Recognizing Spoken Words: The Neighborhood Activation Model

Luce, Paul A.; Pisoni, David B.

Author Information
Language Perception Laboratory, Department of Psychology and Center for Cognitive Science (P.A.L.), University at Buffalo, Buffalo, New York; and Speech Research Laboratory, Department of Psychology (D.B.P.), Indiana University, Bloomington, Indiana and DeVault Otologic Research Laboratory, Department of Otolaryngology-Head and Neck Surgery (D.B.P.), Indiana University School of Medicine, Indianapolis, Indiana.
Address for correspondence: Paul A. Luce, Language Perception Laboratory, Department of Psychology, University at Buffalo, Buffalo, NY 14260.

Outline

Back to Top

Abstract

Objective: A fundamental problem in the study of human spoken word recognition concerns the structural relations among the sound patterns of words in memory and the effects these relations have on spoken word recognition. In the present investigation, computational and experimental methods were employed to address a number of fundamental issues related to the representation and structural organization of spoken words in the mental lexicon and to lay the groundwork for a model of spoken word recognition.

Design: Using a computerized lexicon consisting of transcriptions of 20,000 words, similarity neighborhoods for each of the transcriptions were computed. Among the variables of interest in the computation of the similarity neighborhoods were: 1) the number of words occurring in a neighborhood, 2) the degree of phonetic similarity among the words, and 3) the frequencies of occurrence of the words in the language. The effects of these variables on auditory word recognition were examined in a series of behavioral experiments employing three experimental paradigms: perceptual identification of words in noise, auditory lexical decision, and auditory word naming.

Results: The results of each of these experiments demonstrated that the number and nature of words in a similarity neighborhood affect the speed and accuracy of word recognition. A neighborhood probability rule was developed that adequately predicted identification performance. This rule, based on Luce's (1959) choice rule, combines stimulus word intelligibility, neighborhood confusability, and frequency into a single expression. Based on this rule, a model of auditory word recognition, the neighborhood activation model, was proposed. This model describes the effects of similarity neighborhood structure on the process of discriminating among the acoustic-phonetic representations of words in memory. The results of these experiments have important implications for current conceptions of auditory word recognition in normal and hearing impaired populations of children and adults.

Since the publication of Oldfield's (1966) seminal article, "Things, Words and the Brain," a great deal of attention has been devoted to the structural organization of words in the mental lexicon. Most of this research, however, has focused on the structure of higher level aspects of lexical representations, namely the semantic and conceptual organization of lexical items in memory (e.g., Miller& Johnson-Laird, 1976; Smith, 1978). As a consequence, little attention has been directed to the structural organization of the representations of sensory and perceptual information used to gain access to these higher level sources of information. The goal of the present investigation was to explore in detail this structure and its implications

for perception of spoken words by normal and hearing-impaired listeners.

In the present set of studies, structure will be defined specifically in terms of similarity relations among the sound patterns of words. Similarity will serve as the primary means by which the organization of acoustic-phonetic representations in memory will be investigated. We assume that similarity relations among the sound patterns of spoken words represent one of the earliest stages at which the structural organization of the lexicon comes into play. The precise aim of the present investigation was to gain a detailed understanding of the lower level relations between stimulus input, activation of phonetic representations, and, subsequently, recognition of spoken words.

The identification of structure with similarity relations among sound patterns of words raises the difficult problem of defining similarity. Similarity, although crucial to the present investigation, is an ill-defined concept in research on speech perception and spoken word recognition, and one that deserves considerably more work in these areas of research (see Mermelstein, Reference Note 7). However, similarity can be approximated by both computational and behavioral predictors of confusion, the approach taken here. Thus, similarity will be defined in terms of a specific computational metric for predicting confusions among phonetic patterns as well as a behavioral, or operational, metric based on the results of a series of perceptual experiments.

Having defined structure as the similarity relations among the sound patterns of words, the question arises: Should the structural organization of representations in memory have consequences for spoken word recognition? Consider a content-addressable memory system in which there is no noise either in the signal or the listener (Kohonan, 1980). In such a system, encoding the acoustic-phonetic information in the stimulus word is tantamount to locating the word in memory. In this case, the structural organization of acoustic-phonetic representations in memory would have no consequences for word identification. Instead, the task of spoken word recognition would be identical to phonetic perception, and one need only study phonetic perception to understand how words are recognized. (By phonetic perception we are referring to the perception of individual segments such as consonants and vowels).

It is undeniable that phonetic perception is important in spoken word recognition. It is not undeniable, however, that the human word recognition system operates as a noiseless content-addressable system or that the acoustic-phonetic signal itself is devoid of noise. To begin, the signal is very often less than ideal for the purposes of the listener. Words are typically perceived against a background of considerable ambient noise, reverberation, and the voices of other talkers. In addition, coarticulatory effects and segmental reductions and deletions substantially restructure the phonetic information in a myriad of ways (Luce & Pisoni, 1987). Although such effects may indeed be useful to the listener (Church, Reference Note 3; Elman & McClelland, 1986), they also undoubtedly produce considerable ambiguities in the speech signal, making a strictly content-addressable word recognition system based on phonetic encoding unrealistic. In short, both the noise inherent in the signal as well as the noise against which the signal is perceived make it unlikely that word recognition is accomplished by direct access, based solely on phonetic encoding, to acoustic-phonetic representations in memory.

Not only is the speech signal noisy, so too is the recognition system of the listener. Although the human is clearly well adapted for the perception of spoken language, the system by which language is perceived is by no means a perfect one. In normal listeners, encoding, attentional, and memory demands frequently result in the distortion, degradation, or loss of important acoustic-phonetic information. The data on misperceptions alone attest to the fact that spoken word recognition is less than perfect (Bond& Garnes, 1980). In hearing-impaired listeners, the problems faced by normal listeners are exacerbated by impoverished input representations. Thus, again, a strictly content-addressable system does not suffice as a model of human spoken word recognition.

The alternative to a noiseless content-addressable system is one in which the stimulus input activates a

number of similar acoustic-phonetic representations or candidates in memory, among which the system must choose (Marslen-Wilson, 1989; Marslen-Wilson & Welsh, 1978; Triesman, 1978a, b). In this system, a considerable amount of processing activity involves discriminating among the lexical items activated in memory. Indeed, many current models of word recognition subscribe to the view that word recognition is to a great degree a process of discriminating among competing lexical items (Forster, 1979; Luce, Pisoni, & Goldinger, 1990; Marslen-Wilson, 1989; Marslen-Wilson & Welsh, 1978; McClelland & Rumelhart, 1981; McQueen, Cutler, Briscoe, & Norris, 1995; Morton, 1979; Norris, 1994).

Given that one of the primary tasks of the word recognition system involves discrimination among lexical items, the study of the structural organization of words in memory takes on considerable importance, especially if structural relations can be shown to influence the ease or difficulty of lexical discrimination, and, subsequently, word recognition and lexical access. By the same token, under the assumption that word recognition involves discrimination and selection among competing lexical items, variations in the ease or difficulty of discriminating among items in memory can enlighten us as to the structural organization of the sound patterns of words. In short, lexical discrimination and structure are so inextricably tied together that the study of one leads to a further understanding of the other.

Assuming, then, that structural relations among words should influence spoken word recognition via the process of discrimination, it is important to determine that structural differences among words actually exist. Previous research by Landauer and Streeter (1973) has demonstrated that words vary substantially not only in the number of words to which they are similar, but also in the frequencies of these similar words. These findings suggest that both structural and frequency relations among words may mediate lexical discrimination. Investigation of the behavioral effects of these relations should help us to understand further not only the process of lexical discrimination, but also the organization of the sound patterns of spoken words in memory.

The issue of word frequency takes on an important role in the investigation of the structural organization of the sound patterns of words. Numerous studies over the years (Howes, 1954, 1957; Newbigging, 1961; Savin, 1963; Soloman & Postman, 1952) have demonstrated that the ease with which spoken words are recognized is monotonically related to experienced frequency, as measured by some objective count of words in the language. However, little work has been devoted to detailing the interaction of word frequency and structural relations among words (see, however, Triesman, 1978a, b). If word frequency influences the perceptibility of the stimulus word, it should also influence the degree of activation of similar words in memory. Frequency is important, then, in further specifying the relative competition among activated items that are to be discriminated among.

The goal of the present investigation was, therefore, to examine the effects of the number and nature of words activated in memory on auditory word recognition. Throughout the ensuing discussion, the term similarity neighborhood will be employed. A similarity neighborhood is defined here as a collection of words that are phonetically similar to a given stimulus word.(The term stimulus word will be used to refer to the word for which a neighborhood is computed.) Similarly neighborhood structure refers to two factors: 1) the number and degree of confusability of words in the neighborhood, and 2) the frequencies of the neighbors. This first factor will be referred to as neighborhood density or neighborhood confusability; the second factor will be called neighborhood frequency. In addition to neighborhood structure, the frequency of the stimulus word itself will be of interest.

Back to Top

Previous Research on the Role of Neighborhood Structure and Frequency in Word Recognition
Neighborhood Structure

Little previous research has been devoted to examining the effects of neighborhood structure, primarily

because of the lack of computational tools for determining similarity neighborhoods for a large number of words. One early study of visual word recognition by Havens and Foote(1963) examined the effects of the number of competitors, or neighbors, of words on tachistoscopic identification. The results of this study, although based on a very small number of words and a rather imprecise measure of neighborhood membership, are suggestive. Havens and Foote demonstrated that effects of word frequency could be eliminated if the number of competitors for a given word were controlled. That is, low-frequency words were identified at levels of accuracy equal to those of high-frequency words when the number of competitors was held constant. This result suggests that the effect of frequency on visual word recognition is crucially dependent on the neighborhood in which the word resides in the lexicon.

Similar suggestive evidence was reported in a little known thesis by Anderson (Reference Note 1). In this study, Anderson examined the effects of the nature and number of alternatives on the intelligibility of spoken words. Although the means for determining alternatives were crude, Anderson demonstrated that intelligibility of spoken words was affected by both the number of possible confusors as well as by the frequencies of these confusors. In general, Anderson showed that words with many possible confusors were less intelligible than words with fewer confusors. In addition, he demonstrated that high-frequency confusors tended to depress (i.e., inhibit) identification performance.

Other evidence for the role of neighborhood structure in spoken word recognition was obtained from a reanalysis of a set of data published by Hood and Poole (1980). Hood and Poole examined the intelligibility of words presented in white noise. They found that word frequency failed to correlate consistently with the word intelligibility scores for their data, in apparent contradiction to many previous findings regarding the effects of frequency on noise-masked words (Howes, 1957; Savin, 1963). This finding indicated that factors other than word frequency per se were responsible for the wide range of speech intelligibility scores obtained by Hood and Poole.

To examine the possibility that similarity neighborhood structure was, at least in part, responsible for the differences observed in intelligibility in the Hood and Poole study, we examined their 25 easiest and 25 most difficult words. Similarity neighborhoods were computed (see below) for each of these 50 words. In keeping with Hood and Poole's observation regarding word frequency, no significant difference in frequency was found between the 25 easiest and 25 most difficult words (Pisoni, Nusbaum, Luce,& Slowiaczek, 1985). However, we found that the relationships of easy words to their neighbors differed substantially from the relationships of the difficult words to their neighbors. More specifically, approximately 56% of the words in the neighborhoods of the difficult words were equal to or higher in frequency than the frequencies of the difficult words themselves. For the 25 easy words, however, only approximately 23% of the neighbors of the easy words were of equal or higher frequency. Thus, it appears that the observed differences in intelligibility were due, at least in part, to the frequency composition of the neighborhoods of the easy and difficult words, and were not primarily due to the frequencies of the words themselves.

Taken together, these earlier studies suggest that neighborhood structure may play an important role in word recognition. Furthermore, these studies suggest that the effect of the frequency of the stimulus word itself may be mediated by neighborhood structure of other similar sounding words. The present set of studies was aimed at examining the role of neighborhood structure of spoken word recognition as well as specifying the combined roles of word and neighborhood frequency. Study of the combined effects of neighborhood structure and frequency should, therefore, lead to a better understanding of the effects of both similarity and frequency on word recognition.

Given that so little research has been devoted to these problems, it is hardly surprising that current models of spoken word recognition have had little to say about the structural organization of acoustic-phonetic patterns in the mental lexicon. Only cohort theory (Marslen-Wilson, 1989; Marslen-Wilson & Welsh,

1978) has made any precise claims regarding structural effects, and these have primarily been based on the assumption that words are recognized at the point at which they diverge from all other words in the mental lexicon, a prediction that says little about the structural organization of words in memory. For the most part, therefore, similarity neighborhood structural effects have been ignored in both research and theory on spoken word recognition. As we hope to show, this has been a serious omission in earlier work. Indeed, we will demonstrate here that any adequate theory of word recognition must provide a basic account of the structure of the sound patterns of words in memory as well as how these structural relations affect perceptual processing.

Back to Top

Word Frequency

Although little work has been devoted to the study of neighborhood structure, a voluminous body of data has been published on the effect of word frequency in visual and spoken word recognition (e.g., Glanzer & Bowles, 1976; Glanzer & Ehrenreich, 1979; Gordon, 1983; Howes, 1954, 1957; Howes & Solomon, 1951; Landauer & Freedman, 1968; Morton, 1969; Newbigging, 1961; Rubenstein, Garfield, & Millikan, 1970; Rumelhart & Siple, 1974; Savin, 1963; Scarborough, Cortese, & Scarborough, 1977; Soloman & Postman, 1952; Stanners, Jastrzembski, & Westbrook, 1975; Whaley, 1978). In general, these results have demonstrated numerous processing advantages for high-frequency words. Many theories have also been proposed to account for the advantages associated with increased word frequency. These theories have cited frequency of exposure (Forster, 1976; Morton, 1969), age of acquisition (Carroll & White, 1973a, b), and the time between the present and last encounter with the word (Scarborough, Cortese, & Scarborough, 1977) as the underlying reasons for the processing advantages observed for high-frequency words. Whatever the precise mechanism, it is now widely assumed by many researchers (Broadbent, 1967; Catlin, 1969; Goldiamond & Hawkins, 1958; Nakatani, Reference Note 9; Newbigging, 1961; Pollack, Rubenstein, & Decker, 1960; Savin, 1963; Soloman & Postman, 1952; Triesman, 1971, 1978a, b) that frequency serves to bias, in some manner, the word recognition system toward choosing high-frequency words over low-frequency words. Although the claim that frequency effects arise from biases is not uncontroversial, many theories of the word frequency effect have espoused such a view (see references cited above). Among these theories are sophisticated guessing theory (Neisser, 1967; Newbigging, 1961; Pollack et al., 1960; Savin, 1963; Soloman & Postman, 1952), criterion-bias theory (Broadbent, 1967), and partial identification theory (Triesman, 1978a, b). Although there has been considerable debate among the proponents of each of these theories, all assume that word frequency, by some as yet poorly specified processing mechanism, influences the decisions of the word recognition system via some sort of bias (Gordon, 1983; Norris, Reference Note 10).

Although there is some agreement among researchers as to the means by which processing advantages afforded by high-frequency words arise, there has, as previously mentioned, been little research on the relation between frequency and neighborhood structure, a primary issue in the present set of studies. Only Triesman (1978a, b) has addressed the issue of how neighborhood structure may influence the word frequency effect. The present set of studies is therefore aimed, in part, at examining the role of word frequency in the context of the similarity neighborhoods for words.

Back to Top

Description of the Present Approach

In the present set of studies, similarity neighborhood structure was estimated computationally, using a large, on-line lexicon. This lexicon, based on Webster's Pocket Dictionary (Webster's Seventh Collegiate Dictionary, 1967), contains approximately 20,000 entries. In the version of the lexicon used in the present set of studies, each entry contains: 1) an orthographic representation, 2) a phonetic transcription, 3) a frequency count based on the Kucera and Francis (1967) norms, and 4) a subjective familiarity rating

(Nasbaum, Pisoni, & Davis, Reference Note 11).

The phonetic transcriptions, coded in a computer-readable phonetic alphabet, are based on a general American dialect and include syllable boundary and stress markings. Frequency counts, as noted above, were obtained from an on-line version of the Kucera and Francis (1967) corpus. These counts were based on one million words of printed text. Although the study of frequency effects in spoken word recognition would be best served by a count of spoken words, no such count is available that covers the large number of words in Webster's lexicon. (For discussions of frequency and familiarity estimates of printed versus spoken words, see Gaygen & Luce, in press, and Pisoni& Garber, 1990.) Finally, the subjective familiarity ratings for each of the words were obtained in a large-scale study by Nusbaum et al. (Reference Note 11). In this study, groups of college undergraduates were asked to rate the subjective familiarity of each of the words in Webster's lexicon on a seven point scale, ranging from "don't know the word" (1) to "know the word and know its meaning" (7). The familiarity ratings were obtained from visually presented words.

The general procedure for computing similarity neighborhood structure using the computerized lexicon was as follows: a given phonetic transcription(constituting the stimulus word) was compared with all other transcriptions in the lexicon (which constituted potential neighbors). (The precise methods by which a neighbor was defined varied as a function of the particular experimental paradigm employed. These methods will be described in detail below.) By comparing the phonetic transcription of the stimulus word with all other phonetic transcriptions in the lexicon, it is possible to determine the extent to which a given stimulus word is similar to other words (i.e., neighborhood density or neighborhood confusability). In addition, it is also possible to determine the frequency of the neighbors themselves (i.e., neighborhood frequency), as well as the frequency of the stimulus word. Thus, in each of the studies to be reported, three variables were examined: stimulus word frequency, neighborhood density or neighborhood confusability, and neighborhood frequency.

The effects of similarity neighborhood structure on spoken word recognition were first examined in a large-scale experiment involving the identification of words against a background of white noise. Using confusion matrices for all possible initial consonants, vowels, and final consonants, a rule based on Luce's choice rule (1959) was devised to predict the accuracy of identifying words in noise. Basically, this rule-called the neighborhood probability rule (NPR)-takes into account the intelligibility of the stimulus word (estimated from the confusion matrices), the confusability of the neighbors of the stimulus word (also estimated from the confusion matrices), the frequency of the stimulus word, and the frequencies of the neighbors. The performance of this rule was tested against the data for the words presented for identification in noise. Based on the performance of this rule in predicting identification accuracy, the neighborhood activation model (NAM) of spoken word recognition was proposed to account for the perceptual identification of words in noise.

Two further experiments examined the role of similarity neighborhood structure in spoken word recognition using nondegraded stimuli. In the first of these experiments, words and nonwords varying in frequency and similarity neighborhood structure were presented to listeners in an auditory lexical decision task. In the second experiment, subjects attempted to repeat as quickly as possible auditorily presented words varying in frequency and similarity neighborhood structure. Each of these experiments was performed to answer different questions regarding the role of similarity neighborhood structure in spoken word recognition and to test specific claims of the NAM.

Back to Top

Summary

The present investigation represents an attempt to uncover the precise role of neighborhood structure in spoken word recognition using computational and behavioral techniques. The major hypothesis is that

words are recognized in the context of other words in memory. More precisely, it is predicted that the number of words that must be discriminated among in memory will affect the accuracy and time-course of word recognition. It is furthermore hypothesized that the frequencies of the words activated in memory will affect decision processes responsible for choosing among the activated words. Finally, it is proposed that the well-known word frequency effect may be a function of neighborhood frequency and similarity, and not a simple direct function of the number of times the stimulus word has been encountered.

Back to Top

Experiment 1: Evidence from Perceptual Identification

Approximately 900 monosyllabic words were first presented to subjects for identification and recognition accuracy scores were obtained for each word. Using the computerized lexicon and behavioral measures of similarity based on confusion matrices, NPRs expressing the probability of choosing a word from among its neighbors were computed for each word. The performance of these rules was then evaluated against the obtained identification data.

Back to Top

Method

Stimuli • Nine hundred eighteen words were selected from Webster's lexicon that met the following criteria: 1) All words were three phonemes in length; 2) all were monosyllabic; 3) all were listed in the Brown corpus of frequency counts (Kucera & Francis, 1967); and 4) all words had a rated familiarity of 6.0 or above on a seven point scale. The familiarity ratings were obtained from a previous study by Nusbaum, Pisoni, and Davis (Reference Note 11). In this study, all words from the Webster's lexicon were presented visually for familiarity ratings. The rating scale ranged from "don't know the word" (1) to "recognize the word but don't know the meaning" (4) to "know the word" (7). The rating criterion was established to ensure that the words would be known to the subjects. For the present experiment, only 811 of the 918 words were of interest. These 811 words all had the form consonant-vowel-consonant (CVC). The remaining 137 words, all having forms other than CVC, were included for a separate analysis not directly relevant to the present study (Luce, Reference Note 5).

The 918 words were recorded by a male speaker of a Midwestern dialect. The stimuli were recorded in a sound attenuated booth (IAC model 401A) using an Electro-Voice D054 microphone and an Ampex AG-500 tape recorder. The stimuli were then low-pass filtered at 4.8 kHz and digitized via a 12-bit analog-to-digital converter operating at a sampling rate of 10 kHz. Using WAVES, a digital waveform editor (Luce & Carrell, Reference Note 6), each stimulus was spliced from the entire stimulus set and placed in a separate digital file. After editing, all stimulus files were equated for overall RMS amplitude using the program WAVMOD (Bernacki, Reference Note 2). Equating for RMS amplitude ensured that the stimuli were approximately equal in average intensity.

The 918 stimuli were then randomly partitioned into three stimulus set files consisting of 306 words each. From two of the three stimulus sets, two practice lists of 15 words each were selected and placed in separate stimulus set files.

Screening • Before conducting the identification experiment proper, each of the 918 words was screened to ensure that no clearly anomalous stimuli were included in the final analysis. Each of the three stimulus set files was presented to separate groups of 10 subjects, resulting in ten observations per word. For the screening experiment, each word was presented at 75 dB SPL in the absence of masking noise. Except for the manipulation of signal to noise (SN) ratio, the procedure for stimulus presentation and data collection was identical to that for the identification in noise experiment (see Procedure section below). Only those words identified at a level of 90% correct or above were included in the final analysis. Thirty-six of the original 918 words failed to meet this criterion. Although these words were presented in the identification

experiment to maintain equal numbers of stimuli in each of the three stimulus set files, these words were eliminated from consideration in the final analyses of the data.

Subjects • Ninety subjects participated in partial fulfillment of an introductory psychology course. All subjects were native English speakers, reported no history of speech or hearing disorders, and were able to type.

Design • All stimuli were presented at each of three SN ratios: +15 dB, +5 dB, and -5 dB SPL. SN ratio was manipulated by varying the amplitude of the stimuli against a constant level of white, band-limited, Gaussian noise. The level of the noise was set at 70 dB SPL and was low-pass filtered at 4.8 kHz to match the gross spectral range of the stimuli. The stimuli were presented at 85 dB SPL for the +15 dB SN ratio, 75 dB for the +5 dB SN ratio, and 65 dB for the -5 dB SN ratio. Each of the three stimulus set files was presented to three groups of 10 subjects each. Each group of subjects heard one-third of the stimuli at +15 dB, one-third at +5 dB, and one-third at -5 dB. However, presentation at a given SN ratio varied randomly from trial to trial. For a given stimulus, SN ratio was a between-subjects factor. Altogether, 10 subjects heard each word at each SN ratio.

Procedure • Stimulus presentation and data collection were controlled on-line in real-time by a PDP 11/34 minicomputer. The stimuli were presented via a 12-bit digital-to-analog converter over matched and calibrated TDH-39 headphones. The stimuli and noise were first manually calibrated at 85 dB SPL. Programmable attenuators were then adjusted for each trial to achieve the desired SN ratio.

Subjects were tested in individual booths in a sound-treated room. CRT terminals interfaced to the PDP-11/34 computer were situated in each of the booths. The procedure for an experimental trial was as follows: subjects were first presented with the message "READY FOR NEXT WORD" on their CRT terminals. One sec after the message, 70 dB SPL of white noise was presented over the headphones. One hundred msec after the onset of the noise, a randomly selected stimulus was presented at one of the three attenuation levels. One hundred msec after the offset of the stimulus, the noise was terminated until the beginning of the next trial. After presentation of the stimulus and noise, a prompt appeared on each subject's terminal. Subjects then typed their responses on the terminals and pressed the RETURN key when finished. Subjects were able to see their responses while typing and were able to correct any typing errors before pressing the return key. After each subject had responded, another trial was initiated. In the event that one or more subjects failed to respond, a new trial was initiated within 30 sec of the offset of the noise. Alphanumeric string responses were collected by the PDP-11/34 and stored in disk files for later analysis.

Subjects were instructed to provide their best guesses for each word they heard. They were also instructed to enter no response (i.e., simply press the RETURN key) only in the event that they were completely unable to identify the word. After the instructions, 15 practice words, each at one of the three SN ratios, were presented. None of the 15 practice words were presented in the main experiment. After the practice list, the instructions were summarized and procedural questions were answered. One of the stimulus set files consisting of 306 words was then presented. Three short breaks were given at equal intervals. An experimental session lasted approximately 1 hr.

Data Analysis • The data files were first combined into a master list consisting of the responses from 10 subjects for each SN ratio(resulting in 30 total responses per word) for each of the 918 words. The 36 words failing to meet the criterion established in the screening experiment were marked and excluded from further analysis, leaving data for 882 words. The 811 CVC words were then selected from the remaining 882 words. (The 71 words that were omitted were not CVC words, e.g., "ask," "try," etc.) In total, 24,330 (811 words × 3 SN ratios × 10 observations) subject responses were included in the master data list.

The master list was edited to correct misspellings. Corrections for misspellings were performed by correcting transpositions, deleting single letter insertions, inserting single letter omissions, and correcting single letter substitutions. Single letter substitutions were corrected only when the key of the incorrect letter

was within one key of the target letter on the keyboard or when the correct letter would have been produced if the same keystroke had been performed by the opposite hand. Only responses constituting nonwords were corrected in this manner. Approximately 2.5% of the responses were corrected for misspellings. On completion of the editing of the master list, percentages correct (hereafter, "scores") for each word at each SN ratio were computed. Responses were scored as correct if the phonetic transcription constituted an identical match to the target word or if the response was an inflected form of the target word or a homophone.

Back to Top

Neighborhood Probability Rules (NPRs)

To quantify the effects of similarity neighborhood structure and to devise a single expression that simultaneously takes into account stimulus word intelligibility, stimulus word frequency, neighborhood confusability, and neighborhood frequency, an NPR, based on Luce's general biased choice rule, was devised. However, to devise such a rule for predicting the accuracy of identifying the words in noise, a principled means was required for determining those words in the computerized lexicon that constitute neighbors of a given target stimulus word. This is a particularly important problem for a task involving degradation of words by white noise, given that this noise may differentially mask certain speech sounds (e.g., fricatives) more than others (e.g., vowels). This differential masking may produce confusions(i.e., neighbors) that are dependent, in part, on the spectral properties of the masking noise. Thus, to obtain an independent metric for computing neighborhood confusability, the confusability of individual speech sounds was determined from confusion matrices for all initial consonants, vowels, and final consonants. Details regarding how these confusion matrices were obtained can be found in Luce (Reference Note 5). These confusion matrices were then used to investigate the combined effects of stimulus intelligibility and neighborhood confusability on word identification accuracy.

Having obtained identification scores for the 811 CVC words and confusion matrices for the initial and final consonants and vowels composing these words, the question becomes: How can the segmental intelligibility of the stimulus word, estimated from the confusion matrices, be combined with the segmental confusability of its neighbors, also estimated from the confusion matrices, to provide an index of the identifiability of the stimulus word? One means of accomplishing this goal is to devise an NPR incorporating the probability of identifying the stimulus word and the probabilities of confusing the neighbors with the stimulus word. Stated differently, can an NPR be devised that expresses the probability of identifying the stimulus word given the probabilities of identifying its neighbors?Luce's (1959) choice rule provides a straightforward means of computing such probabilities. Very simply, Luce's choice rule states that the probability of choosing a particular item i is equal to the probability of item i divided by the probability of itemi plus the sum of the probabilities of j other items.

The applicability of Luce's general choice rule to the problem at hand is transparent. Specifically, it provides a means for predicting the probability of choosing a stimulus word from among its neighbors and thus provides the formal basis for devising an NPR. Accordingly, an NPR assumes the following general form: the probability of identifying the stimulus word is equal to the probability of the stimulus word divided by the probability of the stimulus word plus the sum of the probabilities of identifying the neighbors. Thus: Equation 1 where $p(ID)$ is the probability of correctly identifying the stimulus word, $p(S)$ is the probability of the stimulus word, and $p(N_j)$ is the probability of the jth neighbor.

Stimulus Word Probabilities (SWPs) • The data from the confusion matrices can be used to compute the probability of the stimulus word and the conditional probabilities of its neighbors. The probability of the stimulus word was computed as follows: for each phoneme in the stimulus word, the conditional probability of that phoneme given itself can be obtained from the confusion matrices. Assuming independent probabilities, the obtained conditional phoneme probabilities can be multiplied. This product

renders a SWP based on the probabilities of the individual phonemes of the stimulus word. Note that the probability for the stimulus word is based on the product of the conditional probabilities of identifying each phoneme independently, obtained from the confusion matrices. Thus, the SWP can be computed as follows: Equation 2 where $p(PS_i|PS_i)$ is the conditional probability of identifying the ith phoneme of the stimulus word given that phoneme, and n is the number of phonemes in the word. For example, the SWP of the word /d[Latin small letter open o]g/ ("dog") is: Equation 3 where, again, the conditional probabilities of the individual phonemes are determined from the confusion matrices for the initial consonants, vowels, and final consonants. Note that the SWP of /d[Latin small letter open o]g/ can be construed as the conditional probability of the word /d[Latin small letter open o]g/ given /d[Latin small letter open o]g/, or p(d[Latin small letter open o]g/d[Latin small letter open o]g).

Neighbor Word Probabilities (NWPs) • In this manner, conditional probabilities for each neighbor of the stimulus word can also be computed. Thus, the NWP can be computed by finding the conditional probabilities of each of the phonemes of the neighbor given the stimulus word phonemes. Multiplying these probabilities renders an index of the probability of the neighbor, or the NWP. NWP can be computed as follows:Equation 4 where PN_i is the ith phoneme of the neighbor, PS_i is the ith phoneme of the stimulus word, and n is the number of phonemes.

To return to the example of the stimulus word /d[Latin small letter open o]g/ given above, the NWP for the neighbor /tæg/ can be computed as:Equation 5 which also can be construed to be the probability of identifying /tæg/ given /d[Latin small letter open o]g/, or p(tæg/d[Latin small letter open o]g).

Frequency-Weighted Neighborhood Probability Rule (FWNPR) • Given these designations of stimulus word and NWPs, the appropriate substitutions of terms in Equation 1 render an NPR based on the general choice rule: Equation 6 where PS_i is the probability of the ith phoneme of the stimulus word, PN_{ij} is the probability of the ith phoneme of thej th neighbor, n is the number of phonemes in the stimulus word and the neighbor, $Freq_s$ is the frequency of the stimulus word, $Freq_{Nj}$ is the frequency of the jth neighbor, and nn is the number of neighbors. This rule will be referred to as the FWNPR.

A number of properties of the NPR are worthy of mention. First, the intelligibility of the phonemes of the stimulus word itself will determine, in part, the role of the neighbors in determining the predicted probability of identification. Stimulus words with high phoneme probabilities (i.e., words with highly intelligible phonemes) will tend to have neighbors with low phoneme probabilities, owing to the fact that all probabilities in the confusion matrices are conditional. Likewise, stimulus words with low phoneme probabilities (i.e., those with less intelligible phonemes) will tend to have neighbors with relatively higher phoneme probabilities. However, the output of the NPR is not a direct function of the SWP. Instead, the output of the rule is dependent on the existence of lexical items that contain phonemes that are confusable with the phonemes of the stimulus word. For example, a stimulus word may contain highly confusable phonemes. However, if there are few actual lexical items (i.e., neighbors) that contain phonemes confusable with those of the stimulus word, the sum of the NWPs will be low. The resulting output of the NPR will, therefore, be relatively high. Likewise, if the phonemes of the stimulus word are highly intelligible, but there are a large number of neighbors that contain phonemes that are confusable with the stimulus word, the probability of identification will be reduced. In short, the output of the NPR is contingent on both the intelligibility of the stimulus word and the number of neighbors that contain phonemes that are confusable with those of the stimulus word. Thus, intelligibility of the stimulus word, confusability of the neighbors, and the nature of lexical items act in concert to determine the predicted probability of identification. In addition, the frequencies of the stimulus word and the neighbors will serve to amplify to a greater or lesser degree the word probabilities of the stimulus word and its neighbors. Note that frequency in this rule is expressed in terms of the relation of the frequency of the target word to the

frequencies of its neighbors. Thus, the absolute frequency of the stimulus word may have differential effects on predicted identification performance depending on the frequencies of the word's neighbors. For example, given two stimulus words of equal frequency, the stimulus word with neighbors of lower frequencies will produce a higher predicted probability than the stimulus word with neighbors of higher frequencies. The degree to which the frequencies of the neighbors will play a role in determining predicted identification performance will, of course, depend on the NWPs. The frequencies of the neighbors with low probabilities of confusion will play less of a role than those with high probabilities of confusion. Simply put, this rule predicts that neighborhood structure will play a role in determining predicted identification performance in terms of the combined effects of the number and nature of the neighbors, the frequencies of the neighbors, the intelligibility of the stimulus word, and the frequency of the stimulus word.
Back to Top

Predicting Identification Performance Using the NPR

Computation of the Rule • To evaluate the success of the proposed NPR, the data from the perceptual identification study were analyzed in terms of the frequency-weighted rule using the confusion matrix data. Given that confusion matrices were obtained only for consonants occurring in initial and final position, only those words of the form CVC from the original data set were analyzed. As stated earlier, 811 CVC words were analyzed. In addition, to simplify the computational analysis, only monosyllabic words contained within the Webster's lexicon were used to compute the NPR. Inspection of the error responses revealed that this was not an unreasonable simplification, given that a significant majority of the error responses were, in fact, monosyllabic words, which indicates that subjects typically perceived monosyllabic words. The restriction of monosyllabic neighbors was necessitated by the particular procedure used to determine NWPs (see below).

The general method for computation of the NPR was as follows: the SWP was first determined for a given stimulus word. This probability was computed from the confusion matrices using Equation 2. After computation of the SWP, the transcription of the stimulus word was compared with the transcriptions of all other monosyllabic words in Webster's lexicon with a familiarity rating of 5.5 A cutoff familiarity rating of 5.5 was imposed on the possible neighbors to ensure that most words in the lexicon unknown to subjects would be excluded from consideration as neighbors. The value of 5.5 was chosen based on inspection of the familiarity ratings of the error responses.

To compute the NWPs, the vowel of the stimulus word was first aligned with the vowel of the neighbor being analyzed. The conditional probabilities of the vowel and the consonants flanking the vowel for the neighbor were then determined from the appropriate confusion matrices. In the event that the neighbor was a CVC word, the neighbor word phoneme probability was computed using Equation 4. That is, the conditional probability of the initial consonant of the neighbor given the initial consonant of the stimulus word was determined, as were the conditional probabilities of the vowel and the final consonant.

A problem arises as to the treatment of neighbors containing either initial consonant clusters, final consonant clusters, or both. In these cases, the transcriptions for the stimulus word and the neighbor were aligned at the vowel. However, when initial and/or final clusters were present in the neighbor word, those consonants not immediately adjacent to the vowel fail to overlap with anything in the stimulus word. For example, if the stimulus word was /k[LAMBDA]t/ and the neighbor was /skId/, the /k/ of the stimulus word would align with the /k/ of the neighbor, /[LAMBDA]/ would align with /I/, and /t/ would align with /d/. However, the /s/ of the neighbor /skId/ would align with no phoneme in the stimulus word. In this event, the probability of the phoneme /s/ was determined by finding the conditional probability of /s/ given the null phoneme from the confusion matrix for initial consonants, or p(s|0), where "0" is the null phoneme. In essence, this is the probability of perceiving /s/ when in fact no phoneme has been presented. The conditional probabilities for the neighbor /skId/ would thus be: p(s|0), p(k|k), p(I|[LAMBDA]), and p(d|t).

The procedure for final consonant clusters was identical, except that the probability of the neighbor phoneme given the null phoneme was computed from the final consonant confusion matrix. This method of dealing with initial and final clusters in the neighbor word makes the simplistic assumption that clusters are phonetically and acoustically equivalent to the sum of their constituent phonemes. However, in the absence of confusion matrices for all possible clusters as well as single-tons, this simplification appeared reasonable.

A similar problem arises when the neighbor is shorter than the stimulus word. In these cases, however, the solution is much more straightforward. The stimulus word and neighbor are once again aligned at the vowel. The empty slot in the neighbor is then assumed to contain the null phoneme, in which case the conditional probability for that phoneme can be easily determined from the appropriate confusion matrix. For example, if the stimulus word is/k[LAMBDA]t/ and the neighbor /æt/, the neighbor /æt/ is assumed to have the transcription /0æt/. The conditional probabilities for the neighbor phonemes would then be: p(0lk), p(ælæ), and p(tlt). (See Luce, Reference Note 5, for further discussion regarding the applicability of the rule to items other than CVC words.)

In this manner, NWPs were determined and the NPR was computed for each of the 811 CVC stimulus words. The NPRs were computed separately for each SN ratio for each word using the confusion matrices appropriate to that SN ratio. Thus, when computing the NPR for a given SN ratio, the confusion matrix obtained at that SN ratio was used to determine the SWPs and NWPs. That is, three predicted identification scores were computed for each word, one for each of the three SN ratio.

Correlation Analysis • The outputs of the FWNPR was combined with the scores for the 811 CVC words and submitted to correlation analyses. The correlations between identification accuracy and the FWNPR for each of the three SN ratios are shown in Table 1. For comparison, the correlations between identification accuracy and word frequency are also shown. (Note that word frequency is generally considered to be one of the most powerful single predictors of word recognition performance.) All correlations were significant beyond the 0.05 level. The correlation between FWNPR and identification performance was highest at the intermediate level of stimulus degradation, probably because there was simply more variance to account for at this SN ratio. Furthermore, the FWNPR was superior to word frequency at all but the lowest SN ratio, where performance of the two variables was virtually identical. (The difference between the correlations at the -5 SN ratio was not significant, $p > 0.05$.) Given that overall performance was only approximately 14% correct at this SN ratio, there was simply little variance in accuracy to account for by FWNPR.

The FWNPR produced significant, positive correlations at each of the SN ratios, demonstrating that identification of spoken words in noise is a function of the number and nature of items activated in the similarity neighborhood. A number of factors make the success of the rule impressive. First, the obtained correlations were computed for a large number of stimuli($N = 811$). The number of stimuli in fact virtually exhausts the entire population of highly familiar, CVC words in English. The success of the FWNPR is thus rather remarkable given the large number of stimuli examined in this study.

The performance of the rule is even more striking given that no specific information about the idiosyncratic acoustic-phonetic structures of the individual stimuli was included in the rule. Only information concerning the relative intelligibility and confusability of the individual segments was included, and this information was obtained from an independent source of data, namely, the confusion matrices obtained in a separate experiment with different subjects. Thus, the rule was able to achieve the obtained level of performance in the absence of specific measurements of the spectral, durational, and amplitude characteristics of the specific phonetic segments of the individual words.

The FWNPR also incorporates three sources of information that may introduce considerable noise in

predicting identification. The first comes from the confusion matrices themselves. The confusion matrices were obtained from a separate pool of subjects and were based on CV and VC syllables. The pattern of confusions obtained from the CV and VC syllables may differ in fundamental ways from the pattern of confusions produced by real words. In particular, response biases frequently observed in confusion matrices of this type may introduce significant sources of noise in predicting confusions among real words (Klatt, 1968; Miller& Nicely, 1955; Wang & Bilger, 1973). Obtaining confusion matrices for individual segments in a task requiring absolute identification of these segments in nonsense syllables may, therefore, reflect biases that may be inappropriate for predicting confusions among segments in real words. However, this does not mean that the use of confusion matrices to determine stimulus and NWPs was misguided (Moore, Reference Note 8). Indeed, the use of confusion matrices provides the only independent means of assessing stimulus intelligibility and confusability. However, in assessing the performance of the rule, it must be kept in mind that the confusion matrices provide less than perfect estimations of intelligibility and confusability of real words. In light of these observations, then, the performance of the NPR is even more impressive.

A second source of possible noise introduced in FWNPR arises from the lexicon used to compute neighborhood structure. Despite the controls placed on the inclusion of words in the neighborhoods of the stimulus words, the lexicon used may tend either to underestimate or overestimate the actual mental lexicons of the subjects themselves. The computer-based lexicon serves as only a very general model of the mental lexicon of the subject, thus introducing a potentially large source of noise in the estimation of neighborhood structure. However, in the absence of well-controlled techniques for estimating the nature and number of lexical items in the mental lexicon of a particular subject, the lexicon used in the present study provides an invaluable tool for determining neighborhood structure (Lewellen, Goldinger, Pisoni, & Greene, 1993). Indeed, before the availability of computerized lexicons containing phonetic transcriptions, such estimations of neighborhood structure would have been nearly impossible without these computational techniques.

A third source of noise in predicting identification may have arisen from the use of the Kucera and Francis frequency counts. These counts are not only somewhat dated, having been obtained in the 1960's, but they are also based on printed text. However, given that frequency counts were required for a large number of words, the use of available counts of spoken words was not feasible. Thus, the Kucera and Francis counts, although problematic, provided one of the single best estimates of word frequency available for a large number of stimuli.

Once these factors are taken into consideration, the performance of the FWNPR proves to be more than adequate. The results, therefore, clearly demonstrate the role of neighborhood structure in word identification.

Qualitative Analysis • Recall that the FWNPR predicts that identification performance is a function of the intelligibility of the stimulus word, the confusability of its neighbors, and the frequencies of the stimulus word and its neighbors. According to this view, frequency thus serves to bias the choice of a word from its neighborhood. Note that the effects of frequency are contingent on the nature of the words residing in the similarity neighborhood. As in Triesman's (1978a, b) partial identification theory, frequency effects are assumed in the rule to be relative. For example, high-frequency stimulus words residing in neighborhoods containing high-frequency neighbors are predicted by the rule to be identified at approximately equal levels of performance to low-frequency words residing in low-frequency neighborhoods, assuming that stimulus intelligibility and neighborhood confusability are held constant. That is, the frequency of the stimulus word alone will not determine identification performance. Instead, stimulus word frequency must be evaluated in terms of the frequencies of the neighbors of the stimulus word, as well as the confusability of the neighbors. Thus, the rule implies a complex relation between the stimulus word and its neighbors, such that stimulus frequency, neighbor frequency, stimulus intelligibility, and neighborhood confusability all act in

combination to determine identification performance.

The rule, therefore, makes a number of important predictions depending on the SWP and the sum of the neighbor probabilities. For simplicity, we will define the sum of the frequency-weighted NWPs, i.e., Equation 7 as the overall frequency-weighted neighborhood probability, or FWNP. Inspection of Equation 6 reveals that if the frequency-weighted stimulus word probability (FWSWP) (i.e., SWP*Freq$_S$, or the numerator of Equation 6) is held constant, as FWNP (the right-hand term of the denominator of Equation 6) increases, predicted identification will decrease. Likewise, if FWNP is held constant, then increases in FWSWP will result in corresponding increases in predicted identification. The interesting cases arise, however, when both the FWSWP and FWNP are allowed to vary. Consider the four cases in which the FWSWP and FWNP can take on either high or low values: 1) FWSWP high-FWNP high, 2) FWSWP high-FWNP low, 3) FWSWP low-FWNP high, and 4) FWSWP low-FWNP low. The predictions of the NPR for these four cases is shown in Table 2.

As shown in Table 2, the rule predicts best performance for those words with high FWSWPs and low FWNPs. These are high-frequency words that, in a sense, "stand out" in their neighborhoods. The lowest performance is predicted for words with low FWSWPs and high FWNPs. These are low-frequency words that are least distinguishable in their neighborhoods. Interestingly, however, the rule predicts intermediate levels of performance for the remaining two cases. That is, words with high FWSWPs and high FWNPs are predicted to show approximately equal levels of identification performance to words with low FWSWPs and low FWNPs. Thus, the rule does not always predict an advantage for high-frequency words over low-frequency words. In addition, the rule predicts that words matched on FWSWP may show differential levels of performance depending on the FWNP, or frequency-weighted neighborhood structure. To determine weather the general pattern of predictions outlined in Table 2 holds for the present set of identification data, the following analyses were performed: for the 811 words, median values for the FWSWPs and FWNPs collapsed across SN ratio were determined. These median values were then used to divide the stimulus words into classes having high and low FWSWPs and high and low FWNPs. Altogether, four cells were analyzed(two levels of FWSWP by two levels of FWNP). Mean identification scores, collapsed across SN ratio, were then computed for words falling into each of the four cells. The results for the classification of word scores by FWSWP and FWNP are shown in Table 3. Predicted levels of performance are shown in parentheses.

As shown in this table, the pattern of results predicted by the NPR was clearly present in the identification data. As predicted, words with high FWSWPs and low FWNPs were responded to with the highest levels of accuracy; words with low FWSWPs and high FWNPs were responded to with the lowest levels of accuracy. The remaining two cases, as predicted, showed intermediate and nearly identical levels of identification performance. Note that words matched on FWSWP were responded to quite differently depending on the FWNP, demonstrating that stimulus word frequency is a direct function of the neighborhood in which the stimulus word occurs. This is also demonstrated by the cases showing intermediate levels of performance. Although the words in these cells differ substantially in their FWSWPs, they show nearly identical levels of identification performance, owing to the composition of their similarity neighborhoods. In short, the present analysis provides further empirical support for the hypothesis that spoken word recognition is the result of a complex interaction of stimulus word intelligibility, stimulus word frequency, and neighborhood confusability and frequency.

Back to Top

Discussion

Back to Top

Neighborhood Activation Model (NAM)

The NPRs developed above provided the ground-work for the development of a model of spoken word

identification that we call the NAM. The basic postulate of the model is that the process of word identification involves discriminating among lexical items in memory that are activated on the basis of stimulus input. This is a fundamental principle in almost every current model of word recognition (e.g., Forster, 1976, 1979; Marslen-Wilson & Welsh, 1978; Paap, Newsome, McDonald, & Schvaneveldt, 1982). Indeed, one of the most important issues in spoken word recognition concerns the processes by which discrimination among lexical items in memory is achieved (Pisoni & Luce, 1987). The present model attempts to specify those factors responsible for the relative ease or difficulty of recognizing words arising from the processes involved in discrimination among sound patterns of words. Thus, a second fundamental principle of the model is that discrimination is a function of the number and nature of lexical items activated by the stimulus input. The "nature" of lexical items refers specifically to the acoustic-phonetic similarity among the activated lexical items as well as their frequencies of occurrence. The model is, therefore, concerned with the long-standing fundamental issue of word frequency. However, characterizing the effects of word frequency is only one aspect of the present model. More centrally, the model focuses primarily on structural issues concerning the process of lexical discrimination. Word frequency is important in the model only as a factor affecting the structural relationships among lexical items.

A flow chart of the NAM is shown in Figure 1. On presentation of stimulus input, a set of acoustic-phonetic patterns are activated in memory. It is assumed that all patterns are activated regardless of whether they correspond to real words in the lexicon or not, an assumption required by the fact that listeners can recognize the acoustic-phonetic form of novel words and nonwords. As in Triesman's (1978a, b) partial identification theory, the acoustic-phonetic patterns are assumed to be activated in a multidimensional acoustic-phonetic space in which the perceptual dimensions correspond to phonetically relevant acoustic differences among the patterns. Specification of the nature of these dimensions poses an important problem for any complete theory of speech perception and spoken word recognition (Luce & Pisoni, 1987). However, the present model is neutral with respect to the dimensions of the space. The only requirement of the model is that the dimensions of the space produce relative activation levels among the acoustic-phonetic patterns that are isomorphic with the dimensions of similarity to which subjects are sensitive.

The acoustic-phonetic patterns then activate a system of word decision units tuned to the patterns themselves (Morton, 1969). A diagram of a single decision unit is shown in Figure 2. Only those acoustic-phonetic patterns corresponding to words in memory will activate a word decision unit. Neighborhood activation is assumed to be identical to the activation of the word decision units. Once activated, these decision units monitor the activation levels of the acoustic-phonetic patterns to which they correspond. After activation of the word decision units, these units then begin monitoring higher level lexical information relevant to the words to which they correspond. Word frequency is included in this higher level lexical information. In addition to monitoring higher level lexical information in long term memory, the decision units are also assumed to monitor any information in short term memory that is relevant to making a decision about the identity of a word.

The system of word decision units is a crucial aspect of the NAM. These units serve as the interface between acoustic-phonetic information and higher level lexical information. Acoustic-phonetic information drives the system by activating the word decision units, affording priority to bottom-up information, as in cohort theory (Marslen-Wilson & Welsh, 1978). Higher level lexical information such as frequency is assumed to operate by biasing the decision units. These biases operate by adjusting the activation levels of the acoustic-phonetic patterns represented in the decision units. The biases introduced volitional control nor need they be conscious (Smith, 1980). Instead, these biases are assumed to be a fundamental aspect of word perception that enable optimization of the word recognition process via the employment of a priori probabilities and contextual information.

Each word decision unit is, therefore, responsible for monitoring two sources of information, acoustic-

phonetic pattern activation and higher level lexical information. In addition, the decision units are assumed to be interconnected in such a way that each unit can monitor the overall level of activity in the system of units, as well as the activity level of the acoustic-phonetic patterns to which the units correspond (Elman & McClelland, 1986; McClelland & Elman, 1986). As analysis of the stimulus input proceeds, the decision units continuously compute decision values. These values are assumed to be computed via a rule of the type described by the NPR. In the NPR, the SWP corresponds to the activation level of the acoustic-phonetic pattern. The sum of the neighbor word probabilities(NWP_js) corresponds to the overall level of activity in the decision system. Frequency information serves as a bias, as in the FWNPR, by adjusting the activation levels of the acoustic-phonetic patterns represented in the word decision units.

As processing of the stimulus input proceeds, information regarding the match between stimulus input and the acoustic-phonetic pattern accrues. The activation levels of similar patterns drop and the decision values computed by the word decision unit monitoring the pattern of the stimulus steadily increase. Once the output of a given decision unit reaches criterion, all information monitored by that decision unit is made available to working memory. "Word recognition" is accomplished once the word decision unit for a given acoustic-phonetic pattern surpasses criterion (i.e., the acoustic-phonetic pattern is recognized). "Lexical access" occurs when higher level lexical information (i.e., semantic, syntactic, and pragmatic information) is made available to working memory. The term lexical access is actually somewhat misleading in the context of the NAM. Lexical information is monitored by the word decision units once these units are activated. However, this information is used only in the service of choosing among the activated acoustic-phonetic patterns and is, therefore, not available to working memory. Lexical access in the NAM is thus assumed to occur when lexical information is made available for further processing. The word decision units in the model, therefore, serve as gates on the lexical information available to the system (Morton, 1979). In so doing, the units prevent the cognitive system from "over-resonating," making information available only once a decision is made as to the identity of the stimulus input.

The postulation of a system of word decision units is based on the finding that the FWNPR adequately predicted identification performance. Indeed, the system of word decision units is simply a processing instantiation of the NPR. However, the NAM, by instantiating the NPR in a system of decision units, makes a number of important claims. First, it is assumed that the word recognition system is, at least initially, completely driven by the stimulus input. Frequency information is thus assumed only to bias the decision units and not to affect the initial encoding of the acoustic-phonetic patterns. Thus, frequency information is not assumed to be an intrinsic part of the activation levels of the acoustic-phonetic patterns, but is assumed to be a bias that must be interpreted in the context of the frequencies of all other words. If frequency information were assumed to be intrinsic to the activation levels of the acoustic-phonetic patterns and no decisions were made based on the total activity of the system, low-frequency words would be responded to less accurately than high-frequency words regardless of their neighborhood structures, which is clearly in contradiction to the data reported above (Luce, Reference Note 5).

Having laid out a framework for interpreting neighborhood structural and frequency effects, we will now turn to a discussion of how the NAM accounts for the results of the perceptual identification study. Recall that the NAM, under normal circumstances, recognizes a word once the decision value for a given word exceeds criterion. It is assumed that stimulus degradation affects the word recognition system by impeding complete processing of the stimulus input. That is, only so much information can be obtained from the stimulus input when it is masked by noise. Given imperfect information, then, it is assumed that, in the long run, no decision unit will reach criterion, and a decision will thus be forced based on the available information. The "available information" is the state of the decision system at the point at which processing of the acoustic-phonetic information is completed. In a perceptual identification task, therefore, a response is made on the basis of the values of the decision units at the point at which processing is completed. Thus,

the NPR developed above expresses the probability of choosing the stimulus word actually presented. If the stimulus input results in a large number of highly confusable, high-frequency neighbors, the probability of actually recognizing the stimulus word will be low. Likewise, if the stimulus input results in only a few confusable, low-frequency neighbors, the probability of identification will be high.

Note that because the decision units monitor both the activation levels of the acoustic-phonetic patterns to which they correspond as well as the overall activation of the decision system, probability of identification will not depend solely on the intelligibility of the stimulus word nor on neighborhood confusability. Words of low intelligibility with few confusable neighbors are predicted by the model to be equivalent to words of high intelligibility with many confusable neighbors. Indeed, as shown above, this prediction was borne out empirically. In short, perceptual identification is a function of the values of the decision units computed at the completion of stimulus processing. Furthermore, the role of stimulus degradation (whether arising from a noisy signal or a degraded input representation due to an impaired sensory apparatus) is assumed to be one of impeding complete processing of the stimulus input.

Back to Top

Summary

The NAM provides a conceptual and theoretical framework for instantiating the NPR developed here. To the extent that the NPR predicts identification performance, the model can be deemed an adequate account of the word identification process. Indeed, this rule was shown to make a number of precise predictions about the relative effects of stimulus intelligibility and neighborhood structure that were borne out by the data. In particular, the NPR predicts a complex interrelationship between the stimulus word and its neighbors. In addition, the FWNPR was able to account for the stimulus word frequency effect in terms of the frequency relationships between the stimulus word and its neighbors. The picture that begins to emerge from the present findings is one that emphasizes the complementary roles of discrimination and decision in spoken word recognition. Finally, both the data and the NAM emphasize the degree to which the structure of the mental lexicon influences word identification: precise accounts of the process of spoken word recognition are crucially tied to detailed accounts of the structural relationships among lexical items in memory.

Back to Top

Experiment 2: Evidence from Auditory Lexical Decision

The purpose of the present study was to explore further the effects of neighborhood structure on spoken word recognition. In particular, the lexical decision paradigm was employed to examine these effects. In the lexical decision paradigm, a subject is presented with a real word or a nonsense word, or nonword. The subject's task is to decide as quickly and as accurately as possible whether a given stimulus item is a word or nonword. The lexical decision task has proven quite useful in visual word recognition research in examining the effects of such variables as word frequency (Stanners, Jastrzembski, & Westbrook, 1975; Whaley, 1978; Forster, 1979). In general, it has been shown that high-frequency words tend to be classified as words more quickly than low-frequency words. Indeed, this has been a very robust finding in the literature, although there are numerous, and often times conflicting, accounts of frequency effects in lexical decision (Balota & Chumbley, 1984; Glanzer & Ehrenreich, 1979; Gordon, 1983; Paap, McDonald, Schvaneveldt, & Noel, 1986). A spoken analog of the visual lexical decision task thus presents a useful means of examining word frequency effects and the effects of neighborhood structure on spoken word recognition.

The use of the lexical decision task is also attractive for two other reasons. First, investigation of the process of spoken word recognition can be carried out in the absence of stimulus degradation. Although the perceptual identification experiment provided useful data regarding the effects of stimulus word frequency

and neighborhood structure, a more robust test of these effects hinges on the demonstration that neighborhood structural effects can be demonstrated in the absence of stimulus degradation. In other words, it is important to demonstrate that the effects of neighborhood structure generalize beyond words that are purposefully made difficult to perceive.

The second advantage of the spoken lexical decision task is the ability to collect reaction time data. The reaction time data may aid in uncovering some of the temporal aspects of the effects of neighborhood structure on spoken word recognition. Furthermore, it is of crucial importance to demonstrate that neighborhood structure affects not only the accuracy of word recognition, but also the time course. Thus, the auditory lexical decision task provides a useful means of corroborating and generalizing the findings from the previous perceptual identification study.

The approach taken in the present study is similar to that in Experiment 1. Similarity neighborhood statistics, computed on the basis of Webster's lexicon, served as independent variables. The statistics of interest were again: 1) the number of words similar to a given stimulus word, or neighborhood density; 2) the mean frequency of the similar words or neighbors; and the 3) frequency of the stimulus word itself. Neighborhood density and neighborhood frequency were also manipulated for a set of specially constructed nonwords

The means of computing similarity neighborhood structure employed in the present study was somewhat different than in Experiment 1. Because the estimates of similarity of neighbors to their stimulus words in Experiment 1 was based on confusion matrices for consonants and vowels presented in noise, these measures of similarity are inappropriate for stimuli presented in the clear. Thus, a strictly computational means for estimating similarity neighborhood structure was employed. In this method, a given phonetic transcription (constituting the stimulus word) was again compared with all other transcriptions in the lexicon (which constituted potential neighbors). However, in this method of computing similarity neighborhood structure, a neighbor was defined as any transcription that could be converted to the transcription of the stimulus word by a one phoneme substitution, deletion, or addition in any position. For example, among the neighbors of the word /kæt/ would be /pæt/, /kIt/, and /kæn/, which each are derived on the basis of one phoneme substitutions in any position. Also included as neighbors would be the words /skæt/ and /æt/, derived on the basis of one phoneme additions or deletions. (Plurals and inflected forms of the stimulus were not included as neighbors.) The number of such neighbors constitutes the variable of neighborhood density. Neighborhood frequency again refers to the average of the frequencies, based on the Kucera and Francis counts, of each of the neighbors. And, stimulus word frequency refers to the frequency of the stimulus word for which the neighbors were computed (e.g., /kæt/).

This particular algorithm for computing neighborhood membership was based on previous work by Greenberg and Jenkins (1964) and Landauer and Streeter (1973), Sankoff & Kruskall (1983). Clearly, this method of computing neighborhood membership makes certain strong assumptions regarding phonetic similarity. In particular, it assumes that all phonemes are equally similar and that the similarities of phonemes at a given position are equivalent. However, this method of computing similarity neighborhoods provides a computationally simple means of estimating the number and nature of words similar to a given stimulus word presented in the absence of noise.

Because decision units only correspond to words actually occurring in memory, lexical decision in the context of the NAM can only be achieved by accepting or rejecting words. In particular, lexical decisions are assumed to be based-in a majority of the cases-on one of two criteria being exceeded (Coltheart, Develaar, Johansson, & Besner, 1976). A word response is executed if a decision unit determines that the activation level of the pattern it is monitoring exceeds the criterion, the normal procedure for recognizing a word and depositing its lexical information in working memory. However, the procedure for executing a nonword response is contingent on surpassing a lower level criterion. A non-word response is executed if

the total activation level monitored by the decision units falls below a lower level criterion, indicating that no word is consistent with the stimulus input. According to the NAM, therefore, the primary means of word-non-word classification is based on the activity within the decision system surpassing or falling below one of two criteria, which can be referred to as the "word" and "nonword" criterion levels.

Under circumstances in which subjects are required to classify a clearly presented word or non-word under no time constraint, it is assumed that classification accuracy will be near perfect. Exhaustive analysis of the stimulus input would result in few, if any, errors in classification. However, once time constraints are imposed by instructions to respond as quickly as possible, accuracy levels will vary as a function of the amount of stimulus processing carried out before the response. It is assumed that subjects will attempt to classify an item before a self-imposed reaction time deadline. The assumption of a reaction time deadline is motivated by the fact that subjects are attempting to respond as quickly as possible and will allow only a given amount of time to pass before executing a response, regardless of the processing achieved at that moment in time. This assumption has its precedent in earlier models of visual lexical decision (Coltheart et al., 1976). The notion of a self-imposed deadline leads to a further assumption regarding subjects' behavior in the lexical decision task, which we will refer to as the accuracy assumption. The accuracy assumption states that differences in classification accuracy will only be observed when the response time deadline has been exceeded. If stimulus processing is completed before the deadline, few errors in classification should be observed. However, if stimulus processing is incomplete at the expiration of the deadline, subjects will be forced to execute a response based on only partial information, thus producing significant errors in classification. In this case, lexical decisions for both words and nonwords will be based on the total activity level of the system. After expiration of the reaction time deadline, word decisions will not depend on the recognition criterion being exceeded, but will instead be forced based on the total lexical activity in the system. Simply put, according to the accuracy assumption, only those stimuli requiring processing times exceeding the response time deadline should produce errors in classification. In addition, if certain stimuli consistently require processing times that exceed the deadline, reaction times to these stimuli will tend to be equal to the deadline itself. Again, if the deadline is reached before exceeding either the word or nonword criteria, classification is based on the overall level of activity in the decision system. If this activity is high, a word response will be executed; otherwise, a nonword response will be made.

To summarize, lexical decisions in the NAM are assumed to be made via the decision units for words. Word responses are executed once a decision unit surpasses the upper level, word criterion. Nonword responses are executed once the total level of activation for words falls below the lower level, nonword criterion. Furthermore, if neither of these criteria are exceeded before a self-imposed reaction time deadline, a decision is forced based on the overall activity in the decision system, which may lead to erroneous classification responses.

Specifically, it is predicted that, in the long run, high-frequency words will tend to reach criterion before the response time deadline. Thus, effects of neighborhood structure should be observed on classification times. However, decisions regarding low-frequency words may be forced at the response-time deadline, thus producing neighborhood structural effects on classification accuracy for low-frequency words. Furthermore, it is predicted that decisions at the deadline will be based on the total level of activity in the system. Therefore, a decision regarding a word made at the deadline may be more accurate if the overall activity of the system favors a word(i.e., if a word from a dense neighborhood has been presented) than if the overall activity of the system is low (i.e., if a word from a sparse neighborhood has been presented). It is furthermore predicted that the time to respond to a nonword will be a function of the number and frequency of words activated by that nonword. That is, nonwords with many word neighbors should be responded to more slowly than nonwords with few word neighbors. Likewise, it is predicted that nonwords with high-frequency word neighbors will be responded to more slowly than nonwords with low-frequency word neighbors. Finally, it is predicted that the accuracy of classifying nonwords at the response time

deadline with also reflect the total activation of the system. If this activation is high, accuracy for responding "nonword" should be low, and vice versa.

Back to Top

Method

Stimuli • The same 918 words used in the perceptual identification study were used in the auditory lexical decision experiment. These words were partitioned into eight cells, created by orthogonally combining two levels of stimulus word frequency (high and low), two levels of neighborhood density (high and low), and two levels of mean neighborhood frequency (high and low). The method of computing the similarity neighborhood statistics was described above. The partitioning of the stimuli was achieved by performing median splits on the values of each of the three independent variables: stimulus word frequency, neighborhood density, and mean neighborhood frequency. That is, the median frequency of the stimulus words was first determined and words falling above the median were coded as high-frequency words; those equal to or less than the median were coded as low-frequency words. The same procedure was applied to assign words to high density neighborhoods or low density neighborhoods and to high-frequency neighborhoods or low-frequency neighborhoods. The resulting high-frequency words had a mean raw frequency of 254.12; low-frequency words had a mean raw frequency of 5.22. High density neighborhoods contained an average of 21.92 neighbors; low density neighborhoods contained an average of 11.07 neighbors. The mean raw frequency of high-frequency neighborhoods was 370.32; the mean raw frequency of low-frequency neighborhoods was 46.29. The method of stimulus preparation was described in Experiment 1.

To construct a list of phonotactically legal nonwords matched in phoneme length to the word stimuli, a lexicon of nonwords was constructed in the following manner: for all three phoneme words in Webster's lexicon, all initial two-phoneme sequences and all final two-phoneme sequences were determined. That is, for all three phoneme words, P1-P2-P3, all P1-P2 sequences and all P2-P3 sequences were determined. All initial and final sequences sharing P2 were then combined. Three phoneme sequences not containing a vowel were excluded. Also excluded were any sequences corresponding to a real word in Webster's lexicon. Because both the initial and final biphones of the nonwords actually occurred in real words, the resulting nonword lexicon thus contained three phoneme sequences that strongly followed the phonotactic constraints of real words. Altogether, 3123 nonwords were constructed.

Similarity neighborhood statistics for each of the 3123 nonwords were then computed. The similarity neighborhood statistics were computed by comparing each nonword with each word in Webster's lexicon. Similarity neighborhoods were once again computed on the basis of one phoneme substitutions, additions, and deletions. For the nonwords, two variables were of interest: 1) neighborhood density, or the number of words similar to a given nonword, and 2) mean neighborhood frequency, or the mean frequency of the words similar to a nonword. Note that the neighborhood statistics for the nonwords were computed based on words only.

Three hundred four stimuli were selected from the nonword lexicon that fell into one of four cells, resulting in 76 nonwords per cell. The four cells were produced by crossing two levels of density (high and low) with two levels of mean neighborhood frequency (high and low). These cells were: 1) high density-high neighborhood frequency, 2) high density-low neighborhood frequency, 3) low density-high neighborhood frequency, and 4) low density-low neighborhood frequency. Selection of the 76 nonwords for each of the four cells was achieved via an algorithm that first rank-ordered each of the 3123 nonwords on each of the two independent variables. A method of minimizing and maximizing squared deviations of successively ranked nonwords was then employed to ensure that cells that were matched on a given variable (e.g., both high density) were maximally alike and that cells intended to differ on a given variable (e.g., one high and one low density) were maximally different. Nonwords occurring in high density neighborhoods had an

average of 17.78 neighbors; nonwords occurring in low density neighborhoods had an average of 8.10 neighbors. The mean raw frequency of high-frequency neighborhoods was 156.96; the mean raw frequency of low-frequency neighborhoods was 11.84.

The nonwords were recorded by the same male speaker who produced the words. The method for recording and digitizing the nonwords was identical to that for the words (see Experiment 1). Overall RMS amplitude for the nonwords was equated to the overall amplitudes of the words.

Subjects • Thirty subjects participated in partial fulfillment of an introductory psychology course. All subjects were native English speakers and reported no history of speech or hearing disorders.

Design and Procedure • Three stimulus set files were constructed by combining each of the three set files of words with the set file containing the nonwords. Each of the three set files thus contained 306 words and 304 nonwords, producing 610 stimuli. (Three hundred four nonwords, instead of 306, were selected to enable equal partitioning of the nonwords into the four cells. It was assumed that the slight discrepancy between the number of words and nonwords would have little or no effect on the obtained results given the large number of words and nonwords used.) Each set file was presented to a total of 10 subjects. Thus, 10 observations were obtained for each word, whereas 30 observations were obtained for each nonword, given that the same set of nonwords were presented to each group of subjects.

Stimulus presentation and data collection were controlled by a PDP-11/34 minicomputer. The stimuli were presented via a 12-bit digital-to-analog converter at a 10 kHz sampling rate over matched and calibrated TDH-39 headphones at a comfortable listening level of 75 dB SPL.

Groups of six or fewer subjects were tested in a sound-treated room. Each subject sat in an individual booth equipped with a two-button response box. The button on the left-hand side of the response box was labeled "WORD"; the button on the right-hand side of the response box was labeled "NONWORD." A small light was located above each button for feedback. In addition, a cue light was located at the top of the box to warn the subject that a stimulus was about to be presented. Subjects were instructed that they would hear real words in English and nonsense words, or nonwords. They were instructed that after presentation of each stimulus, they were to respond whether they heard a word or a nonword by pressing the appropriately labeled button on the response box. The instructions stressed both speed and accuracy.

A given trial proceeded as follows: the cue light at the top of the response box was illuminated for one second to warn the subject that a stimulus was about to be presented. Five hundred msec after the offset of the cue light, a randomly selected auditory stimulus was presented. Immediately after the subject responded word or nonword, the light above the button that should have been pressed for a correct response was illuminated for one second. Reaction times were recorded from the onset of the auditory stimulus to the response. After each subject had responded, a new trial was initiated. If one or more subjects failed to respond within 4000 msec of the onset of the auditory stimulus, incorrect responses for those subjects were tallied and a new trial was initiated. An intertrial interval of 500 msec elapsed between the end of one trial and the beginning of the next. The 610 experimental trials were preceded by 30 practice trials consisting of an equal number of randomly presented words and nonwords. None of the words or nonwords presented in the practice phase of the experiment were presented in the experiment proper. An experimental session lasted approximately 1 hr.

[Back to Top](#)

Results

Analysis of Word Responses • To factor out the effect of stimulus word duration on reaction times, the duration of each stimulus word in msec was subtracted from each subject's reaction time to that word. The adjusted reaction times for correct word responses were then entered into a stimulus word-by-subjects array. Means and standard deviations for each stimulus as well as each subject were then computed. Those

reaction times falling above or below 2.5 standard deviations of both the subject and stimulus means were deleted and replaced according to the procedure suggested by Winer (1971). Reaction times and percentages correct for each subject were averaged across words within a cell and submitted to analyses of variance. Because the entire set of words was split into thirds and presented to separate groups of equal numbers of subjects, a grouping factor was included in the analysis, producing a 2 (stimulus word frequency)× 2 (neighborhood density) × 2 (neighborhood frequency) × 3(groups) analysis of variance.

Accuracy • Analysis of the percentages correct revealed significant effects of stimulus word frequency, $F(1,27) = 135.94$; $p < 0.05$, neighborhood density, $F(1,27)= 39.39$; $p < 0.05$, and mean neighborhood frequency,$F(1,27) = 4.93$; $p < 0.05$. No effect of groups was obtained, $F(2,27) = 3.13$; $p > 0.05$. A significant interaction of stimulus word frequency and neighborhood density was also obtained, $F(1,27) = 17.07$; $p < 0.05$. No other interactions were significant. Mean percentages correct and standard deviations for each cell, collapsed across groups, are shown in Table 4.

On the average, high-frequency words were responded to 7.39% more accurately than low-frequency words. Words in high density neighborhoods were responded to 3.38% more accurately than words in low density neighborhoods. And, words occurring in low-frequency neighborhoods were responded to 1.39% more accurately than words occurring in high-frequency neighborhoods. Although words in high density neighborhoods were classified correctly more often than words in low density neighborhoods, the significant interaction of stimulus word frequency and density indicates differential effects of neighborhood density as a function of word frequency.

Separate analyses based on the interaction of word frequency and density revealed significant effects of word frequency at both levels of density. High-frequency words were responded to 4.48% more accurately than low-frequency words in high density neighborhoods, $F(1,27) = 40.48$; $p < 0.05$. High-frequency words were also responded to 10.31% more accurately than low-frequency words in low density neighborhoods, $F(1,27) = 81.51$; $p < 0.05$. Thus, significant effects of frequency were observed at each level of neighborhood density.

Separate analyses revealed no significant effect of density for high-frequency words for the accuracy data, $F(1,27) < 1.0$. However, a significant effect of density was observed for low-frequency words, $F(1,27) = 6.29$; $p < 0.05$. For low-frequency words, words in high density neighborhoods were responded to 6.29% more accurately than words in low density neighborhoods. Thus, an effect of density on classification of words was observed only for low-frequency words.

Reaction Times • Analysis of the reaction time data revealed significant main effects of word frequency, $F(1,27) = 70.39$; $p < 0.05$, neighborhood density, $F(1,27)= 14.32$; $p < 0.05$, and neighborhood frequency,$F(1,27) = 14.15$; $p < 0.05$. No effect of groups was obtained, $F(2,27) = 2.32$; $p > 0.05$. A significant interaction of word frequency and neighborhood density was also obtained,$F(1,27) = 14.15$; $p < 0.05$. Means and standard deviations for each cell are shown in Table 5.

Overall, high-frequency words were responded to 55 msec faster than low-frequency words. Words occurring in high density neighborhoods were responded to 13.5 msec lower than words in low density neighborhoods. And, words in high-frequency neighborhoods were responded to 17.5 msec slower than words in low-frequency neighborhoods. Although there was an overall 13.5 msec advantage for words in low density neighborhoods over words in high density neighborhoods, the significant interaction of word frequency and density indicates differential effects of one or both of these variables.

Separate analyses based on this interaction revealed significant effects of word frequency at each level of density. In high density neighborhoods, high-frequency words were responded to 47.5 msec faster than low-frequency words, $F(1,27) = 60.35$; $p < 0.05$. In low density neighborhoods, high-frequency words were responded to 62 msec faster than low-frequency words, $F(1,27) = 54.45$; $p < 0.05$. Significant effects of

neighborhood density were obtained only for high-frequency words for the reaction time data. For high-frequency words, words, in low density neighborhoods were responded to 21 msec faster than words in high density neighborhoods, $F(1,27) = 18.33; p < 0.05$. No effect of density was observed for low-frequency words, $F(1,27) = 1.75; p > 0.05$.

Analysis of Nonword Responses • Before analysis of the nonword response data, stimulus durations were subtracted from the correct nonword reaction times and outliers were eliminated and replaced according to the procedure described for the word response data. The accuracy and reaction time data were then submitted to analyses of variance.

Accuracy • A two-way repeated measures analysis of variance (neighborhood density × neighborhood frequency) on the accuracy data for the nonwords revealed significant main effects of neighborhood density, $F(1,29) = 26.54; p < 0.05$, and neighborhood frequency, $F(1,29) = 17.68; p < 0.05$. In addition, the interaction of neighborhood density and neighborhood frequency was significant, $F(1,29) = 24.75; p < 0.05$. Means and standard deviations for the accuracy data for each cell are shown in Table 6.

Nonwords occurring in high density neighborhoods were responded to 3.34% worse than non-words occurring in low density neighborhoods. Non-words occurring in high-frequency neighborhoods were responded to 3.02% worse than nonwords occurring in low-frequency neighborhoods. Separate analyses based on the significant neighborhood density-by-neighborhood frequency interaction revealed a significant effect of density only for nonwords in high-frequency neighborhoods, $F(1,29) = 51.03; p < 0.05$. In high-frequency neighborhoods, nonwords having many word neighbors were responded to 5.53% worse than nonwords having few word neighbors. In addition, a significant effect of neighborhood frequency was observed only for nonwords occurring in high density neighborhoods, $F(1,29) = 30.49; p < 0.05$. In high density neighborhoods, nonwords with high-frequency neighbors were responded to 5.22% worse than nonwords with low-frequency neighbors. Both of these effects appear to be due to the lower mean percent correct for nonwords occurring in high density, high-frequency neighborhoods.

Reaction Time Data • For the reaction time data for correct nonword classifications, significant main effects were obtained for neighborhood density, $F(1,29) = 60.81; p < 0.05$, and neighborhood frequency, $F(1,29) = 5.39; p < 0.05$. The interaction of neighborhood density and neighborhood frequency was not significant, $F(1,29) < 1.0$. Means and standard deviations for the reaction times for each cell are shown in Table 7.

Nonwords occurring high density neighborhoods were classified 39.5 msec slower than nonwords occurring in low density neighborhoods. In addition, nonwords occurring in high-frequency neighborhoods were classified 11.5 msec slower than nonwords occurring in low-frequency neighborhoods.

Back to Top

Discussion

Back to Top

Word Responses

Three main effects were observed for the word response data. High-frequency words were classified more quickly and more accurately than low-frequency words. Words in low-frequency neighborhoods were classified more quickly and more accurately than words in high-frequency neighborhoods. Finally, words occurring in high density neighborhoods were classified more slowly but more accurately than words in low density neighborhoods. Although this last result suggests a speed-accuracy trade-off, the significant interactions of neighborhood density and word frequency for both reaction times and accuracy revealed differential effects of density on accuracy and reaction time as a function of word frequency. High-frequency words in high and low density neighborhoods were classified equally as accurately. However, classification time was slower for high-frequency words in high density neighborhoods. A different pattern

of results was observed for the low-frequency words. No reaction time differences were observed as a function of neighborhood density. However, low-frequency words in high density neighborhoods were classified more accurately than low-frequency words in low density neighborhoods.

Before considering the interesting interactions of density and word frequency for the accuracy and reaction time data, we consider first how the NAM explains the effects of word frequency and neighborhood frequency on word classification responses. Recall that word frequency increases the activation level of the acoustic-phonetic patterns represented in the decision units. Thus, high-frequency words will tend to have higher levels of activation in the decision system than low-frequency words. These higher levels of activation thus lead to faster word responses for high-frequency words given that the word criterion is surpassed more quickly as stimulus processing proceeds. Thus, high-frequency words show faster reaction times than low-frequency words. And, given the accuracy assumption stated above, slower processing times associated with low-frequency words will result in higher error rates.

Neighborhood frequency affects classification times for words by slowing the time for a decision unit to reach criterion. Recall that the decision units monitor overall activity in the decision system. Because high-frequency neighborhoods result in overall higher activity levels, the time for a given decision unit to surpass the criterion will be extended. Thus, high-frequency neighbors serve as stronger competitors by virtue of the fact that they raise the overall level of activity within the decision system.

Effects of neighborhood density can be explained by the same basic principle used to explain the effects of neighborhood frequency. In particular, heightened overall activity in the decision system extends the time needed for a given decision unit to surpass the criterion. Words with many neighbors produce high levels of activity in the decision system, thus slowing response time. Such an effect was observed for high-frequency words. No effect of density was observed for the accuracy data for high-frequency words because decisions for high-frequency words, in the long run, were made before the response-time deadline, as predicted. However, no effect of density on reaction time was observed for low-frequency words. The failure to observe reaction time differences for low-frequency words may have arisen from the fact that, on the average, decision units failed to surpass the criterion level for the low-frequency words by the time the response-time deadline had expired. Thus, reaction times for the low-frequency words simply reflect a forced decision at the deadline. According to the accuracy assumption, then, density effects should be observed only for the accuracy data. Indeed, we found that low-frequency words in high density neighborhoods were classified more accurately than low-frequency words in low density neighborhoods. Given that decisions that are forced at the response time deadline are based on the overall level of activity in the decision system, low-frequency words with many word neighbors would have higher levels of overall activity and would thus be classified more accurately as words. The seemingly counterintuitive finding that low-frequency words with many neighbors were classified more accurately as words than low-frequency words with few neighbors, therefore, can be accounted for by the NAM.

[Back to Top]

Nonword Responses

The results for the reaction time data for the nonwords are also easily accounted for by the model. Recall that it was assumed that a nonword response is executed whenever the overall activity in the decision system falls below a lower level criterion. Thus, any factor that slows the time for the activity level to drop in the decision system should slow the time to correctly classify a nonword pattern. Indeed, the results for the nonwords showed significant main effects of neighborhood density and neighborhood frequency. Nonwords with many neighbors were responded to more slowly than nonwords with few neighbors. Because the activity level in the decision system takes longer to decay when there are many similar words activated by the nonword stimulus, nonword classification times were longer for nonwords in high density neighborhoods than for nonwords in low density neighborhoods (Forster, 1976; Rubenstein, Richter, &

Kay, 1975). The same reasoning applies to nonwords in high-frequency neighborhoods. Given the overall higher activity level associated with high-frequency neighborhoods, nonwords with highly frequent neighbors were classified more slowly than words with low-frequency neighbors.

For the accuracy data for the nonwords, an interaction of density and neighborhood frequency was observed. This interaction was due to one cell, namely, the cell containing nonwords occurring in high density, high-frequency neighborhoods. The accuracy levels for all other nonwords were approximately equal. The mean reaction time for nonwords occurring in high density, high-frequency neighborhoods was also the longest, and was approximately equal to the maximum reaction time observed in the experiment as a whole. Thus, under the accuracy assumption, it is possible that the reduced accuracy for these nonwords was due to response-time deadline expiration. In the case in which a nonword stimulus activates a set of high-frequency similar words, the overall activity level in the decision system may have at times failed to drop below the criterion for a nonword response. In this case, the overall activity level in the system would be examined to determine a response. Given that the activity level would tend to be high for nonwords with many high-frequency word neighbors, errors in nonword classification would be expected to arise in this case.

Back to Top

Summary and Conclusions

The data from both the word and nonword responses revealed significant effects of neighborhood structure on lexical decision classification time and accuracy. The effects of neighborhood structure on spoken word recognition are, therefore, not restricted to degraded stimuli. In addition, the reaction times for the word and nonword responses support the conclusions drawn from the previous perceptual identification study. Although the effect of density on classification accuracy for low-frequency words ran counter to the results observed in the identification study, the NAM can, in fact, account for this result via the same mechanisms invoked to account for the identification data.

The NAM thus provides a coherent framework for interpreting the effects of neighborhood structure on word and nonword classification times. Although the interpretation of the effects of neighborhood structure on both word and nonword responses relies, in part, on a small number of assumptions regarding the behavior of subjects in the lexical decision task, the NAM permits a unified account of the data observed in the lexical decision task. Many of the particulars of the model, and the characterization of subjects' behavior in the lexical decision task, necessarily require further independent testing based on the predictions of the model. Nevertheless, the results show explicable and consistent effects of neighborhood structure. Furthermore, the data obtained from the lexical decision task reveal that neighborhood structure is an important determinant of the speed and accuracy with which words are recognized in this experimental paradigm.

Back to Top

Experiment 3: Evidence from Word Naming

The present experiment attempts to gather further converging support for the NAM developed in the previous studies by examining the effects of neighborhood structure in another experimental paradigm. Specifically, the paradigm used in the present study is word naming (Andrews, 1982; Forster, 1981; Forster & Chambers, 1973; Frederiksen & Kroll, 1976). In the word naming task, a subject is presented with a spoken word and is required to repeat or pronounce the word as quickly as possible. The dependent variable is the time required to initiate the naming response.

The use of the naming task is motivated by a number of factors. First, as shown previously, auditory lexical decision proves somewhat problematic in examining neighborhood density and frequency effects by virtue of the fact that the task requires discrimination among both word and nonword patterns. Although the

results from the previous lexical decision study provided evidence for the effects of neighborhood structure, these effects may have been attenuated or altered because subjects were required to consider both words and nonwords in making their responses. In the absence of any systematic control for the nonword patterns that may be activated in memory by subjects, manipulation of neighborhood structure on the basis of words alone makes precise control of neighborhood structure difficult in a task requiring discrimination among both words and nonwords. Thus, the naming task provides a means of collecting reaction times to word stimuli without presenting nonwords at the same time.

A second motivation for using the naming task comes from findings in the visual word recognition literature on the role of word frequency in the naming task. Balota and Chumbley (1984) have presented evidence that word frequency effects are severely reduced in the visual word naming task, as compared with the lexical decision task. In addition, there researchers have argued that the small word frequency effects obtained in the naming task are due to factors related to the pronunciation of the visually presented item and not to the frequency of the word itself (Balota & Chumbley, 1985). Finally, Paap et al. (1986) have argued that the visual word naming task circumvents lexical access and thus attenuates effects of frequency. Paap et al. argue that naming a visually presented word simply requires grapheme-to-phoneme conversion with no access to the representations stored in the mental lexicon.

These findings suggest an interesting test of the NAM. Recall that the model proposes that acoustic-phonetic patterns similar to the stimulus input are activated in memory. These patterns then activate decision units corresponding to words that monitor the acoustic-phonetic patterns as well as higher level lexical information, which includes word frequency. The decision units continuously compute probability values based on the activation level of the acoustic-phonetic patterns they monitor and the overall level of activity within the system. Frequency information is assumed to bias the decision units by adjusting the activation levels of the acoustic-phonetic patterns represented in the units.

The system of decision units is, therefore, driven by the activation of acoustic-phonetic patterns, giving acoustic-phonetic pattern similarity priority in the decision system. Word frequency, on the other hand, serves as a biasing factor that may or may not come into play in the decision-making process, depending on the requirements of the task situation. Thus, in the model, similarity and frequency effects arise from two distinct sources and operate in fundamentally different ways in influencing decisions. If the visual and auditory word naming tasks are sufficiently similar, auditory word naming should not be sensitive to the biasing properties of word frequency information. The model predicts, however, that pattern similarity is fundamental to the system of decision units and cannot be bypassed, at least in situations involving an open-response set. Therefore, if the predictions of the NAM are correct, robust effects of neighborhood density should be observed on naming times regardless of whether frequency information acts to bias the decision units. In particular, high density neighborhoods should produce longer naming times than low density neighborhoods. However, if the auditory naming task does not invoke the biasing properties of word frequency information, as predicted by the visual word naming studies, no effects of stimulus word frequency or neighborhood frequency should be observed. The auditory naming task may, therefore, help to dissociate the effects of similarity and bias (i.e., frequency), providing further support for the fundamental predictions of the NAM.

Back to Top

Method

Stimuli • Four hundred CVC words were selected from the 811 words used in the perceptual identification and auditory lexical decision experiments. The words were chosen to construct eight cells with 50 words per cell. The eight cells were constructed by orthogonally combining two levels(high and low) of each of the following independent variables: stimulus word frequency, neighborhood density, and neighborhood frequency. The neighborhood density and neighborhood frequency variables were computed for each word

in the same manner as described in Experiment 2. The phonetic transcriptions of each stimulus word were compared with each monosyllabic word in Webster's lexicon having a familiarity rating of 5.5 or above. A neighbor of the stimulus word was defined as any word that could be converted to the stimulus word via a one phoneme addition, substitution, or deletion in any position. Neighborhood density again refers to the number of neighbors of a given stimulus word and neighborhood frequency refers to the mean frequencies of words in a neighborhood.

Selection of the 50 words for each cell was achieved via an algorithm that first rank-ordered each of the 811 words on each of the three independent variables. A method of minimizing and maximizing squared deviations of successively ranked words was then employed to ensure that cells that were matched on a given variable (e.g., both high density) were maximally alike and that cells intended to differ on a given variable (e.g., one high and one low density) were maximally different. In addition, words were chosen such that the mean stimulus durations for each cell were not significantly different. High-frequency words had a mean of 145.95; low-frequency words had a mean of 4.33. Words occurring in high density neighborhoods had an average of 22.12 neighbors; words occurring in low density neighborhoods had an average of 11.44 neighbors. The mean frequency of high density neighborhoods was 245.17; the mean frequency of low-frequency words was 60.50. Details of the preparation of the auditory stimuli are given in Experiment 1.

Subjects • Eighteen subjects participated as paid volunteers. Subjects received $3.50 for their participation. All subjects were native English speakers and reported no history of speech or hearing disorders.

Back to Top

Design and Procedure

The 400 stimulus words were combined in a single stimulus set file. Each subject heard each of the 400 stimulus words in a different random order. In addition, subjects were given 30 practice trials before the experiment proper on a separate set of words. Stimulus presentation and data collection were controlled by a PDP 11/34 minicomputer. The stimuli were presented on matched and calibrated TDH-39 headphones at a comfortable listening level of 75 dB SPL.

Each subject was run individually in a sound treated room. An Electro-Voice D054 microphone was situated immediately in front of the subject. The microphone was connected to a voice key that was interfaced to the PDP 11/34. The voice key registered a response at the onset of the subject's naming response. The subject was positioned such that his/her lips were approximately 12 inches from the microphone. The subject was instructed to maintain the 12 inch distance at all times. In addition, each subject was instructed to avoid unnecessary noise.

The subjects were instructed that they would hear words over their headphones that they were to repeat back or name as quickly and as accurately as possible. The subjects were told that the microphone would register when they began speaking and that the time it took them to name the stimulus would be recorded by the computer. The experimenter was seated in a booth next to the subject and monitored a CRT terminal. On each trial, the stimulus word appeared on the terminal and the experimenter listened to the subject's naming response. After the subject's response, the experimenter would indicate on the computer terminal whether the subject had responded correctly or incorrectly. If an incorrect response was made, the experimenter typed the mistake on the terminal.

A given trial proceeded as follows: subjects first saw the prompt "GET READY FOR NEXT TRIAL" on a CRT screen situated above the microphone. One second after the prompt, a word was presented over the headphones. The experimenter then scored the subject's response and initiated a new trial. Reaction times were measured from the onset of the auditory stimulus to the onset of the subjects response. A given experimental session lasted approximately 1 hr.

Back to Top

Results

Reaction times were first entered into a stimulus word-by-subject array and means and standard deviations were computed for each word and each subject. Any reaction time falling 2.5 standard deviations above and below both the subject and stimulus means was eliminated and replaced according to the procedure suggested by Winer (1971). Mean reaction times were then computed for each subject for each cell. Preliminary inspection of the data revealed correlations of reaction time with the identity of the initial segment of the stimulus word. In particular, fricatives were associated with longer reaction times, presumably due to the differential sensitivity of the microphone and voice key to the identity of the initial segment of the naming response. Because the identity of initial segments was not evenly distributed across cells, it was deemed necessary to factor out the effects of the initial segment. To do this, the number of initial segments falling in each of six manner classes for each cell was tallied. The six manner classes were: stops, strong fricatives, weak fricatives, nasals, liquids and glides, and affricates. These numbers were then entered as covariates for the reaction times in a repeated measures analysis of variance. In addition to the reaction time data, percentages of correct responses were computed for each subject for each cell and submitted to analysis of variance.

Accuracy Data • A 2 (stimulus word frequency) × 2(neighborhood density) × 2 (neighborhood frequency) repeated measures analysis of variance was computed on percent correct responses. Main effects of stimulus word frequency, $F(1,17) = 24.73$; p< 0.05, and neighborhood density, $F(1,17) = 4.89$;p < 0.05, were obtained. No effect of neighborhood frequency was observed, $F < 1.0$. In addition, none of the interactions were significant. Means and standard deviations are shown in Table 8 for each cell.

Inspection of Table 8 reveals that the significant effects of stimulus word frequency and neighborhood density on accuracy were extremely small. High-frequency words were responded to 0.89% better than low-frequency words. Words occurring in high density neighborhoods were responded to 0.28% worse than words occurring in low density neighborhoods. In addition, it should be noted that the accuracy levels overall were very high, the lowest cell percentage being 96.78%. Thus, naming an auditory stimulus appears to be quite easy for subjects to perform.

Back to Top

Reaction Times

A 2 × 2 × 2 repeated measure analysis of covariance was performed on the reaction times. Recall that the covariates were the number of initial segments in each cell falling into one of the six manner classes. Only the main effect of neighborhood density was observed for the reaction times, $F(1,11) = 5.10$; p < 0.05. Neither stimulus word frequency, $F(1,11) = 1.71$; p < 0.05, nor neighborhood frequency, $F(1,11)<1.0$, reached significance. In addition, no significant interactions were observed. Means and standard deviations are shown in Table 9.

Overall, words occurring in high density neighborhoods were responded to approximately 102 msec slower than words in low density neighborhoods. The effect of density was consistent across word frequency and neighborhood frequency in all cases but one. Virtually no reaction time differences between words in high and low density neighborhoods were observed for low-frequency words occurring in high-frequency neighborhoods. However, the three-way interaction suggesting a lack of statistical significance for this cell was far from significant, $F(1,11) = 0.64$. Thus, although there was clearly a reduction of the effect of density for low-frequency words in high-frequency neighborhoods, there was no statistical support for any differential effects of density across word frequency or neighborhood frequency.

In summary, small effects of word frequency and neighborhood density were observed for the accuracy data. Although both effects were in the predicted direction, the magnitude of the differences were so small as to be almost negligible. However, a large effect of density was observed for the reaction time data.

Words occurring in low density neighborhoods were named an average of 102 msec faster than words in high density neighborhoods.

[Back to Top](#)

Discussion

The results of the naming study lend further support to the hypothesis that the neighborhood structure of words in the mental lexicon strongly affects spoken word recognition. The reaction time data demonstrate that words with many neighbors are named more slowly than words with few neighbors. Perhaps the more interesting and crucial finding, however, is that no frequency effects on reaction times were observed either in terms of stimulus word frequency or neighborhood frequency. This is in contrast to the findings from the perceptual identification and auditory lexical decision studies, in which consistent effects of frequency were observed. Although the present study examined only a subset of the words used in the previous studies, the number of stimuli was still quite large, and the difference in frequency between the high- and low-frequency words was substantial (mean for high-frequency words = 145.95; mean for low-frequency words = 4.33), as was the difference in mean frequency between high- and low-frequency neighborhoods (mean for high-frequency neighborhoods = 245.17; mean for low-frequency words = 60.15). There is no reason to expect that the use of a subset of stimuli was responsible for the lack of frequency effects. Instead, the failure to observe word frequency and neighborhood frequency effects may lie in the attenuation of bias effects in the naming task. As previously discussed, a precedent for the finding that frequency does not affect naming times can be found in the literature on visual word recognition.Balota and Chumbley (1984) compared lexical decision and naming times for high- and low-frequency printed words and found a marked reduction in the frequency effect for naming as compared with lexical decision. In a later study, Balota and Chumbley (1985) found that frequency effects in the naming of visually presented words may be due to the structure of the words themselves. They argued that any frequency effects observed in the naming task are due to factors correlated with frequency that affect articulation. The main conclusion to be drawn from these visual word recognition studies is that the naming task fails to reveal frequency effects that are unrelated to differences in the articulation of high- and low-frequency words.

Paap et al. (1986) have argued that the naming task for visually presented words does not require lexical access. Instead, they claim that a visually presented word may be named via a grapheme-to-phoneme route that bypasses the lexicon. The authors assume that frequency effects are only apparent once lexical access has been achieved, at which time frequency information is made available to the processing system. Paap et al. in fact showed that when the naming task is modified to require multiple lexical decisions before the naming response, large and consistent effects of frequency are observed. The effect of the lexical decisions, therefore, is presumably to force subjects to access the lexicon, thus gaining access to frequency information.

Although the present results corroborate the finding that frequency effects on reaction times are not obtained in the naming task (when highly controlled stimuli are used), the explanation put forth by Paap et al. is not sufficient to account for the finding that neighborhood density affects naming time. As previously mentioned, Paap et al. argue that visually presented words can be named by a direct mapping of graphemes onto phonemes; units corresponding to words need not be activated to name the visual stimulus. If one assumes that an auditorily presented word can be named by some sort of process that maps phoneme-to-phoneme, again circumventing activation of units corresponding to words, no effect of neighborhood density should be observed. This prediction was not supported by the results obtained in the present study, in which a large effect of neighborhood density was found.

It seems more reasonable to assume that the naming task requires activation of units corresponding to words, namely, the decision units in the NAM, but that word frequency information does not bias these decision units. The crucial question arises, then, as to why the naming task does not invoke the biasing

properties of word frequency. One explanation may lie in the nature of the response required by the naming task. In this task, subjects are simply required to repeat back the stimulus word. No explicit decision is required regarding the lexical status of the item, as in the lexical decision task. Frequency biases in lexical decision may help to optimize decision times by allowing certain items to surpass the word or nonword criterion faster than would be expected with no bias. Nor is an explicit decision required as to the identify of a degraded item, as in perceptual identification. In the perceptual identification task, subjects may optimize performance by choosing words of higher probabilities of occurrence in the face of incomplete stimulus information. In contrast, no higher level lexical information is required to make a naming response. Behavior in this task is optimized by simply deciding on the acoustic-phonetic identity of the stimulus word. Because the naming response requires a precise analysis of the acoustic-phonetic properties of the stimulus word to build an articulatory plan for executing a response, biases not based on the acoustic-phonetics themselves (e.g., frequency biases) may indeed hinder response generation. Given the response required by the naming task, therefore, subjects may optimize performance by focusing on discriminating among the acoustic-phonetic patterns and ignoring higher level lexical information. Thus, frequency effects would not be expected to affect naming times. However, because an acoustic-phonetic pattern must be isolated to make the naming response, neighborhood density, or the number of similar acoustic-phonetic patterns corresponding to words in memory, would be expected to influence the time needed to generate a naming response. Indeed, precisely this pattern of results was obtained in the present study.

The results of the present study, therefore, demonstrate that neighborhood density effects are clearly separate from frequency effects. Thus, the present study demonstrates that stimulus similarity and decision biases are separate factors that have differential effects on the levels of processing within the word recognition system. The present study also confirms the prediction of the NAM that stimulus word frequency and neighborhood frequency effects must occur, or not occur, in tandem. That is, the model does not predict stimulus word frequency effects when no neighborhood frequency effects are present, and vice versa. Thus, the absence of one effect requires the absence of the other.

In summary, the results of the auditory naming task provide additional support for the basic predictions of the NAM, in particular, the claim that density effects are distinct from the effects of frequency. The results also provide further support for the proposal that neighborhood structure affects spoken word recognition. Increasing the size of the neighborhood increases the time needed to discriminate among items in memory. Finally, once again, the results from the present study demonstrate that effects of neighborhood structure can be obtained even when stimulus words are not degraded. Thus, the results from the present auditory naming task mesh well with the fundamental predictions of the NAM and provide further support for the claim that a word is recognized in the context of similar sounding words in memory.

Back to Top

General Discussion

The goal of the present investigation was to examine how the structural organization of the sound patterns of words in memory influences spoken word recognition. In addition to structural relations among words, the relative effects of word frequency were examined in the context of similarity neighborhoods. It was hypothesized, and subsequently confirmed, that stimulus word frequency effects are a function of the neighbors of the stimulus word as well as the frequencies of these neighbors. In particular, in Experiment 1, we demonstrated that accuracy of identifying words in noise was best accounted for by simultaneously taking into account word frequency and neighborhood confusability. In Experiment 2, we demonstrated that neighborhood variables have demonstrable effects on processing times for spoken stimuli that were not degraded by noise. Specifically, items in dense and/or high-frequency neighborhoods tended to be

processed less quickly and/or less accurately than items in more sparsely populated, lower frequency neighborhoods. Finally, in Experiment 3, we observed effects of neighborhood density on naming times in the absence of effects of word frequency and neighborhood frequency. This latter finding supports the hypothesis that effects of word and neighborhood frequency may operate via perceptual biases.

Back to Top

Summary of the NAM

The NAM was initially developed to explain the results of the perceptual identification experiment and extended to account for the findings from the auditory lexical decision and auditory word naming studies. Basically, NAM assumes that a set of similar acoustic-phonetic patterns are activated in memory on the basis of stimulus input. The activation levels of these patterns are assumed to be a direct function of their similarity to the stimulus input. Words emerge in NAM when a system of word decision units tuned to the acoustic-phonetic patterns are activated. Once activated, the decision units monitor a number of sources of information. The first source of information is the activation of the acoustic-phonetic patterns, which have previously served to activate the decision units themselves. The word decision units also monitor the overall level of activity in the decision system itself, much like processing units monitor the net activity level of the system in the TRACE model of speech perception (Elman & McClelland, 1986; McClelland & Elman, 1986). Finally, the decision units are tuned to higher level lexical information, which includes word frequency. This information serves to bias the decisions of the units by weighting the activity levels of the acoustic-phonetic patterns by the frequencies of the words to which they respond. The values that serve as the output of the decision units are assumed to be computed via a rule similar to the FWNPR discussed above.

Back to Top

Comparison of the NAM to Other Models of Spoken Word Recognition

As previously mentioned, NAM bears a strong resemblance to other models of spoken word recognition, and many of the concepts incorporated in the model have precedents in previous accounts of word recognition. However, as will be argued below, the model makes certain predictions that are inconsistent with current models of spoken word recognition, in particular with regard to the roles of frequency and similarity. We now turn to a discussion of some of the more influential models of word recognition to highlight the fundamental differences and similarities between NAM and these models.

Back to Top

Logogen Theory

Morton (1969, 1979) has proposed a model of word recognition based on a system of "logogens" that monitor bottom-up sensory information and top-down contextual and lexical information. Information from either of these sources serves to drive the logogens toward threshold. Once a threshold is reached, the information to which the logogen corresponds is made available to the processing system and a word is said to be recognized and accessed. Morton accounts for word frequency effects in the logogen model by assuming that high-frequency words require less evidence than low-frequency words for crossing threshold. Morton thus refers to logogen theory as an evidence-bias model.

The resemblance between Morton's system of logogens and the system of word decision units in the NAM is quite strong. Both logogens and word decision units monitor top-down and bottom-up information. In addition, both logogens and word decision units are assumed to prohibit information from becoming available to the general processing system until a decision regarding the identity of the word has been made. However, word decision units differ from logogens in a number of important ways. Perhaps the most crucial difference between logogens and the word decision units hinges on the problem of accounting for neighborhood structural effects. Logogens are assumed to be independent processing units with no interconnections to other units. The lack of cross-talk among logogens makes it difficult to account for the

findings that words in dense or confusable neighborhoods take longer to respond to than words in less dense or less confusable neighborhoods. Because logogens are independent processing units, stimulus input should push a given logogen over threshold at the same point in time, regardless of whether the stimulus input activates many or few logogens. Granted, accuracy differences between dense and sparse neighborhoods may arise because there is a higher probability that logogens corresponding to similar words may surpass threshold before the logogen corresponding to the stimulus input. It is not so clear, however, how logogen theory would account for neighborhood density effects on reaction times. When presented with clearly specified acoustic-phonetic information, as in auditory lexical decision or auditory word naming, the logogen corresponding to the stimulus input should always cross threshold at the same point in time, regardless of the activity levels of other logogens, assuming that word frequency is held constant. The results for the high-frequency words in the auditory lexical decision task contradict this prediction, as do the results for the auditory word naming task.

A fundamental problem that the present set of results poses for logogen theory concerns the robust findings that frequency effects are dependent on the neighborhood structure of the stimulus word. In the perceptual identification study, it was shown that under certain circumstances, high- and low-frequency words may be responded to at equal levels of accuracy. Because logogens corresponding to high- and low-frequency words are assumed to have differing thresholds, low-frequency words should always require more evidence (i.e., more stimulus input) than high-frequency words to cross threshold. Because a single logogen has no knowledge of the activation levels of other logogens in the system, it is difficult to explain within logogen theory how the frequencies of neighbors could influence recognition of the stimulus word. One could again assume that the effects of neighborhood frequency and density on accuracy reflect incorrect logogens surpassing threshold. That is, it is possible that both neighborhood density and neighborhood frequency increase the probability of incorrect logogens reaching threshold, thus depressing accuracy of identification for words occurring in high density and/or high-frequency neighborhoods. However, such an account does not explain the effects of neighborhood density and neighborhood frequency on reaction times observed in our auditory lexical decision experiment. The time for a given logogen to reach threshold cannot be influenced by the activations of other logogens, and thus logogen theory fails to account adequately for the present set of results that demonstrate that spoken words are recognized in the context of phonetically similar words activated in memory.

Finally, logogen theory has no mechanisms for explaining the results of the naming study. Recall that in the naming study we argued that word units must have been accessed by subjects to produce the effect of neighborhood density. However, no effects of word frequency or neighborhood frequency were observed. It is perhaps possible that the thresholds for logogens corresponding to high- and low-frequency words were temporarily equated as a result of some unspecific property of the naming task. However, not only is this solution inelegant and unparsimonious, it calls into question logogen theory's fundamental claim that thresholds are intrinsic to the logogens themselves and arise over time as a function of degree of exposure to words.

A final problem for logogen theory concerns the nonword data from the auditory lexical decision experiment. Coltheart et al.(1976) have proposed that nonword decisions in the logogen model can be made in a similar manner to nonword decisions in the NAM. Specifically, a nonword decision is executed when no logogen fires. However, because the activation levels within the logogens are not available for inspection (i.e., logogens are either above or below threshold), it is difficult to account for the finding that the number and nature of words activated by the nonword stimulus influence reaction time. As logogen theory stands, there is no means for evaluating the overall level of activity in the logogen system, and there is, therefore, no mechanism for making faster decisions to nonwords with fewer neighbors or lower frequency neighbors. The nonword data from the auditory lexical decision experiment thus prove problematic for a system of independent processing units that respond only on surpassing an intrinsic

threshold.

NAM, on the other hand, provides a coherent description of the present set of results by assuming that the decision units are interconnected and that frequency effects arise from biases stemming from higher level sources of information. Modifications of logogen theory may be possible to account for the present results, but it is very likely that the resulting model would bear a strong resemblance to the major theoretical assumptions embodied in NAM. Nonetheless, there are important similarities between the NAM and logogen theory, owing to the fact that the present model incorporates many ideas from logogen theory. In particular, the NAM assumes a system of word decision units that serve as the interface between the acoustic-phonetic input and higher level information, as proposed by logogen theory. However, because of the interconnectedness of the system of word decision units, the NAM is able to account for the effects of neighborhood structure, whereas logogen theory apparently is not.

Cohort Theory • Perhaps the most influential of current models of auditory word recognition is cohort theory, proposed by Marslen-Wilson (1984, 1987, 1989; Marslen-Wilson & Welsh, 1978; Marslen-Wilson & Tyler, 1980). According to this theory, a "cohort" of words is activated in memory on the basis of the initial acoustic-phonetic input of the stimulus word. Words in the cohort are then eliminated by two sources of information: continued bottom-up acoustic-phonetic input and top-down contextual information. That is, words in the cohort are ruled out or deactivated by continued processing of the stimulus information as well as by inconsistent contextual information. A given word is recognized when it is the only word remaining in the cohort.

Cohort theory has provided a number of valuable insights into the temporal processing of spoken words. In previous versions of the theory, however, no attempt was made to account for word frequency effects. In a more recent version of the theory, though, Marslen-Wilson (1987) has incorporated a mechanism for accounting for word frequency effects by assuming words in a cohort have differing levels of activation depending on their frequencies of occurrence. Words with higher levels of activation take longer to eliminate from the cohort than words with lower levels of activation, thus affording at least an initial advantage to high-frequency words. Because the latter version of cohort theory represents a significant improvement over the initial formulation of the theory, only this version will be considered in the present discussion.

Cohort theory and NAM are similar in several respects because both models assume bottom-up priority in the activation of items in memory. Furthermore, both models assume that a set of items is activated and processed in parallel. In addition, both models state that items receive reduced levels of activity as disconfirming acoustic-phonetic information is presented. Unlike cohort theory, however, the NAM at this stage of formulation has little to say about the time-course of effects in the word recognition system, primarily due to the fact that the model was developed on the basis of data from very short words. Indeed, as stated earlier, some of the aspects of cohort theory may have to be incorporated into the NAM to account for the recognition of longer words. Nonetheless, cohort theory and NAM make fundamentally different predictions, at least for short stimuli.

Marslen-Wilson argues that because cohort theory is realized as a parallel system, no effects of set size should be observed on recognition. Words in a cohort are assumed to be activated at no cost. The NAM is also realized as a system of parallel processing units, but the fundamental claim of the NAM is that the nature and number of items activated in memory influence the accuracy as well as the speed of recognition. This prediction stems from the claim that the word decision units are sensitive to the overall level of activity in the decision system and are, therefore, influenced by the number and nature of competing items. Evidence to support this claim was provided by each of the three experiments previously reported.

Marslen-Wilson argues that set size has no effect on recognition performance on the basis of a set of

experiments examining lexical decisions for nonwords. He claims that if nonwords are matched according to the point at which they diverge from words, no effect of set size is observed on reaction times. This is in contradiction to the findings of our lexical decision experiment reported earlier in which large effects of neighborhood density (i.e., set size) and neighborhood frequency were observed for nonwords. Note that because of the manner in which these nonwords were constructed, each of the nonwords diverged from words at the third phoneme(see Experiment 2). Thus, set size effects were demonstrated even when divergence points were equated. Given that Marslen-Wilson's claim of no effects of set size are based on null results with nonwords, the positive findings reported for the nonwords seriously call this claim into question.

Indeed, each of the experiments reported previously fail to support the notion that the number of items activated in memory has no influence on recognition performance. Although Marslen-Wilson may object to the results from our perceptual identification study, claiming that the use of "noisy" stimuli induce postperceptual processes, the results from the lexical decision study taken together with the auditory naming study clearly contradict a fundamental claim of cohort theory. Indeed, it is not even clear that the postulation of some vague "postperceptual" processes indicts the results from the perceptual identification study, which showed significant effects of neighborhood structure on identification performance. In short, the results of the present set of studies considered together refute several fundamental claims of cohort theory.

The results of the naming study also provide counter evidence to cohort theory's treatment of word frequency. All words used in the naming study can be assumed to have had approximately equal divergence points or isolation points by virtue of their short length (Luce, 1986). Indeed, it has yet to be shown for short words that divergence points influence on-line recognition. Thus, one can safely assume that these stimuli did not differ in important ways in terms of their isolation points. However, despite equivalent isolation points, high-frequency words were named no faster than low-frequency words, in contradiction to the predictions made by the most recent version of cohort theory (Marslen-Wilson, 1987). In addition, because there was a strong effect of density, it cannot be assumed that lexical items were bypassed in the generation of the naming response. Thus, the current version of cohort theory also fails to account for the results obtained in the present investigation.

As previously argued, an adequate model of spoken word recognition cannot assume differing inherent activation levels or thresholds for the units monitoring high- and low-frequency words. Instead, the effects of frequency are best described as biases on the decision units responsible for choosing among activated lexical items. By treating the effects of frequency as biases on the decision process, one can account for results demonstrating the lability of the frequency effect depending on task requirements (e.g.,Pollack, Rubenstein, & Decker, 1959) and higher level sources of information (Grosjean & Itzler, 1984). Thus, the instantiation of frequency in the latest version of cohort theory is difficult to countenance. The NAM, however, provides a more principled explanation of the effects of word frequency on both the stimulus word and its neighbors.

As it stands, cohort theory is clearly inconsistent with a number of findings from the previous studies. Although cohort theory still makes a number of important claims regarding the temporal course of processing longer spoken words, which the present model has yet to address, certain fundamental aspects of the cohort theory appear at this point to be mistaken.

Interactive-Activation Models • Interactive-activation-or connectionist-models of visual (McClelland & Rumelhart, 1981) and spoken word recognition (Elman & McClelland, 1986; McClelland & Elman, 1986) have been increasingly popular in recent years, primarily due to the breadth of phenomena accounted for by these models. Basically, interactive-activation models assume a set of primitive processing units that are densely connected to one another. In models such as TRACE (Elman & McClelland, 1986; McClelland &

Elman, 1986), processing units or nodes have excitatory connections between levels and inhibitory connections among levels. These connections serve to raise and lower activation levels of the nodes depending on the stimulus input and the activity of the overall system. As noted earlier, the system of decision units proposed in the NAM may very well be realized as an interacting set of processing units. The NAM may, therefore, turn out to be virtually indistinguishable from an interactive-activation model. Indeed, the decision values computed by the decision units in the present model may very well arise in a system of interconnected nodes having excitatory and inhibitory connections. McClelland and Rumelhart (1981) in fact discuss the possibility that neighborhood structure (both density and frequency) may be accounted for by their model, although they admit that further work is necessary to confirm their speculations.

At present, little can be said regarding the effects of neighborhood structure in an interactive-activation model short of actually conducting the simulations required for testing the model. One potential problem is the treatment of frequency within the interactive-activation framework as an inherent component of the activation levels of words in memory. As previously argued, the present data strongly suggest a labile frequency bias, and not an inherent threshold or activation level. Thus, it is unclear that an interactive-activation model that assumes higher levels of activation for high-frequency words can account for the present set of data. Nonetheless, the interactive-activation approach is suggestive of an interesting means of further specifying the NAM, although additional research and model development is clearly required.

Shortlist • Norris (1994) has recently proposed a model of spoken word recognition called Shortlist. The Shortlist model, like TRACE, is a connectionist model of spoken word recognition. In the first stage of the model, a "short list" of word candidates is derived that consists of lexical items that match the bottom-up speech input. In the second stage of processing, the short list of lexical items enters into a network of word units, much like the lexical level of TRACE. Lexical units at this second level of processing compete (via inhibitory links) with one another for recognition.

The Shortlist model is attractive for two primary reasons: first, it is able to simulate effects of subsequent context on the recognition of spoken words in fluent speech, effects that have heretofore been ignored in such models as cohort and logogen theory. Second, Shortlist improves on the highly unrealistic architecture of TRACE, in which single words are represented by a plethora of identical nodes across time (see Elman & McClelland, 1986, for a more complete description of TRACE and its assumption). To date, Shortlist is the most attractive connectionist model of spoken word recognition.

NAM and Shortlist share a number of features: both posit a two-stage process of activation and decision and a competitive lexical level. Shortlist makes explicit claims about inhibitory links among lexical items at its second stage of processing, whereas NAM assumes that competition among lexical items is sufficient for predicting effects of neighborhood density and frequency. (The assumption of inhibitory links is not, however, at odds with the fundamental principles of NAM and may eventually prove necessary in simulating various effects of lexical competition.) In addition, neither models claims that lexical nodes affect processing at the first stage of pattern activation.

It is at present unclear how Shortlist could account for the malleable frequency effects observed in our experiments. As in most connectionist models, frequency is probably an inherent component of the activation process. Thus, circumventing the effects of word and neighbor frequency, although preserving effects of neighborhood density, may be problematic. Nonetheless, the Shortlist model constitutes a promising new connectionist model of spoken word recognition with which NAM is fundamentally compatible.

Summary of Word Recognition Models • The previous discussion of current models of word recognition suggests that many of the current models fail to adequately account for the neighborhood structural effects observed in the present experiments. In particular, logogen theory and cohort theory appear to be the most

difficult to reconcile with the present data, although this may simply be a function of the degree of specificity of description provided by each of these models. The interactive-activation approach may ultimately prove most successful in accounting for the present set findings, although it was again argued that any model not treating frequency as primarily a bias effect on decisions cannot adequately account for the present data. In short, the NAM appears at present to provide the most consistent account of the effects observed, primarily due to the interactive nature of the word decision units and the biasing effect of frequency on these units.

Extensions and Application of the NAM: Normal Hearing Adults• Over the last few years, various constructs in the NAM have received support and elaboration in experiments on normal hearing adults. For example,Goldinger, Luce, and Pisoni (1989; see also Goldinger, Luce, Pisoni, & Marcario, 1992; Luce et al., 1990) provided further evidence for NAM using a form-based priming technique. They presented two spoken words on a given trial. The first word was a prime and the second a target. The prime was either a phonetic neighbor of the target word or an unrelated item. The subjects' task was to identify the target word embedded in noise. Goldinger et al. predicted that residual activation from the prime should increase the overall activity of words in the neighborhood and thus reduce the accuracy of identification of the stimulus word itself, relative to an appropriate baseline condition (a phonetically unrelated prime-stimulus word pair). That is, priming with a neighbor should actually result in reduced identification of the stimulus word, because the residual activation from the neighbor prime will result in increased competition with the stimulus word for identification.

Goldinger et al.'s data confirmed their prediction. Target words following phonetically related primes were identified less accurately than target words following unrelated primes, as predicted by the model. Moreover, Goldinger et al. found that primes that were high in word frequency had less of an effect on identification than low-frequency primes. This latter result is consistent with the model's claim that word frequency affects primarily decision processes and not activation levels. (See Goldinger et al., 1989; Luce et al., 1990, for a more detailed exposition of this latter finding.)

The implications of neighborhood activation for multisyllabic words was examined by Cluff and Luce (1990; see also Charles-Luce, Luce, & Cluff, 1990). They asked subjects to identify bisyllabic words (e.g., jigsaw) and nonwords (e.g., manbeat) embedded in noise. The component syllables of the target stimuli were either "easy" or "hard". Easy syllables were high in frequency and resided in low density, low-frequency neighborhoods. Hard syllables were low in frequency and resided in high density, high-frequency neighborhoods. (Note that, unlike the stimuli in Experiments 1 to 3 of the present study, frequency and neighborhood structure were confounded.) Their results confirmed the predictions of NAM: easy syllables were identified more accurately than hard syllables. Moreover, overall performance on the bisyllabic words was a function of the combined effects of the neighborhood structures of the component syllables. Very simply, words with two hard syllables were identified least accurately; words with two easy syllables were identified most accurately. Words consisting of a combination of easy and hard syllables produced intermediate levels of identification performance. Cluff and Luce's results demonstrate that effects of similarity neighborhoods are not restricted to simple monosyllabic words.

More recently, Newman, Sawusch, and Luce (1997) demonstrated that similarity neighborhood composition also has demonstrable effects on phoneme identification. They presented subjects with nonsense words that varied on frequency-weighted neighborhood structure. In certain conditions of their experiment, the initial phonemes of the nonsense words were digitally edited to make their identify ambiguous. In these cases, Newman et al. found that subjects were more likely to label ambiguous phonemes as belonging to nonsense words in dense, high-frequency neighborhoods than in sparse, low-frequency neighborhoods. This finding further demonstrates that the spoken word recognition system operates by activating multiple representations in memory and that this multiple activation has effects even

at the level of segmental phonetic perception under certain task conditions.(For a detailed discussion of why dense neighborhoods facilitate identification in this task, see Newman et al., 1997.)

Developmental Implications • The concept of similarity neighborhoods may also provide important insights into the acquisition of words and the development of the lexicon in young children. Employing a computational analysis of young children's lexicons, Charles-Luce and Luce (1990, 1995; Logan, Reference Note 4) demonstrated that similarity neighborhoods of spoken words are relatively sparsely populated in very young children, perhaps more so than would be expected by smaller vocabulary sizes alone. Charles-Luce and Luce suggested that the structure of young children's similarity neighborhoods may enable them to adopt recognition strategies as they are acquiring the language that are different from those of adults. In particular, Charles-Luce and Luce argued that young children may be able to recognize words based on more holistic or global recognition strategies, given that the fine-grained discrimination processes necessary for deciding among words in densely populated similarity neighborhoods are not as necessary as they are for adult perceivers whose neighborhoods tend to be tightly packed (Logan, Reference Note 4).

Further research by Jusczyk, Luce, and Charles-Luce(1994) has demonstrated that children as young as 9 mo of age are sensitive to differences in neighborhood structures of spoken nonsense words. Jusczyk et al. created sets of nonwords that had high and low density neighborhoods. (Jusczyk et al. actually manipulated the probabilistic phonotactics of the nonword stimuli. However, phonotactics and neighborhoods are highly correlated, such that high and low phonotactics correspond to high and low density neighborhoods.) Using the headturn preference procedure, they found that 9-mo-old infants prefer to listen to lists of nonwords occurring in high density similarity neighborhoods. Because nonwords occurring in high density neighborhoods are-by definition-more "word-like"(Vitevitch, Luce, Charles-Luce, & Kemmerer, 1997), the Jusczyk et al. study demonstrates that very young children are developing a sensitivity to what constitutes a likely word in their native language, a sensitivity that may facilitate their acquisition of a lexicon of spoken words. The Jusczyk et al. study also suggests that early in development, the structure of similarity neighborhoods may play a role different than the one it eventually assumes in adults. In particular, relatively more densely populated neighborhoods in the young child may facilitate learning and recognition by providing a larger set of examples of what constitutes possible words in the ambient language. (Note that the finding that children may have more sparsely populated neighborhoods than adults is not inconsistent with the demonstration that children prefer words in densely populated neighborhoods.)

Clinical Applications • The NAM has also provided an important new conceptual framework for research of hearing impairment. Because NAM makes explicit claims about the importance of such factors as lexical discrimination and frequency biases, the model has enabled researchers and clinicians to view spoken word recognition not only from the perspective of the fairly low-level mechanisms involved in hearing and speech perception, but also in the broader context of perceptual and memory processes that also subserve the recognition process. Indeed, one of the major theoretical claims of the model is that spoken word recognition entails much more than the recognition of the individual phonetic segments that comprise words. Thus, when evaluating hearing-impaired listeners, the model emphasizes that simple tests of speech pattern discrimination and phonetic feature discrimination will grossly underestimate the complex task that faces hearing impaired and normal listeners in understanding spoken words in naturalistic settings.

The concept of similarity neighborhoods has already engendered considerable research on word recognition in hearing-impaired populations.Sommers, Kirk, and Pisoni (1997; see also Kirk, Pisoni, & Osberger, 1995; Kirk, Pisoni, Sommers, Young, & Evanson, 1995) examined normal-hearing, noise-masked normal-hearing, and cochlear implant subjects' spoken word identification performance in open- and closed-set tests. Sommers et al. presented subjects with "easy" (high-frequency words in sparse, low-frequency neighborhoods) and"hard" (low-frequency words in dense, high-frequency neighborhoods)

words. The crucial finding was that in a closed-set format-like that used for assessment in many clinical settings-the authors observed no effects of easy versus hard words. However, robust effects of lexical difficulty were obtained in the open-set format: hard words were identified less accurately than easy words. Sommers et al. concluded from these findings that closed-set formats do not allow an adequate evaluation of the full range of complex perceptual and cognitive processes necessary for recognizing spoken words. The absence of frequency and density effects in closed-set speech discrimination tests demonstrates that these tests are not measuring word recognition or lexical access. Instead, these tests measure speech pattern discrimination, which can be accomplished without accessing words from the lexicon.

Whereas the Sommers et al. study demonstrates that the concept of similarity neighborhoods may provide important insights into clinical assessment, the study also supports a fundamental prediction of the NAM. Recall that the model states that frequency effects in spoken word recognition arise-at least in part-from biases on perceptual decision processes. This claim leads to the prediction that effects of word and neighbor frequency may be malleable depending on the task demands. The finding by Sommers et al. supports this claim. They demonstrated that in closed set formats, effects of lexical difficulty (which includes effects of stimulus word and neighbor frequency) are attenuated. This result is easily accounted for by NAM: if the set of alternatives among which the subject must discriminate is known in advance, frequency biases may be equated across the set of alternatives. Moreover, because words in the similarity neighborhood that do not constitute alternatives in the closed set format are not viable responses, the subject may (implicitly) set their biases to zero, functionally ruling out their involvement in the perceptual decision processes responsible for choosing among words in the neighborhood. Thus, NAM predicts that neighborhood effects (i.e., effects of lexical difficulty) will be neutralized in closed-set formats, just as Sommers et al. found in their study.

Kirk, Pisoni, and Osberger (1995; see also Miyamoto, Kirk, Robbins, Todd, Riley, & Pisoni, 1997) have also demonstrated the utility of the concept of similarity neighborhood structure in research on pediatric cochlear implant users. Kirk et al. presented "easy" and "hard" words to children with multichannel cochlear implants and measured identification accuracy. Their results confirmed that this population of subjects exhibits demonstrable effects of lexical difficulty. More important, however, was the finding that the easy-hard manipulation affected word identification performance but not phoneme recognition. This result is consistent with another claim of NAM, namely, that word recognition involves a two-stage process of activation(phonetic processing) followed by lexical discrimination (although there may be effects of neighborhood activation on segment identification under certain circumstances; see Newman et al., 1997). Furthermore, it is becoming increasingly clear that clinical assessments that fail to consider the second stage of processing (i.e., lexical selection) required for successful word recognition will seriously underestimate the subject's capabilities in understanding spoken words. The findings of Kirk et al. led them to develop a new speech discrimination test, the Lexical Neighborhood Test, for measuring word recognition in clinical populations (Kirk, Diefendorf, Pisoni, & Robbins, 1997). By incorporating recent insights and motivations from principled, theoretically based research on spoken word recognition, this test should prove invaluable in assessing the full range of processes necessary for recognizing spoken words in normal and hearing-impaired populations.

Finally, Sommers (1996) has demonstrated that NAM provides a useful framework for understanding age-related changes in spoken word recognition. He examined young and older adults' perception of "easy" and "hard" words and found that normal-hearing older adults are more strongly affected by similarity neighborhood composition than young adults. In particular, older adults show more of a decrement in identification accuracy for hard words than do younger adults, suggesting that a portion of age-related difficulties in spoken word recognition may stem from factors other than peripheral hearing loss, such as the reduced ability to process multiply activated words in memory.

Back to Top

Summary

NAM provides a new theoretical framework for understanding the complex processes involved in recognizing spoken words among a variety of pediatric and adult normal and clinical populations. We believe this model represents a significant advance in our understanding of how words are represented in the lexicon and how listeners gain access to this structural information. As a consequence, NAM should help basic researchers and clinicians alike in understanding and evaluating the complex processes involved in the recognition of spoken words and processing of spoken language.

Back to Top

Conclusion

The structure of the mental lexicon has been a long-standing and important issue in research on spoken language comprehension. To date, however, little attention has been directed to specifying the structural organization of the acoustic-phonetic patterns in memory that are used to gain access to the mental lexicon. The present investigation serves as an initial attempt at characterizing that structure and its effects on spoken word recognition. The picture that begins to emerge from the results reported in the present studies is one of a perceptual and cognitive system optimized for the recognition of words under a wide variety of circumstances. This optimization is achieved by a simultaneous activation of alternatives based on the stimulus input and by a sophisticated system that attempts to maximize decisions among these alternatives. The fact that the spoken word recognition system is capable of considering numerous alternatives in parallel helps to assure the best performance in the face of less than perfect sensory input from speech signals that are often impoverished, degraded, or poorly specified. However, as the present set of experiments have shown, this optimization is not without its processing costs. Both the number and nature of words activated by the initial acoustic-phonetic stimulus input affect not only the accuracy of spoken word recognition, but also the time required to decided among the activated candidates. Nevertheless, such processing costs subserve the ultimate goal of the human speech processing system, namely to maximize the speed and accuracy with which words are recognized in real-time. In short, the study of the structural organization of the neighborhoods of words in the mental lexicon has provided deeper insights into one important aspect of the fundamentally and uniquely human capacity to communicate using spoken language.

Back to Top

Back to Top

References

Andrews, S. (1982). Phonological recoding: Is the regularity effect consistent? Memory and Cognition, 10, 565-575. [Context Link]

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. Journal of Experimental Psychology: Human Perception and Performance, 10, 340-357. [Context Link]

Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task:

Lexical access and/or production frequency? Journal of Verbal Learning and Verbal Behavior, 24, 89-106. [Context Link]

Bond, Z. S., & Garnes, S. (1980). Misperceptions of fluent speech. In R. A. Cole (Ed.), Perception and production of fluent speech (115-132). Hillsdale, NJ: Erlbaum. [Context Link]

Broadbent, D. E. (1967). Word-frequency effect and response bias. Psychological Review, 74, 1-15. Bibliographic Links [Context Link]

Carroll, J., & White, M. (1973a). Age-of-acquisition norms for 220 picturable nouns. Journal of Verbal Learning and Verbal Behavior, 12, 563-576. [Context Link]

Carroll, J., & White, M. (1973b). Word frequency and age of acquisition as determiners of picture naming latency. Quarterly Journal of Experiment Psychology, 25, 85-95. [Context Link]

Catlin, J. (1969). On the word-frequency effect. Psychological Review, 76, 504-506. [Context Link]

Charles-Luce, J., & Luce, P. A. (1990). Some structural properties of words in young children's lexicons. Journal of Child Language, 17, 205-215. [Context Link]

Charles-Luce, J., & Luce, P. A. (1995). An examination of similarity neighbourhoods in young children's receptive vocabularies. Journal of Child Language, 22, 727-735. Bibliographic Links [Context Link]

Charles-Luce, J., Luce, P. A., & Cluff, M. S. (1990). Retroactive influences of syllable neighborhoods. In G. Altmann (Ed.),Cognitive models of speech perception: Psycholinguistic and computational perspectives. Cambridge, MA: MIT Press. [Context Link]

Cluff, M. S., & Luce, P. A. (1990). Similarity neighborhoods of spoken bisyllabic words. Journal of Experiment Psychology: Human Perception and Performance, 16, 551-563. [Context Link]

Coltheart, M., Develaar, E., Johansson, J. T., & Besner, D. (1976). Access to the internal lexicon. In S. Dornic (Ed.),Attention and performance VI. Hillsdale, NJ: Erlbaum. [Context Link]

Elman, J. L., & McClelland, J. L. (1986). Exploiting lawful variability in the speech waveform. In J. S. Perkell & D. H. Klatt(Eds.), Invariance and variability in speech processing (pp. 360-385). Hillsdale, NJ: Erlbaum. [Context Link]

Forster, K. I. (1976). Accessing the mental lexicon. In R. J. Wales & E. Walker (Eds.), New approaches to language mechanisms. Amsterdam: North Holland. [Context Link]

Forster, K. I. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & E. C. T. Walker(Eds.), Sentence processing: Psycholinguistic studies presented to Merrill Garrett. Hillsdale, NJ: Erlbaum. [Context Link]

Forster, K. I. (1981). Frequency blocking and lexical access: One mental lexicon or two? Journal of Verbal Learning and Verbal Behavior, 20, 190-203. [Context Link]

Forster, K. I., & Chambers, I. M. (1973). Lexical access and naming time. Journal of Verbal Learning and Verbal Behavior, 12, 627-635. [Context Link]

Frederiksen, J. R., & Kroll, J. F. (1976). Spelling and sound: Approaches to the internal lexicon. Journal of Experiment Psychology: Human Perception and Performance, 2, 361-379. [Context Link]

Gaygen, D. E., & Luce, P. A. (in press). Effects of modality on subjective frequency estimates and processing of spoken and printed words. Perception and Psychophysics. [Context Link]

Glanzer, M., & Bowles, N. (1976). Analysis of the word-frequency effect in recognition memory. Journal of Experimental Psychology: Human Learning and Memory, 2, 21-31. [Context Link]

Glanzer, M., & Ehrenreich, S. L. (1979). Structure and search of the internal lexicon. Journal of Verbal Learning and Verbal Behavior, 18, 381-398. [Context Link]

Goldiamond, I., & Hawkins, W. F. (1958). Vexierversuch: The logarithmic relationship between word-

frequency and recognition obtained in the absence of stimulus words. Journal of Experiment Psychology, 56, 457-463. [Context Link]

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. Journal of Memory and Language, 28, 501-518. [Context Link]

Goldinger, S. D., Luce, P. A., Pisoni, D. B., & Marcario, J. K. (1992). Form-based priming in spoken word recognition: The roles of competitive activation and response biases. Journal of Experimental Psychology: Learning, Memory, and Cognition, 18, 1210-1237. [Context Link]

Gordon, B. (1983). Lexical access and lexical decision: Mechanisms of frequency sensitivity. Journal of Verbal Learning and Verbal Behavior, 18, 24-44. [Context Link]

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

Grosjean, F., & Itzler, J. (1984). Can semantic constraint reduce the role of word frequency during spoken-word recognition. Perception and Psychophysics, 22, 180-182. [Context Link]

Havens, L. L., & Foote, W. E. (1963). The effect of competition on visual duration threshold and its independence of stimulus frequency. Journal of Experimental Psychology, 65, 6-11. Bibliographic Links [Context Link]

Hood, J. D., & Poole, J. P. (1980). Influence of the speaker and other factors affecting speech intelligibility. Audiology, 19, 434-455. Bibliographic Links [Context Link]

Howes, D. H. (1954). On the interpretation of word frequency as a variable affecting speech of recognition. Journal of Experimental Psychology, 48, 106-112. Bibliographic Links [Context Link]

Howes, D. H. (1957). On the relation between the intelligibility and frequency of occurrence of English words. Journal of the Acoustical Society of America, 29, 296-305. [Context Link]

Howes, D. H., & Soloman, R. L. (1951). Visual duration threshold as a function of word probability. Journal of Experimental Psychology, 41, 401-410. Bibliographic Links [Context Link]

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. Journal of Memory and Language, 33, 630-645. [Context Link]

Kirk, K. I., Diefendorf, A. O., Pisoni, D. B., & Robbins, A. M. (1997). Assessing speech perception in children. In L. L. Mendel & J. L. Danhauer (Eds.), Audiologic evaluation and management and speech perception assessment. San Diego: Singular. [Context Link]

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

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, & Laryngology, 104, 300-303. [Context Link]

Klatt, D. H. (1968). The structure of confusions in short-term memory. Journal of the Acoustical Society of America, 44, 401-407. [Context Link]

Kohonan, T. (1980). Content-addressable memories. New York: Springer-Verlag. [Context Link]

Kucera, F., & Francis, W. (1967). Computational Analysis of Present Day American English. Providence, RI: Brown University Press. [Context Link]

Landauer, T. K., & Freedman, J. (1968). Information retrieval from long-term memory: Category size and recognition time. Journal of Verbal Learning and Verbal Behavior, 7, 291-295. [Context Link]

Landauer, T. K., & Streeter, L. A. (1973). 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. [Context Link]

Lewellen, M. J., Goldinger, S. D., Pisoni, D. B., & Greene, B. G. (1993). Lexical familiarity and processing efficiency: Individual differences in naming, lexical decision and semantic categorization. Journal of Experimental Psychology: General, 122, 316-330. [Context Link]

Luce, P. A. (1986). A computational analysis of uniqueness points in auditory word recognition. Perception and Psychophysics, 39, 155-158. [Context Link]

Luce, R. D. (1959). Individual choice behavior. New York: Wiley. [Context Link]

Luce, P. A., & Pisoni, D. B. (1987). Speech perception: Recent trends in research, theory, and applications. In H. Winitz (Ed.),Human communication and its disorders (1-87). Norwood, NJ: Ablex. [Context Link]

Luce, P. A., Pisoni, D. B., & Goldinger, S. D. (1990). Similarity neighborhoods of spoken words. In G. Altmann (Ed.),Cognitive models of speech perception: Psycholinguistic and computational perspectives. Cambridge, MA: MIT Press. [Context Link]

Marslen-Wilson, W. D. (1984). Function and process in spoken word recognition: A tutorial review. In H. Bouma & D. G. Bouwhuis(Eds.), Attention and performance X: Control of language processes. Hillsdale, NJ: Erlbaum. [Context Link]

Marslen-Wilson, W. D. (1987). Parallel processing in spoken word recognition. Cognition, 25, 71-102. [Context Link]

Marslen-Wilson, W. D. (1989). Access and integration: Projecting sound onto meaning. In W. D. Marslen-Wilson (Ed.), Lexical access and representation (pp. 3-24). Cambridge, MA: Bradford. [Context Link]

Marslen-Wilson, W. D., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. Cognition, 8, 1-71. Bibliographic Links [Context Link]

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. Cognitive Psychology, 10, 29-63. [Context Link]

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. Cognitive Psychology, 18, 1-86. Bibliographic Links [Context Link]

McClelland, J. L., & Rumelhart, D. E. (1981). And interactive activation model of context effects in letter perception: Part 1. An account of basic findings. Psychological Review, 88, 375-407. [Context Link]

McQueen, J. M., Cutler, A., Briscoe, T., & Norris, D.(1995). Models of continuous speech recognition and the contents of the vocabulary. Language and Cognitive Processes, 10, 309-331. [Context Link]

Miller, G. A., & Johnson-Laird, P. N. (1976). Language and perception. Cambridge, MA: Harvard University Press. [Context Link]

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. [Context Link]

Miyamoto, R. T., Kirk, K. I., Robbins, A. M., Todd, S., Riley, A., & Pisoni, D. B. (1997). Speech perception and speech intelligibility in children with multichannel cochlear implants. In I. Honjo& H. Takahashi (Eds.),Cochlear implant and related sciences update. Advances in oto-rhino-laryngology. Switzerland: Basel, Karger. [Context Link]

Morton, J. (1969). Interaction of information in word recognition. Psychological Review, 76, 165-178. [Context Link]

Morton, J. Word recognition. (1979). In J. Morton & J. D. Marshall (Eds.), Psycholinguistics 2: Structures and processes(pp. 107-156). Cambridge, MA.: MIT Press. [Context Link]

Neisser, U. (1967). Cognitive Psychology. New York: Appleton-Century-Crofts. [Context Link]

Newbigging, P. L. (1961). The perceptual reintegration of frequent and infrequent words. Canadian Journal of Psychology, 15, 123-132. Bibliographic Links [Context Link]

Newman R. S., Sawusch, J. R., & Luce, P. A. (1997). Effects of lexical neighborhood density on phoneme perception. Journal of Experimental Psychology: Human Perception and Performance, 23, 873-889. [Context Link]

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. Cognition, 52, 189-234. [Context Link]

Oldfield, R. C. (1966). Things, words and the brain. Quarterly Journal of Experimental Psychology, 18, 340-353. Bibliographic Links [Context Link]

Paap, K. R., McDonald, J. E., Schvaneveldt, R. W., & Noel, R. W. (1986). Frequency and pronounceability in visually presented naming and lexical-decision tasks. In Attention and performance XII: The psychology of reading. Hillsdale, NJ: Erlbaum. [Context Link]

Paap, K. R., Newsome, S. L., McDonald, J. E., & Schvaneveldt, R. W. (1982). An activation-verification model for letter and word recognition: The word superiority effect. Psychological Review, 89, 573-594. Bibliographic Links [Context Link]

Pisoni, D. B., & Garber, E. E. (1990). Lexical memory in visual and auditory modalities: The case for a common mental lexicon. In H. Fujisaki (Ed.), Proceedings of the 1990 International Conference on Spoken Language Processing. Tokyo: Acoustical Society of Japan. [Context Link]

Pisoni, D. B., & Luce, P. A. (1987). Acoustic-phonetic representations in word recognition. Cognition, 25, 21-52. Bibliographic Links [Context Link]

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. [Context Link]

Pollack, I., Rubenstein, H., & Decker, L. (1959). Intelligibility of known and unknown message sets. Journal of Acoustical Society of American, 31, 273-259. [Context Link]

Pollack, I., Rubenstein, H., & Decker, L. (1960). Analysis of incorrect responses to an unknown message set. Journal of the Acoustical Society of America, 32, 340-353. [Context Link]

Rubenstein, D. E., Garfield, L., & Millikan, J. (1970). Homographic entries in the internal lexicon. Journal of Verbal Learning and Verbal Behavior, 9, 487-494. [Context Link]

Rubenstein, H., Richter, M. L., & Kay, E. J. (1975). Pronounce-ability and the visual recognition of nonsense words. Journal of Verbal Learning and Verbal Behavior, 14, 651-657. [Context Link]

Rumelhart, D. E., & Siple, P. (1974). Process of recognizing tachistoscopically presented words. Psychological Review, 81, 99-118. Bibliographic Links [Context Link]

Sankoff, D., & Kruskall, J. B. (1983). Time warps, sting edits, and macromolecules: The theory and practice of sequence comparison. London: Addison-Wesley. [Context Link]

Savin, H. B. (1963). Word-frequency effect and errors in the perception of speech. Journal of the Acoustical Society of America, 35, 200-206. [Context Link]

Scarborough, D., Cortese, C., & Scarborough, H. (1977). Frequency and repetition effects in lexical memory. Journal of Experimental Psychology: Human Perception and Performance, 3, 1-17. [Context Link]

Smith, E. E. (1978). Theories of semantic memory. In W. K. Estes (Ed.), Handbook of learning and cognitive processes: Linguistic functions in cognition theory (pp. 1-56). Hillsdale, NJ: Erlbaum. [Context Link]

Smith, J. E. K. (1980). Models of identification. In R. Nickerson (Ed.), Attention and Performance VIII.

Hillsdale, NJ: Erlbaum. [Context Link]

Soloman, R. L., & Postman, L. (1952). Frequency of usage as a determinant of recognition thresholds for words. Journal of Experimental Psychology, 43, 195-201. [Context Link]

Sommers, M. S. (1996). The structural organization of the mental lexicon and its contribution to age-related changes in spoken word recognition. Psychology and Aging, 11, 333-341. [Context Link]

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

Stanners, R. F., Jastrzembski, J. E., & Westbrook, A.(1975). Frequency and visual quality in a word-nonword classification task. Journal of Verbal Learning and Verbal Behavior, 90, 45-50. [Context Link]

Triesman, M. (1971). On the word frequency effect: Comments on the papers by J Catlin and L. H. Nakatani. Psychological Review, 78, 420-425. Bibliographic Links [Context Link]

Triesman, M. (1978a). A theory of the identification of complex stimuli with an application to word recognition. Psychological Review, 85, 525-570. [Context Link]

Triesman, 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. [Context Link]

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. Language and Speech, 40, 47-62. [Context Link]

Wang, M. D., & Bilger, R. C. (1973). Consonant confusions in noise: A study of perceptual features. Journal of the Acoustical Society of America, 45, 1248-1266. [Context Link]

Webster's Seventh Collegiate Dictionary. (1967). Los Angeles: Library Reproduction Service. [Context Link]

Whaley, C. P. (1978). Word-nonword classification time. Journal Verbal Learning and Verbal Behavior, 17, 143-154. [Context Link]

Winer, B. J. (1971). Statistical principles in experimental design. New York: McGraw-Hill. [Context Link]

Back to Top

Reference Notes

1 Anderson, D. C. (1962). The number and nature of alternatives as an index of intelligibility. Unpublished doctoral dissertation, Ohio State University. [Context Link]

2 Bernacki, B. (1981). WAVMOD: A program to modify digital waveforms. Research on Speech Perception Progress Report No. 7. Bloomington, IN: Speech Research Laboratory, Psychology Department, Indiana University. [Context Link]

3 Church, K. W. (1983). Phrase-structure parsing: A method for taking advantage of allophonic constraints. Bloomington, IN: Indiana University Linguistics Club. [Context Link]

4 Logan, J. (1992). A computational analysis of young children's lexicons. Research on Speech Perception Technical Report No. 8. Bloomington, IN: Speech Research Laboratory, Psychology Department, Indiana University. [Context Link]

5 Luce, P. A. (1986). Neighborhoods of words in the mental lexicon. Research on Speech Perception Technical Report No. 6. Bloomington, IN: Speech Research Laboratory, Psychology Department, Indiana University. [Context Link]

6 Luce, P. A., & Carrell, T. D. (1981). Creating and editing waveforms using WAVES. Research on Speech Perception Progress Report NO. 7. Bloomington, IN: Speech Research Laboratory, Psychology

Department, Indiana University. [Context Link]

7 Mermelstein, P. (1976). Distance measures for speech recognition-psychological and instrumental. Status Report on Speech Research SR-47. New Haven, CT: Haskins Laboratories. [Context Link]

8 Moore, R. K. (1977). The evaluation and optimization of a basic speech recognizer. Unpublished masters thesis, University of Essex, Colchester, England. [Context Link]

9 Nakatani, L. H. (1969). A confusion-choice stimulus recognition model applied to word recognition. Unpublished doctoral dissertation, University of California, Los Angeles. [Context Link]

10 Norris, D. (1982). Word recognition: Context effects without priming. Unpublished manuscript, University of Sussex, Brighton, England. [Context Link]

11 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, Psychology Department, Indiana University. [Context Link]