RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 21 (1996-1997)
Indiana University

Acoustic, Psychometric and Lexical Neighborhood Properties of the Spondaic Words: A Computational Analysis of Speech Discrimination Scores[1]

Ted A. Meyer,[2] David B. Pisoni, Paul A. Luce,[3] and Robert C. Bilger[4]

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

---

[1] Research supported by NIH/NIDCD Grant DC00111 and Training Grant DC00012. The authors would like to thank Linnette A. Caldwell and Darla J. Sallee for clerical assistance with this project. The authors would like to thank Drs. Mario A. Svirsky, Karen I. Kirk, and Steven B. Chin for their comments on an earlier version of the manuscript.

[2] Indiana University School of Medicine, Department of Otolaryngology, Indianapolis, IN 46202

[3] State University of New York at Buffalo, Department of Psychology, Buffalo, NY 14260

[4] University of Illinois, Department of Speech & Hearing Science, Champaign, IL 61820

Acoustic, Psychometric and Lexical Neighborhood Properties of the Spondaic Words: A Computational Analysis of Speech Discrimination Scores

Abstract. Luce and Pisoni (in press) developed a model of spoken word recognition based on neighborhood probability characteristics of monosyllabic words. This model, known as the Neighborhood Activation Model (NAM), assumes that words are recognized in the context of other similar sounding patterns in the lexicon. NAM predicts that word recognition is dependent on how easy or difficult it is to confuse the stimulus word with phonetically similar words in a lexical neighborhood. Cluff and Luce (1990) demonstrated that spondaic word recognition is also dependent on the confusability of the individual syllables. Spondaic words with two easy to recognize syllables are identified more accurately than spondaic words with two hard to recognize syllables. Recently, Bilger et al. (submitted) analyzed the 36 spondaic words [Tillman et al. (1963) recording of the Hirsh et al. (1952) spondaic words] currently used to determine the speech reception threshold (SRT) (American Speech-Language-Hearing Association, 1988). They found that the thresholds and slopes of psychometric functions for the spondaic words were not equivalent, and that the spondees were not equally intelligible. In this study, we computed the lexical neighborhood characteristics (Pisoni et al., 1985) of the individual syllables in these spondaic words and compared these values to psychophysical and acoustic measures obtained from Bilger et al. Although spondaic word thresholds were not related to any of the lexical neighborhood measures of the words, the slopes of the psychometric functions were negatively correlated with the neighborhood densities of the spondaic words. This finding demonstrates that the rate at which a word becomes intelligible is inversely related to the confusability of the word with its lexical neighbors in memory. Implications for the development of new tests for speech discrimination are discussed.

Introduction

Recent work by Bilger (Bilger, Matthies, Meyer, & Griffiths, submitted; Meyer & Bilger, 1997) has examined the utility of spondaic words (words with two syllables having approximately equal stress) in speech discrimination tests. The spondaic words used to determine thresholds for speech were chosen specifically and modified for the original recording to be of equal intelligibility (Hirsh et al., 1952). In subsequent studies, the thresholds for the individual spondees recorded by Hirsh were not equivalent (Bowling & Elpern, 1961; Curry & Cox, 1966; Wilson & Margolis, 1983). Psychometric functions generated for the individual spondaic words recorded by Tillman (Tillman, Carhart & Wilber, 1963) also clearly demonstrate that the spondaic words are not equally intelligible over a range of presentation levels and that the slopes of psychometric functions for these words are not equivalent (Bilger et al., submitted; Young, Dudley & Gunter, 1982). These findings pose a serious problem for clinicians who routinely derive speech reception thresholds using the spondaic words as stimuli (Penrod, 1994). One possible solution to this problem is to use spondaic words with similar acoustic properties. However, Bilger et al. also found poor correlations between gross acoustic measures (durations and amplitudes) and psychophysical measures (thresholds and slopes of psychometric functions) of the individual spondaic words. Their findings suggest that other structural factors related to acoustic-phonetic discriminability may play a role in determining spondaic word recognition.

For the past ten years, Luce and his colleagues (Luce, 1986; Luce & Pisoni, in press; Luce, Pisoni, & Goldinger, 1990) have been developing a model of spoken word recognition known as the Neighborhood Activation Model (NAM). This model attempts to account for the effects of lexical activation and competition on speech perception and spoken word recognition. In particular, this model assumes that words are recognized relationally in the context of other phonetically similar words in the lexicon. A particular sound pattern activates multiple lexical items in memory and this pattern of activation affects lexical discrimination and subsequent recognition performance.

NAM predicts that when a spoken word is perceived, it activates a set of representations of similar sounding words in memory. This set of phonetically similar words is called a "lexical neighborhood." Once the neighborhood of the stimulus is activated, the word recognition system must then discriminate among the activated neighbors and decide which of the neighbors best matches the stimulus. The lexical discrimination and decision process is influenced by three factors: (1) the frequency of the target word; (2) the number of phonetically similar words (neighbors) activated in memory by the stimulus pattern; and (3) the frequencies of occurrence of the neighbors in the similarity neighborhood. Because of increased competition among activated items in memory, words from densely populated lexical neighborhoods are recognized more slowly and less accurately than words from sparsely populated neighborhoods. In addition, the presence of high frequency neighbors produces increased competition among words that are phonetically similar to the target word, thereby also reducing recognition performance.

According to NAM, identification performance can be predicted by the following rule:

$$p(ID)=\frac{p(Stim)*Freq_s}{p(Stim)*Freq_s+\sum_{j=1}^{n}p(Neighbor_j)*Freq_j}$$

where p(ID) is the probability of correction identification, $p(Stim)*Freq_s$ is the frequency-weighted probability of identifying the stimulus word based on acoustic-phonetic information, and $p(Neighbor_j)*Freq_j$ is the frequency-weighted probability of identifying a neighbor. According to this rule, NAM predicts that stimulus words with few low-frequency neighbors will be identified most accurately (because of the relatively small denominator in Equation 1, which indexes the degree of competition among neighbors activated in memory). NAM also predicts that stimulus words with many high-frequency neighbors will be identified least accurately (large denominator in Equation 1).

Luce and Pisoni (in press) performed numerous regression analyses to evaluate the effects of several variables in NAM on speech recognition scores for monosyllabic words at various signal-to-noise (S/N) ratios. Across all S/N ratios, Luce and Pisoni found that variables from the model explain between 3% and 23% of the variance in the subject response. They also found that as the signal was degraded (lower S/N ratio), the structure of the lexical neighborhood of the stimulus becomes more important (i.e., explained more variance). The relation between S/N ratio and the lexical neighborhood of the stimulus may be of particular importance in examining the signals used in speech discrimination tests in which the stimuli are presented at or near threshold.

In examining how NAM applies to more longer and more complex words, Cluff and Luce (1990) assessed the contribution of the individual syllables to spondaic word recognition. Cluff and Luce categorized the individual syllables of the spondaic words used in their experiment according to the lexical neighborhood properties of the syllables. An "easy" syllable was a high-frequency word in a low-frequency, low-density neighborhood, and a "hard" syllable was a low-frequency word in a high-frequency, high-density neighborhood. The frequencies of the words were obtained from the Kucera and Francis (1967) word counts. Cluff and Luce found that spondees with an "easy-easy" syllable pattern were the easiest to identify, whereas spondees with a "hard-hard" syllable pattern were the most difficult to identify. Identification of spondees with a single "easy" and a single "hard" syllable was of intermediate difficulty. Further, spondees with a "hard-easy" syllable pattern were easier to identify than spondees with an "easy-hard" syllable pattern. Cluff and Luce attributed this asymmetry in performance primarily to a "retroactive" pattern of influence in spoken word recognition, a "hard" second syllable influenced perception more than an "easy" second syllable. In other words, compound words in English are not recognized strictly left to right in serial order.

In addition to providing a better fundamental understanding of how listeners with normal hearing are able to recognize spoken words, the fundamental assumptions of NAM also have several important clinical implications in terms of measuring speech perception in listeners with hearing impairments. For example, Kirk and her colleagues (Kirk, Pisoni, & Osberger, 1995; Kirk, Pisoni, Sommers, Young, & Evanson, 1995) have recently developed a new class of speech recognition tests based on assumptions of NAM for subjects with impaired hearing and cochlear implants (CIs). Their findings show that hearing-impaired listeners and patients with CIs also recognize words in the context of other phonetically similar words in their lexicons. And, moreover, they recognize words by reference to "similarity neighborhoods" through a two-step process of bottom-up "activation" followed by "lexical discrimination," as assumed by NAM (see Luce & Pisoni, in press). Differences in speech perception performance between listeners with normal hearing and impaired hearing can be accounted for by NAM which makes specific predictions about the effect of frequency and acoustic-phonetic similarity and how these factors influence recognition performance.

In this paper, we report the results of several analyses that compared the acoustic and psychophysical measurements on the Tillman et al. (1963) recording of spondaic words (Bilger et al., submitted; Meyer & Bilger, 1997) to the lexical properties of the individual syllables of the spondaic words. We also examine the relation between threshold measures of intelligibility (Bowling & Elpern, 1961; Curry & Cox, 1966; Wilson & Margolis, 1983) of the Hirsh et al. (1952) recording of the spondaic words and the lexical neighborhood properties of the individual syllables of the words. Using computational analyses of these sound patterns, we hoped to gain a better understanding of the structural factors that influence spondaic word recognition near threshold.

Methods

Thresholds of Intelligibility

Hirsh et al. (1952) recorded 36 spondaic words to be used as stimuli for obtaining speech reception thresholds (SRTs). The words were intended to be of equal stress on the two syllables, and of equal intelligibility. Bowling and Elpern (1961) and Curry and Cox (1966) measured the average sensation level at which the individual words first became intelligible to a large group of normal-hearing listeners. Both studies determined that the spondaic words were not of equal intelligibility and suggested that smaller "more homogeneous" sets of spondaic words should be used for determining SRTs. Bowling and Elpern (1961) suggested a shorter list of 23 spondees, whereas Curry and Cox suggested that 27 of their words were "equally intelligible." In these two studies, the words in the "equally intelligible" lists had thresholds that were within one standard deviation of the mean threshold of the entire list of spondees. Wilson and

Margolis (1983) constructed a separate list of 21 "equally intelligible" spondaic words by averaging the results from Bowling and Elpern and Curry and Cox and adding a correction factor to account for the different presentation levels of the different words. In the Wilson and Margolis study, the words in the "equally intelligible" list also had thresholds that were within one standard deviation of the mean threshold of the entire list of spondees.

In the present study, we examined the relation between the individual word thresholds from the entire spondaic word lists from the three studies mentioned above and the lexical neighborhood properties of the words. As was stated earlier, NAM predicts that the lexical neighborhood of a word is of greater importance to word recognition as the S/N ratio becomes more degraded. Our goal, in this study, was to determine whether any of the lexical neighborhood properties of the words might play a role in predicting the performance of subjects in identifying a word when it first becomes intelligible. In other words, do "easy" words (high-frequency words from sparse neighborhoods) have lower thresholds than "hard" words (low-frequency words from dense neighborhoods)?

$$P(C)=\frac{1}{e^{(-mX_i-b)}}$$

Psychometric Functions

Young et al. (1982) generated psychometric functions for the individual spondaic words (Tillman et al., 1963 recording). They fit lines to the middle (linear) portion of the psychometric functions to estimate the thresholds (50% points) and slopes of the functions. Bilger et al. (submitted) also generated psychometric functions for the spondaic words from the data of Young et al. (1982), but in a slightly different manner. In Bilger et al., the data were fitted to the logit function described as follows:

in which *P(C)* is the probability of a correct response, and *m* is the slope of the ogive at its midpoint, *b* [P(C) = .50] (Lord, 1980). This function was fit to a least squares criterion using the logit transformation of proportion of correct responses as one variable in the calculation of a linear regression with intensity level. In the present study, the thresholds [P(C) = .50] and slopes of the psychometric functions of the individual spondaic words estimated by Bilger et al. (submitted) were compared to acoustic and lexical neighborhood measures of the individual syllables of the words.

Acoustic Measures

Bilger et al. (submitted) also measured acoustic parameters of the spondaic words. A tape-recorded version of the Tillman et al. (1963) recording of the spondaic words from CID W-1 (Hirsh et al., 1952) was digitized with 16-bit intensity resolution at a rate of 10-kHz (Data Translation, Model 2823). To eliminate artifacts, the stimuli were played through a bandpass filter (100-4000 Hz) (Wavetek-Rockland, Model 751-A) prior to the A/D conversion. The digitized spondees were then analyzed using Interactive Laboratory System (ILS-PC) software. The measures obtained were the RMS amplitude of each syllable and the overall RMS amplitude of each spondee (in dBV), as well as the duration of each syllable and the total duration of each spondee (in ms). These values were compared to the psychometric and lexical neighborhood measures of the individual syllables of the spondaic words.

Lexical Neighborhood Measures

Computational analyses of the individual spondaic words were carried out using a computerized database. The lexical database consists of 20,000 words from the *Webster's Pocket Dictionary*. Details of the development of the database are described elsewhere (Luce, 1986; Luce & Pisoni, in press; Nusbaum, Pisoni, & Davis, 1984; Pisoni, Nusbaum, Luce, & Slowiaczek, 1985). The word frequencies were obtained from Kucera and Francis (1967). Lexical neighborhoods for a given word were computed by using a single-phoneme substitution, addition, or deletion rule (Greenberg & Jenkins, 1964; Landauer & Streeter, 1973). For example, if the target word is "art," the lexical neighborhood of "art" would include words such as "ark" (substitution), "dart" (addition), and "are" (deletion). In the present study, we obtained the frequency of occurrence for the individual syllables of the spondaic words as well as the frequency of occurrence and density of the lexical neighborhoods of the syllables. The obtained values for the two syllables were then averaged to obtain a measure for the entire spondaic word. These values were compared to the psychometric and acoustic properties of the spondees.

Statistical Analysis

The SigmaStat statistical software package was used to perform a multiple correlational analysis between the psychometric, acoustical, and lexical neighborhood properties of the spondaic words. The results are reported below.

Results

Thresholds of Intelligibility

The thresholds obtained from the Hirsh et al. (1952) recording of the spondaic words (Bowling & Elpern, 1961; Curry & Cox, 1966; Wilson & Margolis, 1983) are shown in Table I. For the three different studies, the range of the levels at which the individual spondaic words became intelligible was quite large (10.0 dB - BE, 8.1 dB - CC, and 7.6 dB - WM) suggesting that the individual spondaic words are not equally intelligible. The correlations between the spondaic word thresholds obtained from the three studies are quite high (BE x CC, $\underline{r} = .769$, $\underline{p} < .001$; BE x WM, $\underline{r} = .763$, $\underline{p} < .001$; CC x WM, $\underline{r} = .611$, $\underline{p} < .001$) (see Table IV) and suggest that the word thresholds are strongly related. Although the correlations between the levels at which the spondaic words first become intelligible in the three studies are significantly greater than zero, the three "homogeneous lists" generated by the three studies described earlier cannot be considered "equivalent" lists of spondees (for complete details see Bilger et al., submitted). The three short lists appear quite similar, but appearances can be deceptive. The appropriate statistic to describe the similarity between two word lists (nominal variables) is the contingency coefficient r (Stevens, 1951; Welkowitz, Ewen, & Cohen, 1991). Only 20 of the words in the 23-word list generated by Bowling and Elpern appear in the 27-word list of Curry and Cox ($\underline{r} = .367$, $\underline{p} = .0275$), only 17 of Bowling and Elpern's 23 "homogeneous" words appear in the 21-word list of Wilson and Margolis ($\underline{r} = .420$, $\underline{p} = .012$), while only 17 of the words in the Curry and Cox list appear in the Wilson and Margolis list ($\underline{r} = .293$, $\underline{p} = .0790$). Although the contingency correlations between the BE and CC and the BE and WM lists are significantly greater than 0.0 ($\underline{p} < .05$), the values of r are also well below any reasonable criterion for equivalence. In addition, only 14 of the 36 spondees appear on all three "homogeneous lists" (see Table I). Thus, the individual word thresholds are more closely related than the "homogeneous" lists created from these thresholds. The relation between the individual word thresholds and the lexical neighborhood values will be discussed later.

Table I.

Spondaic word thresholds of intelligibility (in dB re the lowest level at which a word was intelligible) from Bowling and Elpern (1961), Curry and Cox (1966), and Wilson and Margolis (1983) (Hirsh as talker). Superscripts after the words indicate inclusion in a "homogeneous" list, 1-BE, 2-CC, 3-WM.

Word	BE	CC	WM
AIRPLANE^{1,2}	4.7	4.2	3.06
ARMCHAIR^{1,2,3}	7.5	4.4	8.25
BASEBALL	3.0	3.0	4.18
BIRTHDAY^{1,2}	6.3	5.6	3.87
COWBOY^{1,2,3}	4.3	5.0	5.13
DAYBREAK^{2}	8.3	6.6	9.20
DOORMAT^{2}	8.3	6.8	10.61
DRAWBRIDGE^{2,3}	8.2	4.8	6.43
DUCKPOND^{3}	8.3	10.1	4.78
EARDRUM1^{2,3}	6.9	6.1	6.70
FAREWELL^{1,2,3}	6.3	8.0	7.03
GRANDSON	12.2	8.4	10.62
GREYHOUND^{1,2,3}	7.0	6.0	7.12
HARDWARE^{2,3}	3.5	4.9	4.54
HEADLIGHT	8.3	8.6	9.52
HORSESHOE^{1,3}	7.8	8.6	7.07
HOTDOG^{1}	5.0	3.8	2.98
HOTHOUSE	12.7	10.9	9.87
ICEBERG^{1,2}	4.8	4.0	3.01
INKWELL^{1,2,3}	8.8	7.0	7.52
MOUSETRAP^{1,2,3}	7.7	7.7	7.95
MUSHROOM^{2}	10.3	7.7	9.81
NORTHWEST^{1,2}	5.2	6.2	3.86
OATMEAL^{1,2,3}	7.0	7.6	5.82
PADLOCK^{2}	8.3	8.0	8.93
PANCAKE^{1,3}	7.7	8.0	7.32
PLAYGROUND^{1,2}	4.5	4.6	3.91
RAILROAD^{1,2,3}	4.8	6.5	6.02
SCHOOLBOY^{2,3}	8.5	6.9	6.95
SIDEWALK^{1,2,3}	6.0	6.5	5.31
STAIRWAY^{1,2,3}	6.3	8.0	8.71
SUNSET^{1,2,3}	6.2	6.2	5.32
TOOTHBRUSH^{1,2,3}	7.5	8.0	7.35
WHITEWASH^{1,2,3}	7.2	7.8	5.19
WOODWORK^{1,2,3}	4.3	4.8	7.66
WORKSHOP	2.7	2.8	3.82
Mean	6.84	6.50	6.54
S. D.	2.27	1.91	2.25

Psychometric Functions

Midpoints (thresholds in dB SPL) and slopes (in %/dB) of the psychometric functions from the Bilger et al. (submitted) analysis of the spondaic words from Young et al. (1982) are shown in Table II. Thresholds ranged from 15.0 to 21.6 dB SPL with a mean of 18.9 dB SPL. The slopes of the 36 psychometric functions ranged from 6.8 to 16.9 %/dB with a mean of 10.7 %/dB. Although the spondees are supposed to be "equally intelligible," the average thresholds span a range of nearly 7 dB, and the rate at which the words become intelligible varies over a range of 10%/dB. Below we examine whether the acoustic measures of the words relate to these psychometric measures and determine whether NAM plays a role in the identification of the spondees at threshold.

Acoustic Measures

Acoustic measures of overall amplitude and duration of the spondaic words are also shown in Table II. The duration of the spondaic words ranged from 723 to 992 ms with a mean of 847 ms. The mean duration of the first syllable was shorter (315 ms) than the mean duration of the second syllable (486 ms), and 18 of the 36 spondees have silent periods (closures) between the syllables in the Tillman recording. The RMS levels of the spondaic words ranged from -13.5 to -11.3 dBV with a mean of -12.3 dBV. Although shorter in duration, the first syllables had a greater mean amplitude (-11.2 dBV) than the second syllables (-12.7 dBV). The small ranges and standard deviations of the amplitudes of the individual syllables as well as the entire words suggest that by monitoring their utterances with a VU meter, Tillman et al. were able to produce a list of spondees that were quite similar in overall amplitude. We will examine how these acoustic measures relate to psychometric and the lexical neighborhood characteristics of the spondaic words.

Lexical Neighborhood Measures

Lexical neighborhood measures for the 36 spondaic words based on computational analyses of the similarity spaces from the lexical database are shown in Table III. Included in Table III are the frequency of occurrence for the individual syllables as well as the frequency of occurrence of the entire word. The frequency of occurrence of the words in the individual syllable's lexical neighborhood as well as the density of the individual syllable's lexical neighborhood are also included. As described above, values for the spondaic words were generated by computing the arithmetic average of the measures for the individual syllables. Several of the individual syllables that make up the spondees are "easy" words, others are "hard" words, and still others are of intermediate difficulty. That is, some words come from lexical neighborhoods where there is little competition from other phonetically similar words (i.e., "easy" words), whereas other words come from lexical neighborhoods where there is much more competition from other phonetically similar words (i.e., "hard" words). As shown in Table III, there is a great deal of variability in the lexical neighborhood measures for the different spondees.

Statistical Analysis

A multiple correlational analysis was performed to assess the relations between the different lexical neighborhood, acoustic, and psychophysical measures of the spondaic words described above. The correlations are shown in Table IV. Several interesting findings emerged from the correlational analysis. First, the spondaic word thresholds (levels at which the words first became intelligible) obtained from the studies with Hirsh as talker (Bowling & Elpern, 1961; Curry & Cox, 1966; Wilson & Margolis, 1983) were not significantly correlated with any of the lexical neighborhood measures for the individual syllables

Table II.

Midpoints [dB SPL for a P(C)=50%] and slopes (%/dB) from two-parameter logistic model fitted to the data of Young et al. (1982) (Tillman as talker), and word duration and RMS levels (from Bilger et al., submitted). 1st Syl - first syllable, 2nd Syl - second syllable, dBV = dB re 1V.

WORD	Midpoint [P(C) = 50%] (dB SPL)	Slope (%/dB) Total	Duration			RMS level		
			1st Syl (ms)	2nd Syl (ms)	Total	1st Syl (dBV)	2nd Syl (dBV)	(dBV)
AIRPLANE	18.1	10.8	858	262	506	-12.4	-10.7	-12.8
ARMCHAIR	17.7	9.8	858	378	461	-11.8	-11.8	-11.8
BASEBALL	17.4	11.5	858	307	416	-12.6	-12.1	-11.6
BIRTHDAY	18.8	9.0	813	243	416	-11.3	- 9.9	-10.8
COWBOY	18.0	11.8	723	307	416	-11.5	-11.2	-11.8
DAYBREAK	20.7	8.6	666	378	288	-11.6	-11.7	-11.4
DOORMAT	19.7	13.2	640	352	288	-11.5	-10.9	-12.4
DRAWBRIDGE	18.5	12.1	884	378	506	-12.5	-11.5	-13.4
DUCKPOND	18.9	10.4	896	154	550	-12.7	- 9.0	-12.8
EARDRUM	17.7	7.0	884	378	506	-12.4	-11.9	-12.9
FAREWELL	20.8	7.7	922	397	525	-11.7	-11.1	-12.2
GRANDSON	19.2	9.2	940	442	416	-11.6	-10.1	-11.9
GREYHOUND	21.7	9.1	992	397	595	-12.9	-12.4	-13.3
HARDWARE	17.3	11.6	819	333	486	-12.7	-12.7	-12.7
HEADLIGHT	18.6	7.8	768	243	525	-12.7	-10.5	-14.3
HORSESHOE	17.5	13.0	813	307	506	-12.2	-11.3	-12.9
HOTDOG	15.0	10.9	896	198	506	-12.8	- 9.9	-12.9
HOTHOUSE	18.4	10.1	858	243	570	-13.3	-10.7	-14.6
ICEBERG	16.8	15.0	940	333	525	-12.7	-12.2	-12.4
INKWELL	18.5	6.8	877	243	570	-11.8	-10.3	-12.0
MOUSETRAP	20.3	9.5	896	378	461	-12.5	-11.6	-12.7
MUSHROOM	20.0	7.8	877	333	506	-12.6	-11.9	-12.8
NORTHWEST	18.8	10.8	941	307	550	-13.1	-11.4	-13.6
OATMEAL	21.6	10.4	813	198	486	-11.7	- 8.8	-12.4
PADLOCK	19.7	10.3	793	307	486	-12.3	-11.4	-13.2
PANCAKE	20.1	9.3	858	378	461	-12.3	-11.5	-12.9
PLAYGROUND	17.8	9.9	973	378	595	-12.6	-11.9	-13.1
RAILROAD	19.2	8.9	948	442	506	-11.7	-11.6	-11.9
SCHOOLBOY	20.3	16.9	896	480	416	-12.4	-12.7	-12.1
SIDEWALK	19.4	11.8	793	307	486	-12.3	-11.4	-13.0
STAIRWAY	18.8	10.0	839	378	461	-11.8	-11.4	-12.2
SUNSET	21.3	12.4	813	352	461	-12.1	-11.2	-13.0
TOOTHBRUSH	18.9	14.9	723	134	486	-13.5	-12.4	-13.0
WHITEWASH	16.5	11.6	857	224	525	-12.8	-10.1	-13.4
WOODWORK	19.2	11.9	768	262	506	-12.9	-11.3	-14.1
WORKSHOP	17.9	12.8	813	224	525	-12.7	-10.2	-13.8
Mean	18.86	10.68	847	315	486	-12.3	-11.2	-12.7
S. D.	1.47	2.25	80	82	68	.55	.94	.82

Table III.

Lexical neighborhood properties of the spondaic words. Word values were obtained by averaging the word frequencies and neighborhood properties of the individual syllables. Syl1 -1st syllable, Syl2 -2nd syllable, Freq -frequency, Nhbd -neighborhood, Den -density.

Word	Mean Word Freq	Syl1 Freq	Syl2 Freq	Nhbd Freq	Syl1 Nhbd Freq	Syl2 Nhbd Freq	Nhbd Den	Syl1 Nhbd Den	Syl2 Nhbd Den
AIRPLANE	214	266	162	337.02	560.83	113.20	20.0	30.0	10.0
ARMCHAIR	80	94	66	443.09	559.22	326.95	14.5	9.0	20.0
BASEBALL	101	91	110	105.69	41.63	169.75	25.5	27.0	24.0
BIRTHDAY	378	70	686	197.63	67.56	327.70	21.5	16.0	27.0
COWBOY	135	29	242	523.53	121.37	925.69	16.0	19.0	13.0
DAYBREAK	388	686	90	171.74	327.70	15.77	20.0	27.0	13.0
DOORMAT	160	312	8	325.02	14.00	636.03	21.5	13.0	30.0
DRAWBRIDGE	77	56	98	37.75	42.00	33.50	6.0	6.0	6.0
DUCKPOND	17	9	25	27.58	41.88	13.29	16.0	25.0	7.0
EARDRUM	20	29	11	1111.19	1774.39	448.00	20.5	31.0	10.0
FAREWELL	491	84	897	359.81	614.43	105.19	29.5	28.0	31.0
GRANDSON	163	48	278	172.00	37.57	306.42	16.5	7.0	26.0
GREYHOUND	43	80	7	104.72	57.56	151.89	13.5	18.0	9.0
HARDWARE	121	202	39	191.60	22.44	360.76	21.5	18.0	25.0
HEADLIGHT	379	424	333	255.04	354.80	155.29	30.0	25.0	35.0
HORSESHOE	68	122	14	889.80	76.27	1703.33	17.5	11.0	24.0
HOTDOG	103	130	75	147.03	282.18	11.88	18.0	28.0	8.0
HOTHOUSE	361	130	591	202.16	282.18	122.14	17.5	28.0	7.0
ICEBERG	23	45	1	219.00	395.31	42.69	14.5	16.0	13.0
INKWELL	452	7	897	72.24	39.29	105.19	22.5	14.0	31.0
MOUSETRAP	15	10	20	46.64	79.43	13.85	13.5	14.0	13.0
MUSHROOM	193	1	384	43.22	67.40	19.04	19.0	15.0	23.0
NORTHWEST	221	206	235	81.88	87.75	76.00	11.5	4.0	19.0
OATMEAL	15	1	30	479.79	882.36	77.21	26.5	25.0	28.0
PADLOCK	15	8	23	150.36	225.12	75.61	28.5	26.0	31.0
PANCAKE	15	16	13	202.81	301.35	104.27	28.5	31.0	26.0
PLAYGROUND	193	200	186	107.11	75.63	138.60	10.5	16.0	5.0
RAILROAD	127	16	237	34.32	31.39	37.24	32.5	36.0	29.0
SCHOOLBOY	367	492	242	474.22	22.75	925.69	10.5	8.0	13.0
SIDEWALK	240	380	100	113.21	122.35	104.07	19.0	23.0	15.0
STAIRWAY	465	16	913	232.21	27.75	436.67	21.0	12.0	30.0
SUNSET	346	278	414	228.85	306.42	151.28	29.0	26.0	32.0
TOOTHBRUSH	32	20	44	1031.83	2057.86	5.80	9.5	14.0	5.0
WHITEWASH	201	365	37	183.64	302.27	65.00	9.0	11.0	7.0
WOODWORK	1765	2769	760	149.93	252.87	47.00	17.5	15.0	20.0
WORKSHOP	411	760	63	43.69	47.00	40.38	18.0	20.0	16.0
MEAN	233	235	231	263.82	294.51	233.12	19.1	19.2	18.9
S. D.	303	475	281	265.33	449.27	342.83	6.6	8.3	9.5

Table IV.

Correlation matrix between acoustical, psychophysical and lexical neighborhood variables (~ $p < .05$, * $p < .01$). wd - word, s1 - 1st syllable, s2 - 2nd syllable, nd - neighborhood, dur - duration, freq - frequency, den - density, rms - root mean square level, BE - Bowling & Elpern, CC - Curry & Cox, WM - Wilson & Margolis, Bilg - Bilger et al.

	BE	CC	WM	mean Bilg	slope Bilg	wd dur	s1 dur	s2 dur	wd rms	s1 rms	s2 rms	wd freq	s1 freq	s2 freq	nd freq	s1 nd freq	s2 nd freq	nd den	s1 nd den	s2 nd den
BE																				
CC	.769*																			
WM	.763*	.611*																		
mean Bilg		.387~	.440*																	
slope Bilg																				
wd dur																				
s1 dur																				
s2 dur						.649*														
wd rms							.388~													
s1 rms							-.566*													
s2 rms						-.576*		-.578*	.755*											
wd freq																				
s1 freq										-.338~										
s2 freq					-.335~															
wd nd freq																				
s1 nd freq																				
s2 nd freq																				
wd nd den					-.427*				.433*											
s1 nd den					-.348~															
s2 nd den									.560*					.406~						

or with those generated for the words. Second, the thresholds (levels at which the words were intelligible 50% of the time) of the psychometric functions generated by Bilger et al. (submitted) from Young et al.'s (1982) data with Tillman as talker were also not significantly correlated with any of the lexical neighborhood measures. Although NAM predicts that easy words (high-frequency words from sparse lexical neighborhoods) should be easier to recognize than hard words (low-frequency words from dense lexical neighborhoods), this relation does not appear to hold for the spondaic words presented at threshold (defined as either level of initial recognition or the 50% point on a psychometric curve). Third, none of the acoustic measures (RMS level, duration) of the individual syllables or the entire spondaic words were significantly related to the thresholds or slopes of the psychometric functions of the individual words. However, two significant correlations between the acoustic measures of the spondaic words and their lexical neighborhood characteristics emerged from this analysis. First, the overall RMS level of the word was positively correlated to the neighborhood density of the word ($r = +.433$, $p < .01$), and second, the RMS level of the first syllable was negatively correlated to the frequency of the first syllable ($r = -.388$, $p < .05$). These may represent random occurrences, because it is difficult to imagine any systematic relation between these measures especially given that the RMS measures were obtained from single utterances of the spondaic words.

Finally, the slopes of the psychometric functions generated by Bilger et al. were significantly correlated with the lexical neighborhood densities of the first syllable of the spondees ($r = -.348$, $p < .05$) as well as the lexical neighborhood densities generated for the whole spondaic words ($r = -.427$, $p < .01$). These negative correlations suggest that the rate at which a spondaic word (target) becomes recognized is inversely related to the number of words that are phonetically similar to the target. That is, words from dense neighborhoods are recognized at a slower rate than words from sparse neighborhoods. This finding would be anticipated based on the assumptions of NAM. From Equation 1, the probability of correctly identifying a stimulus is inversely related to the density (i.e., similarity) of the lexical neighborhood of the stimulus. If a stimulus lies in a dense lexical neighborhood, the denominator of Equation 1 is relatively large compared to the denominator when the stimulus lies in a sparse lexical neighborhood. Thus, the predicted change in the ability of a subject to correctly identify a stimulus (slope of a psychometric function) should be less when the stimulus is from a dense neighborhood (large denominator) than when the stimulus is from a sparse neighborhood (small denominator). This is precisely the result we observed here.

Discussion

Spoken language is infinitely variable. Each repetition of a word by an individual talker is a unique representation of speech, yet even under listening conditions that are far from ideal, we are able to recognize familiar words given that they are represented in our mental lexicons. Although the precise neural representation of how sound patterns of words are stored in one's mental lexicon is not known, and is currently a topic of great interest, the lexicon plays an important role in speech perception and spoken word recognition (Luce et al., 1990; Miller, 1946; Treisman, 1978a,b). In identifying an incoming stimulus, we compare that stimulus to a multidimensional representation of words stored in a psychological space. Several recent models of spoken word recognition have attempted to explain the process of word recognition. In general, the models compare information from the incoming stimulus to information stored in a subject's mental lexicon (see Lively, Pisoni, & Goldinger, 1994, for a review).

NAM is a model of spoken word recognition containing both bottom-up and top-down processing components. As the incoming stimulus is processed, a set of phonetically similar words, a "lexical neighborhood," or set of word hypotheses is activated in memory. The density and frequency of the words in a lexical neighborhood have been shown to produce strong effects on the identification of the stimulus word (Luce et al., 1990). Furthermore, the lexicon may be activated and comparisons made to the incoming stimulus multiple times depending on the complexity of the stimulus. For example, in a series of experiments in which the stimuli were spondaic words, Cluff and Luce (1990) and Charles-Luce, Luce, and Cluff (1990) demonstrated that word recognition was dependent upon the principles of multiple activation and delayed commitment. They found that the lexical difficulty of the second syllable influenced the identification of the spondaic words (retroactive influence of the second syllable). That is, words with easy second syllables were identified a greater percentage of the time than words with hard second syllables. Also, words with a "hard-easy" syllable pattern were more easily recognized than words with an "easy-hard" syllable pattern. Their findings also showed that the "easy" second syllable allowed the "hard" first syllable to be recognized more readily, whereas the "hard" second syllable interfered with the recognition of the "easy" first syllable.

In addition to the contribution of the lexicon on speech perception, the specific information processing task used to measure speech discrimination has been shown to have a substantial impact on a listener's performance. For example, open-set and closed-set tasks place very different demands on the listener. Closed-set tests are essentially tests of speech pattern discrimination, whereas open-set tests involve activation, search and retrieval of a lexical representation from memory before a response can be made. Recently, Sommers, Kirk, and Pisoni (1997) found that open-set word recognition was more difficult for all their subjects when the stimuli were produced with multiple talkers instead of a single talker. Open-set recognition scores were also lower when the words were lexically "hard" than when the words were lexically "easy". In contrast, closed-set word recognition did not show effects of talker variability or lexical difficulty when words in the set were chosen to be easily confused with the target words. Sommers et al. concluded that closed-set tests may not adequately simulate the cognitive demands that individuals face in everyday listening situations. Furthermore, speech discrimination scores obtained using closed-set tests are not equivalent to scores obtained using open-set tests and therefore may not be able to predict speech perception performance in real-world conversational situations. In closed-set tests, where listeners simply have to discriminate differences among sound patterns, rather than recognize or identify words, the set of potential responses is no longer determined by the structure of the mental lexicon (Black, 1957). Listeners can restrict their lexical search and decision strategies to the available response alternatives provided in the forced-choice set instead of having to search through phonetically similar words in memory. In contrast, in open-set speech discrimination tests, listeners must compare the incoming speech signal to patterns stored in long-term lexical memory. Thus, the organization of words within the mental lexicon can have a significant influence on speech perception and spoken word recognition performance.

Summary and Conclusions

In the present study, we used a correlational analysis to examine relations between the acoustic and psychophysical measures of spondaic words and the lexical neighborhood properties of the individual syllables of these words. The results revealed several interesting findings. First, we found that the slopes of the psychometric functions generated by Bilger et al. (submitted) from the data of Young et al. (1982) were inversely related to the neighborhood densities of the spondaic words. Words from sparse lexical neighborhoods where there is little competition from other phonetically similar words become intelligible at a faster rate than words from dense neighborhoods where there is much greater competition. This pattern of results supports a major prediction of NAM -- that spoken words are recognized in relation to other words in the lexicon that have similar acoustic-phonetic properties.

Second, neither word frequency, neighborhood frequency, neighborhood density, nor any acoustical measures were significantly correlated with the spondaic word thresholds. This finding does not directly support the assumptions of NAM, however, the failure of the lexical neighborhood measures to correlate with spondaic thresholds may indicate that there is simply too little useful acoustic-phonetic information in the signal at threshold to strongly engage lexical representations in the recognition process. Performance at threshold may be more indicative of subjects' off-line guessing strategies based on very impoverished input than of activated lexical representations in memory. Another possibility is that identification of stimuli at threshold levels is almost entirely driven by the minimal amount of acoustic-phonetic information available to the listener. Because the present method of calculating similarity neighborhoods fails to take into account all but categorical changes in phonemes, this metric may fail to capture subtle phonemic confusions that are primarily responsible for spondee identification at threshold (see Luce and Pisoni, in press).

Another explanation for the lack of correlation for the thresholds may simply be that there is little variance to account for in the first place. Indeed, the standard deviation of the thresholds was only 1.47 dB (Bilger et al. psychometric analysis in Table II), which is much smaller than the standard deviation for the slopes (2.25 %/dB), which did produce significant correlations with the lexical variables. Whatever the precise explanation, Luce and Pisoni (in press) obtained similar results for words embedded in noise over a range of S/N ratios: Correlations between the lexical variables and identification accuracy were quite low at the lowest S/N ratios, and lexical effects were only found at signal levels above threshold.

The data from the present study and other recent investigations of the role of the lexicon in speech perception suggest a need for new theoretically-based tests to assess SRTs and other measures of speech recognition in clinical populations. Recent findings suggest that different populations such as the hearing impaired or patients with CIs may not identify spoken words in the same manner as listeners with normal hearing (see Koch, Carrell, Tremblay, & Kraus, 1996). The peripheral encoding of the speech signal is certainly very different for cochlear implant users than for normal-hearing subjects. And, it is very likely that these patients have different lexical representations of words as well. Different groups of subjects may use different equivalence classes than normal-hearing listeners do depending on the nature of their hearing impairments. These differences in their perceived similarity spaces may prove to be an important set of dimensions in assessing how cochlear implants and hearing aids work, and in measuring changes in speech discrimination performance over time using different aural rehabilitation and perceptual training methods.

New speech discrimination tests will need to embrace theoretical concepts and assumptions that specify and describe how normal-hearing listeners recognize spoken words and how they access information about the sound patterns of words from representations in long-term lexical memory (see Luce & Pisoni, in press). The process of "lexical discrimination," that is, how listeners recognize and identify spoken words is not simply equivalent to phoneme or nonsense syllable perception (Bilger & Wang, 1976; Miller & Nicely, 1955; Tyler & Moore, 1992; Wang & Bilger, 1973) where the primary emphasis is on speech feature recognition (see Rabinowitz, Eddington, Delhorne, & Cuneo, 1992). In lexical discrimination tests, listeners are asked to identify a sound pattern as a unique entry from among the words of their language. Successful completion of this task requires not only access to the sensory-based acoustic-phonetic information in the speech waveform but also retrieval of other structural information about the relationship of the sound pattern of the target word to phonetically similar words in the listener's lexicon. According to this approach, spoken words are recognized in the context of other similar words that the listener knows in his/her language. The present findings demonstrate that the large degree of variability typically observed among different words used in speech discrimination tests derives, in part, from the "lexical" and structural properties of the sound patterns of words, not only their acoustic features or correlates. These results therefore provide additional support for several of the major theoretical assumptions of the Neighborhood Activation Model of spoken word recognition. Furthermore, these results suggest a clinically useful line of research for future studies of hearing-impaired listeners where the stimulus materials are constructed in a theoretically motivated way based on what is currently known about the structural acoustic-phonetic properties of spoken words and how they are recognized by normal-hearing listeners. The findings from the present analysis demonstrate that listeners use information about the structural patterns of words in their lexicons to perceive speech over a wide range of S/N ratios. Moreover, the rate at which information about words is encoded or perceived in open-set speech intelligibility or discrimination tests is inversely related to the acoustic-phonetic confusability of the patterns in lexical memory. This important new finding is predicted by NAM which assumes that spoken words are recognized *relationally* in the context of other phonetically similar words in the language.

References

American Speech-Language-Hearing Association. (1988). Guidelines for determining threshold level for speech. *Asha, 30*, 85-88.

Bilger, R. C., Matthies, M. L., Meyer, T. A., & Griffiths, S. K. (submitted). Psychometric equivalence of recorded spondaic words as test items. *Journal of Speech-Language-Hearing Research.*

Bilger, R. C., & Wang, M. D. (1976). Consonant confusions in patients with sensorineural hearing loss. *Journal of Speech and Hearing Research, 19*, 718-748.

Black, J. W. (1957). Multiple choice intelligibility tests. *Journal of Speech and Hearing Disorders, 22,* 213-235.

Bowling, L. S., & Elpern, B. S. (1961). Relative intelligibility of items on CID Auditory Test W-1. *Journal of Auditory Research, 1*, 152-157.

Charles-Luce, J., Luce, P. A., & Cluff, M. S. (1990). Retroactive influence of syllable neighborhoods. In G. T. M. Altmann, (Ed.), *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives* (pp. 173-184). Cambridge: MIT Press.

Cluff, M. S. & Luce, P. A. (1990). Similarity neighborhoods of spoken two-syllable words: Retroactive effects on multiple activation. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 551-563.

Curry, E. T., & Cox, B. P. (1966). The relative intelligibility of spondees. *Journal of Auditory Research, 6,* 419-424.

Greenberg, J. H., & Jenkins, J. J. (1964). Studies in the psychological correlates of the sound system of American English. *Word, 20,* 157-177.

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

Kirk, K. I., Pisoni, D. B., & Osberger, M. J. (1995). Lexical effects on spoken word recognition by pediatric cochlear implant users. *Ear & Hearing, 16*, 470-481.

Kirk, K. I., Pisoni, D. B., Sommers, M. S., Young, M., & Evanson, C. (1995). New directions for assessing speech perception in persons with sensory aids. *Annals of Otology, Rhinology, and Laryngology, 104* (Suppl. 166), 300-303.

Koch, D. B., Carrell, T. D., Tremblay, K., & Kraus, N. (1996). Perception of synthetic syllables by cochlear-implant users: Relation to other measures of speech perception. *Association for Research in Otolaryngology Abstracts*.

Kucera, F., & Francis, W. (1967). *Computational analysis of present day American English*. Providence, RI: Brown University Press.

Landauer, T. K., & Streeter, L. A. (1983). Structural differences between common and rare words: failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning and Verbal Behavior, 12,* 119-131.

Lively, S. E., Pisoni, D. B., & Goldinger, S. D. (1994). Spoken word recognition: Research and theory. In M. Gernsbacker (Ed.), *Handbook of Linguistics* (pp. 265-301). New York: Academic Press.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Luce, P. A. (1986). *Neighborhoods of words in the mental lexicon.* (Research on Speech Perception Technical Report No. 7). Bloomington, IN: Speech Research Laboratory, Indiana University.

Luce, P. A., & Pisoni, D. B. (in press). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*.

Luce, P. A., Pisoni, D. B., & Goldinger, S. D. (1990). Similarity neighborhoods of spoken words. In G. T. M. Altmann, (Ed.), *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives* (pp. 122-147). Cambridge, MA: MIT Press.

Meyer, T. A. & Bilger, R. C. (1997). Effect of set size and method on speech-reception thresholds in noise. *Ear and Hearing, 18,* 202-209.

Miller, G. A. (1946). Some characteristics of human speech. In S. S. Stevens, (Ed.), *Transmission and Reception of Sounds Under Combat Conditions* (pp. 58-68). Washington, DC: Office of Scientific Research and Development.

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America, 27,* 338-352.

Nusbaum, H. C., Pisoni, D. B., & Davis, C. K. (1984). Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words. *Research on Speech Perception Progress Report No. 10.* Bloomington, IN: Speech Research Laboratory, Indiana University.

Penrod, J. P. (1994). Speech threshold and word recognition/discrimination testing. In J. Katz, (Ed.). *Handbook of Clinical Audiology* (pp. 147-164).

Pisoni, D. B., Nusbaum, H. C., Luce, P. A., & Slowiaczek, L. M. (1985). Speech perception, word recognition and the structure of the lexicon. *Speech Communication, 4,* 75-95.

Rabinowitz, W. M., Eddington, D. K., Delhorne, L. A., & Cuneo, P. A. (1992). Relations among different measures of speech reception in subjects using a cochlear implant. *Journal of the Acoustical Society of America, 92,* 1869-1881.

Sommers, M. S., Kirk, K. I., & Pisoni, D. B. (1997). Some considerations in evaluating spoken word recognition by normal-hearing and cochlear implant listeners I: The effects of response format. *Ear and Hearing, 18,* 89-99.

Stevens, S. S. (1951). Mathematics, measurement, and psychophysics. In S. S. Stevens, (Ed.), *Handbook of Experimental Psychology* (pp 1-49). New York: Wiley.

Tillman, T. W., Carhart, R., Wilber, L. (1963). A test for speech discrimination composed of CNC monosyllabic words. *Technical Documentary Report No. SAM-TDR-62-153.* USAF School of Aerospace Medicine, Brooks Air Force Base, Texas.

Treisman, M. (1978a). A theory of the identification of complex stimuli with an application to word recognition. *Psychological Review, 85*, 525-570.

Treisman, M. (1978b). Space or Lexicon? The word frequency effect and the error response frequency effect. *Journal of Verbal Learning and Verbal Behavior, 17*, 37-59.

Tyler, R. S., & Moore, B. C. (1992). Consonant recognition by some of the better cochlear-implant patients. *Journal of the Acoustical Society of America, 92*, 3068-3077.

Wang, M. D., & Bilger, R. C. (1973). Consonant confusions in noise: A study of perceptual features. *Journal of the Acoustical Society of America, 54,* 1248-1266.

Welkowitz, J., Ewen, R. B., & Cohen, J. (1991). *Introductory Statistics for the Behavioral Science.* Orlando, FL: Harcourt Brace Jovanovich.

Wilson, R. H., & Margolis, R. H. (1983). Measurements of auditory thresholds for speech stimuli. In D. F. Konkle & W. F. Rintelmann, (Eds.), *Principles of Speech Audiometry* (pp. 79-126). Baltimore: University Park Press.

Young, L. L., Dudley, B., & Gunter, M. B. (1982). Thresholds and psychometric functions of the individual spondaic words. *Journal of Speech and Hearing Research, 25*, 586-593.