CONSTRUCTION OF AN ARTICULATION CURVE FOR RECORDED SENTENCES

Thesis for the Degree of M. A.

MICHIGAN STATE UNIVERSITY

Kenneth Ray Johnson

1965

ABSTRACT

CONSTRUCTION OF AN ARTICULATION CURVE
FOR RECORDED SENTENCES

by Kenneth Ray Johnson

In the past few years, our literature has expressed dissatisfaction with some of the current methods used in speech audiometry. With this in mind, an articulation curve was constructed for ten lists of ten sentences each. The sentences used in this investigation were developed at the Central Institute for the Deaf. A recording of the sentences was made at Michigan State University on magnetic tape using five male and five female talkers ranging in age from 24-68 years.

In order that they could be played back at specific intensitites, the median intensity of each list was determined by running them through a Bruel and Kjaer sound-level recorder which plotted the intensity of the sentences against time. The amount of the signal (in millimeters) that was above 45, 51, 57, 63, 69, 75, 81, 87, and 93 dB sound pressure level (SPL) was measured and plotted on a graph. One graph was drawn for each of the ten lists. From these, the average intensity of the lists was found to be 69.6 dB SPL. Now these sentences could be presented to listeners at intensities with known parameters.

KENNETH RAY JOHNSON

Thirty normal-hearing listeners heard the sentences in a sound field situation at ten different intensities, ranging from 18 to 54 dB SPL. The sentences and intensities were presented according to a schedule determined from a 10 x 10 x 10 Graeco-Latin square. The average time for presenting and scoring one list of ten sentences was one minute twenty-eight seconds. The scores represented the percentage of 50 keys words per list, correctly repeated by the listener for each list.

Use of different word lists and different orders of presentation were found to be small but significant sources of variance in discrimination scores. Articulation functions were plotted for all lists and a composite articulation function was developed.

CONSTRUCTION OF AN ARTICULATION CURVE

FOR RECORDED SENTENCES

by

Kenneth Ray Johnson

A THESIS

Submitted to
Michigan State University
in partical fulfillment of the requirements
for the degree of

MASTER OF ARTS

Department of Speech

1965

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER I

INTRODUCTION

Speech audiometry plays an important role in routine hearing and hearing-aid evaluations. The objective, and sometimes subjective, assessment of a person's ability to hear and understand speech can help the examiner complete the "picture" of a person's handicap that was begun with pure-tone audiometry. This is a necessity if proper diagnostic and rehabilitative procedures are to be followed.

Pure-tone testing provides information that will aid in diagnosis and in indicating aural rehabilitative needs by showing how well the cochlea is functioning. Through special techniques, information can be obtained which will help to determine if the lesion is retrocochlear. But the biggest interest in terms of audiological rehabilitation is how well the patient hears and understands speech. According to Davis and Silverman, "Practically, the most important thing we want to know about someone's hearing is whether or not he can follow ordinary conversation. Is his hearing socially adequate?"[1]

---

[1]Hallowell Davis and S. Richard Silverman, Hearing and Deafness (New York: Holt, Rinehart and Winston, 1964), p. 191.

To test whether a person is socially adequate or to determine how much speech he hears and understands, it is necessary to administer a number of tests. Hayes Newby recommends the determination of the following measures:

1. Speech-reception threshold (SRT)--the hearing level at which the patient can repeat 50 per cent of spondees correctly.
2. Most comfortable loudness (MCL)--the hearing level at which speech is most comfortable for the patient.
3. Tolerance level--sometimes referred to as the threshold of discomfort (TD). The hearing level at which speech becomes uncomfortably loud.
4. Dynamic range--the limits of useful hearing.
5. Discrimination--a patient's ability to discriminate among similar sounds or among words that contain similar sounds.
6. Social adequacy index--a measure based on the results of speech audiometry which represent the degree of handicap so far as hearing and understanding speech are concerned.[1]

These tests are all administered with a carefully controlled and known intensity via a speech audiometer and are

---

[1]Hayes A. Newby, Audiology (2nd ed; New York: Appleton-Century-Crofts, 1964), pp. 111-116.

presented monaurally through calibrated earphones or in a calibrated free field situation. Probably the two most important of these measures are the SRT and discrimination. The former is a threshold test while the latter is a supra-threshold test.

The SRT, or hearing level for speech, is most popularly determined by presenting a number of "spondees" (two syllable words with equal stress on both syllables) and gradually decreasing the intensity until the patient correctly repeats 50 per cent of them. Another method includes the use of running speech. In this method, either the examiner or the patient decreases the intensity of the running speech until just the general idea of the material can be followed. Another method uses numbers instead of spondees. Still another method makes use of sentences that are in the form questions that require the listener to answer them or repeat the sense of them. The questions are presented in decreasing intensity until the patient's threshold is reached.

The hearing level for speech turns out to be the same for any of these methods, whether it was obtained with live voice or record material, providing all have been correctly calibrated on similar normal populations.[1] The SRT is a measure of how loud speech has to be before

---

[1]Davis and Silverman, Hearing and Deafness, p. 186.

it is understood by the patient 50 per cent of the time. For the normal ear receiving spondaic material, this is about 22 dB sound pressure level (SPL). The SRT can be estimated from the pure-tone audiogram by averaging the obtained thresholds at 500, 1000 and 2000 cps.

The second measure, speech discrimination, is sometimes called the intelligibility or articulation score, or the PB Max (maximum discrimination score). This measure is most popularly determined with the use of phonetically balanced (PB) word lists. PB words are monosyllabic words that contain samples of speech sounds in the same proportion that they occur in English speech. The PB word lists each contain 50 words, and these are presented monaurally or free field well above the SRT. The discrimination score is the percentage of 50 words that is correctly repeated by the patient. Some hearing losses are of such a nature that the patient will not obtain a score of 100 per cent regardless of how intense the sound is presented. Recorded lists of spondees and PB's are available commercially.

With careful interpretation of the SRT and PB scores, an experienced examiner will be able to judge how well the patient is getting along in society as far as hearing and understanding speech is concerned; and many will agree that this is the main purpose of hearing.

These two tests also play an important role in hearing-aid evaluations. By comparing the results obtained from these tests and others with different hearing aids, the examiner will be in a position to help the patient choose the hearing aid that provides the best amplification.

Speech audiometry is also used for diagnostic and rehabilitative purposes. Diagnostically, the results from speech audiometry can, in most cases, be used to confirm the pure-tone findings. Researchers have shown that a person with a hearing loss with a conductive (as opposed to sensorineural) pathology will have an SRT that is higher (i.e., worse) than normal; but the discrimination score will be within normal limits if the PB's are presented at high intensity. On the other hand, a person with a sensorineural hearing loss will have a high SRT and will have lower than normal discrimination regardless of intensity. In fact, in some cases the discrimination score may get worse if the intensity is increased too high.

Persons with a low discrimination score are probable candidates for auditory training and/or speech-reading. These people will be in need of these rehabilitative procedures because their hearing loss is preventing them from hearing enough of the speech sounds to make them socially adequate. When the patient's SRT in the better ear is 30 dB or greater, they are candidates for a hearing aid and possibly auditory training and/or speechreading. In many cases, a hearing aid will be of no assistance because it may; in fact, make the discrimination score much worse.

These tests may also be used to evaluate the results of surgery. By comparing the post-operative results with the pre-operative results, the benefits of surgery, in terms of improving a person's ability to hear and understand speech, may be determined.

Responses to speech audiometry, SRT and discrimination in particular, can be graphically represented by an articulation curve with the abscissa of the graph representing intensity and the ordinate representing percentage of words correctly repeated. Experiments have shown that the articulation curve is, in general, a smooth S-shaped curve that is steepest at, or a little below, its middle; and the upper end levels off rather gradually to a plateau.[1] The articulation curve is constructed by presenting the stimulus at various intensities, marking the per cent correct at the appropriate spot on the graph, and then drawing the curve. As would be expected, the slope of the curve is steeper for sentences than for spondees, which in turn is steeper than the curve for PB's.

---

[1] Ibid., p. 190.

The advantage of the articulation curve is that it provides a complete "picture" of the patient's ability to hear and understand speech. With one look at the curve, the following characteristics can be seen:

1. Threshold of intelligibility--The point where the curve crosses the 50 per cent line.
2. The maximum intelligibility--The highest point on the curve.
3. The threshold of distortion--The point of the curve where it begins to decline. (This is not present on all curves.)

Being able to obtain these three important measures in one test administration would be time saving.

Because we communicate by speaking in sentences, it would seem logical that a valid and reliable test to determine someone's SRT and discrimination would be a test utilizing sentences. Many sentence tests have been devised and will be discussed in the following chapter, but more frequently word lists are used. Ira J. Hirsh has said:

> We usually like to work with words in measuring intelligibility, because we can construct a list of words and present each word to a listener at a given intensity. To be sure, the relation between such lists and the continuous flow of words that we encounter in conversation is not very clear. Instead, therefore, we may attempt to devise a more

valid test by using groups of words that might appear in conversation. One such group is, of course, the sentence.[1]

One of the main reasons sentences have not been used is that the intensity cannot be controlled while maintaining naturalness. A method of overcoming this will be explained in Chapter III.

Recently, many audiologists have been dissatisfied with the 50-word speech discrimination tests and have attempted to shorten the lists to 25 words.[2] Since it takes about four minutes to go through one 50-word, CID W-22 PB word list, these lists become cumbersome when several such lists are used in one session as in hearing aid selection.[3]

Another problem with PB word lists is the level at which they should be administered. Some audiologists suggest SRT plus 40 dB, while still others recommend giving them at the comfort level. With any of these methods, one is not sure if the PB Max has been obtained. By constructing an articulation curve, one can be sure to have obtained the maximum intelligibility.

---

[1] Ira J. Hirsh, <u>The Measurement of Hearing</u> (New York: McGraw-Hill Book Company, Inc., 1952), p. 131.

[2] Patti Grubb, "A Phonemic Analysis of Half-List Speech Discrimination Tests," <u>Journal of Speech and Hearing Research</u>, VI (1962), p. 271.

[3] R. Edwin Shutts, Kenneth S. Burke, and James E. Creston, "Derivation of Twenty-Five Word PB Lists." <u>Journal of Speech and Hearing Disorders</u>, XXIX (1964), p.442.

The purpose of this study is to establish an articulation curve for normal hearing listeners, using recorded sentences that represent everyday American speech. The sentences are spoken by five male and five female talkers between the ages of 24 and 68. By establishing the curve for normal ears, these same sentences can be used on pathological ears and then compared with the normal curve. In administering this one test, the SRT and a measure of intelligibility would be obtained. Because these are "everyday sentences," the derived curve should give a good "picture" of a person's social adequacy. By comparing curves derived while the patient uses different hearing aids, hearing aid selection could be made easier. Diagnostically, the shape and position of the curve should be valuable in comfirming the pure-tone results. By comparing curves obtained before and after a period of speechreading and auditory training, the effectiveness of these procedures could be evaluated. In regard to time, drawing an articulation curve with sentences might prove to be much quicker than using PB and spondee word lists.

Chapter II of the thesis contains a short review of the literature relevant to the present study. Chapter III contains a discussion of the procedures and equipment employed in the study. Chapter VI presents the results of the study. Chapter V, the last chapter, contains a short discussion of the results of the study together with some possible implications for further research along the present lines.

CHAPTFR II

REVIEW OF THE LITERATURE

This review will be concerned only with standardized speech tests. Crude tests of sensitivity such as the conversational voice and whisper test will not be included. While these tests play an important role for those trained to use them, they are of no importance to this study.

Development of Speech Audiometry

The principle of speech audiometry was established by the success of the 4C group audiometer[1] that was used as a screening device for school children. This audiometer was a refinement of the earlier 4A and 4B audiometers. The 4A phonographic audiometer was first marketed between 1924 and 1926. This audiometer could be equipped with up to 40 earphones and was portable from school to school. The 4A consisted of a hand-wound, spring-driven motor and turntable, plus a large magnetic reproducing head or pick-up unit. The phonographic record used with the 4A and 4B audiometer contained a series of two or three digit numbers that were successively attenuated in 3 dB steps. The 4B and 4C audiometers were improvements over this earlier model.

---

[1]Davis and Silverman, p. 181.

The next improvement came out in the Maico RS group audiometer. The main differences in this audiometer were: (1) it provided a fading, selective word, pictorial type test as well as the earlier fading numbers type test, (2) it had a means of being calibrated, and (3) it was electronic and contained a built-in amplifier. This type of testing procedure accurately determined speech reception thresholds, but its failure was the inability to detect high frequency losses.[1] Some further limitations of the 4C were in the limited range covered by the fixed-gain play back equipment and the relatively few speech sounds sampled.[2]

Speech Audiometry Today

Speech audiometry, as it is practiced today, began during World War II in the military aural rehabilitation centers.[3] However, the test material for speech audiometry was borrowed from another source. Working for the Bell Telephone Laboratories, Fletcher and Steinberg developed tests for hearing and understanding speech as a tool of

---

[1]Leland A. Watson and Thomas Tolan, Hearing Tests and Hearing Instruments (Baltimore: The Williams and Wilkins Company, 1949), pp. 239-246.

[2]C. V. Hudgins et al., "The Development of Recorded Auditory Tests for Measuring Hearing Loss of Speech," Laryngoscope, LVII (1947), p. 61.

[3]Newby, p. 66.

measurement and research inprovement of telephone and other communication systems.[1] Other similar test materials were developed under government contract by the Harvard Psycho-Acoustic Laboratory for the purpose of comparing the efficiency of various communications systems in transmitting speech.[2] Since then, these lists have been adapted for the purpose of speech audiometry as we know it today. The most noteworthy of the adaptations were made at the Harvard Psycho-Acoustic Laboratory and at Central Institute for the Deaf, St. Louis. The next section will briefly describe the more familiar of these tests.

Fletcher and Steinberg report that their main test for measuring the articulation of communication systems used nonsense syllables. To develop the test, they drew on the 48 simple sounds of the revised scientific alphabet, 24 consonants, 19 vowels, and 5 dipthongs. The initial consonants were written on cards and then placed in one box; others, upon which the vowel sounds were written, were placed in a second box; and those, upon which the final consonant sounds were written, were placed in a third box. A card was then randomly drawn from each

---

[1]Watson and Tolan, p. 452.

[2]Newby, p. 66.

box, thus forming a con-vow-con (cvc) syllable. By drawing all of the sounds, a list of 22 syllables was formed. This process was repeated three times to obtain a list of 66 syllables. They also made use of sentences that were interrogative or imperative in form, each containing a simple idea. The original list contained 50 sentences.[1]

At the Harvard Psycho-Acoustic Laboratory, a program of audiometric test development was undertaken with three aims in view:

1. To explore further the problems involved in the construction of audiometric tests for measuring directly the hearing loss for speech.
2. To produce a test suitable for precise laboratory measurements of all degrees of hearing loss.
3. To explore by means of verbal tests the possibility of differentiating between high-frequency deafness and deafness which is uniform throughout the audible frequency range.[2]

The results of this program were Auditory Test No. 9 and Auditory Test No. 12. Test No. 9 was a test for the threshold of hearing for words. This test consisted of two lists of 42 disyllabic words of the spondee stress pattern recorded in six scrambled versions with a carrier phrase "Number One," "Number Two," etc. The words were

---

[1]H. Fletcher and J. C. Steinberg, "Articulation Testing Methods," The Bell System Technical Journal, VIII (July, 1929), pp. 806-854.

[2]Hudgins et al., p. 62

divided into seven groups of six words each, and each group was recorded at progressively lower intensity levels, 4 dB apart. Spondees were used because they were found to have a uniformly high audibility; higher than unselected disyallables and monosyllables.

Auditory Test No. 12 consisted of eight lists of short, simple questions, that could be answered with one word. The lists were composed of 28 items divided into seven groups of four. Each group was recorded at an intensity level 4 dB lower than the preceding one, except List One which contained 21 items divided into groups of three and recorded at an intensity level 6 dB lower than the previous group.

These two tests were constructed with these basic criteria in mind: (1) familiarity, (2) phonetic dissimilarity, (3) normal sampling of English speech sounds, and (4) homogeneity with respect to basic audibility.

They were unable to fulfill the third of the three aims listed above as satisfactorily as they would have liked. They concluded that unless a test can be devised which is relatively simple to administer, and at the same time precise, it may be expected to show no special advantage over the pure-tone audiometer as a device for differentiating between uniform and high frequency losses.[1]

---

[1] *Ibid.*, pp. 58-68.

Also at the Harvard Psycho-Acoustic Laboratory, J. P. Egan constructed the well-known PB-50 word lists which were originally constructed as tests for communication systems. These lists were the result of many revisions of previous word lists and every effort was made to make these new lists satisfy the following criteria: (1) monosyllabic structure, (2) equal average difficulty, (3) equal range of difficulty, (4) equal phonetic composition, (5) a composition representative of English speech, and (6) words in common usage. From a sample of 1200 monosyllabic words, 24 lists of 50 words each were constructed. These lists were then revised to insure that the lists would be more nearly phonetically balanced. An unsuccessful attempt was made to satisfy the above criteria with lists of 25 words and that accounts for the number 50 in each list.[1]

Perhaps the most widely used speech audiometric tests are those developed at Central Institute for the Deaf. Audiologists at CID were concerned about these several deficiencies in the Harvard tests: (1) certain records of Auditory Test No. 9 yielded slightly different thresholds from other of these records, (2) the vocabulary was too large for many clinical patients, and (3) recorded

---

[1]J. P. Egan, "Articulation Testing Methods," Laryngoscope, LVIII (1948), p. 963.

versions of the PB lists were not available in suitably standard form. Because of this, they made some modifications and introduced three new lists: (CID Auditory Test W-1 and W-2, spondees, and CID Auditory Test W-22, PB's).

The first of these, W-1, contains six scramblings of a single list of 36 spondaic words. These are recorded at a constant level 10 dB below the level of an introductory carrier phrase. This test is for determination of the SRT. The thresholds obtained with this test on normal ears for experienced and inexperienced listeners was 20 dB and 21 dB re 0.0002 microbar, respectively. The second test W-2 is also used for determination of the SRT. It contains the same words as W-1 but differs in that the intensity of the words is attenuated within each list at the rate of 3 dB every three words. The mean absolute threshold for 14 listeners was 17.7 dB re 0.0002 microbar. The difference between the W-1 and W-2 thresholds may have resulted from presenting all 36 words at a given level instead of only three. The third test, W-22, is for determination of a listener's discrimination. It contains 200 monosyllabic words divided into four phonetically balanced lists of 50 words each and each word is introduced with the carrier phrase, "You will say. . . ." The threshold for this test is 24 dB re 0.0002 microbar. Above 50 dB 100 per cent of the words are heard correctly by the normal ear.[1]

Phonetic Balance

The importance of using phonetic balance in speech discrimination testing has been questioned by several authors. As early as 1929 Flethcer and Steinberg stated that the results obtained on their word lists, that were randomly constructed, were as representative of speech as the results that would be obtained with lists employing particular sound combinations in proportion to their frequencies of occurrence in speech.[2]

Elpern states, ". . .it seems to be the consensus that phonetic balance is not as crucial a factor to the sensitivity of these tests [speech discrimination tests] as it was thought to be at the time the original PB-50 word lists were developed."[3] He found that one-half list material extracted from the W-22 full-lists offered a valid and reliable auditory discrimination score.[4]

---

[1] I. H. Hirsh *et al.*, "Development of Materials for Speech Audiometry," *Journal of Speech and Hearing Disorders*, XVII (1952), pp. 322-332.

[2] Fletcher and Steinberg, p. 834.

[3] B. S. Elpern, "The Relative Stability of One-Half-List and Full-List Discrimination Tests," *Laryngoscope*, LXXI (1961), p. 31.

[4] *Ibid.*, pp. 30-36.

Campanelli used Egan's PB-50 word lists and divided them into two groups of 25 words each by taking the first 25 words for one list and the remaining 25 for another list. He found that the 25-word lists offered a valid and reliable auditory discrimination score. In a statement on phonetic balance he says, "This notion [that the entire phonetic spectrum should be included in order to yield a valid measure of auditory-verbal comprehension] might be realistically questioned, and the concept of the PB-50 list be re-examined."[1]

Resnick[2] also found a high correlation between full PB-50-word lists and 25-word lists derived from them. Several other authors, Lynn[3], Shutts _et al._[4] have also attempted to shorten the time required to test a person's discrimination.

Patti Grubb has written in the defense of phonetic balance. She claims that, "Phonetic balance is a means of

---

[1]P. A. Campanelli, "A Measure of Intra-List Stability of Four PAL Word Lists," _Journal of Auditory Research_, II (1962), pp. 50-55.

[2]D. M. Resnick, "Reliability of the Twenty-Five Word Phonetically Balanced Lists," _Journal of Auditory Research_, II (1962), pp. 5-12.

[3]G. Lynn, "Paired PB-50 Discrimination Test: A Preliminary Report," _Journal of Auditory Research_, II (1962), pp. 34-36.

[4]R. E. Shutts, K. S. Burke, and J. E. Creston, "Derivation of Twenty-Five-Word PB Lists," _Journal of Speech and Hearing Disorders_, XXIX (1964), pp. 442-447.

achieving, at least, face validity."[1] She conducted a study that showed that the PB characteristic of the whole 50-word list is lost when they are split into two halves.[2] In a letter to the editor, she criticizes the use of half-list speech discrimination tests mentioned above. She used three main points in her argument: (1) that high correlations were found because the part is a part of the whole, (2) that the interpretation of the coefficients was faulty, and (3) that 25-word lists lack validity.[3]

Jerry V. Tobias, in answer to Grubb's article and letter states:

> "She wants to maintain phonetic balance despite the overwhelming clinical and experimental experience that indicates phonetic balance to be an interesting but unnecessary component of one of our current audiometric tests."[4]

Tobias feels that phonetic balance does not seem to be a meaningful criterion.[5]

---

[1] P. Grubb, "A Phonemic Analysis of Half-List Speech Discrimination Tests," Journal of Speech and Hearing Research, VI (1963), p. 271.

[2] Ibid., pp. 271-275.

[3] P. Grubb, "Some Considerations in the Use of Half-List Speech Discrimination Tests," Journal of Speech and Hearing Research, VI (1963), pp. 291-297.

[4] J. V. Tobias, "On Phonemic Analysis of Speech Discrimination Tests," Journal of Speech and Hearing Reasearch, VII (1964), p. 99.

[5] Ibid., pp. 95-102.

Social Adequacy Index

In 1948, Hallowell Davis wrote of the "obvious need for a quantitative measure of the social adequacy of hearing."[1] Developing an earlier idea of Walsh and Silverman, he constructed the Social Adequacy Index. This is an average of the percentage of words heard correctly at three (faint, average, loud) levels of speech. "It indicates how well a person hears speech under average everyday conditions."[2] Davis warns that since the Social Adequacy Index was derived from the Harvard PB word lists, caution must be used when working with W-22's since persons will get different scores on the two lists.

Speech Intelligibility

To understand speech audiometry more fully, especially the factors that enable a listener to respond to what he hears, it is necessary to be familiar with some of the factors that make speech intelligible. In this discussion it will be assumed that the following factors are held constant: (1) high quality communication system, (2) listener's familiarity with the language and

---

[1]H. Davis, "The Articulation Area and the Social Adequacy Index for Hearing," Laryngoscope, LVIII (1948) p. 761.

[2]Ibid., p. 776.

speech sounds that are used, (3) a properly functioning central nervous system, and (4) normal hearing.

The sounds of speech are divided into two classes, vowels and consonants. The latter of these two seems to be responsible for most of the intelligibility carried by individual English words.[1] Another important factor of intelligibility is intensity. As intensity of speech increases from a low level to a higher level, intelligibility will increase. The number of syllables in the speech stimulus also plays an important role. The more syllables there are per word at a given intensity, the more intelligible the word.[2] Still another contributor to intelligibility is the amount of context that is present in the sample. One word is not as intelligible as one sentence because of the additional cues offered by the latter, and also because in the sentence the surrounding words cut down the number of possibilities that are available for choice. The fewer number of possibilities, the better the intelligibility.

Davis et al. list the following as factors affecting intelligibility, ". . .variety of voices, voice levels, speech sounds, and acoustic environment."[3]

---

[1]Ira J. Hirsh, The Measurement of Hearing (New York: McGraw-Hill Book Company, Inc., 1952), p. 125.

[2]Ibid., pp. 133-135.

[3]H. Davis, et al., "The Selection of Hearing Aids," Laryngoscope, LVI (1946), p. 89.

Miller, Heise, and Lichten, in an article on the intelligibility of speech say, "The ease with which a discrimination of speech sounds can be made is limited according to the number of different speech sounds that must be discriminated."[1] They also state that the most important variable in intelligibility is the range of possible alternatives from which a test item is selected. For this reason, words in sentences have a threshold 6 dB lower than the same words presented in isolation. They also found that digits had the lowest threshold with sentences and nonsenses syllables following in that order.[2]

This does not exhaust the factors that affect intelligibility, but it certainly covers some of the more important ones.

Using a recorded sample of actual everyday connected descourse, Falconer and Davis found they could establish a listener's speech reception threshold by allowing the listener to determine it for himself. The listener would attenuate the speech until he could just hear and understand what was being said. They labeled the result the Threshold of Intelligibility for Connected

---

[1] G. A. Miller, G. Heise, and W. Lichten, "The Intelligibility of Speech as a Function of the Context of the Test Materials," Journal of Experimental Psychology, XLI (1951), p. 332.

[2] Ibid., 330-334.

Discourse (TICD). They concluded that the TICD compared very favorably with Auditory Test No. 9 with respect to both dispersion and reliability. The absolute threshold, mean for a normal group, was 23.23 ± 3.77 dB re .0002 dynes/cm2. It was shown that about two minutes of connected discourse or "cold running speech" can be used to establish a speech reception threshold almost as accurately as four or five minutes with spondees or the fading numbers record.[1]

Everyday Sentences

The sentences used in the current investigation were developed at Central Institute for the Deaf. The reasoning and need for developing these sentences was published by Silverman and Hirsh in 1955. They stated that:

> . . .the speech stimulus that is required for the most effective diagnostic differentiation is not necessarily the most representative of everyday speech. We have shown, for example, that relatively short, difficult words are required and, furthermore, that it is necessary that the talker of these words not overarticulate and be not too clear. . . .it seems reasonable to suppose that since everyday hearing is characterized largely by the hearing of speech, that speech may be a more valid predicting test material. But we do not expect, necessarily, that our lists of monosyllabic words, which aid us in diagnosis

---

[1]G. A. Falconer and H. Davis, "The Intelligibility of Connected Discourse as a Test for the 'Threshold for Speech'," Laryngoscope, LVII (1947), pp. 581-591.

would predict this ability very well, since such words can hardly be called representative of everyday speech. Thus, a court of law can justifiably question the validity for rating disability from monosyllabic-word tests or from pure-tone audiograms, until a systematic study of validation has related one or the other to everyday hearing.

We must now forget concepts of hearing tests and approach the problem from the point of view of attempting to characterize samples of everyday speech.

One approach to a solution of this problem is to set up criteria for sampling everyday speech. Such criteria should, at least on the face of it, make sense and indeed such a feature of a test sample is often referred to as 'face validity'.[1]

They then outlined the problem and turned it over to the Armed Forces-National Research Council Committee on Hearing and Bio-Acoustics (CHABA), who in turn appointed a working group to formulate a set of criteria for representing, in a sample, everyday speech. The committee consisted of John W. Black, James F. Curtis, James P. Egan, Harold C. Whitehall, and Grant Fairbanks, Chairman.

One a priori assumption agreed upon was that the sample item should be the sentence. The criteria for this speech material were suggested by the committee and are recorded in Appendix A.[2]

---

[1] S. R. Silverman and I. J. Hirsh, "Problems Related to the Use of Speech in Clinical Audiometry," Annals of Otology, Rhinology and Laryngology, LXIV (1955), p. 1241.

[2] Ibid., pp. 1242-1243.

A set of one hundred sentences was subsequently contructed at the Central Institute for the Deaf. It was decided to use these sentences in the present investigation since they appear to be carefully constructed, appear to have "face validity", and are readily obtainable. The sentences as listed by Davis and Silverman[1] are in Appendix B.

Zerlin and Urban have recognized the need for, ". . .intelligibility test materials which have greater face validity than single words."[2] They mention that the fundamental obstacle in using speech samples longer than single words is defining the energy level. Davis and Silverman state, "The usual convention is to take a sort of running average of the largest of the excursions of the (V.U.) meter as it swings in response to the syllables of the words."[3] The method used in the present study is a modification of that used by Zerlin and Urban of integrating energy over time.[4]

---

[1]H. Davis and S. R. Silverman, Hearing and Deafness pp. 549-552.

[2]S. Zerlin and B. Urban, "The Evaluation of Contextual Speech Materials," Journal of Speech and Hearing Research, VI (1963), p. 291.

[3]Davis and Silverman, p. 182.

[4]Zerlin and Urban, pp. 291-293.

CHAPTER III

PROCEDURES AND EQUIPMENT

Subjects

Thirty subjects were employed in the study. All subjects had normal hearing and were between the ages of 20 and 26 years. For the purpose of the study, a subject was considered to have normal hearing if his hearing loss did not average more than ten decibels overall with no more than 15 dB at 250, 500, 1000, 2000, or 4000 cycles per second. The audiometer was calibrated to ASA specifications (Z24.5-1951). None of the subjects were familiar with the sentences.

Procedures

Each subject listened to ten lists of ten sentences each (see Appendix B). Each list was presented at ten different, predetermined intensities. The experimental design used was a 10 x 10 x 10 Graeco-Latin Square (Appendix C.).[1] Consequently, the subjects were divided into three groups of ten. Therefore, each list

---

[1]R. A. Fisher and F. Yates, Statistical Tables for Biological, Agricultural and Medical Research (6th ed.; London: Oliver and Boyd, 1963), p. 25.

was heard by three listeners at the same intensity and each list was heard at ten intensities.

The ten intensity levels were determined from a pilot study where three listeners required a range of 36 dB to obtain consistently a score from 0 to 100 per cent. This range was then divided into ten equal subdivisions ranging from 8 dB to 44 dB SPL in four dB steps.

Sentences

The sentences used in the study were prepared at The Central Institute for the Deaf, St. Louis. There are 100 sentences divided into ten lists, and each sentence is constructed to be representative of "everyday American speech." The criteria for constructing the sentences were laid down by a Working Group of the Armed-Forces-National Research Council Committee on Hearing and Bio-Acoustics (see Appendix A).

Talkers

Ten untrained talkers were employed to speak the sentences so they could be recorded. There were five male and five female general American talkers whose ages ranged from 24 years through 68 years. There was one male and one female in each of the following age groups: 20-29; 30-39; 40-49; 50-59; and 60-69. An attempt was made to preserve naturalnessof speech in order to represent an everyday speaking situation. The talkers were given the following instructions before the sentences were recorded:

> Here are ten, everyday sentences that I would like to have you read. You will notice that they are common, everyday sentences and that is exactly how I would like you to read them--with your normal, everyday speech. No attempt should be made to make them any clearer then you would speak them to a close friend, in a room at a distance of five feet. I would suggest you read them over once, and then say them twice outloud before we begin recording.

Recording

The sentences were read in an audiometric testing room and the recorder was located in an adjacent control room. The two rooms had a glass window between them. An Electro-Voice model 654 microphone picked up and delivered the sentences to an Ampex PR 10 tape recorder that was running at a speed of seven and one-half inches per second. Scotch Low Noise 202 recording tape was employed. A Bruel and Kjaer Precision Sound Level Meter type 2203 was used to set a 1000 cycle per second calibrating tone at the microphone. The tone was recorded before each talker at a level of 75 dB SPL.

The list given to each talker contained one randomly selected sentence from each of the ten original lists. After the ten randomly selected lists were recorded, they were transferred to a master tape in their original order. The final result was that each original list, A through J. had ten talkers arranged in random order.

The master tape was then run through a Bruel and Kjaer Level Recorder Type 2305 with a writing speed of 400 millimeters per second and a paper speed of 100 millimeters per second. By measuring (in millimeters) the amount of the speech signal of each list that was above the intensities of 45, 51, 63, 69, 75, 81, 87, and 93 dB SPL and plotting these on a graph, the median intensities of each list were found. The graphs had "Reference Intensity in dB re .0002 microbar" as the abcissa and "Per Cent below Reference Intensity" as the ordinate. The median intensity for each list was then found by locating the intensity at which the curve crossed the 50 per cent point on the ordinate. The mean intensity over the ten median intensities was 70 dB SPL and the range of median intensities was from 68 to 72 dB. The graphs are shown in Appendix D.

Testing

The thirty subjects were tested individually in a soundfield situation in the same room where the sentences were recorded. The subjects sat in a chair facing the speaker which was an Electro-Voice SP-12. The master tape was played on the Ampex PR-10 at seven and one-half inches per second and fed into an Allison 20-B speech audiometer, and then into the speaker. The ambient noise level of the room was about 45 dB SPL and at the chair there was a constant 30 dB SPL noise originating in the power amplifier. The intensity of this noise at the chair was very roughly determined by using the inverse square law.

The lists and intensities were presented according to the schedule set up by a 10 x 10 x 10 Graeco-Latin square (see Appendix C.). Each listener was given a pure-tone screening test to detemine if he had normal hearing, as previously defined, before the test began. The instructions read to each listener before the test were as follow:

> You are about to hear 100 sentences divided equally into ten lists. I would like you to repeat each sentence as you hear it. Some lists will be so soft you won't be able to hear all the words, but don't get discouraged; it's not your fault. Don't be afraid to guess. If you only hear one or two words repeat them. The sentences are what you would expect to hear in common, everyday American speech. There is a green light in that window. When it goes on, you should listen because a sentence will follow. When the light goes off, you are to repeat the sentence. Do you have any questions?

The signal light was activated by a 1000 cycle per second tone that was recorded on track B of the tape and was picked up by a three transistor amplifier. The amplified tone operated the relay, which in turn operated the light. The tape was stopped after each sentence so the light was also off.

The VU meter was adjusted to read -5 when the calibrating tone before each list was presented. Because the median intensity of each list was 70 dB SPL and the calibrating tone was originally recorded at 75 dB SPL, the playback system was now reproducing the sentences 10 dB more intense than their median intensity. Therefore, in order to play the sentences back at levels relative to their recorded intensity, 10 dB was subtracted from the attenuator reading. If the VU meter had been set to read 0 on the calibrating tone, much of the speech signal would have been distorted. By calibrating the VU meter to -5 and subtracting ten dB from the attenuator reading, two things were accomplished: (1) there was no distortion in the playback system, and (2) the lists were played back relative to their recorded intensity.

The scores were tallied on a special score sheet devised for the study (see Appendix E). For each list, the total number of words repeated correctly out of a possible 50 was multiplied by two. This gave the per cent correct for each list at its corresponding intensity.

The average time required to administer and score one list of ten sentences was one minute twenty-eight seconds.

CHAPTER IV

RESULTS AND DISCUSSION

Each of the thirty subjects obtained ten scores--one score for each of the ten conditions to which he was exposed (see Appendices C and F). The scores are expressed in terms of percentage of 50 key words correctly repeated from each list. Each row of Appendix F represents one subject and each column represents the trial (or order of presentation) on which the particular score was obtained. The eleventh and twenty-first rows are the beginning of the second and third replications, respectively, of the Graeco-Latin square.

Figures 1, 3, and 4 graphically represent the mean scores for sound pressure levels, presentations, and lists respectively. The ordinate in all three tables is the percentage of 50 key words correctly repeated. Figure 4 is presented as a bar graph because the lists are on a nominal scale.

Figure 1 shows a very definite upward trend for intelligibility plotted as a function of sound pressure level. This is the curve which is of primary concern in this study. The small circles represent the mean articulation score at each corresponding intensity level.

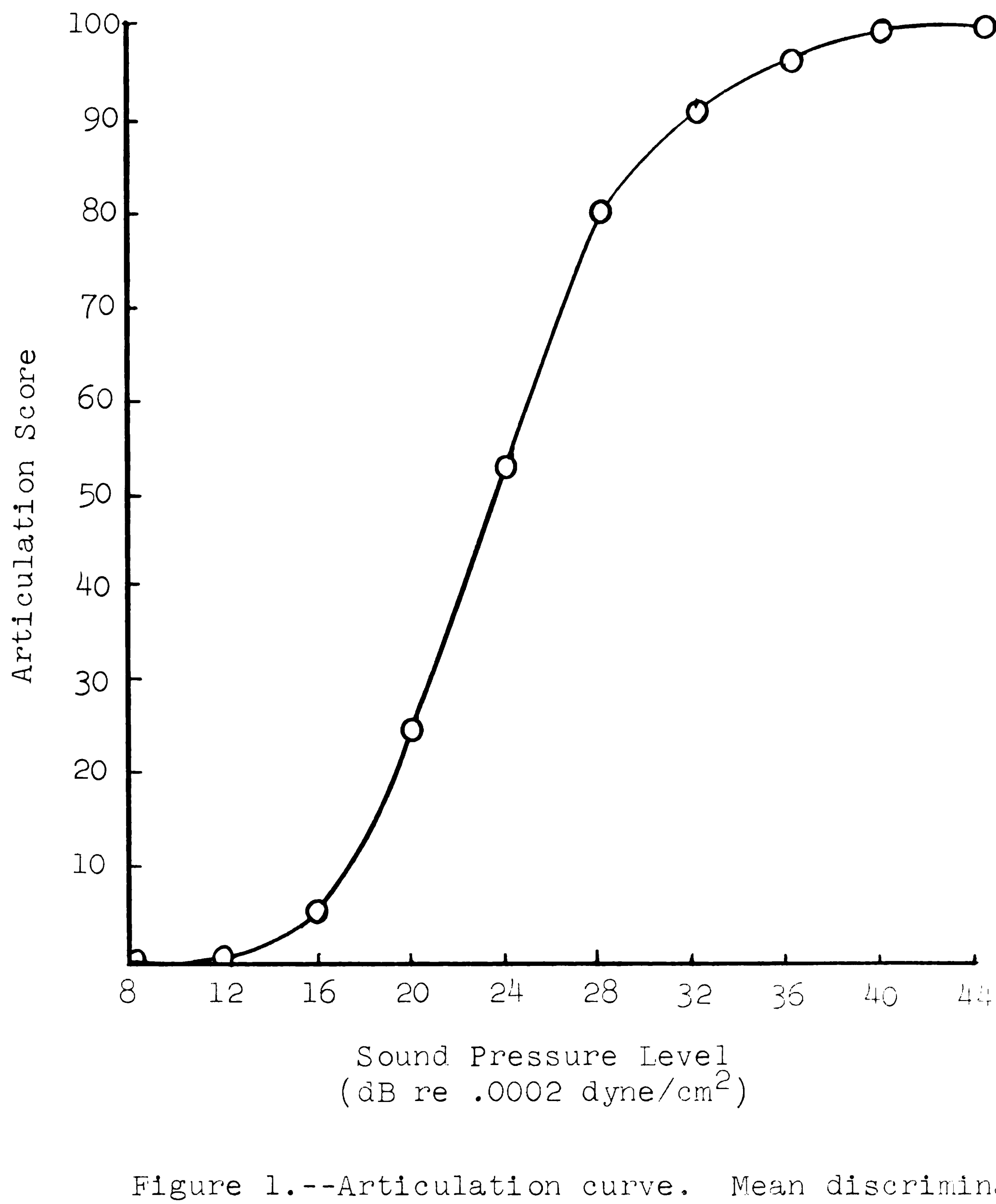


Figure 1.--Articulation curve. Mean discrimination score is plotted as a function of the sound pressure level of the stimulus.

A French Curve was employed to connect the mean articulation scores. The curve is a smooth S shape with a steep slope of about seven per cent per dB at the center. This general shape is not unlike those reported in Stevens[1] and in Hirsh[2] for spondees, monosyllables and PB's. In fact, it corresponds very closely with the curve for the CID Auditory Test W-1 reported by Hirsh *et al.*[3]

In the sentence articulation curve presented in Figure 1, there is an increase from 20 to 80 per cent within a range of 9 dB and throughout this range the slope or rate of rise in score is about seven per cent per dB. The rate of rise of the curve tapers off above 80 percent and does not reach the 100 per cent point until about 16 dB above threshold. Hirsh says, "Below threshold the words drop out very quickly and there is little if any 'tail' at this end of the curve."[4]

The articulation curve obtained by Hirsh had absolute thresholds for spondees at 20 dB and 21 dB

---

[1]S. S. Stevens (ed.), *Handbook of Experimental Psychology* (New York: John Wiley and Sons, Inc., 1951), P. 1046.

[2]Ira J. Hirsh, *The Measurement of Hearing* (New York: McGraw-Hill Book Company, 1952), p. 135.

[3]Ira J. Hirsh *et al.*, "Development of Materials for Speech Audiometry," *Journal of Speech and Hearing Disorders*, XVII (1952), pp. 325-326.

[4]*Ibid*.

re .0002 microbar for experienced and inexperienced listeners respectively, listening monaurally. He explains though that in clinical use "the expected threshold will be more nearly 18 dB."[1] For PB lists he obtained a threshold at 24 dB re .0002 microbar. The threshold for the sentence curve of this study is 23.5 dB re .0002 microbar.

In Chapter II it was stated that the intelligibility for spondees is greater than the inteligibility of PB's and the intelligibility of sentences is greater than the intelligibility of spondees. Licklider and Miller say, "Sentence scores are usually higher than words or syllable scores obtained under identical conditions."[1] The curve presented in Figure 1 does not completely conform to this when it is compared with the reported curves for spondees and PB's. Of course, this sentence curve was not constructed under the identical conditions as those employed by Hirsh and others. While the slope for this curve is about seven per cent per dB, Hirsh found the slope for the spondees of CID Auditory Test W-1, to be eight per cent dB, and about 3.7 per cent per dB for the PB words of CID Auditory Test W-22. Figure 2 shows the relation between Auditory Test W-1, W-22 and the present sentence curve that was taken from Figure 1. Why does the above

---

[1]Stevens, p. 1046.

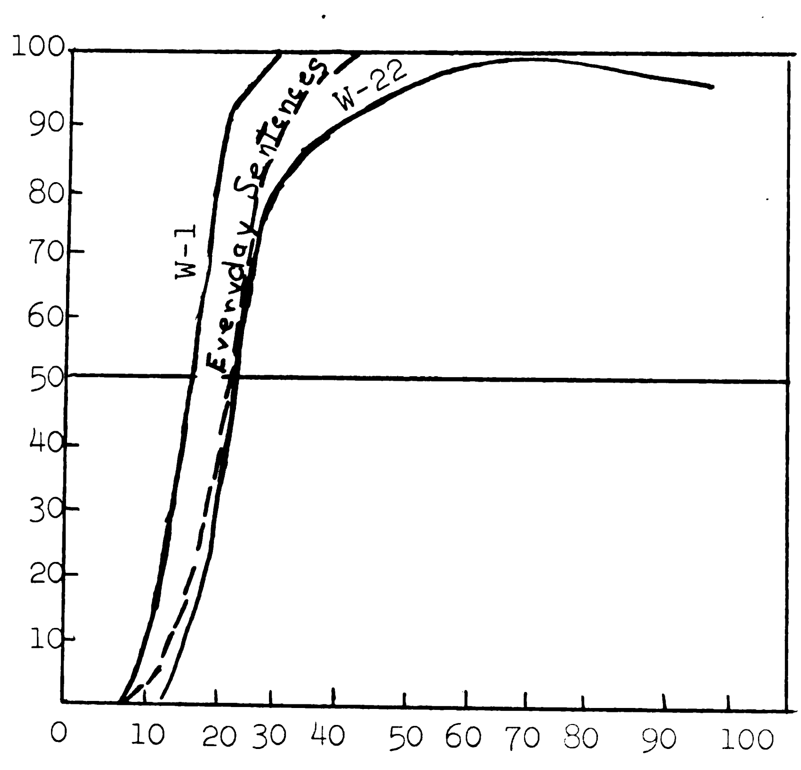


Figure 2.--Relations between Auditory Test W-1, W-22 and the recorded sentences from Figure 1. (from Hirsh *et al.*)[1]

---

[1]Hirsh *et al.*, p. 334.

curve for sentences fall between the curves for spondees and PB's when actually it might be expected to be found to the left of the W-1 curve? One possible reason, other than the previously mentioned one of equal recording conditions, is in the selection of speakers. Hirsh employed one trained, male speaker who carefully monitored his speech with a VU meter. The ten talkers used in the present study were not trained speakers, their ages ranged from 24-68 years, and they were not instructed to speak clearly. In a discussion on the construction of articulation tests, Licklider and Miller state, "Talker, rather than listener, variability is a more important source of instability."[1] Variation in speakers, voices, intensities, inflections, ages, etc., could possibly explain the unexpectedly high threshold obtained in this study.

Figure 3 does not appear to present any great learning or fatigue effect due to practice as the test progresses. Figure 4 likewise appears to show no great differences between the ten lists.

The variances among the ordinate values for each of Figures 1, 3, and 4 were tested for statistical significance by analysis of variance. A summary of the analysis is presented in Table I.

---

[1] Stevens, p. 1046.

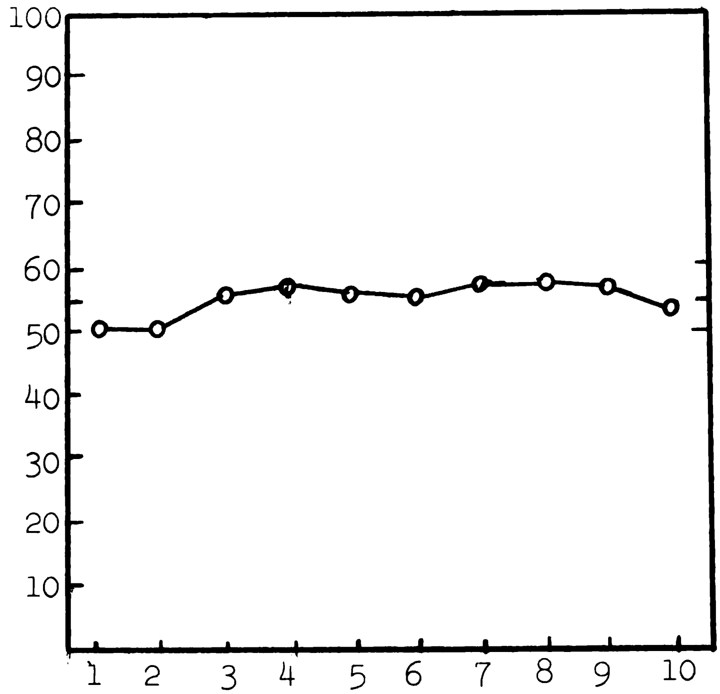


Figure 3.--Practice Effect. Mean discrimination score is plotted as a function of successive trials.

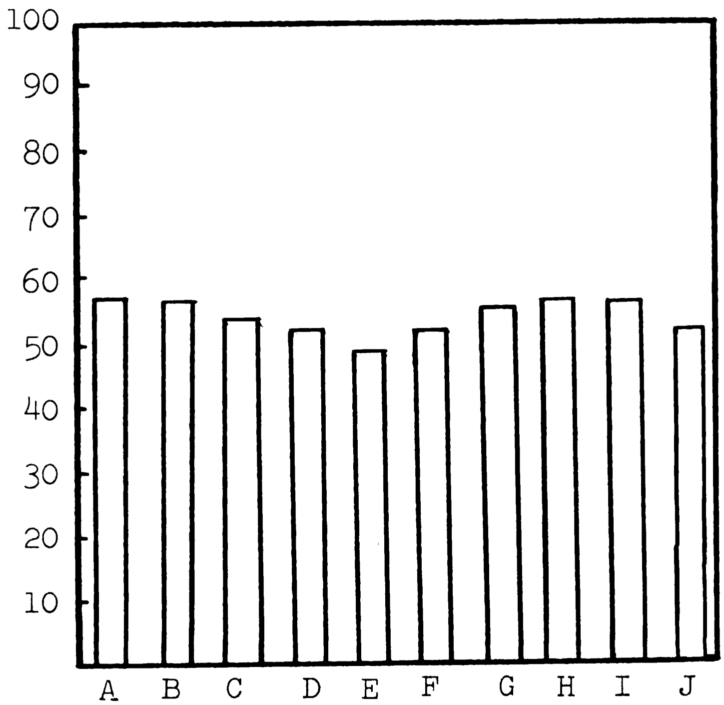


Figure 4.--Difference among Lists. Each bar represents the mean discrimination score on one list of the test.

TABLE I.--Summary for Analysis of Variance

Source of Variation	Sum of Squares	df	Mean Square	F
Ordering of Presentation	1925.19	9	213.91	.66
Subjects Within Orderings	6414.40	20	320.72	
Within Subjects				
Intensity	501703.59	9	55744.843	664.23
Lists	2132.65	9	236.961	2.8235
Order	2394.52	9	266.057	3.1702
Error with Subjects	20393.637	243	83.924	
TOTAL	534963.987			

The results indicate that the sources of variance due to intensity, lists, and order are significant. As would be expected, intensity was highly significant as a source of variation. Lists and order are small but significant sources of variation. It is believed that these two are significant because of the extreme sensitivity of the experimental design.

CHAPTER V

SUMMARY AND CONCLUSIONS

Summary

Determination of the social function of hearing plays an important role in audiology. The evaluation of the social aspects of a hearing impairment is carried out through speech audiometry. Currently, word lists are used for the purposes of determining a person's ability to hear and understand speech. Since we do not usually speak in single words, it was thought that perhaps sentences would offer a more reliable means of testing the social function of hearing. However, before a test can be utilized, it must have a standard to which the results may be compared, hence the purpose of this study.

An articulation curve was constructed for 100 sentences using 30 normal-hearing subjects. The sentences were recorded by five male and five female talkers. The resulting articulation curve was not greatly unlike other articulation curves that have been constructed for word lists. Statistically, the lists and the order of presentation were found to have a small but significant effect on the results.

The threshold of intelligibility for these sentences was found to be slightly higher than would be expected, but the extreme variation of speakers is believed to be a possible explanation for this.

Conclusions

The purpose of this study was to record lists of sentences representative of everyday speech and to determine a normative articulation function for the recorded lists. In accomplishing that purpose, the intensity parameters of each sentence list were graphed. The following conclusions were reached at the end of the investigation.

1. The intensity parameters of the ten CID everyday sentence lists as recorded for this study are as shown in Appendix D.
2. The articulation curve for the ten CID lists as recorded for this study is as shown in Figure 1. The differences among mean articulation scores for different intensities are statistically significant.
3. There appear to be small but signigicant differences among the mean articulation scores due to the use different lists.
4. There appear to be small but significant differences among the mean articulation scores due to order of presentation of the ten CID lists

recorded for this study. In other words, learning or fatigue seemed to play a small but significant role in determining articulation scores in this study.

Implications for Further Research

During the present study, many questions have arisen that should lead to more research utilizing these same sentences. Some of the questions are listed as follows:

1. What is the difference in articulation curves drawn for each of the ten talkers used in this study?
2. Can the Social Adequacy Index be applied to the scores derived from these sentences?
3. What shape does the articulation curve of these sentences take when they are presented to pathological ears?
4. How would the curves established in Number 3 compare with results obtained with currently used word lists?
5. Can these sentences be employed as a valid and reliable tool in hearing aid evaluations?
6. In what way would this articulation curve be displaced if only one speaker was used to speak the sentences?

Now that a norm has been established for these sentences, they are ready to be utilized on a trial basis in a clinical situation.

BIBLIOGRAPHY

Books

Davis, Hallowell and Silverman, S. Richard. Hearing and Deafness. New York: Holt, Rinehart, and Winston, Inc., 1960.

Fisher, R. A. and Yates, R. Statistical Tables for Biological, Agricultural and Medical Research. London: Oliver and Boyd, 1963.

Hirsh, Ira J. The Measurement of Hearing. New York: McGraw-Hill Book Company, 1952.

Newby, Hayes A. Audiology. New York: Appleton-Century Crofts, Inc., 1964

Stevens, S. S. (ed.) Handbook of Experimental Psychology. New York: John Wiley and Sons., Inc., 1951.

Watson, Leland A. and Tolan, Thomas. Hearing Tests and Hearing Instruments. Baltimore: The Williams and Wilkins Company, 1949.

Periodicals

Campanelli, P. A. "Measure of IntraList Stability of Four PAL Word Lists," Journal of Auditory Research, II (1962), pp. 50-55.

Davis, Hallowell. "The Articulation Area and the SAI for Hearing," Laryngoscope, LVIII (1948), pp. 761-778.

Davis, Hallowell, Hudgins, C. V., Marquis, R. J., Nichols, R. H., Peterson, G. E., Ross, D. A., and Stevens, S. S. "The Selection of Hearing Aids," Laryngoscope, LVI (1946), pp. 85-115.

Egan, James, P. "Articulation Testing Methods," Laryngoscope, LVIII (1948), pp. 955-991.

Elpern, B. S. "The Relative Stability of Half-List and Full-List Discrimination Tests," Laryngoscope, LXXI (1961), pp. 30-36.

Falconer, G. A. and Davis, H. "The Intelligibility of Connected Discourse as a Test for the 'Threshold of Speech'," Laryngoscope, LVII (1947), pp. 581-595.

Fletcher, H. and Steinberg, J. C. "Articulation Testing Methods," Bell System Technical Journal, VIII (1929), pp. 806-854.

Grubb, Patti. "A Phonemic Analysis of Half-List Speech Discrimination Tests," Journal of Speech and Hearing Research, VI (1963), pp. 271-275.

Grubb, Patti. "Some Considerations in the Use of Half-List Speech Discrimination Tests," Journal of Speech and Hearing Research, VI (1963), pp. 294-297.

Hirsh, Ira J., Davis, Hallowell, Silverman, S. Richard, Reynolds, Elizabeth G., Eldert, Elizabeth, and Benson, Robert W. "Development of Materials for Speech Audiometry," Journal of Speech and Hearing Disorders, XVII (1952), pp. 321-337.

Hudgins, Clarence, et al. "The Development of Recorded Auditory Tests for Measuring Loss for Speech," Laryngoscope, LVII (1947), pp. 57-89.

Lynn, G. "Paired PB-50 Discrimination Test: A Preliminary Report," Journal of Auditory Research, II (1962), pp. 34-36.

Miller, G. A., Heise, G. and Lichten, W. "The Intelligibility of Speech as a Function of the Context of the Test Materials," Journal of Experimental Psychology, XLI (1951), pp. 329-335.

Resnick, D. M. "Reliability of the Twenty-Five Word Phonetically Balanced Lists," Journal of Auditory Research, II (1962), pp. 5-12.

Shutts, Edwin R., Burke, Kenneth S., and Creston, James E. "Derivation of Twenty-Five-Word PB Lists," Journal of Speech and Hearing Disorders, XXIX (1964), pp. 442-447.

Silverman, S. Richard and Hirsh, Ira J. "Problems Related to the Use of Speech in Clinical Audiometry," Annals of Otology, Rhinology and Laryngology, LXIV (1955), pp. 1234-1244.

Tobias, Jerry V. "On Phonemic Analysis of Speech Discrimination Tests," Journal of Speech and Hearing Research, VII (1964), pp. 98-100.

Zerlin, Stanley and Urban, Beth. "The Evaluation of Contextual Speech Materials," Journal of Speech and Hearing Research, VI (1963), pp. 291-293.

APPENDICES

APPENDIX A

Criteria for Sampling Everyday Speech as Suggested by CHABA[1]

1. The level should be specified in terms of relative frequency, age level, or educational level. There are numerous sources. The level should be of high frequency. The words should be common so that the test in no part depends upon vocabulary. The words should not be selected informally on the basis of personal estimate, but should appear in some specific fist [sic].

2. Within these objective limits, the vocabulary range should be fully exploited so that as many different words occur as possible.

3. Proper names and proper nouns should be excluded. They are unnecessary and unpredictable as to effect on validity.

4. Word length, measured in syllables, should be controlled. The ultimate test as a whole should have a distribution of good fit to the distribution of the vocabulary pool specified.

5. In the matter of syllabic stress, in so far as this is inherent in words, free variation is suggested. Patterns and unusual departures from live speech should be avoided.

6. Contractions should be used freely and frequently. As a principle, they should be used whenever possible.

7. The ultimate test as a whole should have a phonetic frequency distribution that does not differ significantly from that of language, and this control should be demonstrated objectively by comparison to an existent criterion.

---

[1]S. Richard Silverman and Ira J. Hirsh, "Problems Related to the Use of Speech Audiometry," _Annals of Otology, Rhinology and Laryngology_, LXIV (1955), p. 1242-1243.

Sentence Structure

1. The phonetic structure of a given sentence should be as such to avoid "loading," or unnaturally high frequency of occurrence of any one element, such as characterizes certain tongue-twisters. Within each sentence, the phonetic distribution should be at random.

2. To avoid testing memory span, an upper limit of sentence length should be set at 12 words. The lower limit should be fixed at two words, considering that many sentences are of this length and that one-word sentences should be avoided as duplicative of word tests. The distribution over this range, expressed in proportional parts, should be as follows:

Sentence length	Parts
2 - 4	1
5 - 9	2
10 - 12	1

3. On the grounds that it will increase variety and interest, has face validity, and may be important (although the latter we do not know), the sentence form should be controlled as follows:

Sentence Form	Parts
Declarative	8
Imperative	1
Rising Interrogative	1/2
Falling Interrogative	1/2

This distribution appears not to depart far from that of American English in general.

4. Grammatical structure should vary freely and widely, and should avoid sterotyped forms.

5. Common, non-slang idioms should be used freely and it is desirable that they be numerous.

6. Redundancy should be high. An important aspect of validity is inference of unheard or incompletely heard material from fragments. In other words, it would not be good practice to build items all of which demand that every word be heard.

7. Sentence content should be appropriate for adults.

8. Levels of abstraction should be low to avoid the factors of intelligence, etc.

APPENDIX B

Test Sentences[1]

List A

1. Walking's my favorite exercise.
2. Here's a nice quiet place to rest.
3. Our janitor sweeps the floors every night.
4. It would be much easier if everyone would help.
5. Good morning.
6. Open your window before you go to bed!
7. Do you think that she should stay out so late?
8. How do you feel about changing the time when we begin work?
9. Here we go.
10. Move out of the way!

List B

1. The water's too cold for swimming.
2. Why should I get up so early in the morning?
3. Here are your shoes.
4. It's raining.
5. Where are you going?
6. Come here when I call you!
7. Don't try to get out of it this time!
8. Should we let little children go to the movies by themselves?
9. There isn't enought paint to finish the room.
10. Do you want an egg for breakfast?

List C

1. Everybody should brush his teeth after meals.
2. Everything's all right.
3. Don't use up all the paper when you write your letter.
4. That's right.

---

[1]Davis and Silverman, *Hearing and Deafness*, pp. 548-552.

5. People ought to see a doctor once a year.
6. Those windows are so dirty I can't see anything outside.
7. Pass the bread and butter please!
8. Don't forget to pay your bills before the first of the month.
9. Don't let the dog out of the house!
10. There's a good ballgame this afternoon.

List D

1. It's time to go.
2. If you don't want these old magazines, throw them out.
3. Do you want to wash up?
4. It's a real dark night so watch your driving.
5. I'll carry the package for you.
6. Did you forget to shut off the water?
7. Fishing in a mountain stream is my idea of a good time.
8. Fathers spend more time with their children than they used to.
9. Be careful not to break your glasses!
10. I'm sorry.

List E

1. You can catch the bus across the street.
2. Call her on the phone and tell her the news.
3. I'll catch up with you later.
4. I'll think it over.
5. I don't want to go to the movies tonight.
6. If your tooth hurts that much, you ought to see a dentist.
7. Put that cookie back in the box!
8. Stop fooling around!
9. Time's up.
10. How do you spell your name?

List F

1. Music always cheers me up.
2. My brother's in town for a short while on business.
3. We live a few miles from the main road.
4. This suit needs to go to the cleaners.
5. They ate enough green apples to make them sick for a week.

6. Where have you been all this time?
7. Have you been working hard lately?
8. There's not enough room in the kitchen for a new table.
9. Where is he?
10. Look out!

List G

1. I'll see you right after lunch.
2. See you later.
3. White shoes are awful to keep clean.
4. Stand there and don't move until I tell you!
5. There's a big piece of cake left over from dinner.
6. Wait for me at the corner in front of the drugstore.
7. It's no trouble at all.
8. Hurry up!
9. The morning paper didn't say anything about rain this afternoon or tonight.
10. The phone call's for you.

List H

1. Believe me!
2. Let's get a cup of coffee.
3. Let's get out of here before it's too late.
4. I hate driving at night.
5. There was water in the cellar after that heavy rain yesterday.
6. She'll only be gone a few minutes.
7. How do you know?
8. Children like candy.
9. If we don't get rain soon, we'll have no grass.
10. They're not listed in the new phone book.

List I

1. Where can I find a place to park?
2. I like those big red apples we always get in the fall.
3. You'll get fat eating candy.
4. The show's over.
5. Why don't they paint their walls some other color?
6. What's new?
7. What are you hiding under your coat?

8. How come I should always be the one to go first?
9. I'll take sugar and cream in my coffee.
10. Wait just a minute!

List J

1. Breakfast is ready.
2. I don't know what's wrong with the car, but it won't start.
3. It sure takes a sharp knife to cut this meat.
4. I haven't read a newspaper since we bought a television set.
5. Weeds are spoiling the yard.
6. Call me a little later!
7. Do you have change for a five-dollar bill?
8. How are you?
9. I'd like some ice cream with my pie.
10. I don't think I'll have any dessert.

APPENDIX C

GRAECO-LATIN SQUARE FOR TEST PRESENTATIONS*

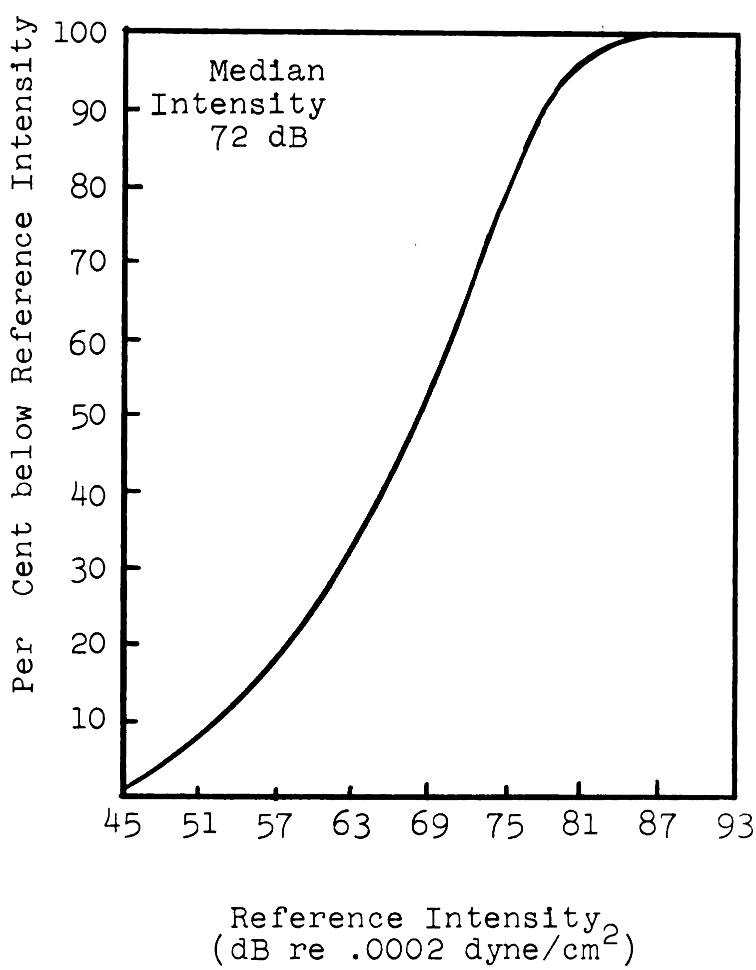
Subject	Order of Presentations									
	1	2	3	4	5	6	7	8	9	10
1	G-32	J-24	I-36	C-28	F- 8	B-20	A-44	D-40	H-12	E-16
2	F-28	I-20	H-32	E- 8	B-44	A-40	J-36	G-12	D-16	C-24
3	D-20	G-40	F-24	A-32	H-28	E-12	I-16	C-44	J- 8	B-36
4	C-16	A-12	B- 8	I-44	J-20	D-24	E-28	F-32	G-36	H-40
5	H-36	D-28	J-40	F-16	C-32	G- 8	B-24	A-20	E-44	I-12
6	E-24	H-44	H-28	B-40	A-36	I-32	F-12	J-16	C-20	D- 8
7	B-12	C- 8	A-16	H-20	I-24	J-28	D-32	E-36	F-40	G-44
8	J-44	F-36	E-20	G-24	D-12	H-16	C-40	I- 8	B-32	A-28
9	I-40	E-32	D-44	J-12	G-16	C-36	H- 8	B-29	A-24	F-20
10	A- 8	B-16	C-12	D-36	E-40	F-44	G-20	H-24	I-28	J-32
11	G-32	J-24	I-36	C-28	F- 8	B-20	A-44	D-40	H-12	E-16
12	F-28	I-20	H-32	E- 8	B-44	A-40	J-36	G-12	D-16	C-24
13	D-20	G-40	F-24	A-32	H-28	E-12	I-16	C-44	J- 8	B-36
14	C-16	A-12	B- 8	I-44	J-20	D-24	E-28	F-28	G-36	H-40
15	H-36	D-28	J-40	F-16	C-32	G- 8	B-24	A-20	E-44	I-12
16	E-24	H-44	G-28	B-40	A-36	I-32	F-12	J-16	C-20	D- 8
17	B-12	C- 8	A-16	H-20	I-24	J-28	D-32	E-36	F-40	G-44
18	J-44	F-36	E-20	G-24	D-12	H-16	C-40	I- 8	B-32	A-28
19	I-40	E-32	D-44	J-12	G-16	C-36	H- 8	B-28	A-24	F-20
20	A- 8	B-16	C-12	D-36	E-40	F-44	G-20	H-24	I-28	J-32
21	G-32	J-24	I-36	C-28	F- 8	B-20	A-44	D-40	H-12	E-16
22	F-28	I-20	H-32	E- 8	B-44	A-40	J-36	G-12	D-16	C-24
23	D-20	G-40	F-24	A-32	H-28	E-12	I-16	C-44	J- 8	B-36
24	C-16	A-12	B- 8	I-44	J-20	D-24	E-28	F-32	G-36	H-40
25	H-36	D-28	J-40	F-16	C-32	G- 8	B-24	A-20	E-44	I-12
26	E-24	H-44	G-28	B-40	A-36	I-32	F-12	J-16	C-20	D- 8
27	B-12	C- 8	A-16	H-20	I-24	J-28	D-32	E-36	F-40	G-44
28	J-44	F-36	E-20	G-24	D-12	H-16	C-40	I- 8	B-32	A-28
29	I-40	E-32	D-44	J-12	G-16	C-36	H- 8	B-28	A-24	F-20
30	A- 8	B-16	C-12	D-36	E-40	F-44	G-20	H-24	I-28	J-32

*Letter represents respective lists; number represents intensity in sound pressure level.

APPENDIX D

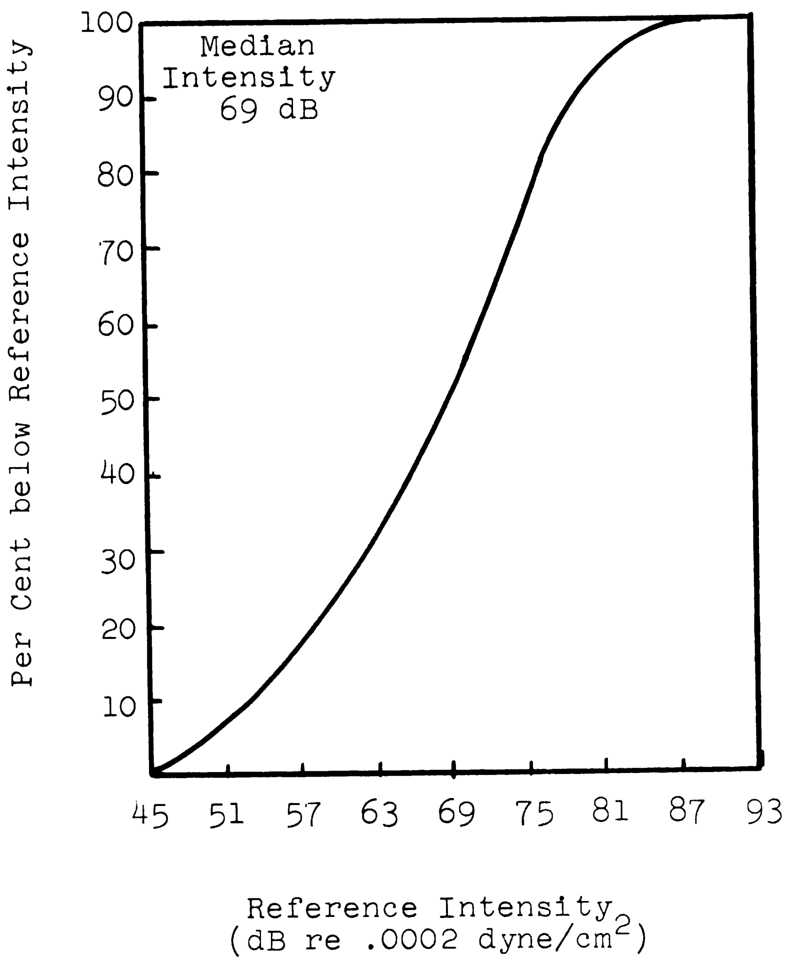
GRAPHIC REPRESENTATION OF THE INTENSITIES FOR EACH LIST OF SENTENCES

List A



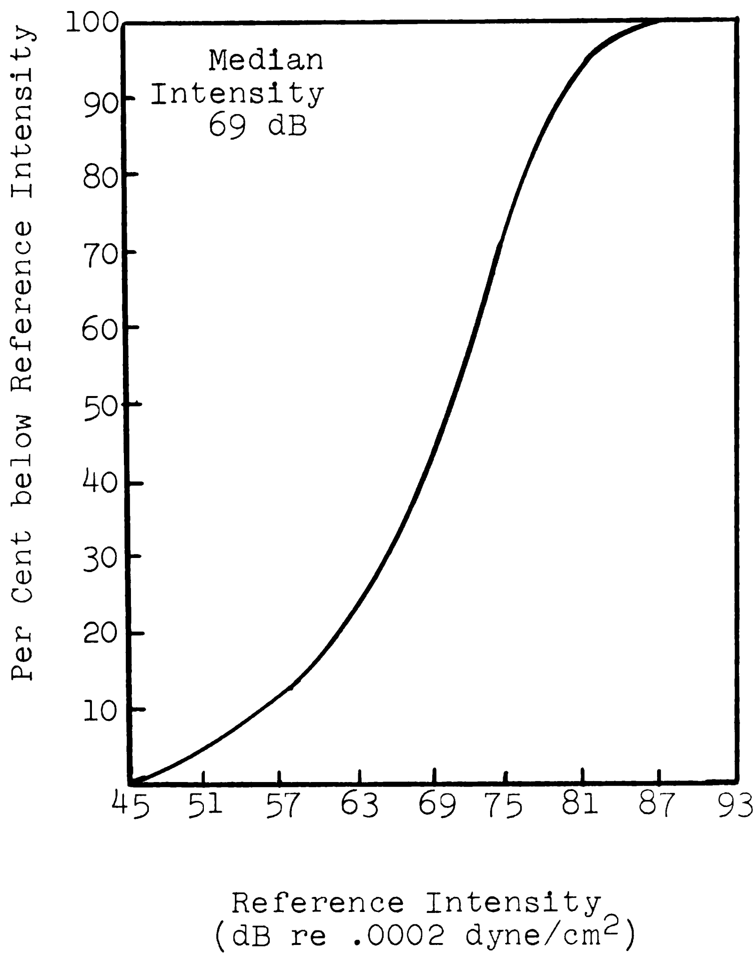
APPENDIX D--Continued

List B



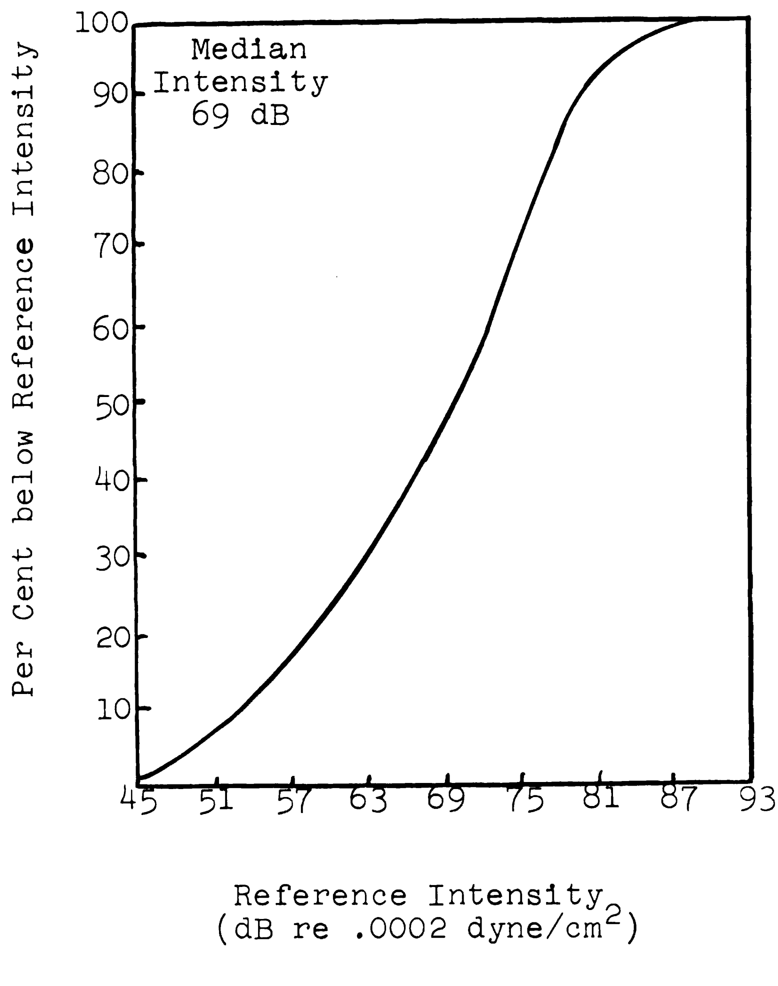
APPENDIX D--Continued

List C



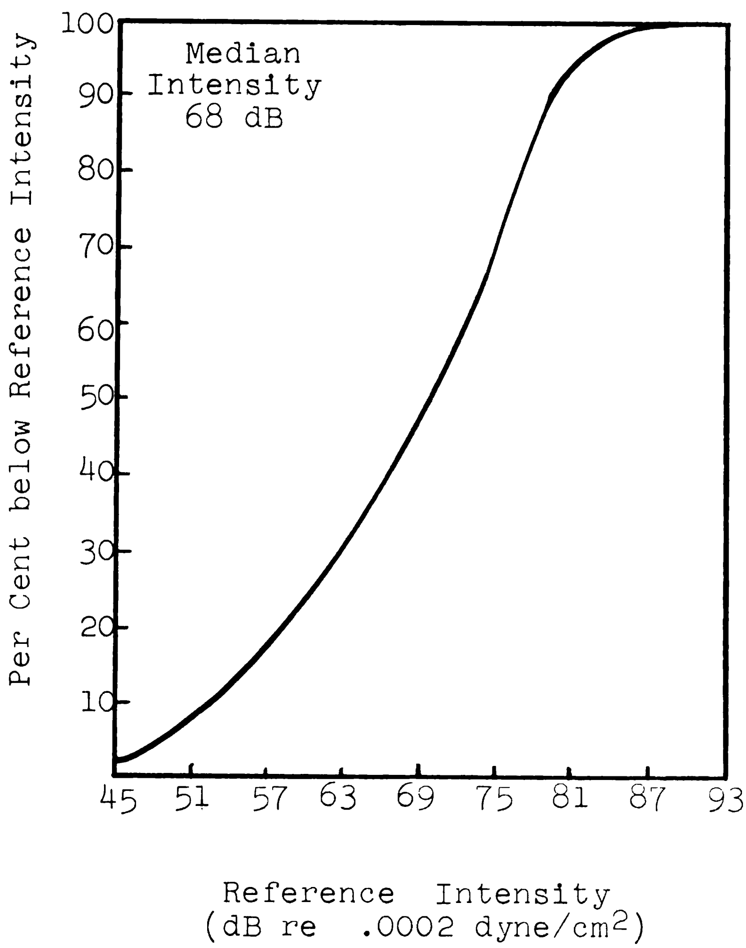
APPENDIX D--Continued

List D



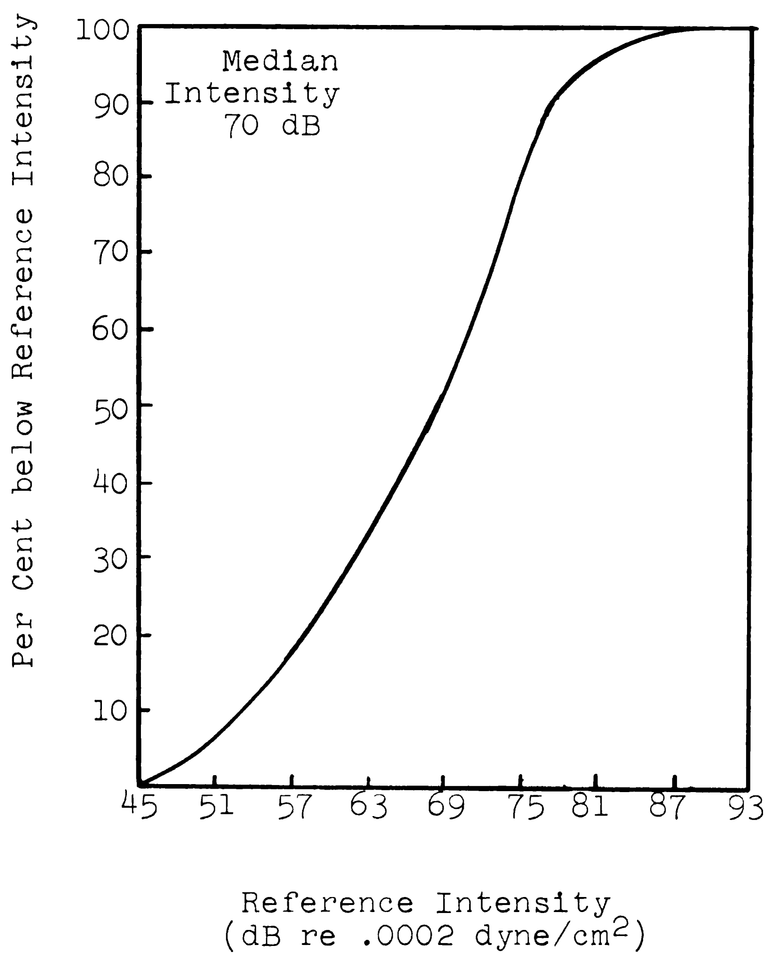
APPENDIX D--Continued

List E



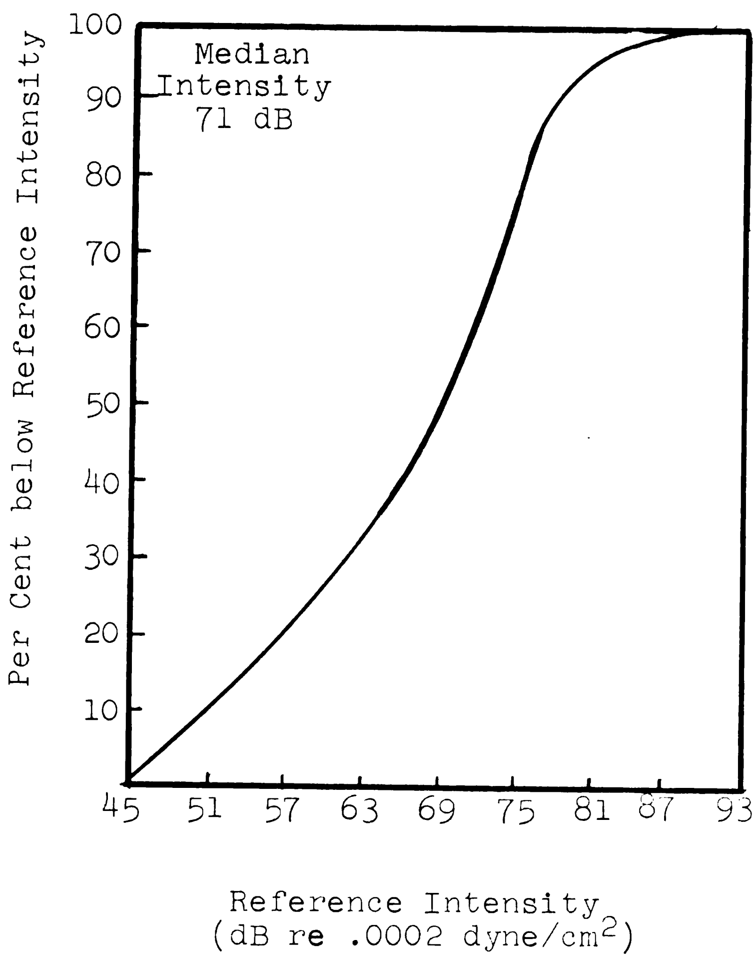
APPENDIX D--Continued

List F



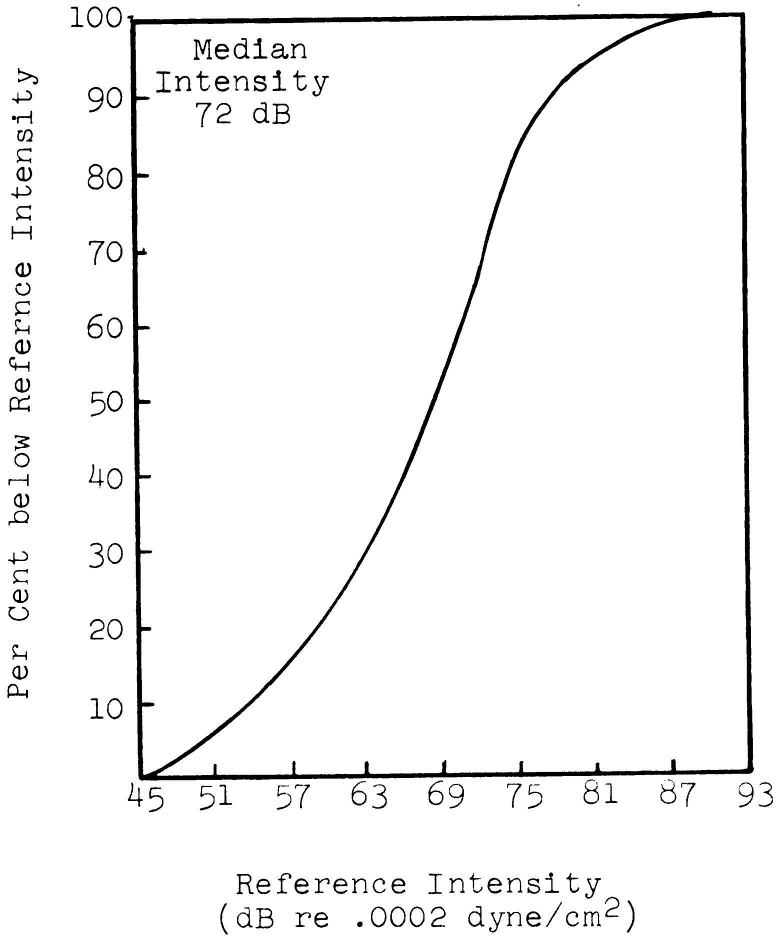
APPENDIX D--Continued

List G



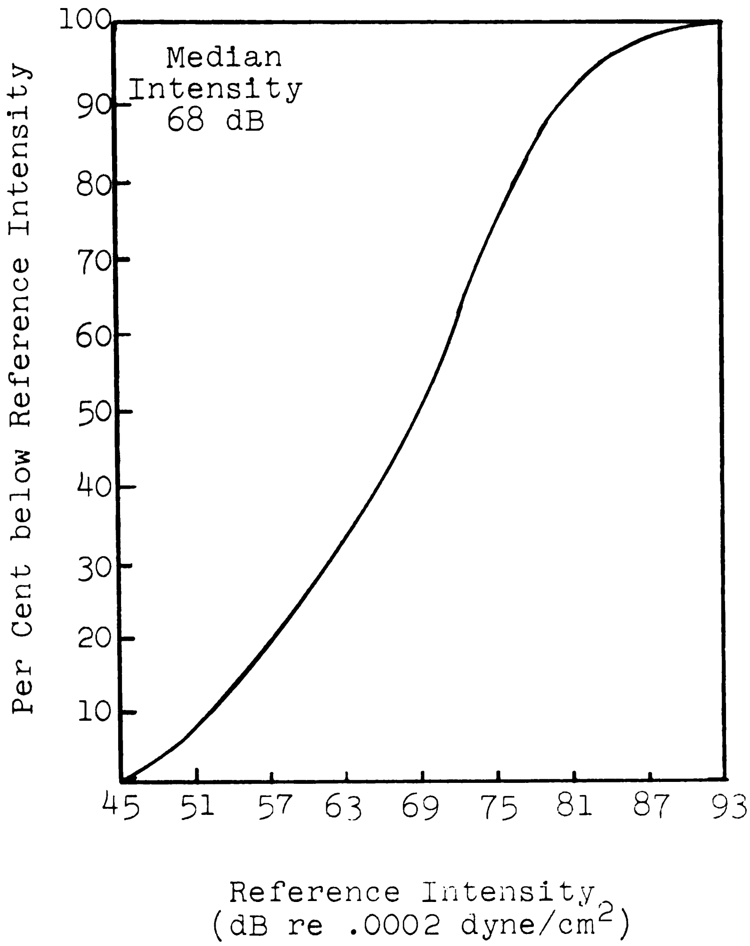
APPENDIX D--Continued

List H



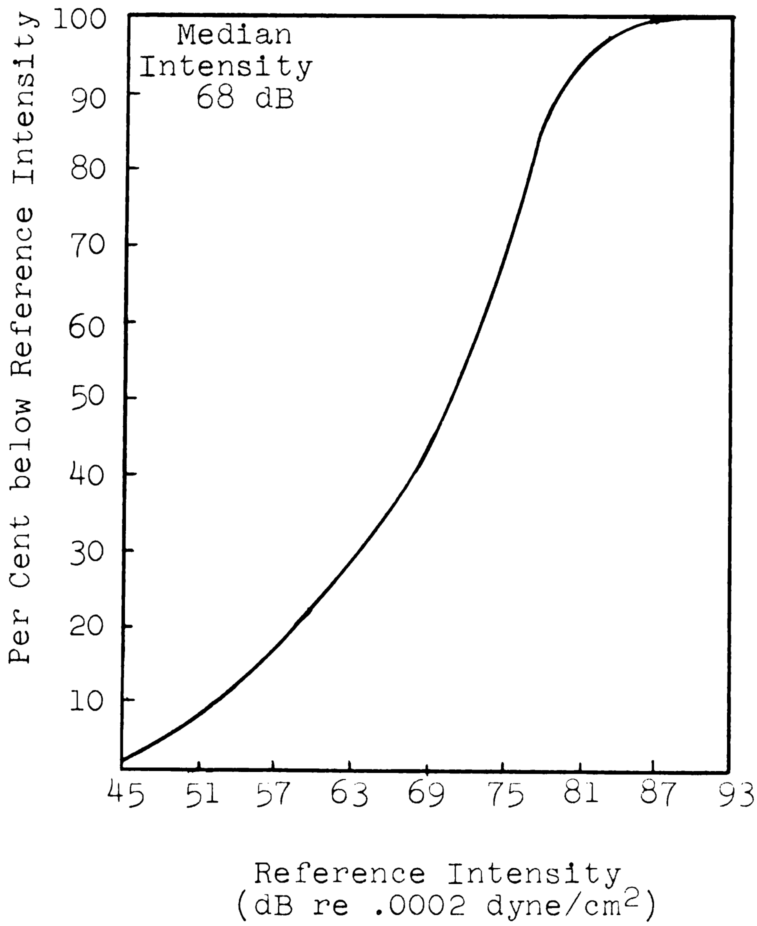
APPENDIX D--Continued

List I



APPENDIX D--Continued

List J



APPENDIX E

SAMPLE TEST FORM (LIST A)

C. I. D. EVERYDAY SPEECH

Name____________________

Date____________________

Playback intensity__________

Listening Condition:

Right ear________________

Left ear_________________

Binaural phones__________

Sound field______________

	Number right
1. Walking's my favorite exercise.	________
2. Here's a nice quiet place to rest.	________
3. Our janitor sweeps the floors every night.	________
4. It would be much easier if everyone would help.	________
5. Good morning.	________
6. Open your window before you go to bed.	________
7. Do you think that she should stay out so late?	________
8. How do you feel about changing the time when we begin work?	________
9. Here we go.	________
10. Move out of the way!	________

Examiner:

____

Total right ________

SCORE (2 x total right) ________

APPENDIX F

TEST SCORES*

Subject	PRESENTATIONS									
	1	2	3	4	5	6	7	8	9	10
1.	96	34	92	88	0	12	100	100	0	0
2.	56	0	94	0	100	100	98	0	0	30
3.	10	100	30	86	76	0	16	100	0	94
4.	0	0	0	100	0	26	64	82	94	98
5.	96	66	100	8	96	0	54	46	100	0
6.	14	100	80	100	98	90	0	0	4	0
7.	6	0	26	70	66	94	100	100	100	100
8.	100	94	16	52	0	0	90	0	90	74
9.	100	80	100	0	6	98	0	86	66	28
10.	0	10	4	100	100	100	38	76	94	90
11.	98	82	100	100	0	54	100	100	16	12
12.	70	18	100	0	100	100	36	0	10	56
13.	8	100	76	100	88	0	16	100	0	94
14.	4	0	0	100	14	44	88	94	100	100
15.	94	82	100	0	84	0	72	36	100	0
16.	48	98	72	100	96	74	0	0	28	0
17.	0	0	4	34	64	88	86	100	100	100
18.	100	94	2	50	4	12	100	0	92	66
19.	98	98	100	2	4	100	0	100	72	34
20.	0	0	4	100	100	100	42	80	88	84
21.	88	24	94	100	0	42	100	100	0	8
22.	70	54	100	0	100	100	100	0	6	52
23.	8	100	72	100	88	0	28	100	0	100
24.	8	6	0	100	52	84	100	100	100	100
25.	94	46	100	0	100	0	54	50	100	12
26.	38	88	100	100	94	88	0	0	26	0
27.	0	0	0	8	48	70	98	100	96	100
28.	100	84	4	48	0	2	96	0	94	58
29.	98	62	100	0	0	100	0	56	50	4
30.	0	0	0	84	100	98	12	56	86	96

*For information regarding list and intensity see Appendix C.