 monosyllabic Gujrati words were selected from Shukla's (1975) 'The Gujrati Vocabulary of Students of Standard I &V in Surat district. Three lists of fifty PB words each were formed. The speech discrimination scores were obtained at various sensational levels above the SRT. The findings were on par with that of Hirsh et al. (1952) on Auditory Test W-22 and Samuel (1976) on Tamil PB list. The three PB lists were matched and could be used as a valid test for speech discrimination for Gujrati speaking population.

11. High Frequency Speech Identification Test for Hindi and Urdu Speakers

This test was developed by Ramachandra (2001), with the objective to develop a list of meaningful, familiar, high-frequency monosyllabic CVC words common to Hindi and Urdu speakers for the establishment of speech identification scores. A total of eighty-five words were collected from various sources and only fifty high-frequency words which were rated as "most familiar" by thirty native Hindi and thirty native Urdu adult speakers were selected for the study. The list was further divided into two categories. The first category consisted of the high frequency phoneme in the initial consonant position of the word and the second category consisted of the high frequency phoneme in the final consonant position of the word. Two randomized versions of the similar words (categories of high frequency phonemes tested in the initial and final position) were recorded on a digital tape. The recorded speech material was fed to the audiometer.

Thirty-five Hindi and thirty-five Urdu speaking adults with normal hearing were tested in a closed-set condition in order to obtain the Performance-Intensity (P-I) function. The list was presented at 0, 5, 10, 15, 20, 25, 30, 35 and 40 dBSL. 50.8% of the correct response of the Hindi speakers and 50.4% of correct response for Urdu speakers was obtained from the P-I function at 20 dBSL. A test of significance showed speakers was obtained from the P-I function at 20 dBSL. A test of significance showed that there was no significant difference between the two groups of Hindi and Urdu speakers for 0, 5, 10, 15, 20, 25, 30, 35 and 40 dBSL at 0.05 level of significance.

The second objective of the study was to compare High Frequency speech identification scores of closed-set conditions with speech identification scores obtained on common speech discrimination test (SDT) for Indians developed by Mayadevi (1974). Speech identification scores were obtained for fifteen normal hearing adults (Hindi and Urdu speakers) and fifteen adults with high frequency loss (mild loss with dip at 4kHz or 6kHz, mild sloping loss, moderate to moderately severe sloping loss). The fifteen subjects with normal hearing scored 100% on a closed-set, open-set and speech identification scores obtained on the Common SDT for Indians. For the fifteen subjects with a high frequency sloping hearing loss a high positive correlation coefficient was found for

> High frequency speech identification scores in open-set versus closed-set condition.

> High frequency speech identification scores in open-set condition versus speech identification scores obtained on the Common SDT for Indians developed by Mayadevi (1974).

> High frequency speech identification scores in closed-set condition versus scores obtained on the Common SDT for Indians. There was a significant difference between all the three conditions.

This common high frequency monosyllabic word list for Hindi and Urdu speakers was both valid and reliable.

12. A High Frequency Kannada Speech Identification Test (HF-KSIT)

Mascarenhas (2002) developed a speech identification material in Kannada for testing adults with a sloping hearing loss. She also checked the usefulness of the material on a sample of hearing-impaired adults with a sloping high frequency hearing loss. The material for the study was obtained from a book compiled by Ranganath (1982) which contained a list of frequently occurring words in Kannada. Only bisyllabic, and trisyllabic words with phoneme distributed in the frequency range from mid towards high frequencies (i.e., above 1kHz) were chosen. A minimum of 100-150 words were selected which was evaluated for familiarity by fifteen native adult Kannada speakers. The words rated as "most frequently used" and twelve "frequently used" words were chosen. They were compiled to form three subtests of words. Each subtest contained twenty-five words, with all the three subtests having equal distribution of the high frequency consonants. For the sentence subtest, each sentence contained one key word, which was used in the word subtest. Three sets of sentences were constructed with each having nine sentences. All the sentences were in the assertive case, with an average length of 7 ± 2 syllables. The same fifteen subjects also evaluated the sentences for their familiarity and grammatical correctness. Three lists were thus constructed. Each list included a word subtest and a sentence subtest. The recording was done in a computer by a female speaker whose mother tongue was Kannada. A 1 kHz calibration tone was recorded prior to each list. The material was copied onto a compact disc. A Philips CD player was used to present the recorded speech. The output was routed through an audiometer Madsen OB822 and given to TDH-39 headphones. The testing was done in a sound treated double room.

The speech identification tests i.e., the Common Speech Discrimination Test for Indians, Mayadevi (1974) and the HF-KSIT were administered on normal hearing adults at 40 dBSL. An open-set response in the form of an oral response was obtained. For the sentence subtests, the subject had to repeat the entire sentence. Word and phoneme scoring was used for the words in the word subtests and for the keywords in the sentence subtests.

The test material was administered on thirty hearing-impaired adults with a sloping hearing loss. The Common Speech Discrimination Test for Indians (CSDTI) (Mayadevi, 1974) was administered at 40 dBSL with reference to the SRT. For six cases, the presentation level was 30 dBSL (four due to audiometric constraints and two due to lowered uncomfortable loudness levels). Following this the HF-KSIT was administered at the same levels. Further the HF-KIT was administered on a pathological population without and with a suitable hearing aid at 40 dBHL.

The results of the test revealed that all three tests yielded similar scores on normal hearing subjects. There was no significant difference between the following:

> The word subtest and the sentence subtest in the normal hearing group;

> The word scores and phoneme scores in the normals;

> Between the HF-KSIT and CSDTI in the normal hearing population.

The high frequency hearing impaired group obtained poor scores on the word subtest compared to the sentence subtest, indicating that the former was more sensitive to detect their problem. They got poor scores when the word scores were used compared to when the phoneme scores were used. Thus, the word scoring procedure was recommended for individuals with a high frequency hearing loss. There was significant difference in the performance of the subjects for the word and sentence subtests in the unaided versus the aided condition for selection of hearing aids. Hence, HF-KSIT could be used as a part of diagnostic test battery as well as fitting procedure during selection of hearing aids.

All of these tests developed to assess the speech identification ability had some similarities in terms of the type of stimulus used, phonetic balancing, familiarity rating, carrier phrase, instruction, presentation mode, presentation level, response mode and scoring.

From the review of tests developed, it is evident that attempts have been made to develop tests in different regional languages. Though speech tests are available in approximately nine Indian languages, it can be noted that there are several Indian languages that do not have speech tests. Thus there is a need to develop speech recognition threshold tests or speech identification tests in these languages.

METHOD

The present study was canned out in two stages:

Stage I involved construction of the test material for the SRT and SI tests

Stage II involved obtaining normative data for the SRT and SI tests

STAGE I: Construction of the test material

Familiar bisyllabic and monosyllabic words were selected from Kelkar's (1994) book titled "Phonemic and Morphophonemic Frequency Count in Oriya" as well as periodicals, magazines and individuals fluent in the language. Bisyllabic and monosyllabic words, representing all the phonemes of the language were selected.

To ensure familiarity of the words, they were given to ten normal adults in the age range of 20-45 years who were fluent in Oriya. All the subjects knew the dialect as spoken in Bhubaneshwar. The subjects were asked to rate the words on a three-point scale of familiarity (i.e., most familiar, familiar and unfamiliar). Words most familiar to 90% of the subjects were included in the construction of the tests. Using these words, the material for the Speech Recognition Threshold (SRT) and Speech Identification (SI) tests were developed. For the SRT test, bisyllabic words were used and for the SI test, monosyllabic words were used. A linguist who was familiar with Oriya was consulted to confirm whether the words selected were bisyllables or monosyllables.

For the SRT testing two lists (List A and List B) were developed consisting of twenty-five bisyllabic words each (Appendix A). It was ensured that all the phonemes of the language were included in each of the lists.

For the SI testing, two lists (List 1 and List 2) were developed consisting of fifty monosyllabic words each (Appendix B). The phonemic balance was maintained in each of the lists. The first half and the second half of each of the lists were also phonemically balanced. The phonemic balance was done based on the frequency of occurrence of phoneme in Oriya, published by Kelkar (1994).

STAGE II: Obtaining normative data

Forty adults were selected for obtaining normative data. The subjects used in the pilot study were excluded from this group. The subjects selected met the following criteria:

(i) All the subjects had normal hearing.

(ii) They did not have any history of otological problems.

(iii) None of the subjects had any speech problems.

(iv) All the subjects knew the dialect of Oriya as spoken in Bhubaneshwar.

Instrum entation:

A two channel diagnostic audiometer (GSI 61), which was calibrated in accordance with ANSI 1989, was used. The test materials were presented using a monitored live voice (MLV) procedure. The VU meter deflection was maintained at zero. The material was presented by an adult female Oriya speaker who had experience in MLV testing. The output of the audiometer was given to TDH-39 earphones housed in MX-4/A ear cushions.

Test environment:

The test was carried out in a sound treated double room situation. The ambient noise levels were within permissible limits as recommended by ANSI (S.3.1-1991) (cited in Wilber, 2002) standards.

Procedure:

All the subjects were subjected to routine audiological testing by obtaining air conduction and bone conduction thresholds for the frequencies 250-8000 Hz and 250-4000 Hz respectively. Only those who obtained normal hearing were selected for further evaluation. The better ear was considered as the test ear for further evaluation of each subject.

Instruction:

The subjects were given the following instruction in Oriya:

Instruction for SRT testing: "You will hear some words through your headphones. Listen carefully to each word and repeat them. The words will get softer. If you are not sure, you can guess".

Instruction for SI testing: "You will hear some words through the headphones. Listen carefully to each word and repeat them".

Normative data for SRT test material

Using the material developed for SRT each of the subjects were tested for the following:

a. Establishment of SRT

b. Identification of the SRT word lists

a. Establishment of SRT

The SRT method proposed by Martin and Stauffer (1975) was used to evaluate the speech recognition threshold. The procedure was as follows:

The subject was presented one spondee word at 50 dB HTL. If the response was correct, the level was attenuated in 10-dB steps, with one spondee presented at each level, until a word was missed. If an incorrect response was obtained, a second spondee was presented at the same level and attenuation was continued in 10 dB steps until two spondees were missed at one level. If the subject missed the first spondee presented at 50 dB HTL, the level was increased in 20-dB steps until a correct response was obtained. The attenuation in 10-dB steps was then begun. The "start level" was defined as 16 dB above the level at which the subject first missed two spondees. Two spondees were presented at this level and at each 2-dB decrement. A tally was kept of the responses to determine if the subject had repeated five of the first six words correctly. If this condition was met, 6 dB was added to the start level and the 2-dB decrements were then initiated. The decrements were continued until five of the last six words were missed by the subject. The SRT was then determined by subtracting the number of correct responses from the start level and a correction of ldB was added.

The SRT was established for both the lists (List A & List B).

b. Identification of the SRT word lists:

The identification of the SRT words was established at three different intensities, 10 dBSL, 15 dBSL and 20 dBSL. The subjects were instructed to repeat the test words and the responses were noted down. It was ensured that all the subjects heard the three intensities. At each intensity both lists (A & B) were presented. The words in the two lists were randomized in order to avoid a familiarity effect.

Normative data for speech identification material:

Each list (List 1 and List 2) was presented at two different intensities i.e., 20 dB and 40 dB above the SRT. All the subjects were tested with both the lists at both the intensity levels. Each subject was tested in the ear where the SRT was established. As with the SRT lists, the words were randomized to avoid the familiarity of the words affecting the scores.

Scoring of responses:

The responses were recorded on a score sheet (Appendix C) for both the SRT and SI tests by the tester. Each correct response was given a score of 1 and an incorrect response was scored as 0. The number of correct responses were noted down for each of the lists. The results were subjected to statistical analysis.

RESULTS AND DISCUSSION

The present study aimed at constructing a speech recognition threshold test and a speech identification test in Oriya. It also aimed at obtaining normative data on adults. Forty normal hearing adults in the age range of 20-45 years who were fluent in Oriya and who knew the dialect as spoken in Bhubaneshwar were evaluated.

Statistical analysis was carried out to obtain the following information:

A. With respect to the SRT Test

i) Correlation of SRT with pure tone average (PTA) for frequencies 500 Hz, 1 kHz and 2 kHz using Pearson's correlation,

ii) Comparison of the SRTs obtained with two lists using the 't'-test,

iii) Check the equality of the two SRT lists in terms of intelligibility using the 't'-test.

iv) Effect of presentation level on the intelligibility of the SRT lists using the 't'-test.

B. With respect to the SI test

i) Check the equality of the two speech identification lists using the 't'-test.

ii) Check the equality of the half lists using the 't'-test.

iii) Check the effect of presentation level on the intelligibility of the lists using the 't'-test.

A. Speech Recognition Threshold Test

i) Correlation of SRT with PTA

The pure tone average for the frequencies 500 Hz, 1 kHz and 2 kHz was obtained at 8.92 dBHL with a SD of 2.95. The PTA was correlated with the SRT of both the lists.

Table-1: Correlation of SRT with PTA

	PTA	SRT (list A)	SRT (list B)
PTA	1.00	0.79**	0.79**
SRT (list A)	-	1.00	0.96**
SRT (list B)	-	-	1.00

**Correlation is significant at 0.01 level

Pearson's correlation was done which revealed a good correlation between the PTA with SRT (List A and List B). The correlation was significant at 0.01 level (table 1).

ii) Comparison of SRT using the two lists

Table 2 shows the mean SRT of both the lists (List A and List B). The mean SRT using list A was 10.40 dB with a SD of 3.49 and that of list B was 10.65 dB with a SD of 3.41. The above findings are in consonance with Swarnalatha (1972) and Ghosh (1986). They obtained an SRT at 9 dBHL (re: PTA 10 dB) and 12 dBHL respectively. However, Hirsh et al. (1952) obtained SRT at -5.7 dBHL and -2.3 dBHL for CID W-l and CID W-2 respectively. Rajashekhar (1976) obtained an SRT at 1 dBHL for list I and 0.5 dBHL for list II and Tanuza (1984) obtained an SRT at 13 dBHL. The difference in the threshold obtained could be due to differences in the procedure used for obtaining SRT.

Further, paired sample t-test was done to find any significant difference between the SRT obtained using each of the lists. It is evident from table 2 that no significant difference was found between the SRTs of the two lists. This reveals that both the lists are able to establish similar SRTs.

Table-2: Mean, SD and 't' values of SRT for lists A and B

Lists	N	Mean	SD	't'
List A	40	10.40	3.49	-1.75*
ListB	40	10.65	3.41	

*Not significant

iii) Equality of the two SRT lists in terms of intelligibility

The intelligibility of the two SRT lists was established at three intensity levels 10 dBSL, 15 dBSL and 20 dBSL. The equivalence of the lists (A and B) was checked at each of these different intensities. Paired sample statistics was done to determine this. Table 3 shows the mean, SD and 't' values of the two lists at each of the intensities.

Table-3: Mean, SD and 't' values of lists A and B across different presentation levels

Presentation level	Lists	Mean	SD	't'
10 dBSL	List A	22.92	1.75	0.41*
	ListB	22.82	1.58	
15dBSL	List A	23.95	1.23	-0.48*
	ListB	24.05	0.98	
20 dBSL	List A	24.60	0.59	-1.40*
	ListB	24.72	0.45	

*Not significant

Table 3 indicates that the mean values and SD for list A and list B, at each intensity, was more or less the same. The 't'-test revealed no significant difference between the lists at each of the presentation levels. Thus, it can be construed that the two lists are equivalent. Either one of them can be utilized to determine the SRT of a client.

iv) Effect of presentation level on the intelligibility of the SRT lists

The identification of the SRT lists was carried out at three intensities viz., 10 dBSL, 15 dBSL and 20 dBSL. Table 4 gives the comparison of different presentation levels for lists A and B.

Table-4: Mean, SD and 't' values across different presentation levels

Presentation level	Mean		**SD**		't'	
	List A	ListB	List A	ListB	List A	ListB
10 dBSL 15 dBSL	22.92 23.95	22.82 24.05	1.79 1.23	1.58 0.98	-4.55*	-5.74*
10 dBSL 20 dBSL	22.92 24.60	22.82 24.72	1.75 0.59	1.58 0.45	-6.34*	-7.67*
15 dBSL 20 dBSL	23.95 24.60	24.05 24.72	1.23 0.59	0.98 0.45	-3.34*	-4.07

*Significant at 0.01 level

This table shows the mean SD and 't' values across the presentation levels for the two lists. Comparison of the mean scores and SD indicates that the mean scores increased with increase in presentation levels. The SD decreased with increase in intensity indicating that at higher presentation levels, the variance was lesser. The 't'-test revealed a significant difference at the 0.01 level across different intensity levels. This indicates that the intelligibility of the lists improved significantly, with increase in presentation level. A study conducted by Hirsh et al. (1952) also reported that with increase in presentation levels the identification scores for bisyllabic words increase.

B. Speech Identification Test

i) Equality of the two speech identification lists

The equivalence of the two SI lists was checked at different intensities (viz., 20 dBSL and 40 dBSL). Paired sample statistics was done to determine if the two lists (list 1 and list 2) were equivalent at different intensity levels. Table 5 shows the mean, SD and 't'-values of the two lists at each of the intensities. The 't'-test indicated no significant difference between the lists at each of the presentation levels. Thus, the two lists are equivalent. This finding is in consonance with Hirsh et al. (1952), TilVman, Carhart and Wilber (1963), Swarnalatha (1972). Each of these investigations also obtained equivalence between their lists.

Table-5: Mean, SD and 't' values of both the lists across different presentation levels

Presentation level	List	Mean	SD	
20 dB	List 1	41.27	2.27	-0.14*
	List 2	41.05	2.18	
40 dB	List 1	49.90	0.37	0.15*
	List 2	49.85	0.36	

*Not significant

ii) Equality of the half lists

The two lists (List 1 and List 2) were subdivided into two half lists which gave rise to four half lists. The equivalence of the half lists at two presentation levels (20 dB and 40 dB above the SRT) was checked using the 't' test.

Table 6 and 7 shows the comparison of mean, SD and 't' values of the four half lists at 20 dB and 40 dB above the SRT respectively. The 't' values revealed no significant difference between each of the lists. Thus, all the four lists are equivalent at both presentation levels. Similar findings were also reported by Samuel (1976) and Tanuza (1984) in the tests developed by them. Further, it was noted that the half lists were equal to the full lists. This was statistically significant at 0.01 level.

Table-6. Mean, SD and 't' values of four half lists at 20 dB above SRT

Half lists	Mean	SD	t
First Half list (List la)	20.60	1.15	-0.94*
Second Half list (List lb)	20.70	1.18	
First Half list (List la)	20.60	1.15	1.15*
Third Half list (List 2a)	20.47	1.32	
First Half list (List la)	20.60	1.15	0.22*
Fourth Half list (List 2b)	20.57	0.98	
Third Half list (List 2a)	20.47	1.32	-0.78*
Fourth Half list (List 2b)	20.57	0.98	
Second Half list (List lb)	20.70	1.18	1.85*
Third Half list (List 2a)	20.47	1.32	

*Not significant

Table-7: Mean, SD and 't' values of four half lists at 40 dB above the SRT

Half lists	Mean	SD	t
First Half list (List la)	24.92	0.26	-1.43*
Second Half list (List lb)	24.97	0.15	
First Half list (List la)	24.92	0.26	1.00*
Third Half list (List 2a)	24.90	0.30	
First Half list (List la)	24.92	0.26	-0.44*
Fourth Half list (List 2b)	24.95	0.22	
Third Half list (List)	24.90	0.30	-0.81*
Second Half list (List 2b)	24.95	0.22	
Second Half list (List lb)	24.97	0.15	1.77*
Third Half list (List 2a)	24.90	0.30	

*Not significant

iii) Effect of presentation level on the intelligibility of the lists

Table 8 shows the mean, SD and 't' values of lists 1 and 2 at different presentation levels i.e., 20 dB and 40 dB above the SRT. For lists 1 and 2, the mean scores increased as the presentation level increased. The SD was more at the lower sensation level indicating greater dispersion of scores. The SD was lesser for higher sensation level reflecting lesser variance. The results are in consonance with Swarnalatha (1972), Mayadevi (1974), Tanuza (1984) and Ghosh (1986).

The paired 't'-test revealed a highly significant difference between the two intensities viz., 20 dB and 40 dB above the SRT. Thus it is recommended to administer the lists at 40 dB above the SRT.

Table-8: Mean, SD and 't' values of both the lists at different presentation levels

Lists	Presentation level	Mean	SD	't'
List 1	20 dBSL	41.27	2.24	-23.8*
	40 dBSL	49.90	0.37	
List 2	20 dBSL	41.05	3.41	-25.32*
	40 dBSL	49.85		

* Significant at 0.01 level

The conclusion of the study is:

A. With respect to the SRT test

i) Both the SRTs correlated well with PTA at the 0.01 level of significance.

ii) SRT was established at 10.40 for list 1 and at 10.65 for list 2.

iii) The two lists were equivalent in terms of intelligibility.

iv) The intelligibility of the SRT lists increased significantly with increase in presentation level.

B. With respect to the SI test

i) The two speech identification lists were found to be equivalent.

ii) The four half lists were also equivalent.

iii) The intelligibility of the lists increased with increase in presentation level.

SUMMARY AND CONCLUSION

Assessing the speech perception abilities of an individual constitute an important part of hearing evaluation. Two tests which seem to be very sensitive in this regard are the speech recognition threshold (SRT) and the speech identification (SI) tests. A few of the SRT tests developed in India were by Rajashekhar (1976), Tanuza (1984), Ghosh (1986) and Mallikarjuna (1990). The SI tests developed for the adults were by Abrol (1971), Kapur (1971), Swaraalatha (1972), De (1973), Nagaraja (1973), Mayadevi (1974), Samuel (1976), Tanuza (1984), Ghosh (1986) and Mallikarjuna (1990). No speech test material for evaluating the SRT and SI abilities is available in Oriya, hence the present study was undertaken.

The present study was aimed at constructing a speech recognition threshold test and a speech identification test in Oriya. The material for the tests was obtained from Oriya literature. The material was subjected to a pilot study to check for the familiarity of the words. Only words that were highly familiar were selected. Using these words the following were constructed:

a) A Speech Recognition Threshold test containing two lists of bisyllabic words. Each list had twenty-five words representing the phonemes of the language.

b) A Speech Identification test containing two lists of monosyllabic words. Each list had fifty words which were phonemically balanced.

Normative data was obtained on forty normal hearing adults in the age range of 20-45 years, who knew the dialect spoken in Bhubaneshwar.

Using the SRT test, the following was checked:

> Correlation between the SRT and PTA for the frequencies 500 Hz, 1000 Hz and 2000 Hz

> Comparison of the SRTs obtained with the two lists

> Equality of the two SRT lists in terms of intelligibility

> Effect of presentation level on the intelligibility of the SRT lists.

The Speech Identification Test was tested to:

> Check the equality of the two speech identification lists

> Check the equality of the half lists

> Check the effect of presentation level on the intelligibility of the lists

The data collected on the subjects were statistically analysed. For each of the variables measure, the means and standard deviation were computed. To check the significance of difference between the mean scores, paired sample t-test was done.

The results of the present study are as follows:

A. With respect to the SRT test

1) The SRTs obtained using the two lists correlated well with the PTA. The correlation was significant at the 0.01 level.
2) SRT was established at 10.40 for list 1 and at 10.65 for list 2. No significant difference was found between the SRTs of the two lists. This reveals that both the lists are able to establish similar SRTs.
3) The speech identification of the two lists were equivalent. No significant difference was found between the lists at each of the presentation levels.

4) The intelligibility of the SRT lists increased with increase in presentation level. The 't'-test revealed a significant difference at the 0.01 level across the different intensity levels for list A and list B.

B. With respect to the SI test

1) The two speech identification lists were equivalent. The 't'-test revealed no significant difference between the lists at each of the presentation levels.

2) The half lists were also equivalent. No significant difference between each of the half lists was obtained. Also it was found that the scores obtained in the half list was equal to that of the full list. Hence, any of the half lists can be used to determine the speech intelligibility of an individual.

3) The intelligibility of the lists increased with increase in presentation levels. The 't'-test revealed a highly significant difference between the two intensities (20 dB and 40 dB above the SRT) at the 0.01 level. Hence, it is recommended to present the lists at a higher sensation level i.e., at 40 dB above the SRT.

These tests could be used to evaluate the speech perception abilities of individuals who know Oriya. The tests can also be used to assess the utility of different devices or therapy programs.

REFERENCES

Bench, J. (1987). Speech Audiometry in Australia. In M.Martin (Ed.), *Speech Audiometry* (pp.247-254). London: Whurr Publishers Ltd.

Bench, J., Kowal, A., & Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially hearing children. *British Journal of Audiology,* 13, 108-112.

Bench, J., & Bamford, J.M. (1979). *Speech-hearing tests and the spoken language of partially-hearing children.* New York: Academic Press.

Berger, K.W. (1969). Speech discrimination task using multiple-choice key words in sentences. *Journal of Auditory Research, 9,* 247-262.

Berger, K.W. (1978). Speech Audiometry. In D.E. Rose (Ed.). *Audiological Assessment.* New Jersey : Prentice Hall Inc.

Berger, K.W., Keating, L.W. & Rose, D.E. (1971). An evaluation of the Kent State University (KSU) speech discrimination tests on subjects with sensorineural loss. *Journal of Auditory Research,* 11, 140-143.

Bess, F.H., & Humes, L.E. (1995). *Audiology: The Fundamentals* (2nd Ed.). Baltimore: Williams & Wilkins.

Bilger, R.C. (1984). Speech recognition test development. In E. Elkins (Ed). Speech recognition testing by the hearing-impaired. *ASHA Reports,* 14, 2-15.

Bilger, R.C., Neutzel, J.M., Rabinowitz, W.M., & Rzeczowski, C. (1984). Standardization of a test of speech perception in noise. *Journal of Speech and Hearing Research,* 27, 32-48.

Black, J.W. (1957). Multiple-Choice Intelligibility Tests. *Journal of Speech and Hearing Disorders,* 22, 213-235.

Black, J.W. & Haagen, C.H. (1963). Multiple Choice Intelligibility Tests Forms A and B. *Journal of Speech and Hearing Disorders,* 28, 77-86.

Borden, G.J., & Harris, K.S. (1980). *Speech Science Primer: Physiology, Acoustics and Perception of speech.* Baltimore: Williams & Wilkins. A Waverly Company.

Brandy, W.T. (2002). Speech audiometry. In J.Katz (5th Ed.). *Handbook Of Clinical Audiology* (pp. 96-110). Baltimore: Lippincott Williams & Wilkins. A Wolters Kluwer Co.

Carhart, R. (1965). Problems in the measurement of speech discrimination. *Archives of Otolaryngology,* 82, 253-260.

Carhart, R. (1971). Observations on relations between threshold for puretones and for speech. *Journal of Speech and Hearing Disorders,* 36, 476-483.

Causey, G.D., Hood, L.J., Hermanson, C.L. & Bowling, L.S. (1984). The Maryland CNC Test: Normative Studies. *Audiology,* 23, 552-568.

Cox, R.M., Alexander, G.C. & Gilmore, C. (1987). Development of the Connected Speech Test (CST). *Ear and Hearing,* suppl., 8(5), 119S-126S.

Danhauer, J.L., Beck, D.L., Lucks, L.E., & Ghadialy, F.B. (1968). A Sentence Test for Audiological Assessment of Severe and Profound Losses. *The Hearing Journal,* 14, 26-33.

De, N.S. (1973). Hindi PB List for Speech Audiometry and Discrimination Test. *Indian Journal of Otolaryngology, 25,* 64-75.

DeBonis, D.A., & Donohue, C.L. (2004). *A Survey of Audiology: Fundamentals for Audiologists and Health Professionals.* Boston: Pearson Education, Inc.

Dermody, P., & Mackie, P. (1987). Speech Tests in Audiological Assessments at the National Acoustic Laboratories. In M.Martin (Ed.). *Speech Audiometry* (pp.255-278). London: Whurr Publishers Ltd.

Dubno, J.R., & Dirks, D.D. (1982). Evaluation of hearing-impaired listeners using a Nonsense Syllable Test. I. Test reliability. *Journal of Speech and Hearing Research,* 25, 135-141.

Dubno, J.R., Dirks, D.D., & Langhofer, L.R. (1982). Evaluation of Hearing-Impaired listeners using a Nonsense Syllable Test. II. Syllable recognition and consonant confusion pattern. *Journal of Speech and Hearing Reserach,* 25, 141-148.

Edgerton, B.J., & Danhauer, J.L. (1979). *Clinical implications of speech discrimination using nonsense stimuli.* Baltimore: University Park Press.

Elkins, E. (1984). Speech recognition by the hearing-impaired (Ed.). *ASHA Reports,* 14. Rockville, MD: ASHA.

Feeney, M.P., & Franks, J.R. (1982). Test-retest Reliability of a Distinctive Feature Difference Test for hearing aid. *Ear and Hearing,* 3, 59-65.

Foster, J.R., & Haggard, M.P. (1987). The Four Alternative Auditory Feature test (FAAF)- linguistic and psychometric properties of the material with normative data in noise. *British Journal of Audiology,* 21, 165-174.

Gardner, H.J. (1971). Application of high frequency consonant discrimination word test in hearing aid evaluation. *Journal of Speech and Hearing Disorders,* 36, 354-355.

Geffner, D., & Danavon, N. (1974). Intelligibility functions of normals and sensori-neural loss subjects on the W-22 lists. *Journal of Auditory Research,* 14, 82-86.

Ghosh, D. (1986). Development and Standardization of speech materials in Bengali language. *Unpublished Master's dissertation.* University of Mysore, India.

Givens, G.D., & Jacobs-Condit, L. (1981). Consonant identification in quiet and in noise in the normals and the sensorineural hearing-impaired. *Journal of Auditory Research,* 21, 279-285.

Goetzinger, C.P. (1972). Word discrimination testing. In J.Katz (Ed.) (2nd Edn.). *Handbook of Clinical Audiology.* Baltimore: Williams and Wilikins.

Goldinger, S.D, Pisoni, D.B., & Luce, P.A. (1990). Speech Perception and Spoken Word Recognition: Research and Theory. *Research on Speech Perception.* Progress Report No. 16, 3-63.

Griffiths, J.D. (1967). Rhyming Minimal Contrasts: A simplified diagnostic articulation test. *Journal of Acoustical Society of America,* 42, 236-241.

Hagerman, B. (1982). Sentences for testing Speech Intelligibility in Noise. *Scandinavian Audiology,* 11,79-87.

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.

Jerger, J., Speaks, C, & Trammell, J. (1968). A new approach to speech audiometry. *Journal of Speech and Hearing Disorders,* 33, 318-328.

Jirsa, R.E., Hodgson, W.R. & Goetzinger, C.P. (1975). Unreliability of half list discrimination tests. *Journal of American Audiological Society,* 1, 47-49.

Johnson, E.W. (1996). Confirmed retrocochlear lesions: Auditory test results in 163 patients. *Archives of Otolaryngology,* 84, 29-36.

Kalikow, D.N., Stevens, K.N., & Elliot, L.L. (1977). Development of a test of speech intelligibility in noise using sentence materials with connected word predictability. *Journal of Acoustical Society of America,* 61, 1137-1351.

Kasden, S.D. (1970). Speech discriminatin in two groups matched for hearing loss. *Journal of Auditory Research,* 10, 210-212.

Kelkar, A.R. (1994). *Phonemic and Morphophonemic Frequency count in Oriya.* Mysore: CIIL Printing Press.

Kollmeier, B., & Wesselkamp, M. (1997). Development and evaluation of a German sentence test for objective and subjective speech intelligibility assessment. *Journal of Acoustical Society of America,* 102(4), 2412-2421.

Kruel, E.J., Nixon, J.C., Kryter, K.D., Bell, D.W., & Lang, J.S. (1968). *Journal of Speech and Hearing Research,* 11, 536-552.

Kruger, B., & Mazor, R.M. (1987). Speech Audiometry in the USA. In M.Martin (Ed.). *Speech Audiometry* (pp. 207-236). London: Whurr Publishers Ltd.

Mallikarjuna (1990). Spondee words and Phonetically Balanced Monosyllabic words in Gujrati language. In S.K. Kacker & V.Basavaraj (Ed.). *Indian Speech Language Hearing Tests* - The ISHA Battery. AIISH, Mysore.

Martin, F.N. (1991). *Introduction to Audiology.* (4th Ed.). New Jersey. Prentice Hall, Engelwood Cliffs.

Martin, F.N. (1994). *Introduction to Audiology.* (5th Ed.). New Jersey: Prentice Hall, Engelwood Cliffs.

Martin, F.N., & Mussel, S.A. (1979). The influences of pauses in the competing signal on Synthetic Sentence Identification scores. *Journal of Speech and Hearing Disorders,* 44, 282-292.

Martin, F.N., & Stauffer, M.L. (1975). A modification of the Tillman-Olsen method for obtaining the speech reception threshold. *Journal of Speech and Hearing Disorders,* 40, 25-27.

Mascarenhas, K.E. (2002). A High Frequency- Kannada Speech Identification Test (HF-KSIT). *Unpublished Master's dissertation.* University of Mysore, India.

Mayadevi (1974). Development and Standardization of Common Speech Discrimination Test for Indians. *Unpublished Master's dissertation.* University of Mysore, India.

McPherson, D.F., & Pang-Ching, G.K. (1979). Development of distinctive feature discrimination test. *Journal of Auditory Research,* 19, 235-246.

Mendel, L.L., & Danhauer, J.L. (1997). *Audiologic Evaluation and Management and Speech Perception Assessment.* San Diego: Singular Publishing Group, Inc.

Morgan, D.E., Kamm, C.A., & Velde, T.M. (1981). Form equivalence of the speech perception in noise (SPIN) test. *Journal of Acoustical Society Of America,* 69, 1791-1798.

Nagaraja, M.N. (1973). Development of Synthetic Speech Identification Test in Kannada Language. *Unpublished Master's dissertation.* University of Mysore, India.

Nagaraja, M.N. (1990). Testing, interpreting and reporting procedures in speech audiometric tests. In S.K. Kacker and V. Basavaraj (Eds.), *Indian Speech Language Hearing Test* - The ISHA Battery -1990.

Nilsson, M, Soli, S.D., & Sullivan, J.A. (1994). Development of the Hearing In Noise Test for the measurement of speech reception threshold in quiet and in noise. *Journal of Acoustical Society of America*, 95, 1085-1099.

Olsen, W.O., & Matkin, N.D. (1979). Speech audiometry. In W.F. Rintelmann. (Ed.). *Hearing Assessment* (pp. 132-206). Baltimore: University Park Press.

O'Neill, J.J., & Oyer, H.J. (1966). *Applied Audiometry.* NewYork: Dodd, Mead & Company, Inc.

Owens, E. (1961). Intelligibility of words varying in familiarity. *Journal of Speech and Hearing Research, A,* 113-129.

Owens, E., Benedict, M., & Schubert, E.D. (1971). Further investigation of vowel items in multiple-choice speech discrimination testing. *Journal of Speech and hearing research,* 14, 841-847.

Owens, E., Kessler, D.K., Tellen, C.C., & Schubert, E.D. (1981). The Minimal Auditory Capabilities (MAC) Battery. *Hearing Aid Journal, 9,* 32.

Owens, E., & Schubert, E.D. (1968). The development of consonant items for speech discrimination testing. *Journal of Speech and Hearing Research,* 11, 656-667.

Owens, E., & Schubert, E.D. (1977). Development of California Consonant Test. *Journal of Speech and Hearing Research,* 20, 463-474.

Pascoe, D.P. (1975). Frequency responses of hearing aids and their effects on the speech perception of hearing-impaired subjects. *Annals of Otology, Rhinology and Laryngology,* suppl., *23,* 1-40.

Pederson, O.T., & Studebaker, G.A. (1972). A new minimal contrasts closed-response-set speech test. *Journal of Auditory Research,* 12, 187-195.

Peterson, G.E., & Lehiste, I. (1962). Revised CNC Lists for Auditory Tests. *Journal of Speech and Hearing Disorders,* 27, 62-65.

Plomp, R. (1986). A signal-to-noise ratio model for the Speech Reception Threshold of the hearing-impaired. *Journal of Speech and Hearing Research*, 29, 146-154.

Rajashekhar, B. (1976). The development and standardization of a picture SRT test for adults and children in Kannada. *Unpublished Master's dissertation.* University of Mysore, India.

Ramachandra, P. (2001). High Frequency Speech Identification Test for Hindi and Urdu Speakers. *Unpublished Master's dissertation.* University of Mysore, India.

Rintelmann,W.F., Schumaier, D.R., & Burchfield, S.B. (1974). Influence of test forms on speech discrimination scores in normal listeners on NU Auditory Test No. 6. In Rintelmann, W.F., Schumaier, D.R., & Burchfield, S.B.: Six experiments on speech discrimination utilizing CNC monosyllables (NU Test No. 6) *Journal of Auditory Research*, suppl. 2.

Rintelmann,W.F., Schumaier, D.R., & Jetty, A.J. (1994). List equivalency nd reliability for normal listeners on NU Auditory Test No. 6: Comparison with data from original talker. In Rintelmann, W.F., Schumaier, D.R., & Jetty, A.J.: Six experiments on speech discrimination utilizing CNC monosyllables (NU Test No. 6) *Journal of Auditory Research,* suppl. 2.

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

Schultz, M.C., & Schubert, E.D. (1969). A multiple-choice discrimination test (MCDT). *The Laryngoscope,* 79, 382-399.

Schumaier, D.R., & Rintelmann,W.F. (1974). Half list versus full list discrimination testing in clinical testing. In Rintelmann,W.F., Schumaier, D.R., & Jetty, A.J. Six experiments on speech discrimination utilizing CNC monosyllables (NU Test No. 6) *Journal of Auditory Research,* suppl. 2.

Schwartz, D.M., & Surr, R.K. (1979). Three experiments on the California Consonant Test. *Journal of Speech and Hearing Disorders*, 24, 55-60.

Sher, A.E., & Owens, E. (1974). Consonant Confusion associated with hearing loss above 2000 HZ. *Journal of Speech and Hearing Research*, 17(4), 669-681.

Singh, S., & Black, J.W. (1966). Study of twenty-six intervocalic consonants as spoken and recognized by four language groups. *Journal of Acoustical Society of America*, 32, 372-387.

Speaks, C, & Jerger, J. (1965). Method for measurement of Speech Identification. *Journal of Speech and Hearing Research*, 8, 185-194.

Swarnalatha, C.K. (1972). Development and Standardization of speech test material in English for Indians. *Unpublished Master's dissertation.* University of Mysore, India.

Tanuza, E. D. (1984). Development and Standardization of Speech test material in Manipuri language. *Unpublished Master's dissertation.* University of Mysore, India.

Tillman, T.W., & Olsen, W.O. (1973). Speech audiometry. In J.Jerger ((2nd Ed.). Modern developments in Audiology (pp. 37-74). New York: Academic Press.

Townsend, T.H., & Schwartz, D.M. (1981). Error analysis on the California Consonant Test by manner of articulation. *Ear and Hearing,* 2, 108-111.

Versfeld, N.J., Daalder, L., Festen, J.M., & Houtgast, T. (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *Journal of Acoustical Society of America*, 107(3), 1671-1684.

Wilber, L.A. (2002). Calibration: Puretone, Speech and Noise Signals. In J.Katz (Ed.). *Handbook of Clinical Audiology.* Baltimore: Lippincott Williams & Wilkins.

Wilson, R.H., & Antablin, J.K. (1980). A picture identification task as an estimate of word-recognition performance of nonverbal adults. *Journal of Speech and Hearing Disorders,* 45, 223-238.

Wilson, R., Morgan, D., & Dirks, D. (1973). A proposed SRT procedure and its statistical precedent. *Journal of Speech and Hearing Disorders,* 38, 184-191.

APPENDIX -A
BISYLLABIC WORD LISTS

List A

1. ତାଙ୍କୁ
2. ଚାରି
3. ଜଣେ
4. ଟଙ୍କା
5. ଡ଼ାକ
6. ତେଣୁ
7. ଦୁଃଖ
8. ସାଙ୍ଗ
9. ଘଣ୍ଟା
10. ଫଳ
11. ମଧ
12. ଲୋକ
13. ସବୁ
14. ଗତ
15. ଛୁଟି
16. ଝଡେ
17. ଠାରୁ
18. ଢାଳି
19. ଥିଲେ
20. ନାମ
21. ତମ
22. ପାଣି
23. ଭଲ
24. ରହି
25. ହେବ

List B

1. ନିଜ
2. ଚିନ୍ତା
3. ଘର
4. ଝାଡୁ
5. ଠେଲି
6. ପିଲା
7. ଢାଳ
8. ଧୀରେ
9. ଥିବା
10. ଫୁଲ
11. ମୋର
12. ଲୁଣ
13. ହାତ
14. ମାସ
15. ଖାଲି
16. ଛୋଟ
17. ବାଟ
18. ଡାଳ
19. କଣ
20. ଦିନ
21. ସୁଙ୍ଘି
22. ପରେ
23. ଭାଗ
24. କଳା
25. ସେହି

BISYLLABIC WORD LISTS

	List – A		List – B
1.	/t̪anku/	1.	/nidʒɔ/
2.	/tʃari/	2.	/tʃint̪a/
3.	/dʒɔṇe/	3.	/gʰɔɤɔ/
4.	/ṭɔnka/	4.	/dʒʰaḍu/
5.	/ḍakɔ/	5.	/ṭʰeli/
6.	/t̪eṇu/	6.	/pila/
7.	/d̪ukʰɔ/	7.	/ḍʰaḷɔ/
8.	/saŋɔ/	8.	/d̪ʰire/
9.	/gʰɔnṭa/	9.	/t̪ʰiba/
10.	/pʰɔḷɔ/	10.	/pʰulɔ/
11.	/madʰjɔ/	11.	/morɔ/
12.	/lokɔ/	12.	/luṇɔ/
13.	/sɔbu/	13.	/hat̪ɔ/
14.	/gɔt̪ɔ/	14.	/masɔ/
15.	/tʃʰuṭi/	15.	/kʰali/
16.	/dʒʰɔḍe/	16.	/tʃʰoṭɔ/
17.	/ṭʰaru/	17.	/baṭɔ/
18.	/ḍʰaḷi/	18.	/ḍaḷɔ/
19.	/t̪ʰile/	19.	/kɔṇɔ/
20.	/namɔ/	20.	/d̪inɔ/
21.	/t̪ɔmɔ/	21.	/suŋi/
22.	/paṇi/	22.	/pɔɤe/
23.	/bʰɔlɔ/	23.	/bʰagɔ/
24.	/ɤɔhi/	24.	/kɔḷa/
25.	/hebɔ/	25.	/sehi/

APPENDIX -B
MONOSYLLABIC WORD LISTS

List - 1

1.	କର୍	26.	କୁଅ
2,	ଖୋଲ୍	27.	ଖେଳ୍
3.	ଗୁଆ	28.	ଚଟ୍
4.	ଯିଏ	29.	ଢେଉ
5.	ଝିଅ	30.	ଦୋଉ
6.	ଡିଆଁ	31.	ନିଏ
7.	ଥାଉ	32.	ବସ୍
8.	ନେଇ	33.	ସାଇ
9.	ପାଇଁ	34.	ହୋଇ
10.	ବୋଉ	35.	କେଇ
11.	ମିସ୍	36.	ଘାଇ
12.	ହସ୍	37.	ଯାଇ
13.	ହୁଏ	38.	ହାଣ୍
14.	କେଉଁ	39.	ଦିଅ
15.	ଦେଖ୍	40.	ପାଇ
16.	ଚଙ୍ଗ୍	41.	ମାଇଁ
17.	ଯାଉ	42.	ସେଉ
18.	ଠିକ୍	43.	ଖୁବ୍
19.	ତାସ୍	44.	ଛୁଇଁ
20.	ଧର୍	45.	ଯୋର
21.	ନିଆଁ	46.	ତିନ୍
22.	ପିଅ	47.	ଧୂଆଁ
23.	ଭାଇ	48.	ଫେଲ୍
24.	ସିଏ	49.	ମାର
25.	ଲାଉ	50.	ହେଉ

List - 2

1.	କାଉ	26.	କୋଉ
2.	ଧିକ୍	27.	ଘିଅ
3.	ଲେଖ୍	28.	ଝୁଅ
4.	ଚଢ୍	29.	ଟଙ୍ଗ୍
5.	ଯୋଉ	30.	ଦସ୍
6.	ନିଅ	31.	ମାପ୍
7.	ଦାଇ	32.	ମାଲ୍
8.	ଧୋଇ	33.	ସୁଅ
9.	ନାଇଁ	34.	କାଇଁ
10.	ପୁଅ	35.	ଖାଇ
11.	ଦାମ୍	36.	ଯୁଆ
12.	ସେସ୍	37.	ତେଇଁ
13.	ହଉ	38.	ନୂଆ
14.	କିଏ	39.	ବେସ୍
15.	ରଖ୍	40.	ଭୋର
16.	ଗାଇ	41.	ଖୋଜ୍
17.	ଜିଆ	42.	ଛୁଆ
18.	ଟପ୍	43.	ଠିଆ
19.	ଥୁଆ	44.	ଦୁଇ
20.	ବନ୍ଦ୍	45.	ଧୁଅ
21.	ଢାଲ୍	46.	ପାଏ
22.	ପାଉ	47.	ଶୁଣ୍
23.	ଭୁଲ୍	48.	ହଇ
24.	ଫେର୍	49.	ହେଇ
25.	ସେଇ	50.	ଦେଇ

MONOSYLLABIC WORD LISTS

List – 1

1.	/kɔr/	26.	/kuɔ/
2.	/kʰol/	27.	/kʰeḷ/
3.	/gua/	28.	/tʃɔṭ/
4.	/dʒie/	29.	/ḍʰeu/
5.	/dʒʰiɔ/	30.	/d̪ou/
6.	/ḍiã/	31.	/nie/
7.	/t̪ʰau/	32.	/bɔs/
8.	/nei/	33.	/sai/
9.	/pãĩ/	34.	/hoi/
10.	/bou/	35.	/kei/
11.	/mis/	36.	/gʰai/
12.	/hɔs/	37.	/dʒai/
13.	/hue/	38.	/haṇ/
14.	/keũ/	39.	/d̪iɔ/
15.	/d̪ekʰ/	40.	/pai/
16.	/tʃɔŋ/	41.	/mãĩ/
17.	/dʒau/	42.	/seu/
18.	/t̪ʰik/	43.	/kʰub/
19.	/t̪as/	44.	/tʃʰũɔ̃/
20.	/d̪ʰɔr/	45.	/dʒor/
21.	/niã/	46.	/t̪in/
22.	/piɔ/	47.	/d̪ʰuã/
23.	/bʰai/	48.	/pʰel/
24.	/sie/	49.	/mar/
25.	/lau/	50.	/heu/

List – 2

1.	/kau/	26.	/kou/
2.	/d̪ʰik/	27.	/gʰiɔ/
3.	/lekʰ/	28.	/dʒʰuɔ/
4.	/tʃɔḍʰ/	29.	/t̪ɔŋ/
5.	/dʒou/	30.	/d̪ɔs/
6.	/niɔ/	31.	/map/
7.	/d̪ai/	32.	/mal/
8.	/d̪ʰoi/	33.	/suɔ/
9.	/nãĩ/	34.	/kãĩ/
10.	/puɔ/	35.	/kʰai/
11.	/d̪am/	36.	/dʒua/
12.	/ses/	37.	/ḍeĩ/
13.	/hɔu/	38.	/nua/
14.	/kie/	39.	/bes/
15.	/rɔkʰ/	40.	/bʰor/
16.	/gai/	41.	/kʰodʒ/
17.	/dʒia/	42.	/tʃʰua/
18.	/t̪ɔp/	43.	/t̪ʰia/
19.	/t̪ʰua/	44.	/d̪ui/
20.	/bɔ̃d̪/	45.	/d̪ʰuɔ/
21.	/ḍʰaḷ/	46.	/pae/
22.	/pau/	47.	/suṇ/
23.	/bʰul/	48.	/hɔi/
24.	/pʰer/	49.	/hei/
25.	/sei/	50.	/d̪ei/

APPENDIX – C

SCORE SHEET

Name :

Age / Sex :

AUDIOLOGICAL FINDINGS

	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz
AC Rt :						
AC Lt :						
BC :						

Pure tone average :

SRT :

List A :

List B :

IDENTIFICATION OF SRT

Score Key :	Marking	Score
Correct response	✓	1
Incorrect response	x	0

List A			List B		
1.	ଟାଙ୍କୁ	/t̪anku/	1.	ନିଜ	/nidʒɔ/
2.	ଚାରି	/tʃari/	2.	ଚିନ୍ତା	/tʃint̪a/
3.	ଜଣେ	/dʒɔṇe/	3.	ଘର	/gʰɔrɔ/
4.	ଟଙ୍କା	/ṭɔnka/	4.	ଝାଡୁ	/dʒʰaḍu/
5.	ଡାକ	/ḍakɔ/	5.	ଠେଲି	/ṭʰeli/
6.	ଟେଣୁ	/t̪eṇu/	6.	ପିଲା	/pila/
7.	ଦୁଃଖ	/d̪ukʰɔ/	7.	ଢାଲ	/ḍʰalɔ/
8.	ସାଙ୍ଗ	/saŋɔ/	8.	ଧୀରେ	/d̪ʰire/
9.	ଘଣ୍ଟା	/gʰɔnṭa/	9.	ଥିବା	/t̪ʰiba/
10.	ଫଳ	/pʰɔḷɔ/	10.	ଫୁଲ	/pʰulɔ/
11.	ମଧ୍ୟ	/mɔdʰjɔ/	11.	ମୋର	/morɔ/
12.	ଲୋକ	/lokɔ/	12.	ଲୁଣ	/luṇɔ/
13.	ସବୁ	/sɔbu/	13.	ହାତ	/hat̪ɔ/
14.	ଗତ	/gɔt̪ɔ/	14.	ମାସ	/masɔ/
15.	ଛୁଟି	/tʃʰuṭi/	15.	ଖାଲି	/kʰali/
16.	ଝଡେ	/dʒʰɔḍe/	16.	ଛୋଟ	/tʃʰoṭɔ/
17.	ଠାରୁ	/ṭʰaru/	17.	ବାଟ	/baṭɔ/
18.	ଢାଳି	/ḍʰaḷi/	18.	ଡାଳ	/ḍaḷɔ/
19.	ଥିଲେ	/t̪ʰile/	19.	କଣ	/kɔṇɔ/
20.	ନାମ	/namɔ/	20.	ଦିନ	/d̪inɔ/
21.	ତମ	/t̪ɔmɔ/	21.	ସୁଙ୍ଗି	/suŋi/
22.	ପାଣି	/paṇi/	22.	ପରେ	/pɔre/
23.	ଭଲ	/bʰɔlɔ/	23.	ଭାଗ	/bʰagɔ/
24.	ରହି	/rɔhi/	24.	କଳା	/kɔḷa/
25.	ହେବ	/hebɔ/	25.	ସେହି	/sehi/

Speech Identification

Score Key :	Marking	Score
Correct response	✓	1
Incorrect response	x	0

List 1		List 2	
1. କର /kɔr/	26. କୁଅ /kuɔ/	1. କାଉ /kau/	26. କୋଉ /kou/
2. ଖୋଲ /kʰol/	27. ଖେଳ /kʰeḷ/	2. ଢିକ /ḍʰik/	27. ଘିଅ /gʰiɔ/
3. ଗୁଆ /gua/	28. ଚଟ /tʃɔṭ/	3. ଲେଖ /lekʰ/	28. ଝୁଅ /dʒʰuɔ/
4. ଜିଏ /dʒie/	29. ଢେଉ /ḍʰeu/	4. ଚଢ /tʃɔḍʰ/	29. ଟଙ୍ /ṭɔŋ/
5. ଝିଅ /dʒʰiɔ/	30. ଡୋଉ /ḍou/	5. ଜୋଉ /dʒou/	30. ଡସ /ḍɔs/
6. ଡିଆଁ /ḍiã/	31. ନିଏ /nie/	6. ନିଅ /niɔ/	31. ମାପ /map/
7. ଥାଉ /t̪ʰau/	32. ବସ /bɔs/	7. ଡାଇ /ḍai/	32. ମାଲ /mal/
8. ନେଇ /nei/	33. ସାଇ /sai/	8. ଢୋଇ /ḍʰoi/	33. ସୁଅ /suɔ/
9. ପାଇଁ /paĩ/	34. ହୋଇ /hoi/	9. ନାଇଁ /naĩ/	34. କାଇଁ /kaĩ/
10. ବୋଉ /bou/	35. କେଇ /kei/	10. ପୁଅ /puɔ/	35. ଖାଇ /kʰai/
11. ମିସ /mis/	36. ଘାଇ /gʰai/	11. ଡାମ /ḍam/	36. ଯୁଆ /dʒua/
12. ହସ /hɔs/	37. ଯାଇ /dʒai/	12. ସେସ /ses/	37. ଡେଇଁ /ḍeĩ/
13. ହୁଏ /hue/	38. ହାଣ /haṇ/	13. ହଉ /hɔu/	38. ନୁଆ /nua/
14. କେଉଁ /keũ/	39. ଦିଅ /d̪iɔ/	14. କିଏ /kie/	39. ବେସ /bes/
15. ଦେଖ /d̪ekʰ/	40. ପାଇ /pai/	15. ରଖ /rɔkʰ/	40. ଭୋର /bʰor/
16. ଚଙ୍ /tʃɔŋ/	41. ମାଇଁ /maĩ/	16. ଗାଇ /gai/	41. ଖୋଜ /kʰodʒ/

17. ଯାଉ /dʒau/	42. ସେଉ /seu/	17. ଜିଆ /dʒia/	42. ଛୁଆ /tʃʰua/
18. ଠିକ୍ /ʈʰik/	43. ଖୁବ୍ /kʰub/	18. ଟପ୍ /ʈɔp/	43. ଠିଆ /ʈʰia/
19. ଟାସ୍ /ʈas/	44. ଛୁଅଁ /tʃʰuɔ̃/	19. ଥୁଆ /ʈʰua/	44. ଦୁଇ /dui/
20. ଢର୍ /ɖʰɔr/	45. ଯୋର /dʒor/	20. ବଣ୍ଡ୍ /bɔ̃ɖ/	45. ଧୁଅ /ɖʰuɔ/
21. ନିଆଁ /niã/	46. ଟିନ୍ /ʈin/	21. ଢାଳ୍ /ɖʰaɭ/	46. ପାଏ /pae/
22. ପିଅ /piɔ/	47. ଢୁଆଁ /ɖʰuã/	22. ପାଉ /pau/	47. ଶୁଣ୍ /suɳ/
23. ଭାଇ /bʰai/	48. ଫେଲ୍ /pʰel/	23. ଭୁଲ୍ /bʰul/	48. ହଇ /hɔi/
24. ସିଏ /sie/	49. ମାର /mar/	24. ଫେର୍ /pʰer/	49. ହେଇ /hei/
25. ଲାଉ /lau/	50. ହେଉ /heu/	25. ସେଇ /sei/	50. ଦେଇ /ɖei/