SOME QUESTIONS ABOUT SPEECH AUDIOMETRY

Seunghwan Lee

A number of test batteries for speech audiometry were reviewed from the viewpoints of reliability and validity. Included in this review were the following: (1) CID Auditory Test No. 1 (36 spondees for SRT) (Hirsch et al. 1952), (2) CID Auditory Test No. 22 (PB-50 monosyllabic words) (Hirsch et al. 1952), (3) NU Auditory Tests (Tillman et al. 1963, Tillman and Carhart 1966), (4) Multiple Choice Discrimination Test (MCDT) (Schultz and Schubert 1969), (5) The Rhyme Tests (Fairbanks 1958, House et al. 1963, 1965, Griffiths 1967), (6) The K.S.U. Speech Discrimination Test (Berger 1969), (7) The University of Oklahoma Closed-Response Speech Test (OUCRT) (Pederson and Studebaker 1972), (8) The California Consonant Test (CCT) (Owens and Schubert 1977), (9) The Synthetic Sentence Identification (SSI) (Speaks and Jerger 1965, Jerger, Speaks, and Trammel 1968), and (10) The Speech Perception in Noise Test (SPIN Test) (Kalikow, Stevens, and Elliot 1977). It was concluded that the audiological test materials for speech are reliable in the sense that the test results we obtain from them are numerically consistent. The consistencies, however, become meaningless when an attempt is made to interpret them in terms of speech segments (i.e., phonemes and allophones) and rule-governed linguistic structures of English. They are disappointingly devoid of validity.

1. Introduction

Pure-tone audiometry, tympanometry, and speech audiometry are usually included in the routine audiological evaluation in hearing clinics. The audiogram from pure-tone audiometry gives us information on the patient's hearing loss as a function of pure-tone frequencies or frequency ranges; and the tympanogram and the data on the acoustic reflexes, the plight of the middle ear function of the patient. However, when it comes to speech audiometry, some questions arise as to the nature of information we get about the patient's hearing capacity of speech.

In the routine audiological evaluation, two speech audiometric batteries are administered: (1) the 36 spondees for the measurement of speech reception threshold (SRT) and (2) the 50 phonetically or phonemically balanced monosyllabic words for the measurement of speech discrimination (i.e., suprathreshold measurement). The initial question which seems to be worth asking about speech audiometry is: What aspect or aspects of speech or language do these two speech materials test? The argument in this paper is that the speech materials used in the tests are not valid representations of speech in that they do not include some of the basic structural linguistic elements and, furthermore, that the rule-governed features of a particular language have not been taken into consideration. Thus, despite the fact that the tests deal with some segments of speech, the scores or the results from the administration of speech audiometric tests seldom give us systematic information about the hearing impairment in terms of sound segments or other structural parts thereof.

The same questions can also be asked and the same conclusion can be drawn about other speech audiometric materials which are usually considered as 'special tests' in the sense that each of them taps a different aspect of speech. They do not show any specific aspect or aspects of speech which they are purported to test. One of the major reasons for the drawbacks seems to be due to the almost complete neglect of the 'systematic' nature of speech segments or structures of language.

As for the rule-governed aspects of speech or language, I have in mind some of the phonetic/phonological systematicities of language in general and in a particular language, specifically English. Some of the sketchy notions of the systematicities are given in the following paragraphs.

'Speech' is the realization of a specific language. A language is a rule-governed system of knowledge, which we assume all normal speakers of a particular language have internalized. Speech of a particular language, therefore, is the overt representation of the linguistic knowledge one possesses. The medium of representation in the case of speech among other communication mediums possible for human beings consists of speech sounds. Each language employs a finite number of functional speech sounds out of a theoretically infinite number of speech sounds which are possible within the limits of human anatomy and physiology. In addition to the constraint on the number of speech sounds which are functional in a language, the speech sounds in a specific language comply with the rules of distribution (i.e., phonotactics) and of phonetic and phonological interactions which are unique to that language, besides complying with the universal phonological constraints.

English makes use of nine simple vowels, a dozen complex vowels (or diphthongs), 24 consonants, four levels of stress, and four levels of pitch (Bloch and Trager 1942, Trager and Smith 1951, Chomsky and Halle 1968, Liberman and Prince 1977, Schane 1979). Other languages employ different inventories of speech sounds. Even in two languages which contain a number of the same sounds as part of each language's phonological system, the distributions and the function of each sound are rarely the same. Whereas English utilizes only three nasal consonants, Kannada spoken in south India, for example, has four nasal consonant phonemes. English distinguishes /l/ and /r/ as two phonemes, but Japanese and Korean have only one phoneme /l/, which are actualized either as [l] or [r] depending on the positions of occurrence. The velar nasal, /ŋ/, and the voiced alveopalatal fricative, /ž/, in English do not occur in the initial position of a word, which can be compared, in contrast, to other languages in which different distributions of the sounds are found, for instance, the occurrences of /ž/ in the initial position in French and the occurrences of /ŋ/ in the initial position in Vietnamese. The alveolar stop series, for instance, in

Kannada and Hindi shows not only the different inventory of sounds from English, but completely different interrelationships are manifested (Benguerel and Bhatia 1980).

The picture of consonant clusters displays rather more stringent constraints in that the number of consonants and their order are far more different among languages. English allows a maximum number of three consonants in the initial clusters, but four in the final clusters. Korean, on the other hand, has consonant clusters neither in the initial position nor in the final position, but only in the medial position and the number of consonants in the medial cluster is limited to two. In English, moreover, if #CCC- (C stands for a consonant; #, the beginning of a word; '-', the position where the first vowel in the word occurs. Given -CCC#, '-' designates the last vowel in the word; and #, the end of the word) occurs the first C is limited to /s/; the second C, to /p/, /t/, or /k/; and the third C to, /l/ or /r/. From the standpoint of a listener, in consequence, the degree of expectation or the conditional probability is conspicuously different in perceiving or understanding /p/, for example, in /#spr-/ from /#pr-/ or /#p-/, and the expectation or the probability of /p/ in /#pr-/ is different from that in /#p-/.

The phonological system of a language, moreover, is not simply a collection of the speech sounds. Individual sounds show complex interrelations or interactions with other sounds in the language that the function of each sound should be understood in terms of these interrelationships. The [p'] (tense and unaspirated) in the word 'spin' is an independent phoneme if interpreted in the framework of either the Thai or the Korean phonological system, but it is a variant (or allophone) of the phoneme /p/ in English in that /p/ in English is realized as [p^h] (strongly aspirated) in the word initial position and as [p'] if /p/ immediately follows /s/ in /#s-/. The fact that [p'] in English is perceived as a /p/ is due to its interaction with the specific position of occurrence and with the preceding /s/, which are not true in languages such as Thai or Korean. For another example of interaction, the intervocalic /t/, if the preceding vowel receives a stress, becomes a voiced flap, /ɾ/, which shows the interaction of /t/ with the preceding and following vowels and with the prosodic feature of stress.

The internalized linguistic knowledge or rules of normal native speakers of a language, of course, does not stop with the phonological rules. Each of the linguistic levels, i.e., phonology, morphology, syntax, semantics, is a system of rules and, furthermore, rules of different levels interact with each other. Besides, factors external to the linguistic knowledge proper exert significant influence on the processing of language message, both production and comprehension. Current developments in pragmatics in the schemes of both philosophy (Searle 1969) and linguistics (Cole and Morgan 1975), the explorations in cognitive psychology (Rumelhart and Ortony 1977, van Dijk 1977, 1980, Kintsch 1974, Kintsch and van Dijk 1978, de Beaugrande 1980) and artificial intelligence

(Minsky 1975, Schank and Abelson 1977, Winograd 1980) clearly show that the whole system of man's general knowledge and cognitive faculty is at work in the processing of language, with the consequence of the emerging realization of the necessity of effort directed to forming a holistic discipline under the name of 'cognitive science' (Norman 1980, D'Andrade 1981). The findings in these areas all seem to point to the rule-governed nature of language structures, not only of the linguistic factors alone, but also of extra-linguistic knowledge in general.

Returning to speech audiometry, questions arise as to how the current speech audiometric test batteries are constructed and what information about speech they tap in relation to human hearing and its impairments.

The purpose of this paper is to review some of the current major speech audiometric test batteries in light of the rule-governed structural linguistic facts. Included in this paper for review are the following: (1) CID Auditory Test No. 1 (the 36 spondees for SRT) (Hirsch et al. 1952), (2) CID Auditory Test No. 22 (PB-50 monosyllabic words) (Hirsch et al. 1952), (3) NU Auditory Tests (Tillman et al. 1963, Tillman and Carhart 1966), (4) Multiple Choice Discrimination Test (MCDT) (Schultz and Schubert 1969), (5) The Rhyme Tests (Fairbanks 1958, House et al. 1963, 1965, Griffiths 1967), (6) The K.S.U. Speech Discrimination Test (Berger 1969), (7) The University of Oklahoma Closed-Response Speech Test (OUCRT) (Pederson and Studebaker 1972), (8) The California Consonant Test (CCT) (Owens and Schubert 1977), (9) The Synthetic Sentence Identification (SSI) (Speaks and Jerger 1965, Jerger, Speaks, and Trammell 1968), and (10) The Speech Perception in Noise Test (SPIN Test) (Kalikow, Stevens, and Elliot 1977).

The so-called Trager-Smith system of phonemic symbols is used for transcription of utterances in this paper. The segmental vowel phonemes are represented with 9 simple vowels (/i, e, æ, ɨ, ə, a, u, ɔ, o/) and three glides (/-y, -w, -h/). The segmental consonant phonemes are 24 in number (/p, b, m, f, v, θ, ð, t, d, n, s, z, l, r, č, ǰ, š, ž, k, g, ŋ, h-, y-, w-/).

The original sources of the Trager-Smith system of phonemic transcription are Bloch and Trager (1942), and Trager and Smith (1951). A well-organized brief summary of the system is found in Gleason's (1961) Chapters 2 to 4.

2. Questions about Speech Audiometry

2.1. CID Auditory Tests W-1 and W-2 (Thirty-six spondaic words)

Speech audiometry became a concern in audiology with the invention of Western Electric 4-A Audiometer (Fletcher 1929), and the first use of speech materials was carried out by Hughson and Thompson in 1942 (Hughson and Thompson 1942) who utilized Bell Telephone Intelligibility Sentences. It was Hughson and Thompson who coined the term *speech reception threshold (SRT)*. Hughson and Thompson were followed by Hudgins et al. (1947) at the Psycho-

Acoustic Laboratory (PAL) of Harvard University, who devised both sentence materials (PAL Auditory Test No. 12) and spondaic words (PAL Auditory Test Nos. 9 and 14). Thus, began the original use of spondaic words for the measurement of speech reception threshold. The Auditory Test No. 12 was composed of eight lists of short questions, each list with 28 questions each of which can be answered with a single word. However, the sentence materials of Bell Telephone Intelligibility and the PAL Auditory Test No. 12 did not receive subsequent attention. The spondaic words, on the other hand, were followed by further elaborations and revisions and finally became the current CID Auditory Test W-1 and W-2 for SRT measurement.

The criteria on which Hudgins et al. (1947: 58) based their construction of PAL spondaic words were (1) familiarity, (2) phonetic dissimilarity, (3) normal sampling of English speech sounds, and (4) homogeneity with respect to basic audibility.

The criterion of familiarity was included because the purpose of speech audiometry was to test 'the threshold of intelligibility for speech, rather than vocabulary or intelligence' (Hudgins et al. 1947: 58). Phonetic dissimilarity was considered important 'because the presence of similar or rhyming words in the list imposes a type of auditory discrimination which serves no useful purpose in threshold tests. For instance, the inclusion in a test list of words that differ only with respect to a single sound, such as *plowboy* and *cowboy*, or *eyeball* and *highball*, increases its difficulty by demanding a finer discrimination, but does not increase the effectiveness of the test as an instrument for measuring the threshold for common English speech' (Hudgins et al. 1947: 58). The criterion of normal sampling of English sounds, though included in their four criteria, was not thought to be important. The criterion of homogeneity with respect to basic audibility seemed to be the most important criterion for them.

> Homogeneity is important for two reasons: First, it increases the probability that the function relating the percentage of items heard correctly and the intensity levels at which they are presented to the listeners will rise steeply from zero to 100 per cent within a narrow range of intensity. . . . Second, it is desirable to determine the threshold of hearing for speech with the use of as small a number of items as possible. (Hudgins et al. 1947: 59-60)

The criterion of homogeneity and its further elaboration by Hudgins et al, in a very significant sense, characterized the nature of audiometric spondaic words, i.e., the steep PI function and the feasibility in the routine audiological evaluation, which was faithfully adopted by the later researchers. Hudgins et al's further elaborations are as follows:

> The audibility of words or sentences may be defined as the relative ease with which they are understood when they are spoken at a constant level by a normal talker. . . . In order to ensure homogeneity in a word list, two procedures are possible. 1) We can choose for the list only those words which, when spoken in a normal tone of voice, tend

to reach the listener's threshold at the same level of amplification. 2) We can so adjust the level at which the individual words are recorded that they tend all to be heard at the same level of reproduction. Or, as a matter of fact, we can combine these two procedures in order to maximize homogeneity. (Hudgins et al. 1947: 59)

Thus, Hudgins et al. already suggested in 1947 the method of achieving the homogeneity of audibility for the spondaic words. But we notice that they were concerned with the homogeneity between the spondaic words and not with the homogeneity between the two syllables composing a spondaic word.

They constructed two lists of spondaic words, each consisting of 42 spondees, a total of 84 spondees. Each list was scrambled in six different versions and recorded. The PI functions of the spondees along with those of random dissyllabic words and monosyllabic words are given in Figure 1. The significance of the figure was found in the PI function of the selected spondees which showed 'an average slope of 10 per cent per decibel over the range between 20 and 80 per cent. The curve for the monosyllables, on the other hand, has a slope of 4 per cent per decibel over the same range. The curve for the unselected dissyllables falls in between' (Hudgins et al. 1947: 66). Thus, they were able to get the steep PI function for the spondees with which the speech reception threshold could be measured with *precision*.

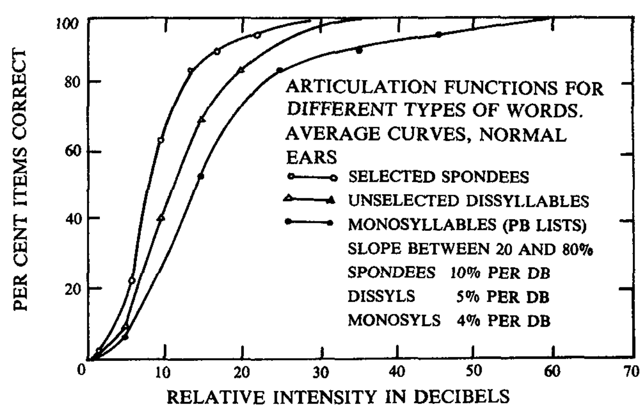

Figure 1. Showing how the steepness of articulation functions depends on the kind of words used in the test. The number of words heard correctly is plotted against the relative level at which they are transduced by the earphones. The threshold of intelligibility, defined as the intensity level at which 50 per cent of the words are heard correctly, is more precisely determined when the articulation function is steep (Hudgins et al. 1947: 65).

Hudgins et al.'s PAL Auditory Test No. 9 and PAL Auditory Test No. 14 are of the same list of 84 spondees, the difference lying only in the different manipulation of recording. That is, the 42 spondees in Test No. 9 consisted of seven groups of six words each and each group was recorded with a progressive 4 dB decrement, resulting in the range of decrement for the whole 42 spondees of 24 dB. The spondees in Test No. 14 were recorded with a constant level of intensity throughout the list.

In 1952, Hirsch, Davis, Silverman, Reynolds, Eldert, and Benson (Hirsch et al. 1952) revised PAL Test Nos. 9 and 14 on the basis of the difficulty of the spondees. Out of the 84 original PAL spondees, Hirsch et al. arrived at 36 spondees, after deleting both extremes of easy and difficult words. For this selection they used six judges.

> In the analysis of this preliminary data an easy word was defined as one missed once or less by all six listeners. A difficult word was one missed five or more times by all six listeners. Words falling in both of these extreme categories were eliminated, and also the words that five of the six listeners found difficult or easy. In the 36 words left, a group of equally intelligible spondees was approximated. (Hirsch et al. 1952: 217)

Hirsch et al. also recorded the 36 spondees in two different ways and named them differentially, CID Auditory Test W-1 and CID Auditory Test W-2, respectively. The spondees in W-1 list were recorded with a constant intensity level, and those of W-2 list were recorded with a 3 dB attenuation for each 3 words, resulting in 33 dB attenuation for the whole list. The CID Auditory Test W-1, which is with the constant recording level, is the one most widely accepted currently by clinicians, the 36 spondees of which are listed in Table 1.

Table 1. Thirty-six Spondees in CID Auditory Test W-1 (Alphabetical order)

1. airplane	10. eardrum	19. iceberg	28. railroad
2. armchair	11. farewell	20. inkwell	29. schoolboy
3. baseball	12. grandson	21. mousetrap	30. sidewalk
4. birthday	13. greyhound	22. mushroom	31. stairway
5. cowboy	14. hardware	23. northwest	32. sunset
6. daybreak	15. headlight	24. oatmeal	33. toothbrush
7. doormat	16. horseshoe	25. padlock	34. whitewash
8. drawbridge	17. hotdog	26. pancake	35. woodwork
9. duckpond	18. hothouse	27. playground	36. workshop

Consonants and vowels which appear in the 36 spondee words are tabulated in Table 2.

Table 2. Consonants and Vowels in the 36 Spondees of CID Auditory Test W-1 ('X' shows the missing phonemes. The complex vowels ending in /-h/ are not checked with 'X' for missing. The counts of consonants are inclusive of both the first and second syllables.)

Vowels:

Simple	i	e	æ	ə	a	u	ɔ	total
First syllable	1	1	3	4	2	1	X	12
Second syllable	1	4	2	5	2	X	3	17

Complex	ih	eh	əh	ah	ɔh	iy	uw	ow	ey	ay	ɔy	aw	Total
First syllable	1	3	1	2	4	X	2	1	5	3	X	2	24
Second syllable		2			2	1	2	1	5	1	2	3	19

Consonants:

p	b	m	w	t	d	n	s	z	l	r	y	f	v	θ	ð	č	ǰ	š	ž	k	g	ŋ	h
7	9	7	11	11	17	10	12	X	10	26	X	1	X	3	X	1	1	5	X	11	5	X	7

A number of questions seem to be in order. First of all, two of the Hudgins et al's criteria for the selection of the spondaic words are in contradiction: 'phonetic dissimilarity' and 'homogeneous audibility.' Basically, if two words are phonetically dissimilar they are not homogeneous in audibility. This has been clearly shown in Olsen and Matkin's (Olsen and Matkin 1979: 140-141) tabulation, in which they listed the 36 spondees in order of the weight of audibility in terms of the "times selected" by six studies. In Table 3 are listed the 36 spondee words in the order as given in Olsen and Matkin, along with the phonemic transcription indicating the occurrences of complex vowels.

Furthermore, audibility is not a phenomenon simply of segmental constituents of a given word. However, Table 3 shows that the complex vowel and its position of occurrence exert a strong influence on the audibility. The first syllable seems to contribute more effectively to audibility. Among the 15 spondees with the range of 4 to 6 'times selected,' 93% (14 out of 15 spondees) contain complex vowels in the first syllable and 67% (10 out of 15 spondees) contain complex vowels in the second syllable. On the other hand, among the 21 spondees with the range of 0 to 3 'times selected,' 48% (10 out of 21 spondees) contain complex vowels in the first syllable and 52% (11 out of 21 spondees) contain complex vowels in the second syllable.

The second question about the W-1 spondees relates to the concept of the spondee itself, which eventually formed the basis of the presentation methods of spondees in the clinic. The following quotations from a number of studies are worth reviewing in this respect.

Table 3. Olsen and Matkin's (1979: 140-141) Ratings of Audibility with Additional Information ('x' shows the complex vowel). The compound noun (i.e., the spondee in English) has dissyllabic stress configuration of / ´ + ˆ / or / ´ ` /.

Spondee	Transcription	Times Selected	Complex vowel 1st syll	2nd syll
birthday	/bəhrθ + dey/	6	x	x
iceberg	/ays + bəhrg/	6	x	x
northwest	/nɔhrθ + west/	6	x	x
railroad	/reyl + rowd/	6	x	x
playground	/pley + grawnd/	5	x	x
stairway	/stehr + wey/	5	x	x
airplane	/ehr + pleyn/	4	x	x
armchair	/ahrm + čehr/	4	x	x
eardrum	/ihr + drəm/	4	x	
farewell	/fehr + wel/	4	x	
hardware	/hahrd + wehr/	4	x	x
mousetrap	/maws + træp/	4	x	
sidewalk	/sayd + wɔhk/	4	x	x
sunset	/sən + set/	4		
whitewash	/wayt + wɔš/	4	x	
cowboy	/kaw + bɔy/	3	x	x
drawbridge	/drɔh + brij̆/	3	x	
greyhound	/grey + hawnd/	3	x	x
horseshoe	/hɔhrs + šuw/	3	x	x
mushroom	/məš + ruwm/	3		x
oatmeal	/owt + miyl/	3	x	x
padlock	/pæd + lak/	3		
toothbrush	/tuwθ + brəš/	3	x	
woodwork	/wud + wərk/	3		
doormat	/dɔhr + mæt/	2	x	
headlight	/hed + layt/	2		x
hotdog	/hat + dɔg/	2		
hothouse	/hat + haws/	2		x
inkwell	/iŋk + wel/	2		
schoolboy	/skuwl + bɔy/	2	x	x
workshop	/wərk + šap/	2		
baseball	/beys + bɔhl/	1	x	x
daybreak	/dey + breyk/	1	x	x
duckpond	/dək + pɔnd/	1		
pancake	/pæn + keyk/	1		x
grandson	/grænd + sən/	0		

> This test consists of two lists of 42 dissyllabic words of the *spondee* stress pattern, *i.e.*, words such as *blackbird, railroad,* in which both syllables are equally accented. . . . The high audibility of the *spondee* as compared to other dissyllabic words, such as the *trochee* or *iamb*, is due apparently to the differences in stress pattern. Since both syllables of the spondee are equally stressed, auditory cues from each syllable are equally available to the ear. In contrast, the single accented syllables of trochees (*fa*ther, *wa*ter) and of iambs (up*on*, e*quip*) are more audible than the weaker, unstressed syllables, and the intelligibility of the words depends in part at least upon the cues from the weaker syllables. (Hudgins et al. 1947: 64-65)

> . . . lists of words of homogeneous audibility. . . . In order to assemble such lists, experiments were conducted in which various types of words were presented to trained listeners under carefully controlled conditions. It was discovered that the class of words having the highest homogeneity contained those dissyllables spoken with equal stress on both syllables. These words are called *spondees.* (Egan 1948: 965)

> Most SRT's are obtained today through the use of spondaic words, often called spondees. A spondee is a word with two syllables, both pronounced with equal stress and effort. While spondees do not occur in spoken English, it is possible, by altering stress slightly, to force such common words as *baseball, hotdog,* and *toothbrush* to conform to the spondee configuration. Whether the spondees are spoken into the microphone or introduced via tape or disc, both syllables of the word should peak at zero VU. While it takes practice for the student to be able to accomplish this equal peaking at zero on the VU meter, it is a knack that most people can acquire in a relatively short period of time. (Martin 1981: 124)

> Words of two syllables with *approximately* equal stress on both syllables. (Italics mine) (Berger 1971: 212, fn. 1)

> In the monitored live voice (MLV) method, the examiner sets the input selector to microphone. With the hearing level dial at zero, he practices several spondees and adjust the calibration knob so that *the first syllable* of the spondee causes the VU meter needle to approximate zero. (Italics mine) (Berger 1971: 222)

> Spondee: a metrical foot consisting of two long syllables or, in English poetry, two heavily accented syllables: most alleged spondees in English really have one secondary accent. (*Webster's New Twentieth Century Dictionary of the English Language,* Unabridged, Second Edition, 1977)

The concept of the spondee came from the poem recitation, in which we distinguish the accentual patterns in terms of trochaic, iambic, and spondaic. One fact about poems is that, in their recitation, we are concerned only with two levels of accent or stress, that is, the syllable which is stressed versus the syllable which is unstressed or weak. Interpreted in the Trager-Smith System, the three stresses, i.e., primary, secondary, and tertiary, are considered as stressed in poems, and the weak in the Trager-Smith System is considered as unstressed. We can, therefore, reasonably assume that when Hudgins et al. (1947) published their article, the linguistic descriptions of English by Bloch and Trager (1942) and by Trager and Smith (1951) were not known to them. Hudgins et al.'s concept of the spondee, therefore, can be presumed not to mean the exact physical equality of stress on both syllables. Moreover, they did not mention that there are no *alleged* spondees in English, which means that they took the secondary stress on the second syllable as being 'equal' to the primary stress on the first syllable. Furthermore, when they tried to 'ensure homogeneity in a word list' (Hudgins et al. 1947: 59), they tried to 'adjust the level at which *the individual words* are recorded that they tend all to be heard at the same level of reproduction' (Italics mine) (Hudgins et al. 1947: 59). If they had adjusted the syllable stress, they might have pointed out the fact instead of saying 'the individual words.' Finally, it is worth noting that Hudgins et al. made it clear that '. . . this paper is concerned with descriptions of the two auditory tests based on *normal, undistorted speech*' (Italics mine) (Hudgins et al. 1947: 64), that suggests that they did not artificially manipulate the stress pattern of the spondee to stress both syllables to make them physically equal.

The quotation from Egan (1948: 965) seems to show that he was also interpreting the stress pattern of the spondee in the framework of the poetic metrical system. Here again, even though he explained that his experiments were done 'under carefully controlled conditions,' he did not specifically mention that he controlled syllabic stresses.

We could then assume that the method currently practiced by clinicians of presenting the spondees to the patients may be due to 'misinterpretation' of the studies in the 1940's. The current method is typically displayed in Martin's (1981) standard textbook of audiology in which he uses the expressions such as 'with equal stress and *effort*' (Italics mine) and 'by altering stress slightly, to *force* such common words' (Italics mine) which eventually forced clinicians to create the distorted and artificially non-existent spondaic words of English. Furthermore, Martin definitely asks clinicians to peak both syllables at zero on the VU meter, which, in actuality, is impossible, unless an unduly longer pause is given between the two syllables, in which case the word ceases to be spondaic and the possibility is that a spondaic word (NP = compound noun) becomes a NP of the adjective + noun structure. The spondaic word or the compound noun in English has the dissyllabic stress configuration of either the primary stress + secondary stress or the primary stress + tertiary stress. For example, 'hotdog' can be pronounced either as /hát + dɔ̂g/ or as /hátdɔ̀g/. If 'hotdog' is pronounced /hàt + dɔ́g/, its meaning is not that of the spondee, but of the construction of 'an adjective and a noun,' that is, the meaning of which becomes 'the dog which is hot.'

It is interesting to note that in Berger's quotation, given above, we find that he says 'approximately equal stress on both syllables' and that he does not force clinicians to peak both syllables at zero on the VU meter, but requires that we practice so that only 'the first syllable of the spondee . . . to approximate zero.' That is, we are allowed to use the natural English spondaic words in the sense defined in the Webster's Dictionary quoted above.

Hirsch et al.'s (1947) third criterion for the spondees was the 'normal sampling of English speech sounds.' Though they considered it not as important as the remaining three criteria, a question can be asked in regard to the nature of the 'normal sampling,' which was not clearly explained by Hirsch et al. On the other hand, they claimed that 'there is ample evidence that a complete representation of English sounds is not essential in threshhold measurement' (Hirsch et al. 1947: 59), which further compounded the problem of representativeness of the spondaic words in speech audiometry. I will return to this question of representativeness in the section on PB-50. Here, I would simply point out that the 36 spondees are not representative of English speech sounds in that not all the English phonemes are represented (See Table 2).

The vowel /u/ does not show up in the second syllable in any of the 36 spondees; nor /ɔ/, in the first syllable. We find six consonants which are not represented in the whole 36 spondees, neither in the first syllable nor in the second. They are /z/, /y-/, /v/, /ð/, /ž/, and /ŋ/. According to Tobias (1959), these phonemes, except for /ž/, show 'significant' relative frequencies of occurrence in everyday English speech. However, regardless of the ratings of frequency of occurrence, to be representative of the speech of a language, the test should include all the structural phonemic segments of that language because we will, later, have to confront a diagnostic situation in which we have to find out what aspect or aspects of speech really have contributed to the hearing loss (i.e., phonemic differentiation).

Another question to be asked about the spondees is the correlation between PTA and SRT. The correlative relationships between PTA and SRT can be traced back to 1929 when Fletcher showed high correlations between PTA and SRT in the range of 512 to 2048 Hz. This finding was confirmed by Hughson and Thompson (1942) and, furthermore, they concluded that 'Comparison of audiometric and speech reception levels shows that frequencies above and below 512 to 2048 have little significance in the subject's ability to understand speech' (Hughson and Thompson 1942: 540). Steinberg and Gardner (1940) showed the agreements between an average of SRT's at 512, 1024, 2048, and 4096 Hz and the levels of hearing impairments. Thus, the concept of speech frequency range emerged, along with the concept of PTA of the three frequencies, i.e., 512, 1024, and 2048 Hz, which eventually developed into the current practice of the measurements of 500, 1000, and 2000 Hz for PTA.

Fletcher in 1950, however, suggested a method of two-frequency average, i.e., the best two frequencies among 500, 1000, and 2000, as a better predictor of hearing loss for those subjects in which one of the three frequencies markedly deviates from the remaining two frequencies. This finding was followed by similar conclusions by Carhart (1971) and Carhart and Porter (1971), in which they recommended that we use the average of two frequencies, 500 and 1000 Hz, for the subjects whose audiograms show either a sharply falling or a sharply rising configuration. Also Siegenthaler and Strand (1964) advocated the superiority of the two-frequency method. In consequence, duplicate information can be found in both PTA and SRT, leading to the criticism of the redundant nature of the SRT (Davis and Silverman 1970). Only justification for the SRT, then, is found in its diagnostic value in the cases of abnormal pure-tone configurations, either sharply falling or sharply rising. Tillman and Olsen (1973: 47), therefore, concluded that '. . . while the justification for measuring the speech-reception threshold is sometimes debated, the reluctance of most clinicians to dispense with this measure attests to its value,' which was simply a begging the question. That is, Tillman and Olsen's argument does not solve the simple logic of redundancy problem. The elevated threshold of SRT does not tell us more than what is contained in the abnormal configuration of the PTA. This is mainly due to the lack of diagnostic power of the SRT via spondees; it does not give us differential diagnostic information as a function of the differential structural impairments of speech, e.g., impairments of specific segmental and/or suprasegmental phonemes and their rule-governed system in a language.

A conclusion, therefore, can be drawn about the CID W-1 spondaic words that the general acceptance of spondees in clinics lies not in its theoretical justification, but in the practical feasibility as was well described by Olsen and Matkin (1979: 136): 'Since tests using spondees were easier and more rapid to administer than were sentence tests, spondaic words gained acceptance for determining speech reception thresholds.' Moreover, there seems to be no justification for the equal stresses on the two syllables of the spondee, except for the steep PI functions of the spondaic words. The steep function is claimed to be required for the precise measurement of SRT; however, it is curious that we can hardly find any study which dealt with the spondaic words with the natural stress configuration, that is, the primary stress on the first syllable and the secondary or tertiary stress on the second syllable. I will return to the significance of the comparative degrees of steepness of the PI functions of different speech audiometric materials in the next section on CID W-22.

2.2. CID Auditory Test W-22 (PB-50)

The first speech audiometric material for the suprathreshold measurement was constructed by Egan (1948) at the Psycho-Acoustic Laboratory of Harvard University, which is known as PAL PB-50. Without defining the source, Egan (1948: 963) said 'From a large vocabulary 1200 monosyllabic words were chosen to make up what was called the RM (revised monosyllabic) lists.' From this sample he constructed 24 lists of 50 monosyllabic words each. The criteria on which he based his constructions of the 24 PAL PB-50 were (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 (Egan 1948: 963).

Egan conducted two difficulty-judgment experiments with six listeners in the first and with 23 listeners in the second experiment. Based on the results of these two experiments, he distributed the words in the 24 lists with 'equal average difficulty' and with 'equal range of difficulty.' By 'equal phonetic composition,' he meant the equal distribution of the different syllabic structures in each list. However, his classification of segments, both vowel and consonants, does not approximate that of the current linguistic analysis. Moreover, we have no basis on which to interpret his consonant categories of 'transitional' or 'semivowel.' He specified, in addition, that there are 12 words in one list which begin with a 'fricative,' but we find, for example, 16 initial fricatives in List 1 of the PAL PB-50.

The fifth criterion of the 'composition representative of English speech' was based on Dewey's (1923) study of the frequency counts of sounds in a sample of 100,000 words. Egan further explained that 'The words were assigned to each list on the basis of the phonetic composition of the first part of the word. No attempt was made to equate these lists with respect to the phonetic structure of the final consonant or consonant compound' (Egan 1948: 963). Each list contained around 10 initial 'compound consonants' (i.e., consonant cluster) and around 10 final compound consonants, but no control was imposed on the types of consonant clusters except for the approximate number of them in each list. A question, thus, arises as to the validity of the representativeness of the sounds for everyday English speech, because Egan based the representativeness only on the initial consonants. To compound the problem, Dewey's (1923) corpus for his study was from the written 'newsprints,' not that of spoken everyday conversations. Egan (1948: 963) finally suggested the reason for the 50 words in each list was that 25 words in a list are not sufficient to meet all the criteria he set up.

The first revision of the PB-50 lists was carried out by Hirsch et al. (1952). The criteria Hirsch et al. used were similar to those of Egan (1948), except for the criteria for difficulty. No mention was made about the difficulty of the words, but other criterial items became a little more detailed or specific. The criteria were '(1) all the words must be one-syllable words with no repetition of words in the different lists, (2) any word chosen should be a familiar word to minimize the effect of differences in the educational background of subjects, and (3) the phonetic composition of each word list should correspond to that of English as a whole as closely as possible' (Hirsch et al. 1952: 221). More emphasis was placed on the familiarity of the words. The basis for the phonetic composition of each word was expanded to include the Bell Telephone Laboratory's (1930) study of business telephone calls in New York City, in addition to the phonetic composition of the newsprint in Dewey's (1923) study. However, they deleted neutral vowels, which created room for the addition of other vowels in such a way that 'The percentage for the other vowels were increased, . . . by an appropriate amount to make up for the absence of the neutral vowel from the distribution' (Hirsch et al. 1952: 223).

The main pool of vocabulary items from which Hirsch et al. (1952) chose words for their PB lists were the 20 PAL PB-50 lists, a total of 1000 monosyllabic

Table 4. CID W-22 PB-50 in List 1 in Alphabetical Order

ace ache an as bathe bells carve chew could dad day deaf earn east felt give high him hunt isle it jam knees law low me mew none not or owl poor ran see she skin stove them there thing toe true twins up us wet what wire yard you

Table 5. Consonants and Vowels in List 1 of CID W-22 PB-50 ('X' shows a missing phoneme. The complex vowels ending in /-h/ and the consonant clusters are not checked with 'X' for missing.)

Vowels:

	i	e	æ	ə	a	u	ɔ	iy	ey	ay	uw	ow	aw	əh	ɔh	uh	eh
	6	5	5	8	1	X	1	5	4	3	4	3	1	1	2	1	1

Consonants:

	p	b	m	w	t	d	n	s	z	l	r	y	f	v	θ	ð	č	ǰ	š	ž	k	g
Initial	1	1	2	3	1	3	3	1	X	2	1	2	1	X	1	2	1	1	1		2	1
Final	1	X	3		4	3	4	2	2	2	4		1	2	X	1	X	X	X	X	1	X

	ŋ	h	tr	tw	st	sk	lz	rv	rn	st	lt	nt	nz
Initial		1	1	1	1								
Final	1						1	1	1	1	1	1	1

words. From this pool, they chose 120 words and added 80 more words from other unspecified sources, the result being 200 monosyllabic words. These 200 words, then, were checked against the frequency counts of occurrence in the Thorndike (1932) list and the Dewey (1923) list. Further check was done in light of French et al.'s (1930) study (cf., Tobias 1959) as regards the distribution of syllabic structures as well as their phonetic composition. Thus was constructed CID Auditory Test W-22, which is composed of 4 lists of PB-50, each of which is referred to as one of CID Auditory Test W-22, List 1 to List 4. These are the currently most widely used phonetically balanced monosyllabic words for the suprathreshold measurement in speech audiometry.

The 50 words in List 1 of the CID W-22 are reproduced in alphabetical order in Table 4, followed by Table 5 of the inventory of vowels and consonants represented in the list.

First of all, as shown in Table 5, the W-22 can be judged extremely unrepresentative of English in that, for example, in List 1 the vowel /u/ is missing. In the inventory of consonants, /z/ and /v/ are not shown in the initial position; /b/, /θ/, /č/, /ǰ/, /š/, /ž/, and /g/ are missing in the final position. This misrepresentation is due to the concept of 'phonetic balance,' in the construction of the PB lists. That is, 'phonetic balance' meant the proportional duplication of the frequencies of occurrence of the sounds on the basis of the Dewey's (1923) and Bell Telephone Laboratory's (1930) studies. As a consequence, 'phonetic balance' carries a meaning no more than a 'face validity.' The structural analysis of the English phonological system was not the concern in the construction of the PB lists. Therefore, again, the PB lists do not inform us of any differential hearing impairments in respect to the specific component or components of the English phonological structure.

The second question concerns the validity of the 'phonetic balance.' For this question, an answer was well presented by Tobias (1964: 99) when he said that 'the overwhelming clinical and experimental experience that indicates phonetic balance to be an interesting but unnecessary component of one of our current audiometric tests.' Relating to the half-list tests, Tobias (1964: 99-100) became more definitive in his conclusion about the phonetic balance: 'Of course, one must ask whether half-list tests measure the same thing as full-list tests. From the literature and from reports of audiologists using half lists, one must conclude that they do. On what grounds then can one insist upon phonetic balance as a criterion of discrimination test validity? There are none.'

The third question about the PB lists concerns the relationship between CID W-1 and CID W-22. Whereas the 36 spondaic words (CID W-1) are claimed to consist of phonetically dissimilar words, but with homogeneous audibility, a list of PB-50 words in CID W-22 are claimed to consist of phonetically similar words, which are heterogeneous in audibility. These two pairs of insoluble contradictions seem to underlie all assumptions in speech audiometry. The homogeneity of audibility in CID W-1 was shown to be ill-founded. See also Curry, Thayer, and Cox (1966) for the question about the homogeneity of audibility of the spondees. The heterogeneity of audibility in CID W-22 is also hard to comprehend in that monosyllabic words, regardless of the different syllabic structures, are more homogeneous in average audibility compared with the average audibility of dissyllabic words and, thus, the variability in audibility among the monosyllable words is less variable than that of the dissyllabic words. What is more difficult to understand is the rationale for the distinction between W-1 and W-22, which is claimed to make it appropriate to use W-1 for the speech threshold measurement and to use W-22 for the suprathreshold measurement.

The three PI functions shown in Figure 2 are of the same nature, the only differences among them being the degree in steepness. The three functions are of the similar curving configuration. The differences at the level of 50% intelligibility span about 10 dB between any two functions.

Figure 2 is from Goetzinger (1978: 150), its origin going back to Hirsch et al. (1952: 226), which compares the PI functions of W-2, W-22, and PB-50.

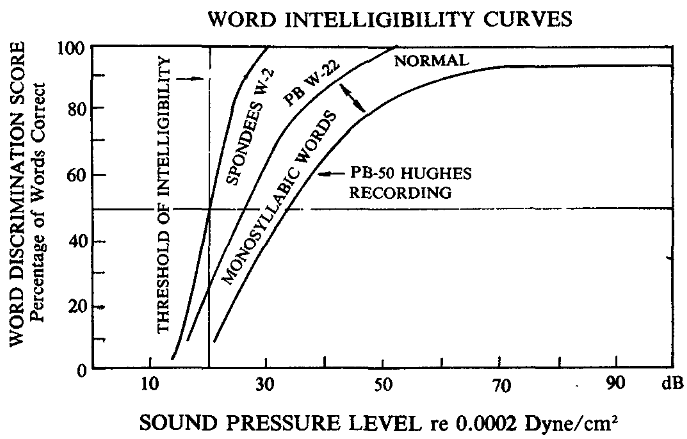

Figure 2. Relations between Auditory Tests W-2, W-22, and the Old PB-50

In connection with the PI functions of W-2 and W-22 and the difference between them, an analogy can be found between the full-list and the half-list of the W-22. That is, the clinical validity of the half-list is the same as that of the full-list (Campanelli 1962, Elpern 1961, Grubb 1963a, 1963b, Resnik 1962, Shutts, Burke, and Creston 1964, Tobias 1964). There is nothing in the speech material itself which makes it mandatory to use either PB list, a half list or the full list, for the measurement of the suprathreshold of speech. This is also true of the relationship between W-2 and W-22 in that it is in the nature of what we artificially manipulate to get, i.e., the SRT and the discrimination score, and not in the distinct linguistic structural characteristics of the two tests that distinguishes the two speech audiometric tests.

2.3. CNC Lists (Lehiste and Peterson 1959), Revised CNC Lists (Peterson and Lehiste 1962), NU Auditory Test Nos. 4 and 6 (Tillman et al. 1963, Tillman and Carhart 1966)

Lehiste and Peterson (1959) constructed 10 lists of CNC PB-50. The source of word selection was Thorndike and Lorge's (1944) 30,000 words, from which they chose 1263 monosyllabic words each occurring at least once in one million words. The phonetic balance was based on the phonetic structures of these 1263 words, which was quite different from the concept of phonetic balance used in PAL PB-50 or CID W-22. Three years later, they revised the original CNC lists, replacing 'unfamiliar words' by more familiar words (Peterson and Lehiste 1962), and this revision came to be known as 'Revised CNC Lists,' again, of 10 lists.

Tillman et al. (1963) and Tillman and Carhart (1966), again, revised the Revised CNC Lists by conforming more closely to the phonemic balance. Two lists were constructed and recorded by a male talker. This version is NU No. 4 (Northwestern University Auditory Test No. 4). Later, they made additional recordings of the two lists of the first version and another two new lists by a different male talker and a female talker, respectively. These later lists and recordings are named NU No. 6-M and NU No. 6-F, where 'M' stands for 'male,' and 'F,' 'female.'

As for the CNC lists and the NU lists, a simple question can be asked about the validity of the representativeness of the 1263 words. The fact that they are words with a frequency of more than one occurrence in one million words may be less reliable for their representativeness than the CID PB-50 words which were based on the studies of Dewey (1923) and the Bell Telephone Laboratory (1930). Moreover, the restriction to the syllabic structure of CNC should be pointed out as a serious deterioration in the CNC and NU lists from the CID PB-50's, in light of both phonetic balance and structural representation of English phonology. For one thing, the requirement of the CNC syllabic structure made them discard all the consonant clusters, which were included in CID PB-50 lists, though inadvertently, in both the initial and the final positions, except for the minor consonant clusters of the form of 'post-vocalic /-r/ + C' in the final position.

2.4. Multiple Choice Discrimination Test (MCDT) (Schultz and Schubert 1969)

Schultz and Schubert (1969) constructed a 'closed-response set' test using W-22 lists. The underlying motivation and the idea for this test is highly commendable in that they tried to obtain differential diagnostic data on the patients' hearing impairments in terms of individual segments or distinctive features, that is, 'to draw reasonable deductions about *which sounds* and *which sound substitutions* contribute to hearing difficulty for a given listener' (Schultz and Schubert 1969: 384).

They came up with four response foils for each of the 50 PB words (CID PB-50), each foil word differing from the test word in only one sound, that is, the consonant either in the intital or in the final position. In the answer sheet, therefore, the subject has five alternative words including the test word. Schultz and Schubert, however, did not explain the principle or theory on which they based their selection of the foil words. They simply explained that 'A panel of experimentalists, familiar with both the literature of speech discrimination testing and with the present study, evaluated all possible responses and discarded those considered inadequate for any reason. The remaining foil choices were assigned in such a manner that each of the 200 stimulus items was assigned four words differing from it by only one sound, . . .' (Schultz and Schubert 1969: 387). A host of questions arise as to the contents of this explanation: for example, what were the criteria for the evaluation? What do you mean by 'inadequate'? What samples were included in the original corpus before the evaluation? How were the consonants distributed in respect to the initial or final position? A sample illustration of the MCDT shows that these questions were not considered in a principled way (cf., Miller and Nicely 1955). In Table 6 are shown a selection from the first test form that uses CID W-22, List 1. The phonemic symbols for the contrasting sounds are added.

Table 6. Samples from MCDT Test 1

1	2	3	4	5
A. ash /-š/	A. jarred /ǰ-/	A. card /-d/	A. fuss /f-/	A. *day* /d-/
B. Al /-l/	B. hard /h-/	B. *carve* /-v/	B. muss /m-/	B. way /w-/
C. as /-z/	C. barred /b-/	C. cars /-z/	C. up /-p/	C. bay /b-/
D. *an* /-n/	D. *yard* /y-/	D. Carl /-l/	D. of /-v/	D. hay /h-/
E. am /-m/	E. lard /l-/	E. carp /-p/	E. *us* /-s/	E. they /ð-/

16	21	49
A. is /-z/	A. twin /-n/	A. hire /h-/
B. ear /ih-r/	B. twinge /-ǰ/	B. *wire* /w-/, /-ə-r/, or /-ayə-r/
C. in /-n/	C. twinned /-d/	C. buyer /b-/
D. *it* /i-t/	D. twigs /-g-/	D. liar /l-/
E. if /-f/	E. *twins* /-nz/	E. wise /-ay-z/

What is meant by a 'principled way' is that there should be certain convincing reason or reasons, for example, for contrasting /-š/, /-l/, /-z/, or /-m/ to /-n/ in Sample 1; why /ǰ-/, /h-/, /b-/, or /l-/ against /y-/, and so forth. Furthermore, is there any cogent reason for contrasting /f-/ or /m-/ to /ϕ-/ (null sound) and for including these initial contrasts in the group where /-p/ or /-v/ is contrasted to /-s/ in the final position in Sample 4? In Sample 16, the test word is 'it,' whose /i-/ and /-t/ are contrasted to /ih-/ and /-r/, respectively. That is, items B and D differ in two sounds, not just one. Therefore, if a subject answered with item B, there is no way to tell which sound is substituted. The test word, 'twins' in Sample 21 contains two contrasting sounds, /-n-/ and /-z/, the first of which is contrasted to /-g-/ in 'twigs' and the second, to /-n/, /-ǰ/, or /-d/. The contrasts in Sample 49 are more confusing. The test word is 'wire,' which is pronounced /wayər/, to which the initial consonant is contrasted in items A, C, and D. But, what contrast are we testing if the subject's answer is 'wise'? Is it the contrast of /-ə-/ to /-z/, of /-r/ to /-z/, or of /-ay-/ to /-ayə-/?

Schultz and Schubert (1969) started out with the distinctive feature analysis of errors in mind, but concluded their study by saying 'the MCDT, using W-22 stimuli in a forced-choice response paradigm, has been constructed to allow systematic appraisal of response errors at the *phonemic level*, . . .' (Italics mine) (Schultz and Schubert 1969: 398).

The basic flaw in MCDT is that the W-22 words are used as test materials. The monosyllabic words in W-22 do not incorporate the phonemic distinctions which the MCDT is attempting to differentiate.

2.5. Rhyme Test (Fairbanks 1958), Modified Rhyme Test (MRT) (House et al. 1963, 1965), and Rhyming Minimal Contrasts Test (Griffiths 1967)

The Rhyme Test by Fairbanks (1958) is another version of the 'closed-set response' test using monosyllabic words, and Fairbanks specifically referred to it as a test for 'phonemic differentiation' (Fairbanks 1958: 596), which shows the realization of the necessity of diagnostic test which can tap hearing impairments in terms of structural linguistic segments. A drawback in the Fairbanks' test, however, was that he limited the test sounds to the initial position of the monosyllabic word. He used the term 'rhyme' for which 'alliteration' might have been more appropriate; we do not rhyme initial parts of words, 'rhyming' applies only to the final parts of words. The drawback in Fairbanks' Rhyming Test was remedied by the Modified Rhyme Test (MRT) (House et al. 1963, 1965) in that in MRT the test sounds were placed in both the initial and final positions in monosyllabic words.

The MRT is composed of 50 ensembles, the choice of the number 50 being due to the 'deference to traditional usage' (House et al. 1965: 159). Each ensemble is made up of six words. In the first 25 ensembles, the final consonants are contrasted, and the initial consonants are varied in the last 25 ensembles. In the selection of test words, they tried to 'exclude exotic and objectionable words, and efforts were made to ensure the inclusion of *representative sounds* from each of the *major categories of speech sounds*. In the final analysis, the nature of the lists was determined largely by this last requirement and by the desire

Table 7. Samples of MRT

	A	B	C	D	E	F
1.	bat	bad	back	bass	ban	bath
2.	bean	beach	beat	beam	bead	beak
3.	bun	bus	but	buff	buck	bug
4.	came	cape	cane	cake	cave	case
5.	cut	cub	cuff	cup	cud	cuss
⋮	⋮	⋮	⋮	⋮	⋮	⋮
26.	led	shed	red	bed	fed	wed
27.	sold	told	hold	fold	gold	cold
28.	dig	wig	big	rig	pig	fig
29.	kick	lick	sick	pick	wick	tick
30.	book	took	shook	cook	hook	look

to equate the representation of sounds from test form to test form' (Italics mine) (House et al. 1965: 160). However, more specific delineations of the concept of 'representative sounds' and 'major categories' seem to be required. In Table 7 are given five ensembles from the first 25 and the last 25 ensembles each (House et al. 1965: 159). Table 8 is a reproduction of House et al.'s (1965: 160) Table 11, with the added markings, X's, for missing segments.

Table 8. Frequency of Consonants in MRT

Consonants:														
	p	b	m	w	t	d	n	s	z	l	r	y	f	v
Initial	11	14	7	9	14	8	5	14	X	7	10	X	12	1
Final	12	6	9		15	11	19	12	4	10	2		4	5
	θ	ð	č	ǰ	š	ž	k	g	ŋ	h	#			
Initial	1	1	X	2	3		9	8		12	2			
Final	4	1	3	1	X	X	18	6	5		3			

The first question which can be raised concerns the different distributional weight assigned to each segment by a different number of segments respresented in the test. The MRT is not based on the Dewey's (1923) frequency counts, which is harmless. Then, what is the source of or rationale for the differential consonant assignments?

The second question is about the representative consonants of the major consonant categories. First of all, we have to know what is meant by 'major category': is it the category in terms of places of articulation, in terms of manners of articulation, in terms of the functional load of individual sounds, or some other properties, such as the 'major distinctive features' (Chomsky and Halle 1968)? There are 14 initial /s-/'s, and none of initial /z-/'s; 12 initial /f-/'s, but only one initial /v-/. Can we infer from these numbers that House et al. took the voiceless fricative as being representative rather than the voiced counterparts in the category of fricative consonants? If that was their criterion, then, the reason for having only one segment for each of /θ/ and /ð/ is unexplainable. Moreover, in the category of affricate consonants, the voiced segment shows up twice, whereas the voiceless counterpart is completely missing. I do not think that House et al. meant by this that /ǰ-/ is more representative of the two affricate consonants in the initial position, but the reverse is true in the final position, because of the fact that there are more final /-č/'s than final /ǰ-/'s. Moreover, I do not think that they thought there was no way to decide on the representative segment among the final alveopalatal fricatives and, therefore, they included neither /-š/ nor /-ž/ in their test materials. But we need an explanation for this skewed distribution of segments.

Table 9. Missing Consonants in Rhyming Minimal Contrasts (The place where the segment is missing is marked with an 'X.' F = final, I = initial. The number in 'Total' shows the number of segments that appear in the test.)

	A		B		C		D		E		Total	
	F	I	F	I	F	I	F	I	F	I	F	I
p											6	8
b											8	7
m								X	X		4	4
w												8
t											9	11
d											10	7
n				X							9	4
s											9	5
z				X		X					6	3
l											5	6
r					X		X		X		2	9
y				X		X		X		X		1
f											6	7
v								X			5	4
θ				X							7	4
ð									X		4	5
č	X	X						X		X	4	2
ǰ		X			X	X					3	3
š									X		4	6
ž					X		X		X		2	
k											8	7
g											8	6
ŋ											6	
h												7

Griffiths' Rhyming Minimal Contrasts (Griffiths 1967) is a further modification of MRT. Its format is the same as that of MRT with 50 monosyllabic words, each of which has six choices from which the subject is to recognize the test word. The consonant contrasts are again distributed in respect to the initial and final positions. Again, the first 25 ensembles are contrasted in the final consonants and the remaining 25 are contrasted in the inital consonants. Table 9 shows the missing segments and the frequency of occurrence of each contrasting consonant, in both final and initial positions.

The same questions which were raised about MRT can also be asked about the Griffiths' rhyming test, even though we grant that Griffiths' consonant distribution better reflects the structural English segments. For instance, the absence of the missing segments in Table 9 should be explained. Weighted to the emphasis that Griffiths placed on the concept of 'minimal feature contrast,' his conclusion on this matter was not clear. In the section on 'Analysis of Confusions,' he pointed out the confusions in the pairs of /v/ and /ð/, and /θ/ and /f/, but no mention is found about the minimal feature contrast, be it in terms of distinctive features or in terms of Miller and Nicely's (1955) articulatory/physiological features. Rather surprisingly, he attributed the confusions to the frequencies of occurrence of the words, and concluded that 'the distinguishing acoustic cues for these pairs of sounds are tenuous at best' (Griffiths 1967: 240) (For the acoustic correlates of distinctive features, see Jakobson, Fant, and Halle 1952, and Chomsky and Halle 1968).

2.6. K.S.U. Speech Discrimination Test (Berger 1969)

Berger's (1969) Kent State University (K.S.U.) Speech Discrimination Test is innovative in the sense that sentences are employed to deliver the test words. The construction of the sentences and the selection of the test words are based on his own study of English conversations (Berger 1967). There are 8 test forms, Forms A through H, each consisting of 13 sentences, the mean length of each sentence being 6.7 words. Each sentence is one of three types: a declarative, an interrogative, or an exclamatory sentence. The 13 sentences in one test are arranged in such a way that each sentence becomes more difficult than the preceding one. Five alternative choices are incorporated into each sentence, each contrasting to the remaining four by either the initial or final consonant.

Table 10 (Berger 1969: 255) shows the Test Form A and Table 11 shows the initial and final consonants of the test words and some vowel segments which need to be brought to our attention.

No consideration at all of the systematic representation of the structural segments has been taken in the construction of the test. The vertical comparisons of consonants of five words, both initial and final, clearly reveal that there is no system. In some, single consonants are contrasted, but in others, single consonants are contrasted to consonant clusters. In Nos. (4), (10), and (11), contrasts are shown in one position: the initial position in (4) and (10), but the final position in (11). In the remaining 11 sentences, constrasts are found in both positions.

Table 10. K.S.U. Speech Discrimination Test, Form A (The test word and the four foil words are given in capital letters in each sentence. The test word is italicized.)

(1) The baby started to BAWL FALL WALK *TALK* CRAWL early.
(2) The farmer put the COW *PLOW* SOW TROWEL BOUGH in the new barn.
(3) That's a strange CULT GULL *SKULL* HULL HULK!

(4) Was that a *CLINK* MINK RINK BLINK LINK?
(5) Our pony CRUNCHED PLUNGED MUNCHED *LUNGED* LUNCHED down the hill.
(6) John couldn't find the GROOM RULE MOON BROOM *ROOM*.
(7) There's a big DOG *FROG* HOG LOG RUG next to the house.
(8) The group BOMBED CAUGHT FOUGHT *CALMED* CALLED the mad man.
(9) Hear the noisy RABBLE BABBLE BATTLE CATTLE *RATTLE* now?
(10) Was I told to BREAK MAKE *BAKE* RAKE TAKE it today?
(11) I'm glad that WAVE WHALE WADE WAIT *WAKE* is behind us.
(12) I hope they don't CATCH SCRATCH *PATCH* PASS MATCH it.
(13) His brown dog will *FIGHT* HIDE BITE GUIDE SLIDE on command.

Table 11. Consonants in the Test and Foil Words (with some vowel segments)

(1)	(2)	(3)	(4)	(5)
/b - l/	/k - /	/k - lt/	/kl - ŋk/	/kr - nčt/
/f - l/	/pl - /	/g - l/	/m - ŋk/	/pl - nǰd/
/w - k/	/s - /	/sk - l/	/r - ŋk/	/m - nčt/
/t - k/	/trawəl/	/h - l/	/bl - ŋl/	/l - nǰd/
/kr - l/	/b - /	/h - lk/	/l - ŋk/	/l - nčt/
(6)	**(7)**	**(8)**	**(9)**	**(10)**
/gr - m/	/d - g/	/bahmd/	/r - bl/	/br - k/
/r - l/	/fr - g/	/kɔht/	/b - bl/	/m - k/
/m - n/	/h - g/	/fɔht/	/b - tl/	/b - k/
/br - m/	/l - g/	/kahmd/	/k - tl/	/r - k/
/r - m/	/r - g/	/kɔhld/	/r - tl/	/t - k/
(11)	**(12)**	**(13)**		
/w - v/	/k - č/	/f - t/		
/w - l/	/skr - č/	/h - d/		
/w - d/	/p - č/	/b - t/		
/w - t/	/p - s/	/g - d/		
/w - k/	/m - č/	/sl - d/		

The subject is to be tested with one form of the test. Then, inevitably, there are many consonants which are not tested. That is, there is quite a number of missing consonants; for example, /n-/, /z-/, /y-/, /v-/, /θ-/, /ð-/, /č-/, /ǰ-/, /š-/, and /g-/ in the initial position, and /-p/, /-b/, /-z/, /-r/, /-f/, /θ-/, /ð-/, /ǰ-/, /š-/, /ž-/, and /ŋ/ in the final position are not found in Form A of the test.

Moreover, no differential diagnosis is possible with this test, because the contrast of a single consonant to a consonant cluster does not allow us to access one-to-one confusion matrix. Item 11 in Table 11 is the only exception.

2.7. The University of Oklahoma Closed-Response Speech Test (OUCRT) (Pederson and Studebaker 1972)

The main motivation behind the construction of the OUCRT is claimed to be the avoidance of the 'contaminating effects,' that is, 'the contaminating influences of word frequency, word familiarity, intra-word context and learning effects' (Pederson and Studebaker 1972: 187).

The OUCRT is a test of closed-response set. It is composed of three subtests: two for consonants and one for vowels. The two consonant subtests are the initial-consonant subtest and the final-consonant subtest.

Pederson and Studebaker's (1972: 188) criteria for the construction of the consonant tests are as follows: '(1) The variable phonemes in each set were selected to vary only in the place of articulation. . . . (2) Identical test-phoneme sets were used in both the initial- and final-position consonant subtests.' They further added that 'Three phonemes with identical manners of production were selected to be evaluated within each test set' (Pederson and Studebaker 1972: 188). Table 12 shows the monosyllabic test words in the two subtests of consonants (Pederson and Studebaker 1972: 188).

Table 12. Words Used in the Consonant Subtests

Initial-consonant subtest			
pair /p-/	tear /t-/	care /k-/	air /ø-/
bail /b-/	dale /d-/	gale /g-/	ail /ø-/
fin /f-/	thin /θ-/	sin /s-/	in /ø-/
vee /v-/	thee /ð-/	zee /z-/	ee /ø-/
stop /st-/	chop /č-/	shop /š-/	hop /h-/
Final-consonant subtest			
pop /-p/	pot /-t/	pock /-k/	pa /-ø/
robe /-b/	rode /-d/	rogue /-g/	row /-ø/
roof /-f/	Ruth /θ/	ruse /-s/	rue /-ø/
live /-v/	lithe /-ð/	lies /-z/	lie /-ø/
least /-st/	leech /-č/	leash /-š/	lee /-ø/

It is extremely difficult to search for the system on which Pederson and Studebaker based their selection of the 15 consonants and one consonant cluster for their consonant tests. Are these the consonant tests only for the non-resonant consonants? Why is the voiced affricate excluded? What is the reason for the inclusion of the single consonant cluster, i.e., /-st/? Is there any specific reason for contrasting the consonant cluster to /č/, /š/, and /h/?

Pederson and Studebaker's (1972: 189) third subtest is the vowel subtest, which is an '8-item, closed-response set.' The 8 items are 'beat' /biyt/, 'bit'

/bit/, 'bet' /bet/, 'bat' /bæt/, 'bait' /beyt/, 'boat' /bowt/, 'but' /bət/, and 'boot' /buwt/. Why does it have to be 8? Why did they exclude /u/ (i.e., the vowel in 'put'), /a/ (i.e., the vowel in 'pot'), and /ɔ/ (i.e., the vowel in 'pond')? Is there any structural reason for their exclusion of diphthongs, such as /aw/, /ay/, and /ɔy/? If any structural factor had been the reason for the selective inclusion of diphthongs, /ay/, /aw/, and /ɔy/ should have been included instead of /iy/, /ey/, /uw/, and /ow/ (Pike 1947, Fires 1945, Hill 1958, Hockett 1955, IPA 1949, Lehitste and Peterson 1962, Ladefoged 1982).

2.8. California Consonant Test (CCT) (Owens and Schubert 1977)

The California Consonant Test (CCT) is the end product of a series of test constructions by Owens and Schubert (Owens and Schubert 1968, Owens, Benedict, and Schubert 1971, Owens and Schubert 1977). The final test form, which is Form 7, was named CCT. This consonant test consists of 100 monosyllabic words, each of which contains a test consonant either in the initial or final position. The test consonants are balanced between the first 50 items and the last 50 items. The subject is tested with an answer sheet in front of him, in which there are four choices of words per each test consonant, the four words rhyming with each other. The subject is instructed to check the word on the answer sheet he thought he heard and speak it out loudly. Table 13 shows the distribution of the consonants in CCT.

Table 13. Distribution of Test Consonants in CCT (I = initial, F = final. The missing consonant is marked by 'X.')

Items 1 - 50																
	p	b	m	w	t	d	n	s	z	l	r	y	f	v	θ	ð
I	2	X	X	X	4	2	X	3	X	X	X	X	X	X	1	X
F	6	I	X		1	X	X	6	1	X	X		3	2	X	X
	č	ǰ	š	ž	k	g	ŋ	h		Total						
I	2	X	2		3	X		X		19						
F	4	3	1	X	3	X	X			31						
Items 51 - 100																
	p	b	m	w	t	d	n	s	z	l	r	y	f	v	θ	ð
I	2	1	X	X	3	X	X	3	X	X	X	X	1	X	1	X
F	4	1	X		3	X	X	5	2	X	X		1	3	1	X
	č	ǰ	š	ž	k	g	ŋ	h		Total						
I	2	1	1		3	X		X		18						
F	4	2	2	X	3	1	X			32						

The total numbers show that they placed more emphasis on the consonant confusions in the final position rather than in the initial position. More attention seems to have been given to the confusions among the voiced consonants rather than among voiceless consonants, because in their earlier studies they found that there were 'relatively few voicing errors' (Owens and Schubert 1977: 465). Furthermore, they found that subjects 'seldom confuse nasal with nonnasal consonants' (Owens and Schubert 1977: 465). However, the complete exclusion of nasal consonants in the test is hard to understand. Even though the nasals are seldom confused with nonnasals, e.g., /m/ vs. /p/ or /b/, no evidence has been presented that the distinctions among the three nasals, i.e., /m/, /n/, and /ŋ/, do not cause problems to hearing-impaired subjects.

Another finding in their earlier studies was that 'confusion between /f/ and /θ/ is such that when either is the stimulus consonant, the other member of the pair should not be included among the foils (the same holds true for /v/ and /ð/)' (Owens and Schubert 1977: 465). In my interpretation, this finding means that the rates of confusion are so high between /f/ and /θ/, and between /v/ and /ð/ that there is no necessity to further test the confusions. And this fact seems to have caused them to exclude completely /ð/ from the test. However, this seems to be too radical a solution. Moreover, they have not given an explanation for the exclusion of /h-/, another fricative.

In addition to the exclusion of the nasal consonants, all other categories of resonant consonants are also excluded: the liquids (i.e., /r/ and /l/) and the semi-consonants (i.e., /w-/ and /y-/), and this has not been justified.

Finally, it is worth noting that the test-retest reliability of the CCT is 0.93, but the correlation between W-22 and the CCT is only 0.35, which was interpreted as meaning that 'a multiple-choice consonant test would probably not be measuring the same thing as a W-22 word discrimination test, . . .' (Owens and Schubert 1977: 465).

2.9. Synthetic Sentence Identification (SSI) (Speaks and Jerger 1965, 1966, Jerger, Speaks, and Trammell 1968)

The SSI is a speech audiometric test with 'artificial synthetic sentences.' The motivation for constructing the synthetic sentence is found in the following remarks: 'In a "real" sentence, "meaning" may be conveyed by only one or two key words' (Speaks and Jerger 1965: 187, also see Jerger, Speaks, and Trammell 1968: 319). That is, the synthetic sentences were primarily motivated by a desire to control 'meaning,' whatever meaning it may be. In the following quotation, we come closer to what is meant by 'meaning,' but still not satisfactorily clear: '. . . it seems desirable that the sentences be of controlled length and controllable relative informational content. "Relative information" is used here in the sense of unspecifiable variations in the amount of information conveyed in a message set as a consequence of variable sequential constraints'

(Speaks and Jerger 1965: 185). Paraphrasing the word 'meaning' into 'information' which is of 'unspecifiable variation' does not help much. Another hint on the meaning of 'information' is found in the expressions, '. . . systematic exploration of temporal processing is desired. Interest in the time domain suggests that the verbal materials should consist of sentences of sufficient duration to permit systematic alteration of temporal characteristics of the speech message' (Speaks and Jerger 1965: 185). It is, however, still unclear what they meant by 'temporal exploration.' A reasonable guess might be that by 'temporal processing' is meant the psychological mental processing in terms of the temporal constratint of short-term memory for the processing of information.

The SSI consists of a pool of 24 10-sentence message sets. Each sentence is one of the three types of sentences of different order: the first-order sentence, the second-order sentence, or the third-order sentence. The words in a first-order sentence are in a linear order in which no word has informational relations with its adjacent word, both preceding and following. In a second-order sentence, each word has linear informational relationships only with the immediately preceding word and with the immediately following word. In a third-order sentence, any word has informational connection with two immediately preceding words. Table 14 is a reproduction of Speaks and Jerger's (1965: 188) table.

Table 14. Typical Examples of Seven-word Sentences Constructed on the Basis of Conditional Probabilities

Order of Approximation	Example Sentence
First	Do mind instead edge drop quickly till.
Second	Laugh long name my french women laugh.
Third	Forward march said the boy had a.

Table 15 is a reproduction of Jerger, Speaks, and Trammell's (1968: 320) table of 10 third-order synthetic sentences.

Table 15. Example Message Set Consisting of 10 Alternative Synthetic Sentences, Constructed as Third-order Approximations to Real Sentences

	Alternative Sentences
1.	Small boat with a picture has become
2.	Built the government with the force almost
3.	Go change your car color is red
4.	Forward march said the boy had a

5. March around without a care in your
6. That neighbor who said business is better
7. Battle cry and be better than ever
8. Down by the time is real enough
9. Agree with him only to find out
10. Women view men with green paper should

This table is followed by the following elaboration which defines the basic characteristic nature of the synthetic sentences in the SSI, which, in turn, hints at what is meant by the 'information content':

> They [the third-order sentences in Table 15] are homogeneous in the sense that all have seven words and a controlled, but not necessarily equal, number of syllables. But the most important feature that determines the degree of homogeneity is that each is based on the word-triplet rule. Expressed differently, the amount of redundancy, related to the dependence of any one word on the other words in the sourrounding context, is relatively similar among the 10 alternatives. (Jerger, Speaks, and Trammell 1968: 320)

The number of words or of syllables in a sentence might contribute to the control of informational content of the sentence, but only to the negligible extent and that only peripherally and negatively. There are infinite cases in which a sentence consisting of one or two words can be expressed in a much longer sentence with the similar informational message. An important question relates to the 'word-triplet rule' which seems to be a rule responsible for the control of redundant information in the sentence. What it presumably suggests is that any three words which are linearly sequenced in a sentence contain the same amount of redundant information. Pushed a little further, it means that any sequences of words of the same length in the number of words (let it be three words in length) in a sentence are of the same amount of redundancy, and, therefore, they are the same in meaning or in informational content. If interpretable in this way, the definition of meaning in the SSI flatly contradicts the concept of meaning on which the synthetic sentences are based, that is, the meaning of a real English sentence 'is often conveyed by only one or two key words, . . .' (Jerger, Speaks, and Trammell 1968: 319)

The meaning of a sentence can sometimes be predicted by all kinds of redundancies, including expectations which are possible due to the speaker's preknowledge of the listener, the environment in which a conversation is taking place, the general human knowledge between the conversants about the society and its permissible set-procedures and behavior patterns, concrete and/or abstract. In addition, a variety of linguistic predictions play important roles in the comprehension of sentences. Word meanings and structural meanings are among them. In short, possible sequences of the same number of words each cannot be taken as equal in informational content. Differences in word meaning can create different expectations and predictions. Apart from word meanings, structural meanings also make a great difference. The tree diagrams given in Figures 3 to 9 show the seven different structures of triplet, 3 to 7 for each of the five triplets in 'Small boat with a picture has become,' and 8 and 9 for the first two triplets in ''That neighbor who said business is better.'

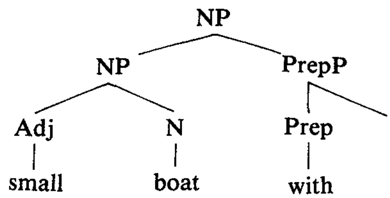

Figure 3. 'small boat with'

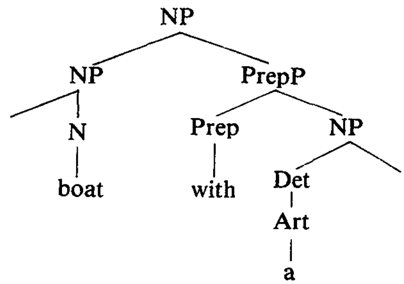

Figure 4. 'boat with a'

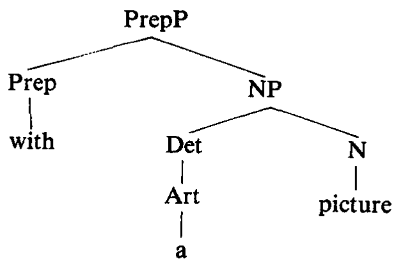

Figure. 5. 'with a picture'

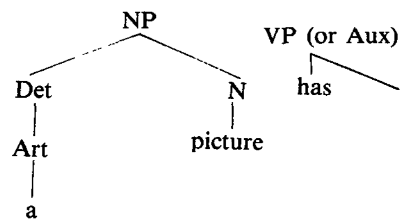

Figure 6. 'a picture has'

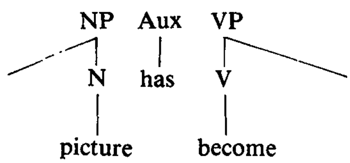

Figure 7. 'picture has become'

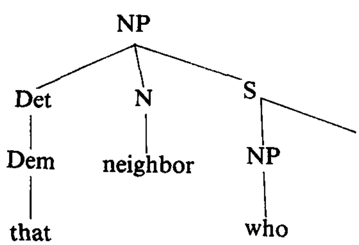

Figure 8. 'that neighbor who'

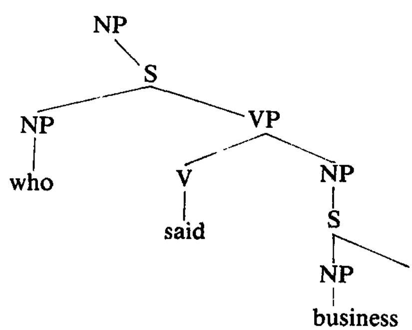

Figure 9. 'who said business'

(Legend: NP = noun phrase, N = noun, PrepP = prepositional phrase, Prep = preposition, Adj = adjective, Det = determiner, VP = verb phrase, V = verb, Aux = auxiliary, Dem = demonstrative, S = sentence)

No two triplets among the 7 shown in Figures 3 to 9 are equivalent in structure. Figure 3 and Figure 4 show that the two expressions are NP's; however, they are different NP's in that 'small boat with' is of the structure of 'Adj + N + Prep,' whereas 'boat with a' is that of 'N + Prep + Art.' The NP in Figure 8 is further different from either NP in Figure 3 or Figure 4. 'That neighbor who' has the structure of 'Dem + N + NP,' in which the NP is from an embedded S which is different from the S which dominates 'that neighbor.' That is, the NP 'who' is from a different 'clause.' A far more different NP is found in Figure 9, in which the whole expression 'who said business' is an embedded structure. 'Business' is further embedded in the S which dominates 'who.' Either triplet in Figure 6 or Figure 7 cannot even be traced to a single node below the highest node S for the whole sentence in which the triplet is a part. More importantly, only one triplet, that is, 'with a picture' shown in Figure 5, is a complete 'construction.' All other triplets shown in Figures 3 to 9 are structurally defective as demonstrated by the stranded lines which do not end up with any 'constituent.' Simply put, a different structure has a different 'meaning' or is different in 'informational content.'

Finally, and fatally, the synthetic 'sentences' are not sentences at all; none of them is a rule-governed whole sentence in English (Radford 1981, Chomsky 1965, 1981).

2.10. The Speech Perception in Noise Test (SPIN Test) (Kalikow, Stevens, and Elliot 1977)

I would like to begin with quotations.

> Basically two kinds of operations are involved in the understanding of sentences. One is the reception and initial processing of acoustic information through the auditory system, and the other is the utilization of linguistic information that is stored in memory. . . . One component in the decoding of a sentence by a listener is the extraction of a partial set of phonetic features from the acoustic signal. The phonetic features are placed in short-term memory, where they are available for further processing. The linguistic information available in the long-term memory of a listener includes knowledge of the phonological, lexical, syntactic, and semantic constraints that occur in language. The more these kinds of information provide a context for a particular utterance, the less it is necessary for the listener to depend on the detailed properties of the acoustic signal in order to understand the utterance. A test of a listener's ability to understand everyday speech must, therefore, assess both the acoustic-phonetic and the linguistic-situational components of the process. (Kalikow et al. 1977: 1337)

> A major objective in developing this test was to produce a measure that would assess utilization of the linguistic-situational information. In contrast to most tests of speech intelligibility which examine only processing of acoustic-phonetic information, we wanted

> to place equal emphasis on examining the contribution of 'cognitive' variables of memory and language competence. (Kalikow et al. 1977: 1339).

The SPIN Test is based on the distinction in processing between acoustic information and linguistic-situational information. The linguistic-situatinal information is claimed to be accessible from long-term memory. But how is 'the extraction of a partial set of phonetic features from the acoustic signal' possible? The acoustic signal is objective with no contamination from a particular language. However, the extraction of phonetic features is a process of structure imposition on acoustic signal, which is only possible through the stored knowledge in long-term memory. Thus, the dichotomy of Kalikow et al. into knowledge-involved information versus non-knowledge information is at fault. The phonetic system is part of 'language competence' which involves higher level 'cognitive' processes.

The SPIN Test consists of ten 50-item tests (in Kalikow et al. 1977: 1348-1349, only 8 tests are given). Each item is a sentence. Twenty-five of the 50 sentences are of 'high predictability' (PH) and the other 25 are of 'low predictability' (PL). The last word, which is monosyllabic, in each sentence is the test word. In a PH sentence, the test word is highly predictable from two or three 'pointer' words which precede the test word in the sentence. The 'pointer' words are so-called content words. The assumption is that the subject can identify the test word in a PH sentence through the cognitive processing of the pointer words in addition to the phonetic information, whereas in a PL sentence the subject has to identify the test word only with the available phonetic information.

In Table 16 are reproduced the first 20 sentences from the test, Form 2.1.

Table 16. The First 20 Sentences in the SPIN Test, Form 2.1. (H = high predictability sentence, L = low predictability sentence)

(H)	1.	The watchdog gave a warning growl.
(H)	2.	She made the bed with clean sheets.
(L)	3.	The old man discussed the dive.
(L)	4.	Bob heard Paul called about the strips.
(L)	5.	I should have considered the map.
(H)	6.	The old train was powered by steam.
(H)	7.	He caught the fish in his net.
(L)	8.	Miss Brown shouldn't discuss the sand.
(H)	9.	Close the window to stop the draft.
(H)	10.	My T.V. has a twelve-inch screen.
(L)	11.	They might have considered the hive.
(L)	12.	David has discussed the dent.
(H)	13.	The sandal has a broken strap.
(H)	14.	The boat sailed along the coast.

(H)	15.	Crocodiles live in muddy swamps.
(L)	16.	He can't consider the crib.
(H)	17.	The farmer harvested his crop.
(H)	18.	All the flowers were in bloom.
(L)	19.	I am thinking about the knife.
(L)	20.	David does not discuss the hug.

One thing we can easily notice is that in the PL sentence the verb which precedes the test word, which is the only source for prediction, is a verb of extremely general meaning: 'discuss,' 'call about,' 'consider,' 'think about,' 'know about,' 'speak about,' 'hear about,' 'be interested in,' 'talk about,' 'ask about,' and 'have a problem with.' Furthermore, the test word which is used with a specific verb is so outlandish that the meaning of the sentence turns out to be odd: 'discuss the sand' (Sentence 8), 'discuss the pine,' 'discuss the yell,' 'had a problem with the bloom.' On the other hand, the pointer words in PH sentences are the lexical items which are so closely related to the test word, not only in the lexical meaning, but also in the structural meanings of a number of different imports.

For one thing, the prepositions used in PH sentences have a very high predicting power: '*with* clean sheets' (Sentence 2), '*in* muddy swamps' (Sentence 15), '*in* bloom' (Sentence 18). For another, a large number of set phrases or idiomatic expressions are found in PH sentences, e.g., 'give a kick,' 'give a hint,' 'a game of cat and mouse,' 'scare out of one's wits,' 'Adam's rib.' The point is that if a test word does not have alternatives, it may cease to be a test word.

The crucial question, as with other speech audiometric materials, is: What are we testing? In the case of a PL sentence, we might be testing the subject's auditory processing of acoustic properties. However, in the case of a PH sentence, we are not only testing the subject's understanding of the 'pointer' content words, but the whole gamut of human knowledge, some of which is auditorily cued, but a large portion of it has nothing to do with the auditory processing, that is, the subject can come up with the correct test word before he has heard the word. The informational gap between PH sentences and PL sentences are so large that it becomes unjustifiable to include the PH sentences in the auditory test.

3. Conclusion

It is only recently that we can find some efforts to incorporate the parameters for linguistic differentiations in speech audiometric materials. However, the two major speech materials currently used for routine audiological evaluation, i.e., CID W-1 and CID W-2, are devoid of any considerations of systematic structures of speech of language.

Despite the recent endeavor to incorporate speech structures in the speech audiometric tests, careful speech or language research is lacking in the assumptions, principles, and procedures of the material construction.

As a consequence, both traditionally and currently, speech audiometric tests can give us reliable correlations and reliable test-retest results; however, they are not valid tests because, first of all, they do not give us information on differential diagnosis as a function of differential structural elements in a language. In the worst cases, the tests do not have speech or linguistic parameters which should be the focal goals of measurement. The two routine speech audiometric materials, the CID W-1 and the CID W-22, belong to the latter worse cases. Tests without validity, despite their reliability, need serious appraisals.

One 'justification' which many audiologists bring up in favor of the W-1 and the W-22 is that the speech audiometric test should be easy to administer with the least training for the examiner and should be short enough not to burden the patient to the extent of fatigue or boredom. But, these reasons are not persuasive, because the incorporation of linguistic structures does not itself make the administration of the test difficult or cause the patient's fatigue or boredom. Viewed from a different perspective, the present speech audiometric tests, in fact, take more of the patient's and the examiner's time, because if a hearing impairment is found in a patient which necessitates an administration of a further 'special' speech test, the time taken by the further test is the extra time. If the test is linguistically structured, this extra time is unnecessary for the patient. It is only the examiner who has to take extra time to analyze the test results. This means that the analysis of the speech audiometric result is limited to the patient who needs further detailed differential diagnosis. The audiologist or the speech pathologist, who is responsible for the analysis of test result, naturally, should be linguistically oriented, if not sophisticated, because language, i.e., a set of linguistic rules, underlies speech, the object of speech audiometry.

REFERENCES

Benguerel, A-P. and T. K. Bhatia (1980) 'Hindi Stop Consonants: An Acoustic and Fiberscopic Study,' *Phonetica* 37, 134-148.

Berger, K. W. (1967) 'Conversational English of University Students,' *Speech Monographs* 34, 65-73.

________ (1969) 'Speech Discrimination Task Using Multiple Choice Key Words in Sentences,' *J. Aud. Res.* 9, 247-262.

________ (1971) 'Speech Audiometry,' in D. E. Rose, ed., *Audiological Assessment,* Prentice-Hall, Englewood Cliffs, N.J.

Bloch, B. and G. L. Trager (1942) *Outline of Linguistic Analysis,* Waverly Press, Baltimore.

Campanelli, P. A. (1962) 'A Measure of Intra-list Stability of Four PAL Word Lists,' *J. Aud. Res.* 2, 50-55.

Carhart, R. (1965) 'Problems in the Measurement of Speech Discrimination,' *Arch. Otolaryngol.* 82, 253-260.

_________ (1971) 'Observations on Relations between Thresholds for Pure Tones and for Speech,' *J. Speech Hear. Disord.* 36, 476-483.

_________ and L. S. Porter (1971) 'Audiometric Configuration and Prediction of Threshold for Spondees,' *J. Speech Hear. Res.* 14, 486-495.

Chomsky, N. (1965) *Aspects of the Theory of Syntax,* The MIT Press, Cambridge, Mass.

_________ (1981) *Lectures on Government and Binding: The Pisa Lectures,* Foris Publications, Dordrecht, Holland.

_________ and M. Halle (1968) *The Sound Pattern of English,* Harper and Row, New York.

Cole, P. and J. L. Morgan, eds. (1975) *Syntax and Semantics, Vol. 3: Speech Acts,* Academic Press, New York.

Conn, M., J. Dancer, and I. M. Ventry (1975) 'A Spondee List for Determining Speech Reception Threshold without Prior Familiarization,' *J. Speech Hear. Disord.* 40, 368-376.

Curry, E. T., A. Thayer, and B. P. Cox (1966) 'The Relative Intelligibility of Spondees,' *J. Aud. Res.* 6, 419-424.

Davis, H. and S. R. Silverman (1970) *Hearing and Deafness,* Holt, Rinehart and Winston, New York.

De Beaugrande, R. (1980) *Text, Discourse, and Process: Toward a Multidisciplinary Sciences of Text,* Ablex Publishing Corporation, Norwood, N.J.

Dewey, G. (1923) *Relative Frequency of English Speech Sounds,* Harvard University Press, Cambridge, Mass.

D'Andrade, R. G. (1981) 'The Cultural Part of Cognition,' *Cognitive Science* 5, 179-195.

Egan, J. P. (1948) 'Articulation Testing Methods,' *Laryngoscope* 58, 955-991.

Elpern, B. S. (1961) 'The Relative Stability of Half-list and Full-list Discrimination Test,' *Laryngoscope* 71, 30-35.

Fairbanks, G. (1958) 'Test of Phonemic Differentiation: The Rhyme Test,' *J. Acoust. Soc. Am.* 30, 596-601.

Fletcher, H. (1929) *Speech and Hearing,* Van Nostrand-Reinhold, Princeton, N.J.

_________ (1950) 'A Method of Calculating Hearing Loss for Speech from an Audiogram,' *J. Acoust. Soc. Am.* 22, 1-5.

French, N. R., C. W. Carter, and W. Koenig (1930) 'The Words and Sounds of Telephone Conversations,' *Bell Syst. Tech. J.* 9, 290-324.

Fries, C. C. (1945) *Teaching and Learning English as a Foreign Language,* University of Michigan Press, Ann Arbor, Mich.

Gleason, Jr., H. A. (1961) *An Introduction to Descriptive Linguistics,* Holt, Rinehart and Winston, New York.

Goetzinger, C. P. (1978) 'Word Discrimination Testing,' in J. Katz, ed., *Handbook of Clinical Audiology*, The Williams and Wilkins Co., Baltimore.

Griffiths, J. D. (1967) 'Rhyming Minimal Contrasts: A Simplified Diagnostic Articulation Test,' *J. Acoust. Soc. Am.* 42, 236-241.

Grubb, P. A. (1963a) 'A Phonemic Analysis of Half-list Speech Discrimination Test,' *J. Speech Hear. Res.* 6, 271-275.

_________ (1963b) 'Some Consideration in the Use of Half-list Speech Discrimination Test,' *J. Speech Hear. Res.* 6, 294-297.

Hill, A. A. (1958) *Introduction to Linguistic Structures: From Sound to Sentence in English,* Harcourt, Brace and World, New York.

Hirsch, I. J., H. Davis, S. R. Silverman, E. Reynolds, E. Eldert, and R. W. Benson (1952) 'Development of Materials for Speech Audiometry,' *J. Speech Hear. Disord.* 17, 321-337.

Hockett, C. F. (1955) *A Manual of Phonology,* Indiana University Press, Bloomington, Ind.

Houghson, W. and E. A. Thompson (1942) 'Correlation of Hearing Acuity for Speech with Discrete Frequency Audiograms,' *Arch. Otolaryngol.* 36, 526-540.

House, A. S., C. E. Williams, M. H. L. Hecker, and K. D. Kryter (1963) 'Psychoacoustic Speech Tests: A Modified Rhyme Test,' Technical Documentary Report ESD-TDR-63-403, U.S. Air Force Systems Command, Hanscom Field, Electronic Systems Division.

House, A. S., C. E. Williams, M. H. L. Hecker, and K. D. Kryter (1965) 'Articulation Testing Methods: Consonantal Differentiation with a Closed Response Set,' *J. Acoust. Soc. Am.* 37, 158-166.

Hudgins, C. V., J. E. Hawkins, Jr., J. E. Karlin, and S. S. Stevens (1947) 'The Development of Recorded Auditory Tests for Measuring Hearing Loss for Speech,' *Laryngoscope* 57, 57-89.

International Phonetic Association (1949) *The Principles of the International Phonetic Association: Being a Description of the International Phonetic Alphabet and the Manner of Using it,* Department of Phonetics, University College, London, International Phonetic Association.

Jakobson, R., G. Fant, and M. Halle (1952) *Preliminaries to Speech Analysis,* The MIT Press, Cambridge, Mass.

Jerger, J., R. Carhart, T. W. Tillman, and J. L. Peterson (1959) 'Some Relations between Normal Hearing for Pure Tones and for Speech,' *J. Speech Hear. Res.* 2, 126-140.

Jerger, J., C. Speaks, and J. L. Trammell (1968) 'A New Approach to Speech Audiometry,' *J. Speech Hear. Disord.* 33, 318-328.

Kalikow, D. N., K. N. Stevens, and L. L. Elliot (1977) 'Development of a Test of Speech Intelligibility in Noise Using Sentence Materials with Controlled Word Predictability,' *J. Acoust. Soc. Am.* 61, 1337-1351.

Kintsch, W. (1974) *The Representation of Meaning in Memory,* John Wiley and Sons, New York.

Kintsch, W. and T. A. van Dijk (1978) 'Toward a Model of Text Comprehension and Production,' *Psychological Review* 85, 363-394.

Kreul, E. J., J. C. Nixon, K. D. Kryter, D. W. Bell, J. S. Lang, and E. D. Schubert (1968) 'A Proposed Clinical Test of Speech Discrimination,' *J. Speech Hear. Res.* 11, 536-552.

Ladefoged, P. (1982) *A Course in Phonetics,* Harcourt, Brace and Jovanovich, New York.

Lehiste, I. and G. E. Peterson (1959) 'Linguistic Considerations in the Study of Speech Intelligibility,' *J. Acoust. Soc. Am.* 31, 280-286.

__________ (1962) 'Transitions, Glides, and Diphthongs,' *J. Acoust. Soc. Am.* 33, 268-277.

Liberman, M. and A. Prince (1977) 'On Stress and Linguistic Rhythm,' *Linguistic Inquiry* 8, 249-336.

Martin, F. N. (1981) *Introduction to Audiology,* Prentice-Hall, Englewood Cliffs, N.J.

Miller, G. A. and P. E. Nicely (1955) 'An Analysis of Perceptual Confusions among some English Consonants,' *J. Acoust. Soc. Am.* 27, 338-352.

Minsky, M. (1975) 'A Framework for Representing Knowledge,' in P. H. Winston, ed., *The Psychology of Computer Vision,* McGraw-Hill, New York.

Norman, D. A. (1980) 'Twelve Issues for Cognitive Science,' *Cognitive Science* 4, 1-32.

Olsen, W. O. and N. D. Matkin (1979) 'Speech Audiometry,' in W. F. Rintelmann, ed., *Hearing Assessment,* University Park Press, Baltimore.

Owens, E. and E. D. Schubert (1968) 'The Development of Consonant Items for Speech Discrimination Testing,' *J. Speech Hear. Res.* 11, 656-667.

Owens, E., M. Benedict, and E. D. Schubert (1971) 'Further Investigation of Vowel Items in Multiple-Choice Discrimination Testing,' *J. Speech Hear. Res.* 14, 841-847.

Pederson, O. T. and G. A. Studebaker (1972) 'A New Minimal Contrasts Closed-Response-Set Speech Test,' *J. Aud. Res.* 12, 187-195.

Peterson, G. E. and I. Lehiste (1962) 'Revised CNC Lists for Auditory Tests,' *J. Speech Hear. Disord.* 27, 62-70.

Pike, K. L. (1947) *Phonemics: A Technique for Reducing Language to Writing,* University of Michigan Press, Ann Arbor, Mich.

Radford, A. (1981) *Transformational Syntax: A Student's Guide to Chomsky's Extended Standard Theory,* Cambridge University Press, Cambridge.

Resnick, D. (1962) 'Reliability of the Twenty-five Word Phonetically Balanced Lists,' *J. Aud. Res.* 2, 5-12.

Rumelhart, D. E. and A. Ortony (1977) 'The Representation of Knowledge in Memory,' in R. C. Anderson, R. J. Spiro, and W. E. Montague, eds., *Schooling and Acquisition of Knowledge,* Lawrence Erlbaum Associates, Hillsdale,

N.J.
Schane, S. A. (1979) 'The Rhythmic Nature of English Word Accentuation,' *Language* 55, 559-602.
Schank, R. C. and R. P. Abelson (1977) *Scripts, Plans, Goals, and Understanding,* Lawrence Erlbaum Associates, Hillsdale, N.J.
Schultz, M. C. and E. D. Schubert (1969) 'A Multiple Choice Discrimination Test (MCDT),' *Laryngoscope* 79, 382-399.
Searle, J. R. (1969) *Speech Acts: An Essay in the Philosophy of Language,* Cambridge University Press, Cambridge.
Shutts, R. E., K. S. Burke, and J. E. Creston (1964) 'Derivation of Twenty-five Word PB Lists,' *J. Speech Hear. Disord.* 29, 442-447.
Siegenthaler, B. M. and R. Strand (1964) 'Audiogram-average Methods and SRT Scores,' *J. Acoust. Soc. Am.* 36, 589-595.
Silverman, S. R. and I. J. Hirsch (1955) 'Problems Related to the Use of Speech in Clinical Audiometry,' *Ann. Otol. Rhinol. Laryngol.* 64, 1234-1244.
Speaks, C. and J. Jerger (1965) 'Method for Measurement of Speech Identification,' *J. Speech Hear. Res.* 8, 185-194.
_________ (1966) 'Performance-intensity Characteristics of Synthetic Sentences,' *J. Speech Hear. Res.* 9, 305-312.
Thorndike, E. L. (1932) *A Teacher's Word Book of the Twenty Thousand Words Found Most Frequently and Widely in General Reading for Children and Young People,* Teachers College, Columbia University, New York.
_________ and I. Lorge (1944) *The Teacher's Word Book of 30,000 Words,* Teachers College, Columbia University, New York.
Tillman, T. W. and R. Carhart (1966) 'An Expanded Test for Speech Discrimination Utilizing CNC Monosyllabic Words: Northwestern University Auditory Test No. 6,' Technical Report, SAM-TR 66-55, USAF School of Aerospace Medicine, Aerospace Medical Division (AFSC), Brooks Air Force Base, Texas.
_________ and L. Wilber (1963) 'A Test for Speech Discrimination Composed of CNC Monosyllabic Words: Northwestern University Auditory Test No. 4,' Technical Report, SAM TDR-62-135, USAF School of Aerospace Medicine, Aerospace Medical Division (AFSC), Brooks Air Force Base, Texas.
Tillman, T. W. and W. O. Olsen (1973) 'Speech Audiometry,' in J. Jerger, ed., *Modern Developments in Audiology,* Academic Press, New York.
Tobias, J. V. (1959) 'Relative Occurrence of Phonemes in American English,' *J. Acoust. Soc. Am.* 31, 631.
_________ (1964) 'On Phonemic Analysis of Speech Discrimination Tests,' *J. Speech Hear. Res.* 7, 98-100.
Trager, G. L. and H. L. Smith, Jr. (1951) *An Outline of English Structure,* Battenberg Press, Norman, Okl.
Van Dijk, T. A. (1977) *Text and Context: Explorations in the Semantics and*

Pragmatics of Discourse, Longman Group, London.
Van Dijk, T. A. (1980) *Macrostructures: An Interdisciplinary Study of Global Structures in Discourse, Interaction, and Cognition,* Lawrence Erlbaum Associates, Hillsdale, N.J.
Wang, W. S-Y. and J. Crawford (1960) 'Frequency Studies of English Consonants,' *Language and Speech* 3, 131-139.
Webster's New Twentieth Century Dictionary of the English Language, Unabridged, Second Edition, William Collins and World Publishing Co., New York, 1977.
Weinhouse, I. and M. H. Miller (1963) 'Discrimination Scores for Two Lists of Phonetically Balanced Words,' *J. Aud. Res.* 3, 9-14.
Winograd, T. (1980) 'What does it Mean to Understand Language?' *Cognitive Science* 4, 209-241.

Department of English
Ewha Womans University
Daehyundong, Suhdaemunku
Seoul, 120
Korea