

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 28 (2007)
Indiana University

Effects of Clustering Coefficient on Spoken Word Recognition¹

Nicholas A. Altieri and David B. Pisoni

*Speech Research Laboratory
Department of Psychological and Brain Sciences
Indiana University
Bloomington, Indiana 47405*

¹ This study was supported by the NIH Grants DC-00111 and DC00012. I would like to acknowledge Thomas Gruenenfelder for helpful comments, ideas and data analysis.

Effects of Clustering Coefficient on Spoken Word Recognition

Abstract. Since the late 1960's, researchers have explored how the structure of the mental lexicon affects spoken word recognition. The proposal that words are recognized relationally in the context of other words in lexical memory has encouraged the use of complex systems to describe connectivity in both the phonological and semantic lexicon. The present study assessed the role of the graph theoretical measure of *Clustering Coefficient* (CC) using two experimental paradigms: same-different discrimination, and perceptual identification to explore how global lexical variables affect spoken word recognition. In Experiment 1, listeners judged whether two words were the same or different. Longer response latencies were obtained for high CC words than low CC words. In Experiment 2, the stimuli were processed with an 8-channel noise vocoder to degrade the signal and a new group of listeners also performed the same-different task. In contrast to findings obtained in Experiment 1, listeners discriminated high CC words more accurately than low CC words. In Experiment 3, an open-set perceptual identification task was carried out to examine correct and incorrect responses using 8, 10 and 12 channels. Listeners identified low CC words more accurately than high CC words in the 10 and 12 channel conditions, but not in the 8-channel condition. Detailed analysis of the incorrect responses revealed that listeners used different perceptual strategies as the number of channels increased. These results suggest that global, emergent factors reflecting the structural organization and connectivity of words in the mental lexicon affect spoken word recognition and should be included in current models of spoken word recognition and lexical access.

Introduction

Structural analyses of the mental lexicon have been used to explore how listeners process and store spoken words. Numerous researchers have focused their efforts on understanding the structural properties and topology of the phonological mental lexicon following the publication of Oldfield's seminal article "Things, Words, and the Brain" (1966). One of the findings that motivated inquiries into the structure of the mental lexicon were word frequency effects; that is, effects demonstrating a systematic relationship between the frequency of a word's occurrence in the language and its intelligibility in noise (Howes, 1957). In order to explain word frequency effects, Oldfield (1966) hypothesized that the lexicon had a structure where words are organized into separate compartments or bins depending on their frequency of occurrence in the language. During the recognition process, the compartments are searched using a binary search algorithm. Oldfield assumed that memory bins containing high frequency words were searched first, which could explain why response latencies were shorter for common words in both *naming* and *same-different* judgment tasks.

Other models of the mental lexicon have also been proposed to account for frequency effects. Morton's Logogen Theory (1979) assumed that high frequency words have lower recognition thresholds than low frequency words in the lexicon. Logogens are not words, but are hypothetical units that act as evidence accumulators during the word recognition process. The term *logogen* literally means "word birth" (*logos* meaning words, *genus* for birth). According to Morton, evidence from both auditory and visual input is analyzed by logogens. Both bottom-up sensory evidence and top-down contextual information interact to bring the evidence accumulators above threshold. In general, the more contextual information that is available, the less bottom-up sensory information is required to bring the logogen above the critical threshold for word recognition. Once sufficient evidence has been accumulated to surpass a specified threshold, word recognition is assumed to occur. Logogen theory accounts for word-

frequency effects by assuming that less sensory evidence is required for high frequency words than low frequency words during the recognition process.

One shortcoming of Logogen theory is its inability to explain how pseudowords and non-words are recognized. In Logogen theory, the lexicon consists only of words with critical threshold values, and counters that accumulate evidence from features. However, the lexicon does not contain information about non-words, and exactly how logogens can accumulate evidence required for the recognition of phonologically possible non-words is not addressed.

In an effort to mathematically formalize predictions regarding the structure of the lexicon with respect to word frequency effects, Treisman (1979) considered three ways words might be represented in memory: as a *tree*, as a *collection*, or as a multidimensional *space*. An example of a tree structure is the way that words are organized in the dictionary. Locating a word in the lexicon involves finding the branch corresponding to the first sound of the word, then the second branch corresponding to the second sound of the word, and so on until a word is identified on a terminal twig. Degraded stimuli might cause the search to finish before ending at a terminal twig corresponding to an entire word in the lexicon because details of certain features might be missing.

Treisman described the “urn model” of Pollack, Rubenstein and Decker (1960) as an analog of a *collection* model of the lexicon. Each ball in the urn represents a word, and the number of balls in an urn that correspond to a word is proportional to the frequency of occurrence of that word in the language. The words organized in a collection have no fixed relation to one another. A model that describes how spoken word recognition occurs over a collection of lexical representations is the *universal forced choice model* (e.g., Luce, 1959). Luce’s choice model assumes that decisions about word recognition are made using information from all of the words represented in the mental lexicon. According to this model, high frequency words will be identified more accurately because the probability of selecting a word is proportional to its frequency of occurrence. However, since high frequency words have more representations in the lexicon, they will also have a higher probability of being generated as error responses than low frequency words. Thus, the model assumes that the correlation between frequency of occurrence in the language and the frequency with which a word is generated as an error, will be strong because both correct and incorrect responses are selected based on a frequency weighted bias.

Words residing in a *multidimensional space*, on the other hand, are classified along continuous acoustical dimensions and are represented as points in the space, whereas non-words are represented as holes in the space. Identifying a word in the continuous space involves finding unique values on the relevant dimensions; if a unique value cannot be identified on a particular dimension, a range of uncertainty remains. Treisman outlined several predictions derived from *partial identification theory*, which hold for a model of the mental lexicon as a multidimensional space, but not a collection or a tree. Partial identification theory follows Luce’s choice rule in assuming that words are identified in a forced-choice decision process, but differs inasmuch as the choice is taken to be limited to a subset of words lying in the “acoustic sub-volume defined by the stimulus” (Treisman, 1979).

The idea of an acoustic sub-volume is similar to the notion of a phonological neighborhood, which is a portion of the subspace of the lexicon containing phonologically similar words. The predictions derived from partial identification theory assert that the correlation between word frequency, and the frequency with which a word is generated as an error will be weak, and that words which are infrequently given as errors will be more easily recognized than words that are more frequently given as errors. To see why this is the case, consider a high frequency word with few similar sounding words residing in its sub-space. Since it has few neighbors, it is less likely to be generated as an error compared to words that have many similar sounding neighbors. This principle becomes even stronger as the signal-

to-noise ratio increases and the sub-volume of similar sounding words becomes smaller and more refined. Treisman provides evidence supporting all three hypotheses and concludes that a multidimensional space is a more accurate representation of the mental lexicon than either a tree or a collection model.

The urn and tree models of the mental lexicon also rely on the principle of “structural equivalence”— which assumes that the phonological characteristics of high and low frequency words are basically the same (Pollack et al., 1960). The principle of structural equivalence assumes that since the phonemic compositions of high and low frequency words are identical, the only difference between them is experienced frequency in the language. If the lexicon is modeled as continuous space with varying degrees of lexical density (the number of non-word holes differs depending on where we are in the space) then the composition of high and low frequency words could differ with regard to the number of similar sounding words in their acoustical subspaces.

Connectivity and Sub-Lexical Components

Landauer and Streeter (1973) assessed the widely held assumption of “structural equivalence” by showing that high and low frequency words differ in the number of similar sounding phonological neighbors they have, and the distribution of component phonemes. In a computational study using 260 high frequency and 260 low frequency words, they found that high frequency words tend to have more lexical neighbors than low frequency words. That is, high frequency words were phonetically similar to many other high frequency words. Landauer and Streeter defined “lexical neighbor” as a word that can be created from a target word by a single deletion, addition or substitution (DAS) of a letter. In a second computational study involving 150 four, five, and six-letter high frequency words, and 150 four, five, and six-letter low frequency words, they also found systematic differences between the frequency distributions for phonemes and letters across different levels of word frequency. For example, the phonemes /n/, /l/, /t/, and /z/ represented a total of 23 percent of the phonemes in high frequency words but only 18 percent of the phonemes in low frequency words. Thus, high and low frequency words not only differ in their experienced frequencies, but they also differ in their structural properties. Landauer and Streeter’s findings influenced subsequent models of spoken word recognition because they showed that sub-lexical segments might affect how words are organized in lexical memory.

Following Landauer and Streeter (1973), several researchers began to study the relations between the phonological properties of words and the structure of the lexicon (Eukel, 1980). Eukel had a group of subjects listen to a recorded list of 25 CCVC nonsense words varying in Greenberg and Jenkins’ measure of phonological similarity, and 58 real words with a wide range of objective frequencies. Greenberg and Jenkins’ metric measures the phonological similarity of nonsense words to real words of English, and constitutes an indirect way of measuring the phonotactic probability of sequences in the lexicon (Vitevitch, Luce, Charles-Luce, & Kemmerer, 1996).

The subjects in Eukel’s study were asked to make subjective judgments about the frequency of occurrence of real words as well as nonsense words. His findings showed that participants subjective judgments of word frequency are highly correlated with both objective measures of experienced word frequency, and Greenberg and Jenkins’ computational measure, indicating that word frequency effects for non-words are due at least in part to probabilistic phonotactics. The finding of subjective frequency effects for non-words also adds converging evidence to Landauer and Streeter’s hypothesis that the principle of “structural equivalence” does not hold, because high and low frequency words differ not only by virtue of how often they occur in language, but also in their the segmental composition.

The idea that words are organized into similarity spaces based on phonotactic probability and word frequency has motivated the assumption that words are recognized relationally in the context of

other similar sounding words rather than in isolation (Luce & Pisoni, 1998; Marslen-Wilson, 1984). The belief that words are recognized relationally has led to several models concerning the way the lexicon might be organized. One important milestone in the field was the development of Cohort theory (Marslen-Wilson, 1984; Marslen-Wilson, 1990). In Cohort theory, word recognition is assumed to take place one phoneme at a time from the beginning of word in real time. During the initial stages of word recognition, a set of potential word candidates, or what Marslen-Wilson called a “word initial cohort”, becomes activated based on information in the stimulus. As additional phonemes are perceived during the recognition process, more words are eliminated from the cohort due to deactivation of words that are no longer compatible with the input signal. This process continues until there is only one possibility. For example, when perceiving an utterance of the word “catapult”, the cohort is first reduced to words beginning with the phoneme /k/, and then to words composed of word initial /ka/ (for example, “can”, “cap”, “catapult”, etc.). Finally, once enough phonemes are perceived to the point where the input diverges from all other possible word candidates, “catapult” is recognized.

A connectionist model of word recognition sharing some design features of Cohort theory is the TRACE model of speech perception (McClelland & Elman, 1986; also see Protopapas, 1999). Two versions of TRACE have been developed to model different data in the literature: TRACE I was the initial implementation of the model and was built to model phoneme perception, while TRACE II was designed to explain data concerning lexical access. The TRACE I model consists of three layers of nodes: the feature, phoneme, and word layer. The representations in TRACE are localist where each node, or independent processing unit, represents a particular unit at each layer. Thus, at the phoneme layer, each node represents a phoneme, and at the word layer, each node represents a word in the lexicon. When input activates a set of features, activation spreads bi-directionally between layers, activating items consistent with the input on relevant dimensions. Activation flows between layers in a process termed “interactive activation”; phonemes activate words, and words containing activated phonemes send the proportional amount of activation back down to the phoneme and feature layers.

As sensory evidence accumulates, nodes begin to inhibit activated items that are mismatched with the input. Unlike activation, inhibition only operates between nodes within a particular layer. In order to build time into the model, the units in TRACE are reproduced and represented multiple times, with one representation at each time slice. As different representations become active at different time slices, the list of possible word candidates changes, which is an important feature shared with Cohort theory—where activation is a key assumption (Marslen-Wilson, 1984).

Using an activation framework that is similar to Cohort theory and the TRACE model, the Neighborhood Activation Model (NAM) (Luce & Pisoni, 1998) describes how words are recognized “relationally.” NAM assumes that similar sounding words compete with or inhibit the input word during the recognition process. The metric for similarity used by NAM differs from Cohort theory. In Cohort theory, perceptually similar words are organized into cohorts based on shared features from the beginning to the end of the word. The metric of phonological similarity used by NAM is the deletion, addition, substitution rule (DAS) originally used by Landauer and Streeter (1973), where two words are “neighbors” if one word can be changed into the other via the deletion, addition, or substitution (DAS) of a single phoneme. Using this metric, “cat” and “cab” are neighbors because they differ only in the coda position and “bat” and “sat” are neighbors because they differ only in the onset position.

NAM assumes that words are organized into similarity spaces in lexical memory, which can be quantified by the number of similar sounding neighbors a word has, referred to as *neighborhood density*. If a word has a large number of phonologically similar neighbors based on the DAS rule, then the word is classified as *high-density* because it resides in a high-density neighborhood. If a word has few phonological neighbors, then it is classified as a *low-density* word. NAM makes several specific

predictions about spoken word recognition including neighborhood density effects: words with many phonological neighbors are inhibited more by their phonological neighbors and consequently are recognized more slowly and less accurately than words with fewer phonological neighbors.

Luce and Pisoni (1998) tested NAM in several experimental paradigms including: lexical decision, word repetition (naming), and perceptual identification. The Hoosier Mental Lexicon, a database of 20,000 words and their phonological transcriptions, was used to compute “similarity neighborhoods” (Nusbaum, Pisoni, & Davis, 1984). Data from lexical decision and word repetition experiments showed that words in *dense neighborhoods* had longer response latencies than words in *sparse neighborhoods*. Data from identification experiments demonstrated that listeners identified low-density words over a range of different signal-to-noise ratio more accurately than high-density words.

Luce and Pisoni’s findings inspired a series of follow-up studies that revealed competitive inhibition at the lexical level and facilitatory probabilistic phonotactic effects at the sub-lexical level (Vitevitch & Luce, 1998; Vitevitch & Luce, 1999). For example, using a same-different discrimination experiment, Vitevitch and Luce (1999) showed that response latencies were longer for high-density words consisting of high phonotactic probability segments (i.e., high probability words) than low-density words consisting of low phonotactic probability segments. The opposite result was observed for non-words; non-words with high phonotactic probability segments were recognize more quickly and more accurately than non-words consisting of low phonotactic probability segments.

Graph Theory and Complex Systems

While Cohort theory, TRACE, and NAM have provided several novel insights about how words are recognized in the context of other similar sounding words in memory, the models are incomplete because they fail to provide a description of how the global properties of the mental lexicon might affect spoken word recognition. If Treisman’s (1979) hypothesis is correct with regard to conceptualizing lexical representations of words as trajectories in a multi-dimensional acoustical space, then it is important to explore the effects of the global topology and connectivity among words in this space. In order to accomplish this goal, new tools and new variables are needed to describe the structure of the lexicon and quantify structural relationships between words.

Recently, several researchers have applied tools used in the analysis of complex systems to the study of the mental lexicon (Gruenenfelder & Pisoni, 2006; Steyvers & Tenenbaum, 2005; Vitevitch, 2004). Complexity theory conceptualizes how the separate parts of a system interact with one another to produce emergent behaviors (Barabási, 1999). Complex systems can be represented graphically with individual components represented as nodes or vertices, and relationships between nodes represented as links. Graph theory provides a means of describing the patterns of connectivity among words in the mental lexicon, and offers a different theoretical approach for modeling lexical growth and development (see Vitevitch, 2004).

The study of complex systems is interdisciplinary, spanning several fields of scientific inquiry. Barabási described the network structure, including the *scale-free* distribution of links, in various complex systems including social networks, the world-wide-web, and the Internet (Barabási, 1999; Albert & Barabási, 2002). The degree distribution of a network refers to the number of links a node has. If a node is randomly selected, the probability that it has k neighbors is $p(k) = c/k^\alpha$, where c and α are constants.² The scale-free degree distribution follows a power law; which is a linear trend if plotted on Log-Log coordinates, and approximates an exponential distribution when plotted on linear coordinates. A

² The degree of the exponent in complex networks following the power law degree distribution of links is between 2 and 3.

“scale-free degree distribution” also means that no single descriptive parameter such as a mean or median can accurately describe the number of links attached to a randomly selected node. One property of scale-free networks is that a few nodes are highly connected hubs, since they contain many links, while the majority of nodes have few links.

Many real world and man made complex systems share the *small world* property (see Barabási, 1999). A short path length where the distance between any two nodes is small (i.e., six degrees of separation) and higher clustering coefficient (CC) than a random graph of comparable size characterize the small-world structure. CC is the probability that any two neighbors of a given node are connected (Watts & Strogatz, 1998). In small world networks, the average CC is several magnitudes larger than the CC of a random graph.

Small World and Scale Free Structure of the Mental Lexicon

In a recent computational study, Steyvers and Tenenbaum (2005) modeled three types of semantic networks as complex graphs. The semantic networks they examined consisted of *word associations*, *WordNet*, and *Roget's Thesaurus*. The word association database consisted of stimulus words given to participants and words they wrote down that were associates of the stimulus. For example, if a subject was given the word “dog”, they might have generated the associative response “fetch”. *WordNet*, the second database, is a modern version of *Roget's* thesaurus consisting of words along with their synonyms and antonyms. Steyvers and Tenenbaum reported that all three networks displayed a small world structure and exhibited a scale-free degree distribution, suggesting that semantic representations in memory can be modeled as a complex system.

Vitevitch (2004) conducted a phonological analysis of the mental lexicon, modeling it as a complex network using standard graph theoretical measures. Vitevitch used the DAS metric to construct the graph. Two words (nodes) were connected in the graph if they differed by a single phoneme. The results of the analysis of 19,340 words in the Hoosier Mental Lexicon database suggested that the lexicon shared a variety of general properties with natural and artificial complex systems. Vitevitch found a short path length and a CC that was several magnitudes larger than would be predicted from a random graph of similar size, as well as a scale-free degree distribution.³ Based on the earlier suggestions of Albert and Barabási, Vitevitch argued that these findings were indicative of a mechanism for “preferential growth” and “attachment”. These terms refer to the notion that during lexical development, words that are already highly connected are more likely to acquire new connections as novel words are learned and added to the lexicon. The more time a word remains in the lexicon, the more connections it forms to other words already in the lexicon. That is, as the mental lexicon grows, children are more likely to learn words that are similar to ones they already know.

Recently, Gruenenfelder and Pisoni (2006) replicated and extended Vitevitch's (2004) work. They conducted a reanalysis of data collected by Luce and Pisoni (1998) in their word repetition and identification studies. Graph theory measures including CC were computed for each word in the database in order to also replicate Vitevitch's findings. When comparing the reaction times with the CC of the stimuli in an open set word repetition task, Gruenenfelder and Pisoni observed a significant positive correlation between CC and response latency. However, no significant effects were observed in their analysis correlating percent correct scores and CC from the perceptual identification experiment.

³ The complex graph constructed by Vitevitch (2004) is problematic because out of nearly 20,000 words, over 10,000 words are isolated hermits without any phonological neighbors. The lack of connectivity, particularly among multi-syllabic words, suggests that the metric used for constructing the graph (i.e., the DAS rule) might be flawed in some respects.

The findings of Gruenfelder and Pisoni (2006) suggest that CC affects spoken word recognition. As with the case of neighborhood density, high CC words appear to inhibit the processing of spoken words producing increased response latencies relative to low CC words, indicating that global variables of the lexicon might have behavioral effects on spoken language processing. Their study, however, has several weaknesses. The study was a post-hoc investigation, and neighborhood density and CC were confounded in the original stimuli used by Luce and Pisoni. There are no behavioral studies where these variables are controlled and analyzed separately from one another.

Given these weaknesses, the present study was carried out to experimentally investigate the effects of CC on spoken word recognition to determine if non-local properties of the lexicon affect the word recognition process. Graph theoretic analyses of the mental lexicon are in their infant stages and behavioral evidence is needed in order to begin to support the hypothesis that the organization of words in memory shares structural properties with other complex systems. In the present study, we replicated the results reported by Gruenfelder and Pisoni (2006) and extended the findings into new domains by using the same-different discrimination, and perceptual identification paradigms.

Models and Hypotheses

What would previous models of word recognition predict about CC effects? Most models do not address how differences in CC would affect spoken word recognition. The “urn model” and Logogen theory, for instance, would not be able to make predictions about the effects of graph theoretical measures like CC, since the structure of these models does not presuppose lexical connectivity. Likewise, since Cohort theory assumes that words are recognized from beginning to end and words not in the cohort are discarded, it would not make predictions about CC.

A connectionist model like TRACE could, in principle, make predictions about the effect of CC on word recognition. Recall that activation in TRACE spreads between layers, while inhibition functions within layers. Thus, words send inhibition to other similar sounding words as activation reaches the word level from the phoneme level. As one word node begins receiving more activation than other words, it begins to inhibit those words—making them less likely to reach threshold. The connectivity pattern through which inhibition is sent is irrelevant. What matters is which word most closely matches the activation of phonemic units. Once inhibition from a more activated word is sent to its neighbors, the connectivity, or similarity between the neighbors (phonemically similar words) should not affect recognition of the stimulus. Therefore, TRACE would predict null results with respect to CC because there are no connections beyond local lexical interactions.

NAM, like the TRACE model of speech perception, also does not make any predictions about the effects of CC on spoken word recognition. While NAM describes how words are related to one another through the DAS metric, and how words in this sub-volume of the lexicon affect recognition, it does not describe how neighborhood connectivity affects response times or identification of words in degraded listening conditions. Thus, NAM assumes that all neighborhoods of a given size and neighborhood frequency have similar effects on spoken word recognition.

The purpose of this study was to determine if the CC effects on spoken word recognition and performance observed by Gruenfelder and Pisoni (2006), i.e., where shorter response latencies were observed for low CC stimuli in word repetition tasks, can be replicated when controlling for word frequency. Based on their earlier findings, we expected to find shorter response latencies for low CC words in a same-different discrimination task and more accurate identification of low CC words relative to high CC words under degraded listening conditions. Observing effects of CC on spoken word recognition in both same-different discrimination and perceptual identification tasks would provide

converging support for the hypothesis that global properties of the mental lexicon affect spoken word recognition processes.

Experiment 1: Effects of CC on Same-Different Discrimination

The purpose of Experiment 1 was to test the hypothesis that differences in CC affect the processing time of spoken words in a same-different discrimination task. An analysis of Vitevitch and Luce's (1999) high probability/high density stimuli revealed that the words they used also had high CC. They found that the reaction times for high probability/high-density words were lower than the reaction times for low probability/low-density words. Since probability and CC were confounded in the stimuli used in their study, the pattern of results means that low CC words were responded to more quickly than high CC words. In order to replicate both Vitevitch and Luce (1999) and Gruenenfelder and Pisoni (2006), we created a new set of stimuli that controlled for phonotactic probability and word frequency while manipulating CC and neighborhood density separately.

In the same-different task used in Experiment 1, listeners were presented with two spoken words over headphones and were asked to determine whether the two stimuli were the same or different words. The independent variables of CC and neighborhood density were moderately correlated ($r \sim .49$) in the Hoosier Mental Lexicon database, but the stimuli used in these experiments controlled for CC and neighborhood density. In addition to the prediction that high CC stimuli would be processed more slowly than low CC stimuli, we also predicted that high-density words would be processed more slowly than low-density words.

Method

Design

A 2 x 2 within subject design with CC and neighborhood density as independent variables was used. The dependent variable was reaction time measured in milliseconds. One hundred and sixty words were evenly divided into four experimental conditions: low-density and low CC; low-density and high CC; high-density and low CC; and high-density and high CC. Table 1 shows an outline of the basic design used here and in the following experiments.

High CC High Density (n = 40)	Low CC High Density (n = 40)
High CC Low Density (n = 40)	Low CC Low Density (n = 40)

Table 1. Experimental design: 2x2 within subject factors. Neighborhood Density and Clustering Coefficient are the independent variables and 40 words are used in each cell.

Participants

All of the participants were native speakers of Midwestern American English, who reported no history of speech and hearing disorders at the time of testing. The nineteen participants were recruited from the undergraduate introductory psychology subject pool at Indiana University in Bloomington. None of the subjects participated in more than one experiment reported here.

Stimulus Materials

All words in this experiment were selected from a subset of 938 monosyllabic words obtained from the Hoosier Mental Lexicon database. The stimuli in this experiment consisted of one hundred sixty words used as same pairs, and three hundred twenty words used in one hundred sixty different pairs. The same pairs consist of two different recordings of the same word. We did this in order to encourage listeners to process the stimuli lexically rather than attend to fine acoustic details.

Three hundred twenty words were selected for the different pairs. The words within each pair were phonological neighbors and differed by only one phoneme. The one hundred sixty pairs were counterbalanced for the location of the phonemic differences. A male talker recorded all of the stimulus tokens. The list of stimulus words was presented to the talker and the words were recorded one at a time on a PC using the SAP program. The words were subsequently digitized and edited into individual files using the PRAAT waveform editor. The Level-12 program was used to level the sound level of all the words at 65 dB.

Clustering Coefficient

The program Pajek was used to compute the CC for each of the stimulus words. Since CC is a measure of probability, it ranges from 0 to 1. We selected words in the upper 40 percent of the probability distribution as high CC words, and the words in the lower 40 percent of the distribution as low CC words. The average CC for high CC stimuli was .235, and the average CC for low stimuli was .126.

Neighborhood Density

We took similar steps in the selection process of high and low-density words. Neighborhood density was defined as the number of neighbors based on the DAS rule computed in the Hoosier Mental Lexicon database consisting of $n = 19,340$ words. High and low density items were selected from the upper and lower 40 percent of the distribution respectively using the same subset of 938 monosyllabic. The average neighborhood density for high-density words was 19.9 neighbors, and 11.1 neighbors for low-density words. Density was not weighted by frequency as we only computed raw scores.

Phonotactic Probability

Phonotactic probability was controlled for across levels of CC. The average phonotactic probability was .15 for the high CC words, and .14 for the low CC words. These probabilities refer to the frequency that a certain phoneme or segment occurs in a word. Thus, the segments in the high and low CC words occur at approximately the same rate in the English language. Since phonotactic probability was correlated with lexical density (Vitevitch & Luce, 1999), phonotactic probability varied across the high and low-density stimuli. High-density words had an average phonotactic probability of .168 and low-density words had an average phonotactic probability of .124.

Word Frequency

Frequency of occurrence in the language obtained from Kucera and Francis (1967) was matched for each of the four conditions. Word Frequency scores were transformed using a logarithmic function in order to compress frequencies at the high end of the distribution. The mean log frequency was 2.37 for the low frequency words, and was 2.32 for the high frequency words ($F(1,79) = .095, p < 1.0$). For density, the mean log frequencies were 2.35 and 2.34 for low and high densities respectively ($F(1,79) = 2.0, p = .665$).

Procedure

Testing took place in individual booths with up to three listeners being tested at a time. Each listener was seated in front of a PC equipped with *Beyer Dynamic DT 100* headphones and a two-button response box. Presentation of stimuli and collection of listener's responses were controlled by the PC. At the beginning of each trial, a light was illuminated at the top of the button box before each pair of words was played. The stimuli were presented over the headphones fixed at a comfortable listening level. There was a delay of 500 ms between the offset of the first word and the onset of the second word in the pair. The next trial began after the listener made a response on the button box. The instructions given at the beginning of the experiment asked listeners to make responses as quickly and as accurately as possible as soon as they were sure of their decision. Listeners responded *same* with their right hand and *different* with their left hand. Same and different trials were randomized for each subject. All listeners received ten practice trials at the beginning of the experiment that were not included in the final data analysis. No feedback was given on any of the trials.

Results

Analysis of Same Pairs

The mean response times for the *same* responses are shown in Figure 1 separately for CC and neighborhood density. Figure 2 shows the results for CC collapsed across neighborhood density. In each figure, mean response time in milliseconds is plotted on the ordinate. A one way ANOVA shows that the observed mean accuracy for the task approached ceiling levels and was not significant across conditions ($(F(1,18) = 1.47$ and $(F(1,18) = 3.99)$ for neighborhood density and CC respectively). Error responses were omitted from the data analysis of the response latencies.

Figures 1 and 2 show that the mean response times on the same trials were faster for low CC words across the two levels of lexical density. A series of ANOVAs were carried out on the response latencies for the *same* trials. F values were computed across both subjects and items; that is, both subjects and stimulus words were treated as random variables (Clark, 1973). We chose a .05 (two-tailed) level of significance for each test.

The results from the ANOVA across subjects demonstrated a significant main effect for CC. High CC words (mean = 987.14 ms) were responded to more slowly than low CC words (mean = 950.71 ms) ($F(1,18) = 10.007, p < .01$). The ANOVA for CC over items also showed a significant main effect for CC ($t(39) = 5.45, p = .022$). A binomial test was also significant; 16 out of 19 listeners showed longer response latencies for the high CC word pairs ($p = .004$). That is, the effects of CC on *same* pairs were robust across listeners.

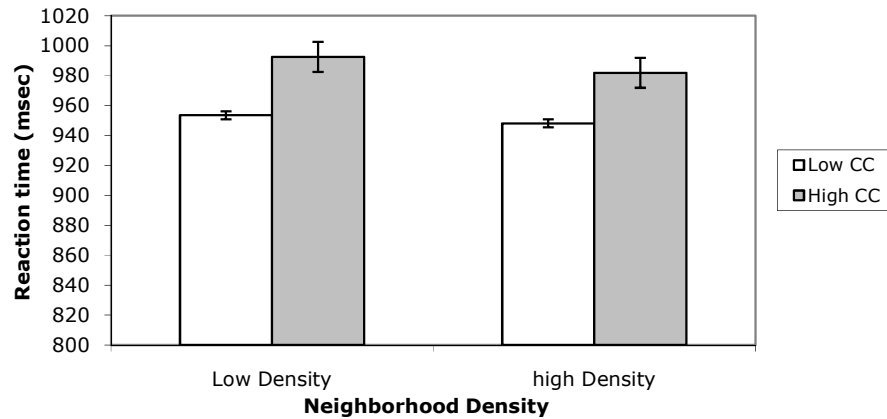


Figure 1. Mean reaction times across density and CC levels in the same-difference matching experiment. The data do not show a significant interaction, or significance across density levels. A significant main effect was observed for CC, where high CC words have longer response latencies than low CC.

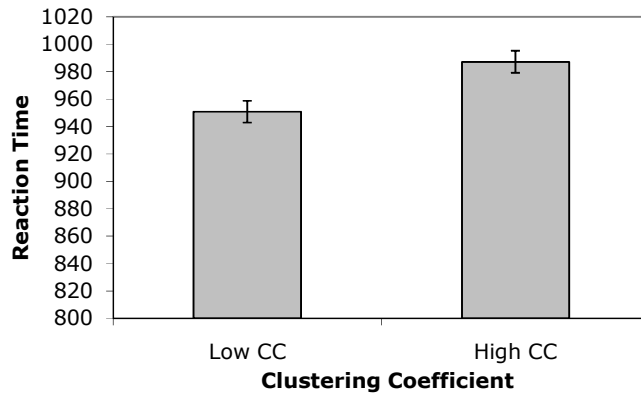


Figure 2. Mean reaction times for CC collapsed across density levels. The results show that high CC words were responded to more slowly than low CC words.

The effect of neighborhood density was not significant (mean HD = 964.9; mean LD = 972.62) ($F(1,18) = .420, p = .525$). The only instance where neighborhood density reached significance was when the high CC low-density and low CC and high-density conditions were compared (mean HCC LD = 992.45 ms; mean LCC HD = 948.03 ms) ($t(39) = 3.091, p < .01$). The interaction between density and CC was not significant ($F(1,18) = .072, p < .720$).

Discussion

The results from the same-different discrimination task confirm the hypothesis that words with high CC produce longer response latencies than words with low CC. The hypothesis about the effects of neighborhood density was that high-density stimuli would show longer response latencies than low-density stimuli. Vitevitch and Luce found longer response latencies for high-density words in a same-

different discrimination task (1999), although they manipulated phonotactic probability directly rather than neighborhood density. One explanation for our failure to find significant effects for neighborhood density is that CC might be responsible for such effects, but in previous studies, neighborhood density was confounded with CC.

In order to understand the effects of CC more completely, it is necessary to obtain converging evidence. The stimuli used in Experiment 1 were spoken in the clear and presented under optimal listening conditions. In Experiment 2, we explored how listeners would respond to words with different levels of CC under degraded listening conditions using a same-different discrimination task. An analysis of the error patterns across experimental conditions was also carried out in order to further understand how CC affects spoken word recognition.

Experiment 2: Same-Different Discrimination. What Errors do Listeners Make?

In the first experiment, we analyzed reaction time data and found a significant main effect for CC. Experiment 2 was designed to investigate the effects of signal degradation on spoken word recognition. The stimuli used in Experiment 2 were degraded by processing them with a noise vocoder, which is used to simulate speech sounds that cochlear implant users are exposed to. The Tiger Speech Cochlear Implant Simulator version 1.01.07 was used to degrade the signal. Filtering the signal into a specific number of frequency bands in the first step in creating vocoded speech—in this experiment we used 8 bands. Once the speech was filtered into a specified number of bands, the amplitude envelope for each band was extracted with a low pass filter. Frequency was then replaced in each band with noise. Shannon, Fang, Kamath, Wygonski, and Ekelid. (1995) found that 4 channels of vocoded speech, using either noise or sine wave carriers, provided sufficient information for word identification in meaningful sentences. Shannon et al.'s findings suggest that this might be the minimal spectral information required for recognition, provided that temporal cues are also present in the signal. In order for subjects to have a better chance of perceiving words in isolation, we used 8 spectral channels instead of 4-channels.

Listeners in Experiment 2 carried out the same discrimination task used in Experiment 1; only now the stimuli were degraded using noise vocoded speech. Recall that error data were analyzed in Experiment 1 for the purpose of ruling out the possibility that significant differences in accuracy were found across conditions, and we observed that listeners made few errors with no difference in accuracy across conditions. Our prediction for Experiment 2 was that low CC pairs of words would be identified by listeners as *same* more accurately than high CC words.

Method

Design

This experiment also used a 2 x 2 design with CC and neighborhood density as independent variables. The dependent variable was percent correct since the degraded signal should decrease response accuracy. The one hundred sixty stimuli were evenly divided into the same four conditions used in Experiment 1.

Participants

The participants were twenty native speakers of Midwestern American English who reported no history of a speech or hearing disorder at the time of testing. The participants were recruited from the undergraduate psychology paid subject pool from Indiana University in Bloomington. Subjects were paid seven dollars and none of them served in the previous experiment.

Stimulus Materials

The same stimuli used in Experiment 1 were used in Experiment 2. In this experiment, both the same and different words were processed using the Tiger Speech Cochlear Implant Simulation version 1.01.07.

Procedure

Listeners were informed that they would hear pairs of English words. This point was emphasized in order to encourage listeners to engage in lexical access and prevent them from interpreting the vocoded speech as noise. Listeners were instructed to respond *same* or *different* as accurately as possible while also moving quickly in the experiment. The procedure was otherwise identical to the task described in Experiment 1.

Results

A series of repeated measures ANOVAs calculated the main effects of CC, neighborhood density, and the interaction between the two variables. A significant main effect was observed for CC (mean percent correct for low CC = 83.7 and the mean percent correct for high CC = 87.6) ($F(1,19) = 19.696$ and $p < .001$). 17 of the 20 listeners recognized high CC words more accurately than low CC words ($p = .004$). The results are shown in Figure 3. The ANOVA of the items analysis showed a trend where high CC words were recognized more accurately than low CC words ($F(1,39) = 3.89$, $p = .056$).

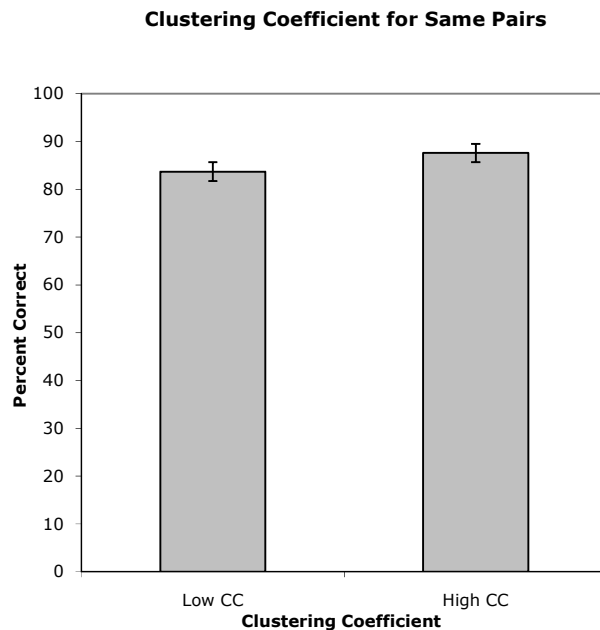


Figure 3. An analysis of the *same* pairs from Experiment 2. We collapsed across the variable neighborhood density and analyzed percent correct as a function of CC. The figure shows that high pairs of CC words were recognized more accurately than low CC pairs of words.

The main effect of neighborhood density showed a marginal but non-significant trend toward low-density words being recognized more accurately than high-density words. The marginal effects of density replicated findings reported in previous studies (Luce & Pisoni, 1998; Vitevitch & Luce, 1998). The mean percent correct for low-density words was 86.625, and the mean percent correct for high-density words was 84.675 ($F(1,19) = 2.144, p = .159$). The marginal main effect for neighborhood density points in the opposite direction as the effect of CC.

We also analyzed the responses to the different pairs. The purpose of this analysis was to determine if listeners were biased to respond *same* or *different*. In other words, how well were listeners able to discriminate the same pairs from the different pairs? To answer this question, the overall d' (sensitivity) for the *same* trials was calculated giving a value of 1.72: where $d' = Z \text{ False Alarm} - Z \text{ Hit}$. The value of 1.72 corresponds to an ROC curve indicating that listeners were able to discriminate the same pairs from the different pairs above chance. We also computed the mean d' for each independent variable across subjects and analyzed this data using one-way ANOVAs. Data showed a significant main effect for CC with a mean d' of 1.711 for low CC words and 1.96 for high CC words ($F(1,19) = 19.381$ and $p < .001$), suggesting that high CC words were more discriminable than low CC words. Effects were non-significant for density ($F(1,19) = 1.094$), and non-significant for the CC x density interaction ($F(1,19) = .634$).

Discussion

The results of manipulating CC in Experiment 2 revealed a different pattern of results from Experiment 1. In Experiment 1, we observed faster reaction times for low CC words relative to high CC words in the same-different task. When we used the 8-channel noise vocoder to degrade the stimuli in Experiment 2, we observed that low CC words were responded to less accurately across same pairs than high CC words, which was the opposite of what we predicted.

One explanation for the surprising and anomalous results from Experiment 2 was that listeners might have used pattern-matching strategies to complete the task without accessing words in their lexicon. That is, listeners carried out the same-different task without recognizing words even though they were told in the instructions that the stimuli consisted of English words. Although the d' analysis showed that listeners could discriminate the same pairs from the different pairs, it is doubtful that they were discriminating them on a lexical basis. In order to test this hypothesis, it is necessary to measure the accuracy of word recognition in a perceptual identification experiment carried out under the same degraded listening conditions. Recall that in Experiment 1, listeners heard words spoken in the clear and were instructed to make same-different judgments based on different tokens of the same word. This provides evidence that listeners were accessing their lexicon in Experiment 1.

In order to analyze listener's error responses while encouraging them to engage in lexical access, we used a perceptual identification experiment with degradation level as a between subjects variable. Unlike the same-different task used in Experiments 1 and 2, perceptual identification provided a more detailed description of how listeners perceived spoken words since they had to access their lexicon and respond with the word they heard. And, since an important aspect of this study is the investigation of the structural relationship of spoken words in the phonological mental lexicon, the perceptual identification paradigm provides a useful means for examining these relationships. Another reason for using perceptual identification in Experiment 3 was that the procedure allowed for the systematic study of listener's error patterns (Savin, 1963), as well as a detailed analysis of the component segments of the correct and incorrect responses.

Experiment 3: CC and Perceptual Identification of Spoken Words

Experiment 2 demonstrated that discriminated high CC words more accurately than low CC words under degraded listening conditions. Because the stimuli were degraded, listeners might not have perceived words and accessed representations from their lexicon, but instead could have relied on an auditory pattern matching strategy. No significant main effect was observed for neighborhood density, although the data indicated a trend toward a higher percentage of correct responses in the low-density condition compared to the high-density condition.

While the data from the same-different discrimination task used in Experiment 1 and 2 provided some indication of error patterns across conditions, they tell us little about how the stimulus words and their sublexical components were perceived by listeners. Which segments do listeners perceive most accurately when exposed to different levels of degraded speech? Also, how does the perception of words and component segments vary as a function of CC and density?

Both correct and incorrect responses were analyzed in Experiment 3 since both types of responses contain phonological and frequency related information. We also measured the word frequency of listener's responses as a function of CC and neighborhood density. Pollack, Rubenstein, and Decker (1960) conducted an analysis of incorrect responses to spoken words presented in noise. Hypothetically, because listeners were more accurate in judging high CC words as *same* in Experiment 2, we expected a similar pattern of results from the identification task in the between subject condition in which listeners were presented with 8-channel vocoded speech. That is, listeners should be more accurate in identifying high CC words relative to low CC words. Another prediction was that since listeners might not have recognized words in Experiment 2, we expected that percent correct identification in the 8-channel condition would be near the floor. Because listeners responded to low CC words more quickly than high CC words in Experiment 1, we expected low CC words to be recognized more accurately as the number of channels increased and the listening conditions improved.

Method

Design

A 3 x 2 x 2 design was used in Experiment 3. The between subject variable was the number of channels (8,10,12) and the within subject variables were CC and neighborhood density. The dependent variables were percent correct response, word frequency of listener's response, and response entropy. The stimuli were evenly divided into four conditions.

Participants

The participants in Experiment 3 were sixty-three native speakers of Midwestern American English, who reported no prior history of speech or hearing disorders at the time of testing. Twenty-one participants were recruited for each of the three between subject conditions from the undergraduate psychology pool at Indiana University in Bloomington. Listeners were either assigned course credit or paid seven dollars for their participation.

Stimulus Materials

The same set of one hundred sixty words used in the *same* pairs in Experiment 1 and 2 were used in Experiment 3. The stimuli were degraded using the Tiger Speech Cochlear Implant simulation version 1.01.07 described under the stimulus materials section under the Experiment 2 heading. The stimuli were degraded using 8-channels, 10-channels, and 12-channels.

Procedure

Experiment 3 used an open-set word identification task. Words were played over *Beyer Dynamic DT 100* headphones connected to a Macintosh computer at a comfortable listening volume. Listeners were instructed to listen to the words and use the keyboard to type in what they thought they heard as accurately as possible. Subjects were also instructed to listen carefully and take their time during the procedure.

Each trial began with the presentation of a plus sign on the center of the screen displayed for 500 milliseconds. After the plus sign disappeared from the screen, a degraded word was played over the headphones at a comfortable volume. After the word finished playing, a dialogue box was displayed on the screen asking the subject to type in what they heard. There was a 1,500 ms pause before the next trial began. The next trial did not begin until the subject finished typing in the response.

Results

In the data analysis, both the target word and response were phonetically transcribed using the alphabet form the CMU dictionary. If the transcription of the target and response matched in the onset, nucleus, and coda positions, the response was scored as correct. If listeners typed in a homonym of a target word, for example, by typing in the word *sea* instead of *see*, the response was scored as correct since the phonetic transcriptions are identical. In the analysis of correct responses, we first analyzed the percentage of words correctly identified as a function of number of channels, CC, and neighborhood density. We also looked at the number of correct responses in the onset, nucleus, and coda position across the two levels of CC and lexical density.

Words

In the first analysis, we computed the percentage of words identified correctly in each condition. The percentage of words identified correctly as a function of channel is shown in Figure 4(a). The overall percentage of words correctly identified as a function of number of channels and CC is plotted in Figure 4(b).

As the level of degradation decreased by increasing the number of channels from 8 to 10 to 12, listeners identified words more accurately. The largest increase in performance occurred when the number of channels was increased from 8 to 10. The ANOVA results show a significant effect of the number of channels ($F(2,60) = 4213.732, p < .001$), clustering coefficient ($F(1,60) = 1399.787, p < .001$), and neighborhood density ($F(1,60) = 4213, p < .001$). There were also significant interactions between CC and the number of channels ($F(1,20) = 1402.87, p < .001$), neighborhood density and CC ($F(1,20) = 769.893, p < .001$), density and the number of channels ($F(1,20) = 1345, p < .001$), and a three way interaction between neighborhood density, CC, and the number of channels ($F(1,20) = 763, p < .001$). The reason for these interactions is that high CC and high-density words were recognized more accurately than low CC and low-density words in the 8-channel condition, while the opposite pattern was the case in the 10 and 12-channel condition.

Significant differences were observed for CC in the 8-channel condition, as indicated in Figure 4(b). The mean percent correct words identified for high CC stimuli was 9.3 percent, and for low CC stimuli, 7.5 percent ($F(1,20) = 5.1, p < .05$). The effects of neighborhood density showed a similar difference where the mean number of correctly identified high-density words was 9.34 percent, and 7.5 percent for low-density words ($F(1,20) = 5.5, p < .05$). The CC x density interaction was also significant ($F(1,20) = 11.182, p < .01$). The data show that the mean percent correct identification for low-density and low CC words was 8.2 percent, low-density and high CC was 6.8 percent. The reverse pattern was true for

high-density stimuli; listeners recognized high CC words more accurately relative to low CC words (11.85 and 6.9 percent respectively). We observed the highest number of words scored correctly when the level of neighborhood density and CC were either both high or both low. A subsequent items analysis revealed no significant differences for CC, density or the interaction—forcing us to question the main effects for the 8-channel condition ($F(1,20) = .015, p < 1.0$).

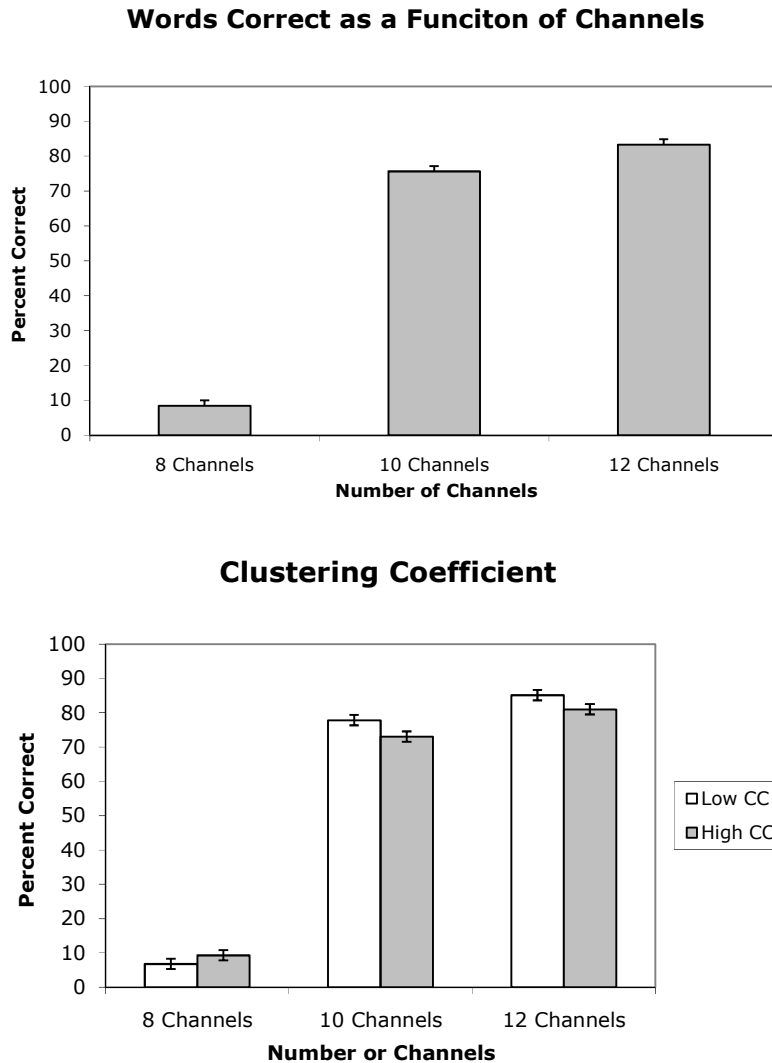


Figure 4. (a) The upper panel of the plot shows percent correct word identification as a function of number of channels. As the signal to noise ratio indicated by the number of channels increases, percent correct increases. Notice that the function is not linear but instead approximates a sigmoidal function. (b) The bottom panel shows percent correct word identification both as a function of CC and number of channels. At the low end of the performance function, listeners identify high CC words more accurately than low CC words. The opposite is true for the 10 and 12-channel conditions.

In the 10-channel condition, we observed an improvement in the percentage of words correctly identified. As predicted, listeners identified low CC words more accurately than high CC words. The mean percent correct for low CC words was 78.30, and for high CC words it was 74.2 percent ($F(1,20) = 21.94, p < .001$). These results confirmed our results from Experiment 1. No significant difference was observed for density ($F(1,20) = .55, p < 1.0$), and no interaction was observed ($F(1,20) = .334, p < 1.0$). An items analysis also showed significant results for CC as predicted ($F(1,20) = 14.8, p < .001$).

For the 12-channel condition, we observed an additional improvement in the percentage of words identified by listeners. As predicted, listeners identified low CC words more accurately, confirming both the results from Experiment 1 measuring reaction time, and the results from Experiment 3 in the 10-channel condition. The mean percent correct for low CC words was 85.12, and for high CC words, 81 percent ($F(1,20) = 18.3, p < .001$). Again, no significant difference was observed for density ($F(1,20) = .45, p < 1.0$), and no interaction was observed ($F(1,20) = 1.2, p < 1.0$). An items analysis showed a marginally significant difference for low and high CC ($F(1,20) = 2.98, p < .10$).

Analysis of Sub-Lexical segments

In another series of tests, we analyzed listener's responses in terms of percent correct of the onset, nucleus, and coda positions. In the first analysis, we examined the percent correct of the onset position as a function of CC and density. These results are shown below in Figure 5 for 8, 10, and 12 channels.

In the 8-channel condition shown in panel (a), the mean percent correct observed was 28.8 percent for high CC words, and was 37 percent for low CC words ($F(1,20) = 39.87, p < .001$). The effects of lexical density were marginal but not significant. The mean percent correct observed was 31.5 percent for high-density words, and 34 was percent for low-density words ($F(1,20) = 3.294, p < .10$). The CC x density interaction was not significant ($F(1,20) = .007, p < 1.0$).

In the 10-channel condition shown in panel (b), we observed significant results for CC and density. We observed a mean percent correct of 90.83 for the onset in low CC words, and 88.87 in high CC words ($F(1,20) = 5.1, p < .05$) (shown in Figure 5). No significant interaction was observed. The mean percent correct for the onset in low-density words was 88.3 and high density was 91.4 ($F(1,20) = 12.69, p < .01$).

A similar trend was observed in the 12-channel data for CC shown in panel (c), although we did not observe significant effects for density, or the CC x density interaction. Listeners responded to the onset cluster correctly 94.12 percent of the time in low CC words, and 91.13 percent of the time for high CC words ($F(1,20) = 19.055, p < .001$). The ANOVA for density and the interaction showed ($F(1,20) = 2.086, p = .164; F(1,20) = .091, p < 1.0$), respectively.

A second set of analyses focused on the nucleus position of the word. The results for the 8-channel condition showed that listeners respond more accurately to the nucleus in high CC words relative to low CC words. The mean percent correct for nuclei in high CC words was 28.3, and for low CC words it was 21.3 percent correct ($F(1,20) = 25.308, p < .001$). Analysis of nuclei also revealed significant results for density (mean HD = 27.9 and mean LD = 21.7 and $F(1,20) = 14.873, p < .001$). The CC x density interaction was non-significant ($F(1,20) = .563, p = .462$).

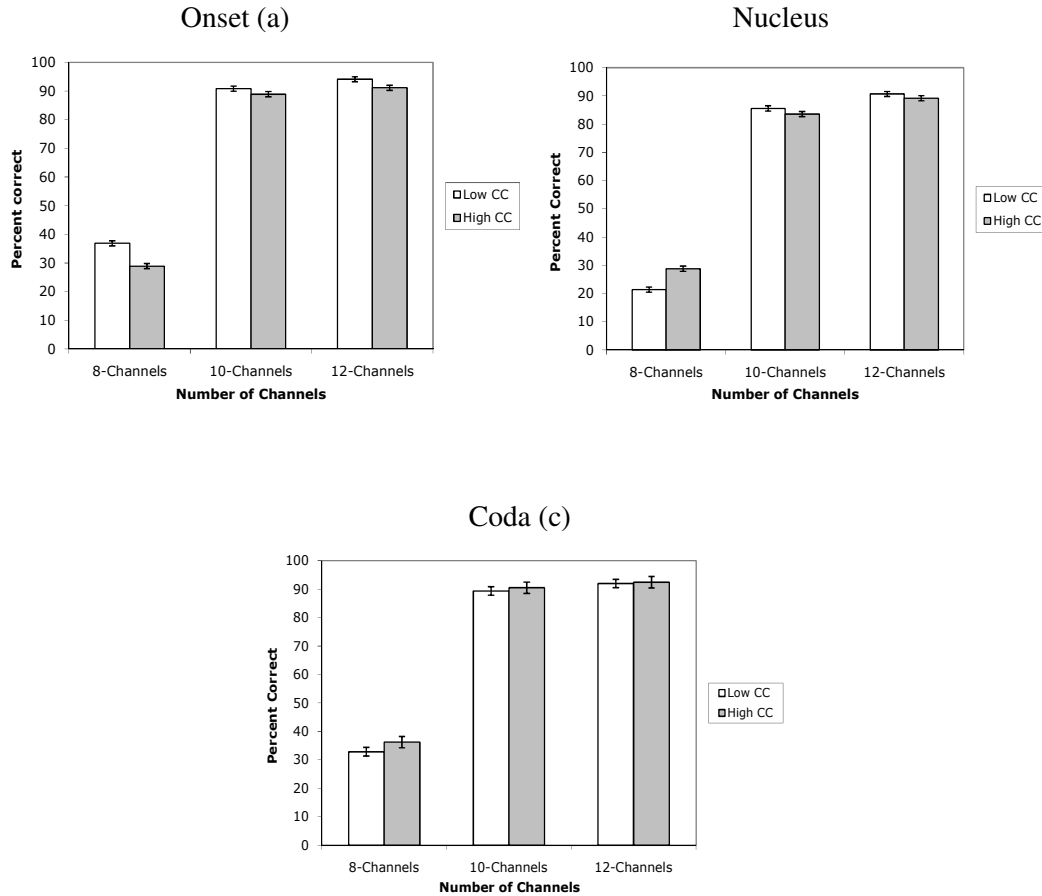


Figure 5. (a) Percent correct identification of the onset position as a function of CC and number of channels. (b) Percent correct identification of the nucleus position as a function of CC and number of channels. (c) Percent correct identification of the coda position as a function of CC and number of channels. Neighborhood density level, rather than CC has a more salient effect on identification in the coda position.

A significant result for CC was observed in the 10-channel condition. Listeners accurately identified the vowel in the nucleus position in low CC words 85.6 percent of the time and 83.6 percent of the time in high CC words ($F(1,20) = 8.157, p < .01$). No significant results were observed for neighborhood density ($F(1,20) = .09, p < 1.0$). Non-significant results were also observed for the CC x density interaction ($F(1,20) = 2.079, p = 165$).

Significant results were not observed for any of the independent variables in the 12-channel condition (mean HCC = 90.7 and mean LCC = 89.23). The mean percent correct for low-density words was 90, and for high-density words it was 89.94. The F value for the CC x density interaction was .005, and $p < 1.0$. Figure 5 (b) shows the percent correct for the nucleus position plotted as a function of CC and number of channels.

In a final analysis of the sub-lexical components of listener's responses, we examined the mean percent correct responses in the coda position. In the 8-channel condition, the data showed non-significant results for CC ($F(1,20) = 1.6, p = .22$). We observed significant effects for density with a mean percent

correct of 41.25 for high-density words, and 29.4 for low-density words ($F(1,20) = 40.675, p < .001$). Even though CC was not significant, the CC x density interaction was ($F(1,20) = 20.965, p < .001$).

In the 10-channel condition, significant results were observed for density with a mean of 88.27 percent correct for low-density words, and 91.55 percent correct for high-density words ($F(1,20) = 12.583$ and $p < .01$). Non-significant results were observed for CC ($F(1,20) = 2.326, p < .143$). We also observed a significant CC x density interaction ($F(1,20) = 8.325, p < .01$). Although significant effects were observed for neighborhood density in the coda position, they went in the opposite direction of what we predicted based on previous studies (Luce & Pisoni, 1998).

The twenty-one listeners in the 12-channel condition provided similar data as the 10-channel condition. The average percent correct for low-density words was 90.2, and 94.27 for high-density words ($F(1,20) = 4.537, p < .05$). No significant results were observed for CC, or the interaction. The percent correct identification across listeners for the coda position as a function of CC and number of channels is shown in panel (c) of Figure 5. Table 2 summarizes the proportion of correct response for high and low CC words in the onset, nucleus and coda positions as the number of channels increases.

Channel	Position	High CC	Low CC
8 C	Onset	.29	.37
	Nucleus	.28	.21
	Coda	.41	.29
10 C	Onset	.89	.91
	Nucleus	.84	.86
	Coda	.92	.88
12 C	Onset	.94	.91
	Nucleus	.90	.90
	Coda	.94	.90

Table 2. Proportion of correct responses across listeners for high and low CC words for number of channels in the onset, nucleus and coda positions.

Analysis of Incorrect Responses

Word Frequency of Incorrect Responses

Does the word frequency of incorrect responses differ across levels of CC and number of channels? Pollack, Rubenstein, and Decker (1960) reported that as the signal-to-noise ratio improved, the word frequency of listener's responses decreased independently of the stimulus word frequency. In a similar study of frequency bias in responses using a visual word recognition task, Goldiamond and Hawkins (1958) found that in the absence of any stimuli, subjects were more likely to respond with items they had been exposed to more frequently during training.

Not long after Pollack et al. (1960) published their results, Gerstman and Bricker (1960) discovered a serious methodological error in their study. Pollack et al. presented listeners with a list of 144 words three times at signal-to-noise ratios of 0, 5, 10, 15, 20, and 25, in that particular order. The observed decrease in the word frequency of incorrect responses was confounded with learning since the words were repeated multiple times across signal-to-noise ratios. In our study, the number of channels

was a between subject condition, where listeners were exposed to each word only once. We investigated whether we could replicate Pollack et al.’s results without the confounding of learning. Additionally, we analyzed incorrect responses to determine whether CC or density affect the word frequency of listener’s responses.

To determine the word frequency of incorrect responses, we calculated the modal incorrect response to each stimulus in each of the three between subject conditions. The modal incorrect responses were then placed into the Washington university online speech and orthographic database (<http://128.252.27.56/Neighborhood/Home.asp>) where an items analysis computed the word frequency score for each word. A necessary condition for a word to be declared a modal incorrect response was that it be given as an incorrect response at least twice.

Figure 6 plots the log frequency of incorrect responses as a function of channels. As predicted by Pollack et al.’s data, the trend indicates that as the number of channels increases, the average word frequency tends to decrease. A pair-wise t-test between the 8 and 10 channel conditions showed a marginally significant trend, where the mean log frequency of incorrect responses for 8-channels was 2.732, and for 10-channels it was 2.45 ($t(17) = 2.01, p = .061$). While the average log frequency for the 12-channel condition was lower than the frequency for 10-channels (mean = 2.39), no significant difference was observed between those two conditions ($t(17) = .777, p < 1.0$). A t-test revealed that the only possible difference was between the 8 and 12-channel conditions (without correcting for multiple comparisons) ($t(17) = 5.0, p < .05$). In short, we observed effects similar to Pollack et al (1960) without confounding the level of stimulus degradation with word repetition.

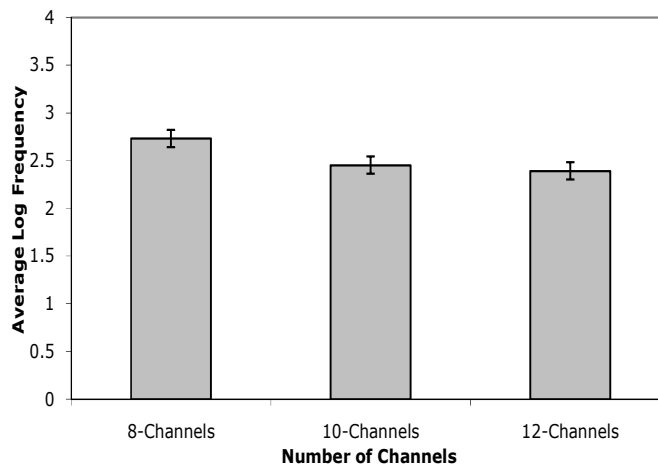


Figure 6. The plot shows the average log frequency of listener’s incorrect responses as a function of the number of channels. The overall mean for each condition is collapsed across CC and neighborhood density.

Next, we analyzed the frequency of incorrect responses as a function of CC and density. In the 8-channel condition, the data showed higher word frequency of incorrect responses for high CC words relative to low CC words (mean HCC = 2.88 and mean LCC = 2.59; $F(1, 39) = 4.775, p < .05$). The effects of density and the CC x density interaction were non-significant ($F(1, 39) = .397, p < 1.0$; $F(1, 39) = .686, p < 1.0$). In Experiment 1 we observed faster reaction times, and in Experiment 3 we observed more accurate responses for low CC words than high CC words. This suggests that low CC words are more

easily recognized than high CC words. This explanation is consistent with Pollack et al.'s earlier result showing a decrease in word frequency of responses as signal-to-noise ratio increases. Only now, we are showing a similar pattern in the error responses for the independent variable CC rather than signal-to-noise ratio.

The effect of CC on the word frequency of incorrect responses was weaker in the 10 and 12-channel conditions—where no significant effects were observed (for CC and 10-channels, $F(1, 39) = .002$, $p < 1.0$; for density and 10-channels, $F(1, 39) = .680$, $p < 1.0$; for CC and 12-channels, $F(1, 39) = 1.72$; for density and 12-channels, $F(1, 39) = 4.35$, $p = .052$; and the interaction, $F(1, 39) = .50$, $p < 1.0$). Only density in the 12-channel condition was (marginally) significant. Figure 7 is a plot of the average log frequency as a function of number of channels, and levels of CC.

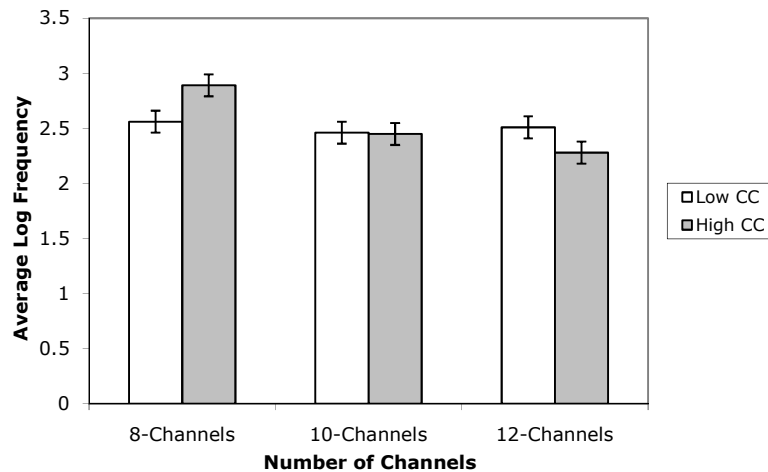


Figure 7. This figure shows the log frequency of listener's incorrect responses as a function of both CC and number of channels. CC is collapsed across neighborhood density.

Diversity of Incorrect Responses

In order to obtain a measure of uncertainty across experimental conditions, we also calculated the number of different incorrect responses given by listeners for each stimuli in the 8, 10, and 12 channel conditions. This analysis was carried out because we hypothesized that different levels of variability or entropy across CC and density might be related to response properties. That is, if listeners give many different responses relative to the total number of incorrect responses, it suggests that they were inconsistent in using incomplete information. A measure of *response entropy* was computed by dividing the number of unique incorrect responses by the total number of incorrect responses given to each stimulus word. We called the measure *response entropy* because it measures variability and uncertainty in the pattern of incorrect responses. The question in the following analyses is whether CC, density, or number of channels, affects the level of entropy in listener's incorrect responses.

Figure 8(a) shows the effect of CC on response entropy across channels, and Figure 8(b) shows analogous results for the effect of lexical density on response entropy.

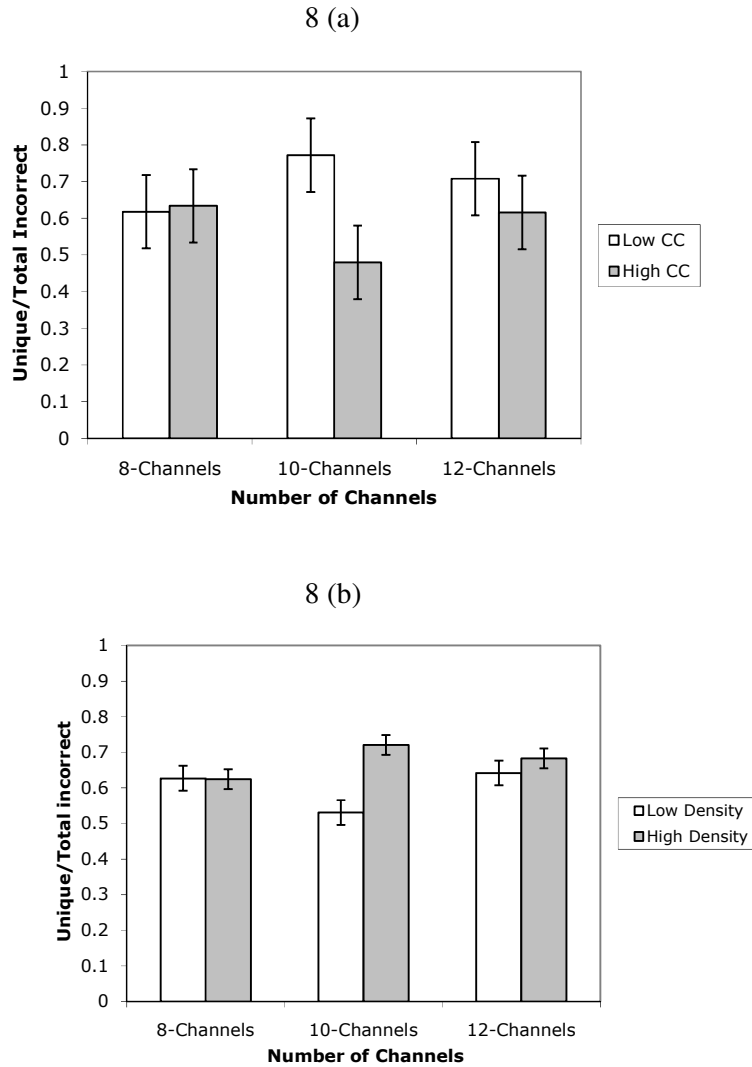


Figure 8. (a) The top panel of the figure shows entropy levels as a function of CC across noise channels. (b) The bottom panel shows entropy as a function of neighborhood density and noise channels.

The results from the ANOVA show that there was not an overall effect of number of channels ($F(1, 39) = 2.501, p = .130$), but there was a significant effect for CC ($F(1, 39) = 11.17, p < .01$) and neighborhood density ($F(1, 39) = .042, p < .05$). We also observed a significant interaction between CC and number of channels ($F(1, 39) = 17.662, p < .001$).

The data suggest that levels of response entropy were unaffected by lexical density or CC in the 8-channel condition as shown in Figure 8. The interaction between density and CC with respect to response entropy was marginally significant ($F(1, 39) = 3.457, p < .10$). The mean level of response entropy for low CC words was .618, and for high CC words it was .634 ($F(1, 39) = .509, p = .480$). The mean entropy for low-density words was .625, and for high-density words it was .627 ($F(1, 39) = .002, p < .1.0$).

Significant effects for response entropy were observed, however, across levels of CC and density in the 10-channel condition. A significant interaction between the two variables was observed as well. As expected, a higher level of response entropy was observed in high-density words compared to low-density words (mean HD = .531, and mean LD = .721; $F(1, 39) = 16.772, p < .001$). These data are consistent with the predictions of NAM (Luce & Pisoni, 1998) where high-density words are inhibited more by similar sounding phonological neighbors than low-density words. The higher level of inhibition and lexical competition in dense neighborhoods could potentially cause listeners to make a wider variety of errors. We observed higher levels of response entropy in Low CC words than in high CC words (mean LCC = .772, mean HCC = .480; $F(1, 39) = 41.591, p < .001$). This result was unexpected considering listeners respond more accurately to low CC words. We also observed a significant interaction between CC and density ($F(1, 39) = 16.13, p < .001$).

A trend similar to the 10-channel condition was observed under 12-channels. Listeners demonstrated higher levels of uncertainty for high-density words relative to low-density words (mean HD = .681 and mean LD = .570; $F(1, 39) = 7.747, p < .01$). Likewise, we observed the opposite trend for CC that we observed in the 10-channel condition (mean HCC = .502, mean LCC = .749; $F(1, 39) = 37.51, p < .001$).

Discussion

Analysis of Words Correct

In the 8-channel condition in Experiment 3, percent correct identification was low, suggesting that listeners were not able to reliably recognize words, but were guessing based on limited phonological information. The finding that listeners used guessing strategies rather than engaging in lexical access could explain why we did not observe the predicted results for CC: i.e., that low CC words would be recognized more accurately than high CC words. It is possible that if listeners had more exposure to the stimuli in the 8-channel condition, the pattern of results might be different and perhaps match the data from the 10 and 12 channel conditions where listening conditions improved. Observing the strategies used for recognizing spoken words after listeners adapt to highly degraded speech would be an interesting direction for future research.

The results from the 10 and 12-channel conditions confirmed the hypothesis that low CC words would be identified more accurately than high CC words. These results were based on predictions from Experiment 1 where listeners responded to low CC words more quickly because there was less competition from similar sounding words in the lexicon than there was for high CC words.

Analysis of Sub-Lexical Segments

The second major set of analyses concerned the response properties of the onset, nucleus, and coda. We examined whether accurate identification for each position differed as a function of CC or density. Another purpose of Experiment 3 was to investigate in which part of the word more errors were made.

We observed more accurate identification scores for onsets in low CC words than high CC words in 8, 10, and 12-channels. The data indicate that a lower degree of connectivity among the neighbors of a word allows listeners to more accurately identify the onset cluster. This reasoning is consistent with the results from Experiment 1 showing faster reaction times for low CC words, and the results from Experiment 3 showing more accurate identification for low CC words.

The data from the responses in the nucleus position differed from the onset position. Listeners correctly identified the nucleus in high CC words more accurately than the nucleus in low CC words in the 8-channel condition. Neighborhood density also significantly affected identification rates, where high-density words were more accurately identified than low-density words. The fact that density had the strongest effect in the coda position, regardless of the number of channels, suggests that when density is a factor in the recognition process, the effect is most salient for deletions, additions, or substitutions at the end of words.

General Discussion

The goal of this project was to investigate how the connectivity of a word's phonological neighbors in its subspace affect the reaction time and accuracy of spoken word recognition. This project extended the assumption of relational word recognition by embedding it within a graph theoretical framework to empirically test the proposal that the representation of words in memory can be modeled as a complex system. In order to obtain support for the hypothesis that the representation of words in memory might share properties with other complex systems, we carried out several behavioral studies to determine whether a word's CC affects spoken word recognition.

The theoretical motivation for the present set of experiments was based on the proposal that the mental lexicon can be viewed as a multidimensional space. Treisman (1979) described *partial identification theory* as a model for how the search of the lexical space is carried out by listeners. Partial identification theory assumes that only a subspace of the lexicon is searched, where the size of the subspace depends on the quality of the listening conditions. Recall that this is what distinguishes partial identification theory from Luce's universal forced choice model (1959). Also, when listening conditions are optimal, the subspace searched by the algorithm is very constrained, perhaps including the stimulus word and its immediate phonological neighbors. Under highly degraded listening conditions, the size of the subspace becomes much larger, and if conditions are degraded enough, the size of the subspace might well include most of the words in the lexicon. Under these listening conditions, the size of the subspace selected by partial identification theory would approximate Luce's universal forced choice model. Thus, one would predict that when listening conditions become severely degraded and listeners begin to use unconstrained guessing strategies, the word frequency of listener's incorrect responses would increase. In short, if listeners were searching a subspace consisting of a significantly large portion of the lexicon, their incorrect responses would generally be high frequency words because words with a higher frequency of occurrence in the language are more likely to be generated as responses to degraded and underspecified signals.

While the primary focus of these experiments was on reaction time and accuracy data, the incorrect responses from Experiment 3 provided information about structure in the lexicon. In the 8-channel condition in Experiment 3, listening conditions were highly degraded. The fact that listeners were biased to generate high frequency words as incorrect responses indicates that they were using a forced-choice decision rule over a very large subspace of the lexicon. The bias of listeners to generate high frequency error responses under degraded listening conditions replicated the previous results of Pollack et al. (1960).

The data from Experiment 3 also show that as the number of spectral channels increased from 8 to 10 to 12, the word frequency of listener's error responses decreased, suggesting that the subspace of the lexicon being searched was more highly constrained because more reliable stimulus information could be obtained from the signal. Therefore, the observation in Experiment 3 that the relation between frequency of the stimulus and frequency of the error response changed as listening conditions improved was

contrary to the predictions made by models of spoken word recognition that assume pure guessing strategies and the underlying assumption of acoustical or structural equivalence.⁴ The results from the 10 and 12 channel conditions in Experiment 3 were consistent with Treisman's partial identification theory, which assumes that as listening conditions improve, the lexical neighborhood or subspace becomes more refined, reducing the bias to generate high frequency words as incorrect responses.

The results obtained in Experiment 1 using a same-different discrimination task suggest that increasing the level of CC among lexical neighbors slows the discrimination process down. Experiment 2, while methodologically flawed in the sense that listeners were not reliably recognizing words, showed that CC and neighborhood properties had different effects under degraded listening conditions, where listeners were using guessing strategies. That listeners were using guessing strategies was evident from the results obtained in the 8-channel condition in Experiment 3 in which the mean percent correct identification of words was under 10 percent. That is, it was likely that listeners were drawing upon a large number of potential lexical candidates in a large subspace of the lexicon based on partial phonological information. We also observed in the 8-channel condition in Experiment 3 that when listeners made errors, their incorrect responses were biased toward high frequency words, a prediction derived from previous studies (Pollack et al., 1960).

The results from the 10 and 12-channel conditions in Experiment 3 replicated the general pattern of results observed in Experiment 1. Putting the results from Experiments 1 and 3 together, it is important to begin considering various models that describe how the lexical effects of CC on the word recognition process operate. As discussed in the introduction, previous models of spoken word recognition have not addressed how the global properties of the mental lexicon affect spoken word recognition. Neither the "urn model" (Pollack et al., 1959; see also Oldfield, 1966) nor Logogen theory (Morton, 1979) explicitly define how words might be related to or connected to one another. Since current models in the field that assume relational word recognition like NAM, TRACE, and Cohort Theory have not described how global lexical variables operate and affect word recognition and retrieval, it is important to begin exploring how graph theoretical variables such as CC, which are integral to complex systems, affect word recognition in the context of a particular model. What type of model could predict these effects of CC, given the computational analyses carried out by Vitevitch (2004), Gruenenfelder and Pisoni (2006), as well as the behavioral results obtained in this current study, including the reaction time data generated in Experiment 1, and the results generated by the perceptual identification Experiment 3? The global structure and topology of the lexicon, including the interconnectivity of words in a lexical subspace is an important structural parameter affecting spoken word recognition.

Summary and Conclusions

Recent research on natural and artificial complex systems provided the theoretical motivation for this study. In the present set of experiments, we found effects of the graph theoretical variable CC on spoken word recognition. The motivation for studying this variable also came from several models that assume that spoken words are recognized relationally, and that the lexicon can be represented as a multidimensional acoustical space. Recent computational studies of the effects of graph theoretic variables on spoken word recognition also suggested the usefulness of this approach. More generally, our goal was to examine how the global structure and topology of words in the mental lexicon affect spoken

⁴ See Treisman (1979) for a discussion on this topic. He argued that sophisticated guessing theory relies on the argument that the acoustical and structural properties of high and low frequency words are identical (Pollack et al., 1960). Since the phonological properties of high and low frequency words were believed to be identical, there was no need for the listener to focus on a particular "subspace" of the lexicon. In the introduction, there was a discussion regarding how Landauer and Streeter (1973) challenged this assumption.

word recognition. The three behavioral experiments analyzing correct and incorrect responses represent preliminary efforts to demonstrate the psychological reality and potential importance of CC as a new variable that affects spoken word recognition and performance.

Because we found that CC affects spoken word recognition in both same-different discrimination and perceptual identification tasks, it is important for future models of spoken word recognition to account for these new results showing that not all phonological neighborhoods of similar size are equal in their effects of the recognition process. Structural properties within a neighborhood including CC affect word recognition as well. Models motivated by the theoretical foundations of complex systems, like the general spreading activation model proposed here, should be developed to account for these findings.

References

- Clark, H.H. (1973). The language-as-fixed fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 12, 335-359.
- Crystal, D. (1980). *A first dictionary of linguistics and phonetics*. London: Andre Deutsch.
- Eukel, B. (1980). A phonotactic basis for word frequency effects: Implications for automatic speech recognition. *Journal of the Acoustical Society of America*, 68, S33.
- Gerstman L.J., & Bricker, P. (1960). Word frequency effects in learning unknown message sets. *Journal of the Acoustical Society of America*, 32, 1078-1079.
- Goldiamond, I. & Hawkins, W.F. (1958). Vexiersversuch: The log relationship between word frequency and recognition obtained in the absence of stimulus words. *Journal of Experimental Psychology*, 56, 457-463.
- Gruenenfelder, T.M. & Pisoni, D.B. (2006). Modeling the mental lexicon as a complex System: Some preliminary results using graph theoretic measures. Unpublished Manuscript.
- Howes, D. (1957). On the relation between intelligibility and frequency of occurrence of English words. *Journal of the Acoustical Society of America*, 29, 296-305.
- Kucera, F., & Francis, W. (1967). *Computational analysis of present day American English*. Providence, RI: Brown University Press.
- Landauer, T.K. & Streeter, L.A. (1973). Structural differences between common and rare words: Failure of equivalence assumptions for common and rare words. *Journal of Verbal Learning and Verbal Behavior*, 12, 119-131.
- Levelt, W.J.M. (1978). A survey of studies of sentence perception. In W.J.M. Levelt & G.B. Flores d'Arcais (Eds.), *Studies in the perception of language*. New York: Wiley. Pp. 1-73.
- Lively, S.E., Pisoni, D.B. & Goldinger, S.D. (1994). Spoken Word Recognition: Research & Theory. In M. Gernsbacher (Ed.), *Handbook of Psycholinguistics*, New York: Academic Press. Pp. 265-301.
- Luce, R. D. (1959). *Individual Choice Behavior*. New York; Wiley.
- Luce, P.A., Pisoni, D.B. (1998). Recognition of spoken words: The Neighborhood Activation Model. *Ear & Hearing*, 19, 1-36.
- Marslen-Wilson, W.D. (1984). Function and process in spoken word recognition. A tutorial Review. In H. Bouma & D.G. Bouwhis (Eds.), *Attention and Performance X: Control of Language Processes*. Hillsdale, NJ: Erlbaum. Pp 125-150.
- Marslen-Wilson, W.D. (1987). Functional parallelism in spoken word recognition. *Cognition*, 25, 71-102.
- Marslen-Wilson, W.D. (1990). Activation, competition, and frequency in lexical access. In G.T.M. Altman (Ed.), *Cognitive models of speech processing: Psycholinguistic Computational Perspectives*. Cambridge, MA: MIT Press. Pp. 148-172
- McClelland, J.L. & Elman, J.L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1-86.

- Morton, J. (1979). Word recognition. In J. Morton & J.C. Marshall (Eds.), *Structures and Processes*. Cambridge: MIT Press. Pp. 108-156.
- Nusbaum, H.C., Pisoni, D.B. & Davis, C.K. (1984). Sizing up the Hoosier Mental Lexicon: Measuring the familiarity of 20,000 words. In *Research on Speech Perception Progress Report No. 10*, (Pp. 357-376), Bloomington, IN: Speech Research Laboratory.
- Oldfield, R.C. (1966). Things, words and the brain. *Quarterly Journal of Experimental Psychology*, 18, 340-353.
- Pollack, I., Rubenstein, H, & Decker, L. (1959). Intelligibility of known and unknown message sets. *Journal of the Acoustical Society of America*, 31, 273-279.
- Pollack, I., Rubenstein, H. & Decker, L. (1960). Analysis of incorrect responses to an unknown message set. *Journal of the Acoustical Society of America*, 32, 454-457.
- Protopapas, A. (1999). Connectionist modeling of speech perception. *Psychological Bulletin*, 125, 410-436.
- Savin, H.B. (1963). Word-frequency effect and errors in the perception of speech. *Journal of the Acoustical Society of America*, 35, 200-206.
- Shannon, R.V., Fan-Gang, Z., Kamath, V., Wygonski, J., Ekelid M. (1995). Speech recognition with primarily temporal cues. *Science*, 270, 303-304.
- Speech and Hearing Lab Neighborhood Database. Retrieved from <http://128.252.27.56/Neighborhood/Home.asp>
- Steyvers, M. & Tenenbaum, J.B. (2005). The large-scale structure of semantic networks: Statistical analysis and a model of semantic growth. *Cognitive Science*, 29, 41-78.
- Trask, R.L. (1996). *A dictionary of phonetics and phonology*. Routledge: London.
- Treisman, M. (1979). Space or lexicon? The word frequency effect and the error response frequency effect. *Journal of Verbal Learning & Verbal Behavior*, 17, 37-59.
- Vitevitch, M.S. & Luce, P.A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science*, 9, 325-329.
- Vitevitch, M.S. & Luce, P.A. (1999). Probabilistic Phonotactics and Neighborhood Activation in spoken word Recognition. *Journal of Memory and Language*, 40, 374-408.
- Vitevitch, M.S., & Luce, P.A. (2004). A Web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavioral Research Methods, Instruments & Computers*, 36, 481-487.
- Vitevitch, M.S. Luce, P.A. Charles-Luce, J., & Kemmerer, D. (1996). Phonotactic and metrical influences on adult ratings of spoken nonsense words. Proceedings from the *Fourth International Conference on Spoken Language*, Volume 1, Issue, 3-6 Oct 1996 Pp. 82 – 85.
- Vitevitch, M.S. (2004). Phonological neighbors in a small world: What can graph theory tell us about word learning? Unpublished Manuscript.
- Watts, D.J. & Strogatz, S.H. (1998). Collective dynamics of small-world networks. *Nature*, 393, 440-442.

