

**RESEARCH ON SPOKEN LANGUAGE PROCESSING**

Progress Report No. 27 (2005)

*Indiana University*

**Modeling the Mental Lexicon as a Complex System:  
Some Preliminary Results Using Graph Theoretic Measures<sup>1</sup>**

**Thomas M. Gruenenfelder and David B. Pisoni**

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

---

<sup>1</sup> The research reported in this paper was supported by NIH Grants DC00111 and DC00012. The authors would like to thank Nick Altieri, Vsevolod Kapatsinski, Shane Mueller, and Mike Vitevitch for valuable discussions of the work reported in this paper.

## Modeling the Mental Lexicon as a Complex System: Some Preliminary Results Using Graph Theoretic Measures

**Abstract.** The mental lexicon used for spoken word recognition was modeled as a complex system using tools of graph theory. Words were represented as nodes in the model, and an edge was placed between two nodes if the corresponding words could be changed into one another via a single phoneme deletion, addition, or substitution. The resulting graph had a small-world, scale-free structure. However, the scale-free property reflected the fact that words have different lengths and are created from a relatively small set of phonemes, rather than reflecting the way the network evolves over time. Various network properties of words were also found to be correlated with listeners' performance in an open-set word identification task and in a word repetition task. Those properties included the number of lexical neighbors of a word at different network distances from that word, the mean shortest distance from a word to all other words within the mental lexicon, and a word's clustering coefficient, a measure of the probability that a word's neighbors are also neighbors of one another. The results suggest that including these new measures did not significantly improve the ability to predict the accuracy with which a word can be recognized in various levels of noise using only the word's neighborhood size. In contrast, repetition latencies tended to be longer for words with higher clustering coefficients. Furthermore, this effect did not appear to be modulated by the word's neighborhood size. Possible reasons for the effects of this non-local, global variable are discussed.

### Introduction

A number of recent studies have modeled a diverse set of complex systems as graphs or networks (see Albert and Barabási, 2002, for a review). These systems include the structure of the Internet (Faloutsos, Faloutsos, & Faloutsos, 1999) and of the World Wide Web (Huberman, & Adamic, 1999; Huberman, Pirollo, Pitkow, & Lukose, 1998; Lawrence & Giles, 1998, 1999), metabolic interactions (Jeong, Tombor, Albert, Oltavi, & Barabási, 2000), protein-protein interactions (Wuchty, 2001), citation patterns in scientific papers (Newman, 2001), neural networks (Achacosa & Yamamoto, 1992), contacts among potential disease carriers (Liljeros, Edling, Amaral, Stanley, & Aberg, 2001), and different aspects of language, including people's representations of word meanings (Steyvers & Markham, 2004) and the co-occurrence of words in sentences (Dorogovtsev & Mendes, 2001; Ferrer & Solé, 2001). The World Wide Web, for example, can be modeled as a graph in which each web site is represented by a *node*. An *edge* between two nodes is created if the web site represented by one node has a link to that represented by the second node. Graphs can be either *directed*, in which each edge has a particular direction, from one node to the other (web site A links to web site B), or *undirected*, in which case each edge has no specific direction (persons A and B are married to one another).

A question typically asked in such studies is whether the system of interest shows a "small world" (Albert & Barabási, 2002; Barabási & Albert, 1999; Watts & Strogatz 1998), "scale-free" (Albert & Barabási, 2002; Barabási & Albert, 1999) structure. In a small-world network, the mean shortest path length between any two arbitrary nodes in the network—that is, the minimum number of edges that must be traversed to get from one of the two nodes to the other—is small relative to the total number of nodes in the network. More precisely, the mean shortest path length grows much more slowly than the number of nodes. Albert, Jeong, and Barabási (1999), for example, found that in the University of Notre Dame intranet, which at the time consisted of over 300,000 documents, any arbitrary document could be

reached from another arbitrary document by traversing on average 11 links. In 1998, in the World Wide Web as a whole, which at the time consisted of over one billion documents, the mean shortest path length between any two documents was estimated to be 19 links (Albert et al., 1999). In a well-known study, Milgram (1967) asked people (the sender) in one part of the United States to forward a letter to another person (the target, who was not known to the sender) in another part of the United States by sending the letter to someone known to the sender who in turn the sender thought might know the target person. This intermediate recipient of the letter, in turn, would either forward it directly to the target, if known, or to another intermediate recipient. Milgram found that for those letters that eventually reached their target, the mean number of intermediate recipients, out of a population at the time of 175 million, was 6. All these values are much smaller than would be predicted for a random network, that is, a network with the same total number of connections, but where the connections were placed between randomly selected pairs of nodes.

Many small-world networks also have a higher than chance clustering coefficient (Albert & Barabási 2002; Watts & Strogatz 1998). The clustering coefficient (CC) is a measure of the probability that two nodes, B and C, are connected, given that a third node A is connected to both B and C. In other words, the CC is a measure of the probability that any two neighbors of a given node are themselves neighbors.

Scale free structure refers to a property of the network's degree distribution (Albert & Barabási, 2002; Barabási & Albert, 1999). A node's degree is the number of edges going into (the in-degree) or out of (the out-degree) the node, or both into and out of the node. The degree distribution is the frequency distribution of node degrees in the network. A scale-free network has a degree distribution characterized by a power law,  $N(k) \sim k^{-\gamma}$ , where  $N(k)$  is the degree distribution,  $k$  is the degree (i.e., the number of edges going into or coming out of the node), and the exponent,  $\gamma$ , is typically between 2 and 3. In other words, on a log-log plot, the degree distribution is a straight line with a slope between -2 and -3. In such a degree distribution, most of the nodes have only a very few edges coming into or going out of them. A small number of nodes, however, frequently referred to as hubs, are connected to a very large number of edges.

Demonstrating that a network has a small-world, scale-free structure is important because it potentially has several implications for how the network developed over time (Albert & Barabási, 2002; Barabási & Albert, 1999; Barabási, Albert, & Jeong, 1999). In particular, Barabási and his colleagues have argued that a small-world, scale-free structure indicates that the network evolved over time (1) by adding new nodes, and (2) through a process called "preferential attachment." When a new node is added to a network, a process must exist for it to form edges to other nodes. In preferential attachment, a new node forms edges with an already existing node with a probability that is proportional to the number of edges that existing node already has, i.e., with a probability proportional to the existing node's degree. Such a growth process results in rich nodes—nodes with a large number of edges—getting richer—getting even more edges, a phenomenon sometimes referred to in the psychological literature as the "rich get richer" principle.

The present paper models the human mental lexicon of spoken words as a complex network. The mental lexicon refers to the representations in our brains of the various characteristics of every word we know, including semantic, orthographic, and acoustic-phonetic characteristics. The focus here is on the last of these properties, the acoustic-phonetic properties, or the word's sound structure. Our long-term lexical knowledge is part of what enables the rapid and efficient recognition of speech under a wide range of listening conditions. By comparing the acoustic input to stored representations in the mental

lexicon, and applying some algorithm or heuristic for selecting the best match, listeners are able to recognize each of the individual words spoken by a talker (Luce & Pisoni, 1998).

Traditionally, linguists describe a word's sound as a sequence of phonemes. A phoneme is the smallest unit of sound that distinguishes meaning within a given language. It is an idealized abstract unit in the sense that it does not distinguish all differences in sound but only those necessary to differentiate meaning. The American English word "cat," for instance, though its exact pronunciation varies from utterance to utterance, can be described as consisting of the three phonemes, /k/, /æ/, and /t/, usually written as /kæt/, which is referred to here as the phonetic transcription.<sup>2</sup> The phoneme /k/ distinguishes it from a number of other American English words, such as "mat" (/mæt/), "pat" (/pæt/), and so on. The phoneme /æ/ distinguishes it from other American English words, such as "kit" (/kIt/), and similarly for the phoneme /t/. In American English there are approximately 12 vowel phonemes and 24 consonantal phonemes, the exact number varying by dialect and phonetician.

More than twenty years ago, Nusbaum, Pisoni, and Davis (1984) created an on-line lexicon of nearly 20,000 American English words, based on Webster's Pocket Dictionary (*Webster's Seventh Collegiate Dictionary*, 1967). This lexicon has become known as the Hoosier Mental Lexicon, or HML. The lexicon contains every word in that dictionary, but with homophones and morphemic derivatives eliminated. For instance, the word "dear" (/dɪr/) appears in the lexicon; the homophone "deer" (also /dɪr/) does not. The word "ask" (/æsk/) appears; the morphemic derivatives "asks," "asking," and "asked" do not. The on-line lexicon contains a phonetic transcription for each word, as well as information such as the word's frequency in printed English (Kucera & Francis, 1967), its orthography (i.e., spelling), its syntactic role(s) (noun, verb, adjective, and so on), and length in number of phonemes.

Following an earlier paper by Vitevitch (2004), an undirected graph was constructed from this lexicon in the following manner. Each word was represented as a node. An edge was created between two nodes if the word represented by one can be turned into the word represented by the other through the deletion, addition, or substitution of a single phoneme. (In the remainder of this paper, this rule is referred to as the Deletion-Addition-Substitution or DAS rule.) Otherwise, no link was placed between the two nodes. This rule has long been used for operationally distinguishing similar sounding words from dissimilar sounding words (Greenberg & Jenkins 1964; Landauer and Streeter, 1973). Two words that can be changed into one another using the DAS rule are referred to as lexical neighbors. A word's total collection of neighbors is referred to as its lexical neighborhood. As an illustration of the rule, the neighborhood of the word "bait" (/bet/) includes the words "late" (/let/), "rate" (/ret/), "bit" (/bIt/), "bail" (/bel/), and "bake" (/bek/), but not the word "sane" (/sen/) or the word "bare" (/ber/).

Two primary questions were then asked about the properties of this graph. Does the structure of the lexicon, modeled this way, show a small-world, scale-free structure? Finding that it does have such a structure has potentially strong implications for theories of how children acquire language (Vitevitch, 2004). In particular, this result would suggest that children acquire new words using a process of preferential attachment, learning words that are neighbors of words that they already know. We are not the first to ask if the mental lexicon shows a small-world, scale-free structure. Vitevitch, using the same base lexicon (Nussbaum et al., 1984) and the same single phoneme DAS rule as we used, modeled the mental lexicon as a graph and concluded that it does in fact follow a small-world, scale-free structure. Thus, our initial work provides an opportunity to replicate Vitevitch.

---

<sup>2</sup> Throughout this paper, the International Phonetic Alphabet (IPA) symbols are used for phonemes. Following the normal conventions in the word recognition literature, phonemic transcriptions are enclosed within forward slashes, /.../, and orthographic transcriptions within quotation marks "..."

The second, and more important, question concerns whether there are global properties of the lexicon, or non-local properties of individual words that affect people's ability to identify particular words? In other words, are there network metrics that correlate with the ease with which individual words are perceived? It is well-known that the larger a word's neighborhood, the more difficult it is to recognize that word (Elman & McClelland, 1986; Luce & Pisoni, 1998; Marslen-Wilson, 1987, 1989; McClelland & Elman, 1986; Norris, 1994; Vitevitch & Luce, 1999). Luce and Pisoni (1998), for example, found that the more neighbors a word had, the more likely was that word to be misperceived in noise, the longer it took for people to discriminate that word from spoken non-words, and the longer it took for people to repeat the word after hearing a spoken version of it.

Luce and Pisoni (1998) developed the Neighborhood Activation Model (NAM) of spoken word recognition to explain these effects as well as other findings in the word recognition literature. In NAM, as in most current models of word recognition, words are assumed to exist in a large multi-dimensional acoustic space (Triesman, 1978a, b). There is no detailed description of how the sound patterns of words in this space are organized or structured. The NAM computes the lexical neighborhood of a word using the DAS rule. When the neighborhood is computed in this way, the resulting similarity space is defined only locally for an individual word based on a one-step distance metric without regard to other perceptually similar words in the lexicon. Thus, in the present architecture of NAM and all other word recognition models, words are not organized in any global structure and consequently there is no consideration of how the word's position in the overall lexicon or even of the structure of its local neighborhood can affect its identification.

Modeling the lexicon as a complex network using graph theory can potentially reveal global structural properties of the lexicon that may influence human word recognition. These properties will then need to be accounted for by current models of word recognition. Two properties were of special concern in the present study, a word's mean shortest distance (or mean shortest path length) to all other words, and a word's clustering coefficient. Word A's shortest distance to Word B is the smallest number of edges that must be traversed to reach Word B from Word A. Word A's mean shortest distance is the mean of that value across all words in the lexicon. Unlike the DAS rule, mean shortest distance takes into account the relative position of a word to all other words in the lexicon, not only to its immediate neighbors. The DAS rule is a coarse measure of phonological similarity that may miss important behavioral consequences. Intuitively, the word "bait" (/bet/) is more similar to the word "bare" (/ber/) than it is to the word "epileptic" (/ɛpələptɪk/), and the more "bare"s there are in the lexicon relative to "epileptic"s, the harder should it be to identify "bait." The DAS rule does not capture this intuition, but a rule based on mean shortest distance does. Similarly, the clustering coefficient provides a measure of the density of interconnections in a word's neighborhood. Typically, models of spoken word recognition posit the initial activation of multiple lexical candidates followed by a competition among those candidates. The competition ultimately leads to the identification of the word. Neural network models, for example, implement this competition with inhibitory connections between candidate words (cf., Elman & McClelland, 1986; McClelland & Elman, 1986). In such models, the overall density of connections in a word's neighborhood, and not just its neighborhood size, may have a significant effect on its recognition.

As already mentioned, we are not the first to model the mental lexicon as a graph, where the DAS rule is used to place edges between nodes that represent words. Here, we attempt to relate network properties of a word to the ease with which it can be recognized by human listeners.

## Method

### Constructing the Network

An Excel based version of the HML database was used as the basis for constructing the network. The neighborhood of each word was determined by finding all other words in the lexicon that could be converted to the target word by deleting one phoneme of the target word, adding one phoneme to it, or changing one phoneme to a second valid phoneme of American English. The software for determining neighbors can be publicly accessed at the Washington University (St. Louis, MO, USA) Speech and Hearing Lab website: <http://128.252.27.56/neighborhood/Home.asp>. The output of this software was converted into a list of pairs of words, where each word pair was a pair of neighbors and the pairs collectively were an exhaustive list of all neighbors in the lexicon. This list was used as input to Version 1.0 Pajek (Batagelj & Mrvar, 1998), a program designed for the analysis of large networks, and available for non-commercial use at <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>. All statistics on the network, unless explicitly stated otherwise, were calculated using Pajek.

The network was constructed as an undirected graph, rather than directed graph, since, given the definition of neighbor, if Word A is a neighbor of Word B, then Word B must also be a neighbor of Word A.

### Behavioral Data

Global properties of the network structure were correlated with two sets of behavioral data: word identification and repetition. The behavioral data were collected as part of an earlier study (Luce & Pisoni, 1998) and reanalyzed here. In the identification task, a digitized recording of a spoken, monosyllabic word was played to a listener at one of three Signal to Noise ratios (SNR): -5 dB SPL, +5 dB SPL, and +15dB SPL. The listener's task was simply to identify the spoken word by typing a response on the computer keyboard. The identification task was open-set. That is, the listener did not choose the correct alternative from a limited set of alternatives, but rather from the entire set of American English words. A total of 90 listeners participated in this experiment in partial fulfillment of an introductory psychology course requirement at Indiana University, Bloomington, IN, USA. Data were collected for a total of 908 monosyllabic words. Because of the large number of words in the study and the use of three SNR, in order to keep experimental sessions to a manageable length, each word was identified by a total of 10 listeners at each SNR. The dependent variable was the percent correct identifications of each word across listeners as a function of SNR.

In the word repetition study, a digitized recording of a monosyllabic word was played to the listener, in the clear (that is, without noise). The listener's task was simply to repeat back the word. Because listeners rarely make errors in this task, the dependent variable of interest was response latency, or the time for the listener to begin repeating the word. Latencies were measured from the offset of the spoken word to the beginning of the listener's utterance. Mean latencies to each word across listeners were analyzed in the present study. Eighteen volunteers from the Indiana University community served as listeners in this experiment. Latencies were collected for 939 words. All of the stimuli were monosyllabic words.

Additional details on the procedure used for collecting both the identification and the repetition data can be found in Luce and Pisoni (1998). The major finding of the earlier Luce and Pisoni study was that percent correct identification of a word decreased as the size of the word's neighborhood increased, while repetition latencies increased with neighborhood size. Luce and Pisoni interpreted these results as

support for the hypothesis that spoken words are recognized relationally in the context of other words in the lexicon.

## Results

### Network Analysis

The column labeled Whole Corpus in Table 1 shows some of the basic properties of the graph constructed from this lexicon. For comparison purposes, the results from Vitevitch (2004) are shown in Table 1 in the column labeled Vitevitch. In the present study, each word had a mean number of 3.18 neighbors. The clustering coefficient was 0.048 and the mean shortest distance from one word to any other word was 6.08. These properties are nearly identical to those reported by Vitevitch, a not surprising finding, given that we started with the same base lexicon as he did.

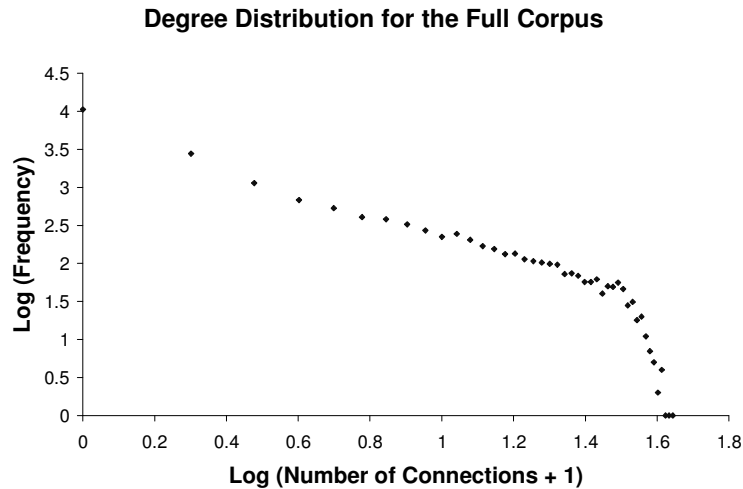
**Table 1.** Parameters of the networks based on the whole HML corpus and on monosyllabic words. The column labeled Vitevitch shows the data from Vitevitch (2004). The column labeled Random Networks is taken from Vitevitch (2004).

Parameter	Whole Corpus	Monosyllabic Corpus	Vitevitch	Random Networks
Number of Nodes (n)	19,587	4110	19,340	19,340
Number of Edges (l)	3.18	11.56	3.23	3.23
Mean Shortest Path Length (l)	6.08	4.66	6.05	8.44
Maximum Path Length (D)	29	13	29	19
Clustering Coefficient (CC)	0.048	0.10	.045	.000162
Degree Exponent ( $\gamma$ )	1.97	n/a	1.96	n/a

Vitevitch (2004) also constructed 10 random networks in which he fixed the number of nodes and the number of edges to be the same as in the network constructed from the lexicon. However, instead of using a rule to determine which pairs of nodes were to be connected by an edge, the two nodes connected by each edge were chosen randomly. The properties of these graphs are shown in Table 1 in the column labeled Random Network. The values in this column are the means across the 10 random networks. Notably, the mean clustering coefficient in the 10 random networks was significantly less than the clustering coefficient observed in the network based on the lexicon, and the mean of the mean shortest path length was significantly longer in the random networks. These two observations—that the mean shortest path length was less than that expected by chance and that the clustering coefficient was greater than that expected by chance—suggest that the lexical network created using the single DAS rule follows a small-world structure (Watts & Strogatz, 1998). Vitevitch drew the same conclusions from these observations.

Given the small-world structure of the lexicon, we next asked if it also follows a scale free structure. Recall that a network that has a scale-free structure is characterized by a degree distribution

that follows a power law, with the power law's parameter,  $\gamma$ , being in the range 2 – 3, and where a node's degree is the number of edges to which it connects. Figure 1 shows the degree distribution on log-log coordinates for the HML lexicon.<sup>3</sup>



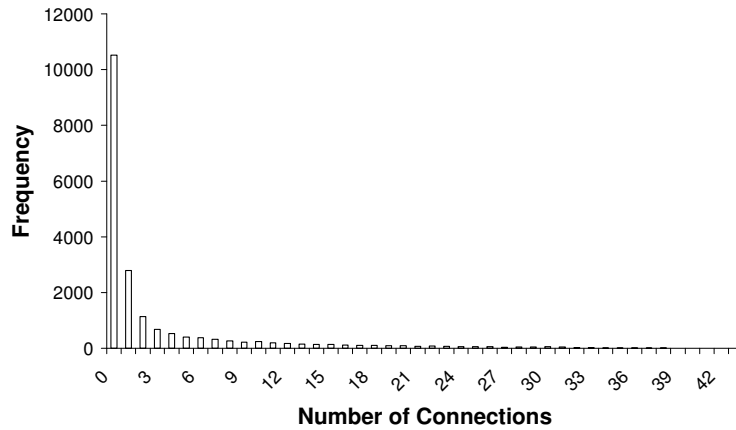
**Figure 1.** Log-log degree distribution (number of occurrences as a function of degree) for the 19,587-word corpus.

Ignoring momentarily the sharp drop in frequency at degrees higher than approximately 36, the degree distribution appears to be linear on a log-log plot. The Pearson product moment correlation between frequency and degree (including the degrees after the sharp drop off evident in the curve) is -0.85 ( $p < .01$ ), indicating that the degree distribution can be reasonably fit with a straight line. The best fitting straight line for that distribution has a slope of -1.97, a value not far out of the range of -2 to -3 characteristic of scale-free distributions (Albert & Barabási, 2002; Barabási & Albert, 1999). This slope is in very close agreement with the slope of -1.96 found by Vitevitch (2004).

Two further observations, however, suggest that the scale-free properties of the lexicon may be an illusion. First, using the DAS rule, most words in the HML lexicon have no neighbors. This fact is evident in Figure 2, which shows the degree distribution on linear coordinates. Note that the plot begins with degree = 0, not 1. Note also the large number of words with a degree of 0. In fact, over 10,500 words, representing roughly half the lexicon have no neighbors at all when the network was created using the DAS rule. Although it is true that had we built our network using a different definition of lexical neighbor, we might not have found such a large number of isolates, it is also true that the definition of neighbor that we did use has been found to be a powerful predictor of performance on a wide variety of experimental tasks involving spoken word recognition (e.g., Goldinger, Luce, & Pisoni, 1989; Luce & Pisoni, 1998; Vitevitch, 2002; Vitevitch & Luce, 1999) and hence seems to be appropriate at least as a starting point for building the graph.

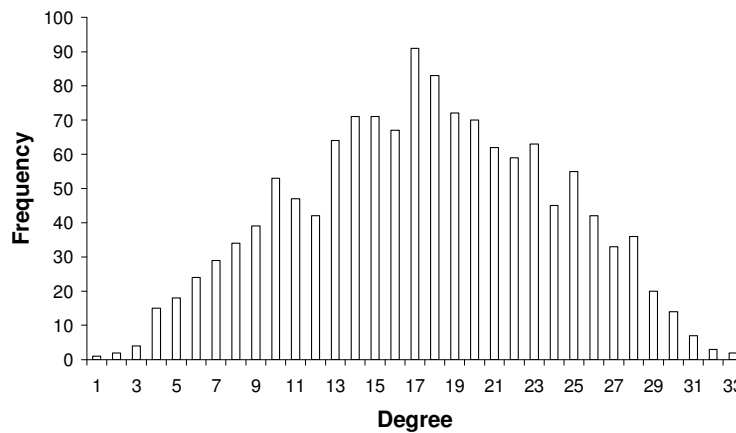
<sup>3</sup> The X axis in Figure 1 is actually the logarithm of the (degree plus 1). As will be discussed in more detail, many words have no neighbors, i.e., have a degree of 0. Since the logarithm of zero is undefined, the adjustment of adding 1 to the degree was made before plotting on log-log coordinates.

**Degree Distribution for the Full Corpus**



**Figure 2.** Degree distribution for the full corpus on linear-linear coordinates.

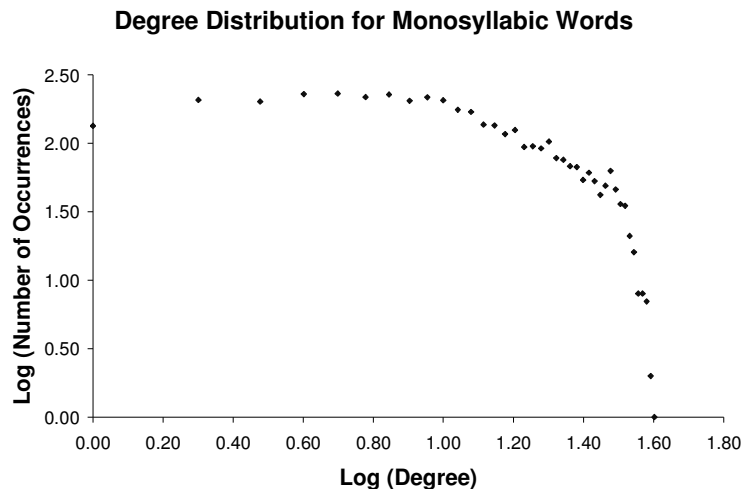
**CVC Degree Distribution**



**Figure 3.** Degree distribution for the CVC network on linear-linear coordinates.

The second and more crucial observation indicating that this network is not scale free involves examination of particular subsets of the data. Figure 3 shows the degree distribution, on linear coordinates, for a network built according to the same deletion-addition-substitution rule as the original network, but including only those 1338 words with a CVC structure, that is, words consisting of an initial consonant, followed by a single vowel, followed by a single final consonant. This distribution clearly does not follow a power law, but more closely resembles a Poisson distribution. This kind of distribution would be expected if links between nodes were placed at random. Figure 4 shows the degree distribution for the subset of the corpus that includes all monosyllabic words, that is words with a single vowel

(CVC, CCVC, CVCC, and so on). This distribution also clearly does not follow a power law. Inspection of Figures 2, 3, and 4 makes it clear that the middle and right hand portions of the overall degree distribution reflect primarily the contribution of monosyllabic words. These words tend to have more neighbors simply because they are shorter in length, not because of how the network grew. The left hand portion of the distribution is determined primarily by multi-syllabic words, which tend to have few neighbors simply because they are longer. Hence, any resemblance of the overall degree distribution of the lexicon to a power law merely reflects differences in the length of words rather than the acquisition and development processes used to create the network. To summarize, there is good evidence that the mental lexicon used for spoken word recognition has a small-world structure, but there is little evidence that it is scale-free.



**Figure 4.** Degree distribution for the monosyllabic network on log-log coordinates.

**Correlations with Behavioral Measures of Spoken Word Recognition.** Our primary concern in this project is not with establishing whether the mental lexicon follows a scale-free structure, but with determining whether there are fundamental properties of the lexicon, when modeled as a complex network graph, that correlate with the processing and perception of spoken words. Accordingly, we correlated several characteristics of a word’s neighborhood in the lexical network with the accuracy with which people could recognize that word in noise, and the speed with which people could repeat that word when it was spoken to them in isolation. It is now well-known that as a word’s “local” neighborhood density increases, where density is the number of immediate neighbors, as defined by the DAS rule, weighted by the frequency of occurrence of each of those neighbors, the accuracy of the identification of that word decreases (Luce & Pisoni, 1998; Vitevitch & Luce, 1998, 1999). Percent correct in a word identification task decreases and latency in a word repetition task increases as density increases (e.g., Luce & Pisoni, 1998), a finding that supports models of spoken word recognition that posit a stage of processing in which candidate words are first hypothesized to be activated by the acoustic-phonetic input, and then compete with one another for recognition through, for example, inhibitory connections in a neural network (cf. Goldinger et al., 1989; Vitevitch & Luce, 1998). In our network, these immediate neighbors are adjacent words, separated by a single link. What about words further removed from one another? Does the number of words separated by two links or three links in the lexical network from a target word also affect the recognition of the target word? More generally, does a word’s average

distance from all other words in the network affect its recognition? And, does the density of the interconnections within a word's neighborhood, as measured by the clustering coefficient, affect that word's recognition?

Our original corpus on which the network was built consisted of over 19,500 words, of which 10,521 had no neighbors. The two most salient characteristics of these 10,500 hermits was that they were longer than words with neighbors in terms of number of phonemes (7 – 8 phonemes per hermit compared to 4 – 5 for words with at least one neighbor) and they were multi-syllabic. Over 98% of the hermits in this network were multi-syllabic words. In contrast, of the 9066 words with at least one neighbor, only 5087, or 56% were multi-syllabic. In addition, the words for which we have behavioral data are all monosyllabic words. In fact, almost all work on spoken word recognition has been done with short, monosyllabic words, at least in part because recognition data for multi-syllabic words can be more easily contaminated by post-perceptual guessing strategies. For these reasons, for our analyses of spoken word recognition data, we rebuilt the lexical network using neighborhoods only for monosyllabic words. The column labeled "Monosyllabic" in Table 1 shows some characteristics of the new network. Of the 4110 monosyllabic words in our original corpus, only 131, or 3.2%, were lexical hermits. The correlations of behavioral data with network structure we report here are based on this new network built from monosyllabic words. The results, however, are qualitatively the same when the network based on the entire lexicon is used in these calculations.

**Word Identification Results.** Table 2 shows Pearson product moment correlations between a word's percent correct identification and various measures of its structure in the network, and with the logarithm of its frequency of occurrence in the language. Correlations are shown for each of the three SNR used. A single asterisk indicates a statistically significant correlation at the  $p < .01$  level. A double asterisk indicates a statistically significant correlation at the  $p < .001$  level. In the table,  $D_m$  refers to the number of words whose shortest path in the network from the target word is exactly  $m$  links. The value  $D_8$  was included as a control condition. Words 8 links removed from a target would not be expected to affect its recognition and hence correlations with  $D_8$  would be expected to be near 0. PL refers to path length. Mean  $PL_{m-n}$  refers to a word's mean shortest path distance from all words whose shortest distance is from exactly  $m$  through and including exactly  $n$  links. Hence, mean  $PL_{1-13}$ , because the maximum shortest distance between any two words in the network was 13 links, is the mean distance of the word from all other words in the network (excluding the 3.2% of the words with no immediate neighbors). Mean  $PL_{1-3}$  is the mean shortest distance from the target word to all other words exactly 1, 2 or 3 edges distant from it. This value goes up as the proportion of words 2 or 3 links, as opposed to 1 link, from the target word increases. The measure  $PL_{1-13}$  was included in order to assess a word's mean distance from every other word on its recognition. The measure  $PL_{2-13}$  was likewise included in order to assess the effects of the word's mean distance from all other words, but without also including effects of the number of its immediate neighbors, an already well-studied variable. The measures  $PL_{1-3}$  and  $PL_{2-3}$  were included in order to assess possible effects of a word's more immediate neighborhood without diluting those effects by simultaneously including effects of far away communities. Finally, CC is the clustering coefficient, a measure of the probability of two word's being neighbors when each is a neighbor of some other third word. The mean proportions correct for the three SNR of +15, +5 and -5 dB were .77, .57, and .17, respectively, suggesting that any correlations of the predictor variables with the percent correct data would not be artificially lowered due to floor or ceiling effects.

**Table 2.** Correlations of percent correct identification with log frequency of occurrence and with network measures of distance, path length, and clustering for monosyllabic words for three different signal to noise ratios. See the text for an explanation of abbreviations used. \*:  $p < .01$ ; \*\*:  $p < .001$ .

	Log Freq	D1	D2	D3	D8	CC	PL1-13	PL1-3	PL2-3	PL2-13
SNR +15	0.18*	-0.03	-0.04	-0.04	0.02	-0.03	0.03	0.00	0.01	0.03
SNR +5	0.23*	-0.11**	-0.13**	-0.14**	0.05	-0.08*	0.12**	0.08*	0.10*	0.12**
SNR -5	0.20*	-0.04	-0.06	-0.05	0.03	-0.03	0.05	0.05	0.06	0.05

Replicating previous findings, the correlation between a word's frequency of occurrence in the language, as measured by Log Freq, and the accuracy with which it is identified, while low, were statistically significant, at least for the SNR +5 and SNR -5 conditions. The correlations between identification accuracy and network measures, on the other hand, were uniformly low and statistically non-significant. The correlations between the accuracy with which a word is identified and the number of neighbors it has at short distances (D1, D2, D3) were negative in sign and hence in the expected direction—as the number of neighbors increases, identification accuracy decreases. However, the correlations were statistically indistinguishable from 0.

**Table 3.** Correlations of percent correct identification with network measures of distance, clustering, and path length for monosyllabic words with a low, medium, and high frequency of occurrence. \*:  $p < .01$ ; \*\*:  $p < .001$ .

	Log Freq	D1	D2	D3	D8	CC	PL1-13	PL1-3	PL2-3	PL2-13
Frequency < 20 (N = 512)										
SNR 15	0.15*	-0.02	-0.03	-0.05	0.02	0.01	0.03	0.04	-0.02	0.03
SNR 5	0.17*	-0.10	-0.12*	-0.12*	0.03	-0.02	0.09	0.05	0.04	0.09
SNR -5	0.15*	-0.04	-0.08	-0.08	0.04	-0.02	0.07	0.01	0.07	0.07
20 <= Frequency <= 99 (N = 228)										
SNR 15	.18*	-.10	-.11	-.10	.09	-.12	.10	.09	.10	.10
SNR 5	.08	-.17*	-.20*	-.21**	.14	-.18*	.20*	.16*	.17*	.20*
SNR -5	.17*	-.11	-.10	-.09	.06	-.14	.09	.17*	.16*	.09
Frequency >= 100 (N=168)										
SNR 15	.04	-.01	.00	.00	-.07	-.08	-.02	.01	.00	-.02
SNR 5	.02	-.13	-.16	-.15	-.05	-.13	.13	.07	.10	.13
SNR -5	-.04	.00	.00	.02	-.03	.04	-.02	-.01	-.01	-.02

Because a word's frequency of occurrence in the language can modulate the effects of other variables on its recognition, we re-analyzed the identification data after dividing the words into low frequency (fewer than 20 occurrences per million), medium frequency (from 21 to 99 occurrences per million), and high frequency (100 or more occurrences per million) words. These correlations are shown

in Table 3. As is the case for the overall analysis, the correlations with network parameters were low and statistically non-significant, with the exception of medium frequency words (and to a lesser extent, high frequency words) presented at a SNR of +5 dB. For these words, identification was less accurate the more neighbors the word had that were separated by 1, 2, or 3 phonemes.

Table 3 also shows that for medium frequency words presented at a Signal-to-Noise ratio of +5 dB SPL, the various  $PL_n - m$  measures also correlate significantly with identification accuracy. As a word's path length, that is, its mean distance to other words in the lexicon, decreases, so does the ability of listeners to correctly identify it. The importance of all these correlations, however, needs to be considered in light of the how the various measures of network structure inter-correlate with one another, in particular with how the  $D = 1$  measure correlated with the  $D = 2$ ,  $D = 3$ , and the  $PL_n - m$  measures. Table 4 shows these inter-correlations for the entire monosyllabic corpus. Note first that the metric  $D = 1$  is the same variable that previous investigators have referred to as neighborhood size or neighborhood density, a variable already known to affect percent correct identification of words spoken in noise (and word repetition latencies). As is evident from Table 4, this measure was strongly correlated with the  $D = 2$ ,  $D = 3$ , and  $PL_n - m$  measures (all  $R^2$ 's  $> .65$ ,  $p < .001$ ) and when the effects of density were partialled out from these other measures, they no longer significantly correlated with percent correct identification. This finding suggests that our observation that a word's mean distance to other words in the lexicon correlated with listeners' ability to identify it can be entirely accounted for by the word's local neighborhood size.

**Table 4.** Inter-metric correlations among distance, clustering, and path length measurements for the monosyllabic corpus.

	Log Freq	D=1	D=2	D=3	D=8	CC	PL1-13	PL1-3	PL2-3	PL2-13
Log Freq	1.00	0.05	0.05	0.04	0.00	0.03	-0.93	-0.02	-0.03	-0.03
D=1		1.00	0.94	0.88	-0.48	-.78	-0.84	-0.81	-0.86	-0.83
D=2			1.00	0.96	-0.53	0.77	-0.91	-0.73	-0.84	-0.91
D=3				1.00	-0.63	0.73	-0.97	-0.56	-0.70	-0.97
D=8					1.00	-0.46	0.78	0.12	0.32	0.78
CC						1.00	-0.73	-0.68	-0.74	-0.73
PL1-13							1.00	0.50	0.66	1.00
PL1-3								1.00	0.96	0.49
PL2-3									1.00	0.65
PL2-13										1.00

Finally, again for medium frequency words presented with an SNR of +5 dB SPL, as shown in Table 3, there was a statistically significant albeit small negative correlation between a word's clustering coefficient and its percent correct identification. As a word's clustering coefficient increased, its percent correct identification decreased. Further consideration of the possible effects of the clustering coefficient is postponed to the description of the results for word repetition latencies, where the effects of the clustering coefficient are more robust.

To summarize, in the present data set, there is little evidence that non-local or global properties of the lexicon improve the ability to predict word identification accuracy above and beyond the ability of word frequency and the single DAS rule. There are several possible reasons why we failed to observe

such an effect. First, the DAS rule is a rather crude measure of the perceptual similarity of two words, as evidenced by the fact that over half the words in our original corpus are, by the DAS rule, hermits—i.e., they have no neighbors. Second, given the method we used to construct the network, two words that differ by two phonemes are separated by a distance of 2 if, and only if, there is a third word that is separated by a single phoneme from each of those two words. If no such word exists, then the two words are separated by a distance of greater than 2 in the network. There seems no *a priori* reason to believe that the two words in the former case are more similar than the two words in the latter case. Hence, a similarity measure more refined than the DAS rule may make the effects of global properties of the mental lexicon on word identification more evident. We are currently exploring this possibility in more detail.

**Word Repetition Latencies.** Pearson product moment correlations of a word's repetition latency and various network properties are shown in Table 5. Latencies were measured from the offset of the spoken word to the onset of the repetition. Hence, they reflect both the time to perceive the word and the time to initiate the motor program for pronouncing the word. As is the case for the identification data, in the table,  $D_m$  refers to the number of words whose shortest path in the network from the target word is exactly  $m$  links. PL refers to path length. Mean  $PL_{m-n}$  refers to a word's mean shortest path distance from all words whose shortest distance is from exactly  $m$  through and including exactly  $n$  links. Finally, CC is the clustering coefficient, a measure of the probability of two words being neighbors when each is a neighbor of some other third word.

**Table 5.** Correlations of word repetition latencies with measures of network distance, path length, and clustering as a function of a word's frequency of occurrence. \*:  $p < .01$ ; \*\*:  $p < .001$ .

Log Freq	D1	PL 1-13	PL 1-3	PL 2-3	PL 2-13	CC
Entire Corpus (N = 939)						
-.07	.26**	-.28**	-.26**	-.27**	-.28**	.20**
Frequency < 20 (N = 467)						
.01	.20**	-.21**	-.20**	-.22**	-.21**	.15**
20 <= Frequency <= 100 (N = 275)						
-.03	.27**	-.30**	-.26**	-.27**	-.30**	.21**
Frequency > 100 (N = 197)						
-.07	.48**	-.48**	-.44**	-.44**	-.48**	.40**

Consistent with the earlier findings reported by Luce & Pisoni (1998), repetition latencies did not correlate with frequency of occurrence in the language. However, they did correlate with several of the network parameters. First, a statistically significant positive correlation was observed between the number of immediate neighbors a word has (D1) and its naming latency. The more local neighbors a word has, the longer was the naming latency. At first glance, the data also suggest that words at distances beyond immediate neighbors also influenced word repetition latencies. Significant negative correlations ( $p < .01$ ) were also observed for naming latency with the mean shortest path to for all words at a distance of 2 to 3 edges (PL 2-3), and for naming latency with average shortest distance to all words in the network (PL 1-13), even when words at a distance of 1 (i.e., immediate neighbors) were excluded (PL 2-13). That is, as mean shortest path lengths increased, the time to repeat a word decreased. These correlations need to be interpreted with some caution, however. Given the methods used to construct the original graph, the number of neighbors a word has at a distance of 2 (and at distance 3, and so forth) must correlate with the number of neighbors it has at distance 1 (because to reach a word in two edges, some other word needs to be reached in one edge), meaning that all  $PL_{n-m}$  measures would also tend to

(negatively) correlate with D1. In fact, as can be seen in Table 4, PL2-3, PL1-13, and PL2-13 all did correlate strongly with D1 (all  $R^2$ 's  $> .75$ ,  $p < .001$ ). When the effects of D1 are partialled out of PL2-3, PL1-13, and PL2-13, the correlations of these variables with repetition latency all become non-significant. Hence, we can conclude that extending distance measures beyond a word's immediate neighbors, as defined by the single phoneme deletion/addition/substitution rule, does little to improve the prediction of repetition latencies.

The results obtained from the analyses of the clustering coefficient (CC) are more revealing. CC correlated significantly with repetition latency,  $r = +0.20$ ,  $p < .02$ . As the CC became larger, naming latencies increased in duration. The effect tended to be stronger as frequency of occurrence in the language increased. The correlation of CC with repetition latency was  $+0.15$  (n.s.),  $+0.21$  ( $p < .02$ ), and  $+0.40$  ( $p < .001$ ) for low, medium, and high frequency words, as defined above.

We also found that CC correlated strongly with D1, the number of immediate neighbors ( $r = +0.76$ ,  $p < .001$ ), leaving open the possibility that the observed correlation between CC and repetition latency was in fact due to the correlation of D1 with repetition latency, or conversely, that the observed correlation between D1 and repetition latency was in fact due to the correlation of CC with repetition latency. To determine whether there are independent effects of CC and D1 on naming latencies, a median-split analysis was performed on the data. Each word for which naming latency data were available was assigned to one of four cells in a 2 X 2 ANOVA, corresponding to whether the word was above or below the median value for CC and above or below the mean value for D1. LowCC/LowD1 words ( $n = 351$ ) were below the median value for both variables, lowCC/HighD1 ( $n = 121$ ) words were below the median CC value and above the median D1 value, HighCC/LowD1 words ( $n = 93$ ) were above the median CC value and below the median D1 value, and HighCC/HighD1 words ( $n = 374$ ) were above the median value for both variables. The means for these four cells are shown in Table 6. An analysis of variance found no significant interaction between D1 and CC ( $F < 1$ ). The main effects, however, of both D1 and CC, though numerically small, were both highly significant, (for D1,  $F(1, 937) = 51.84$ ,  $p < .001$ ; for CC,  $F(1, 937) = 30.80$ ,  $p < .001$ ). The results of this analysis suggest the presence of an effect of clustering coefficient on the latency to repeat a word that is independent of the effect of the word's number of neighbors on the time to repeat that word. Note that the CC is a non-local, global property of the word that appears to affect response times in at least one spoken word recognition task.

**Table 6.** Mean word repetition latencies (ms) as a function of neighborhood density and clustering coefficient.

Density	Clustering Coefficient	
	Low	High
Low	294	304
High	316	326

The present analysis makes two contributions to our understanding of the structure of the mental lexicon and how it affects spoken word recognition. First, replicating Vitevitch's (2004) earlier findings, we found that the mental lexicon, when modeled as a complex graph, displays a scale-free structure. However, rather than reflecting growth through preferential attachment, we suggest that this structure reflects the constraints that all words are constructed from a limited number of phonemes and that words differ in length. Second, our ability to recognize an isolated word is affected not only by the number of

neighbors that word has, but also by how interconnected those neighbors are with one another, as measured by the clustering coefficient of a word. We discuss each of these points in turn below.

### **The Mental Lexicon as a Scale-Free Structure**

As we mentioned in the introduction, Vitevitch (2004) has also recently modeled the mental lexicon as a complex network. Our results for the various network measures we examined closely follow his, a finding that is not surprising given that we both began with the same HML lexical database (Nussbaum et al., 1984), and we both used the same single phoneme DAS rule. Vitevitch concluded that the mental lexicon has a small-world, scale-free structure. He further suggested that his findings indicate that the mental lexicon grows via a process of preferential attachment (Albert & Barabási, 2002; Barabási & Albert, 1999). That is, as a child learns new words in his or her language, he/she adds words that are acoustically similar to those already learned. In fact, Storkel (2004), as noted by Vitevitch (2004), has recently reported that words learned early by a child are words that have many neighbors in the adult mental lexicon, consistent with the notion that the child's lexicon grows through a process like preferential attachment.

Our results suggest an alternative explanation of why a power-law like degree distribution occurs for the lexicon. Words are constructed from a small number of basic sounds, i.e., phonemes or particles (cf. Abler, 1989). Furthermore, words can have different lengths; some have more phonemes than others. As a consequence, and given the DAS rule used to construct the lexical network, short words will have more neighbors than longer words. The overall result will be a degree distribution that looks quite similar to a power law distribution, but which arises primarily from random processes, not preferential attachment.

The results from our analysis of monosyllabic words support this explanation. The degree distribution for these words did not approach anything resembling a power law. Likewise, the degree distribution for multi-syllabic words also does not resemble a power law. The power law degree distribution is the result of averaging the degree distribution for monosyllabic words, which contributes the “middle” and “right hand (hub)” sides of the distribution with that for multi-syllabic words, which contributes the “left hand” side of the overall degree distribution.

The present study is not the first investigation to find power law frequency distributions arising from averaging non-power law distributions. Improvements with practice frequently follow a power law, and a great deal of research has been directed at determining why this phenomenon occurs (e.g., Newell & Rosenbloom, 1981). Anderson (2001) and Brown and Heathcote (2003a, 2003b; see also Newell & Rosenbloom, 1981) have shown that a power law distribution frequently results when a number of underlying distributions, none of which itself is a power law distribution, are averaged together. For example, power law learning curves can result when individual learning curves for a number of experimental subjects are averaged together, where each individual's curve is an exponential, but each with a different parameter. In Vitevitch's (2004) case, and in our analyses, the degree distribution of monosyllabic words was effectively averaged with the degree distribution of multi-syllabic words (or more precisely, degree distributions for words of different lengths were averaged together), with a similar result. The extent to which this effect underlies other observations of power law degree distributions in other analyses of complex systems remains to be seen.

Nevertheless, it does seem to be the case that words that children learn early do have more neighbors in the adult lexicon (Storkel, 2004). This finding suggests that a process akin to preferential attachment operating at the level of words' acoustic patterns is operating during the course of language

development. At the same time, however, semantics also plays an important role in shaping the organization of the lexicon. A child will learn those words that are most important to meeting its perceived needs, independent of the underlying sounds that comprise those words. That is, the meaning of the word and its relevance to the child is at least as likely to influence which words are added to the child's lexicon as the acoustic-phonetic similarity of the word to other already learned words in the lexicon.

In summary, a complex network constructed from the lexicon using the DAS rule does display a scale-free structure. However, we suggest that the scale-free structure is as likely to reflect the averaging of degree distributions across words of different lengths as it is to reflect fundamental underlying language acquisition processes using preferential attachment.

### **Effects of the Clustering Coefficient on Spoken Word Recognition**

The second purpose of the current study was to determine if global measures of network structure would provide novel insights into spoken word recognition processes previously overlooked by more traditional analyses. To this end, we examined the correlations of various network measures, including measurement of distance and clustering, on word repetition latencies. It is already known that as the number of words at a distance of 1 from a given word increases, so does its repetition latency and the accuracy of identifying that word in noise (Luce & Pisoni, 1998), and the current study replicated these findings. However, we also found that taking into account the mean distance of a word from all other words in the lexicon, or even its mean distance to relatively nearby words, adds little additional predictive power to this original measure. Hence, beyond a distance of 1, distance measurements based on the DAS rule, showed little relation to repetition latencies. This finding suggests that similarity effects drop off sharply as similarity decreases. Alternatively, the present findings suggest that network distance, where the network is built using the single DAS rule, is not a very robust measure of phonological similarity. A *prima facie* case could be made for the second of these two alternatives. Consider two words, A and B, which differ by two phonemes. If a third word, C, exists, such that A can be converted to C by the single DAS rule and C to B by the single DAS rule, then A and B would be separated by a network distance of 2 in our model. If no such word C exists, then the network distance between A and B would be greater than 2, a seemingly somewhat artificial situation. Despite the superficial reasonableness of the second of the two above alternatives, we are currently investigating other objective measures of perceptual similarity in an effort to select the best explanation for the process of spoken word recognition.

The findings obtained with the clustering coefficient were different, however. We found that the higher a word's clustering coefficient, the longer its repetition latency. In other words, repetition latencies were longer for words whose neighbors were also neighbors of one another. This correlation appears particularly strong for higher frequency words. The clustering coefficient thus appears to be an example of a "global," non-local, emergent property that affects the recognition of a particular word.

This result is of course *post hoc*, and like any *post hoc* finding, this finding needs to be independently verified in a replication. In addition, the word repetition task involves both perceptual and production processes. Participants must first correctly perceive the word and then execute a motor program for pronouncing the word out loud. The current study does not address the issue of whether the effects of the clustering coefficient are on perceptual processes, production processes, or both. In collaboration with Nick Altieri, we are currently undertaking a series of new studies in our laboratory designed to address two questions. First, is the effect of the clustering coefficient on repetition naming latency reliable? Second, assuming the effect is reliable, to what extent does a word's clustering

coefficient affect perceptual processes and to what extent does it affect production processes? Although these experiments are still in an early stage, preliminary results are encouraging in terms of confirming the reliability of the effect. The results also indicate that at least part of the effect is on perceptual processes used at the time of encoding.

What are the implications of the effects of clustering coefficient on repetition latency for models of spoken word recognition and speech production? To a large extent, until the locus and replicability of the effect are better delineated, any detailed answer to the question is a little premature. However, at least one general comment can be made at this time. As noted several years ago by Goldinger, Luce, Pisoni, and Marcario (1992), most contemporary models of spoken word recognition (Elman & McClelland, 1986; Luce & Pisoni, 1998; Marslen-Wilson, 1987, 1990; McClelland & Elman, 1986; Norris, 1994) are “activation-plus-competition” models. In such models, the acoustic input activates a number of possible lexical candidates that are each consistent with that input. Each candidate then competes with the other lexical candidates for recognition. In connectionist terms, the competition is often modeled as mutual inhibition of acoustically similar words (Elman & McClelland, 1986; McClelland & Elman, 1986). Models with such an inhibitory mechanism might expect a higher clustering coefficient to actually facilitate word recognition, a result opposite in direction to the findings we have observed. Facilitation would occur because a target’s neighbors, when the clustering coefficient is higher, would tend to inhibit one another, reducing the overall activation of the target’s immediate neighbors, thus reducing the inhibition they exert on the target itself.

The negative effects of a high clustering coefficient on spoken word recognition, however, might be more easily understood if viewed from a slightly different perspective. Saying that a word has a high CC is another way of saying that a word’s neighbors are not only similar to the word itself, but that they are also similar to one another. In contrast, the neighbors of a word with a low CC are similar to that word but not to one another. This observation in turn implies that a high CC word is going to share any given phoneme sequence with many other lexical neighbors; if it did not, those neighbors would not be neighbors of one another. For a low CC word, on the other hand, there will be at least some phoneme sequences that are not shared with many neighbors. For instance, the high CC word “boot” (/but/) shares the phoneme sequence /ut/ with root, loot, soot, coot. It shares the “sequence” /b\_t/ with bit, bat, but, bout. Hence, the acoustical evidence for any given sequence is not likely to be very good at discriminating amongst the various alternatives, making it harder to eliminate alternatives. In the case of low CC words, however, having available partial phonological information about a possible word will keep in play fewer neighbors of the target word, since different neighbors of a target word are neighbors for different reasons. In other words, any partial information is more discriminating in the case of low CC words, making overall identification easier.

In addition to these general theoretical issues, we are also interested in using network concepts to explore the lexical organization and processing of spoken words in profoundly hearing-impaired children with cochlear implants (CI). Kirk, Pisoni, and Osberger (1995) found that in an open-set spoken word recognition identification task, deaf children with cochlear implants recognized high frequency words from sparse neighborhoods more accurately than low frequency words from dense neighborhoods. This pattern of results follows that found with normal hearing adult listeners. Based on such results, Kirk et al. suggested that children with CIs may organize their lexicon in a manner similar to normal hearing adults. However, the speech signal received processed by a CI is a highly degraded signal. This degradation could result in the blurring of some phonetic distinctions. The end-result of such blurring may be that the lexical network of CI listeners is much more densely inter-connected than the lexicon of normal listeners. This higher degree of connectivity could contribute to at least some of the difficulty CI listeners experience with spoken word recognition. Higher connectivity, for example, would tend to lead to higher

average clustering coefficients, which in turn, as suggested by the results reported here, may lead to increased difficulty with spoken word recognition in open-set tasks.

To summarize, the present study was designed to investigate whether modeling the lexicon as a complex network, using the tools of graph theory, could provide some new insights into the processes underlying spoken word recognition. In particular, we were interested in whether spoken word recognition is affected by structural properties that go beyond a word's local, immediate lexical neighborhood and reflect more global properties of the mental lexicon, or connectivity patterns of words in the mental lexicon. On the one hand, we found evidence that the size of a word's neighborhood, as originally defined by Landauer and Streeter (1973) (see also Greenberg & Jenkins, 1964, and Luce & Pisoni, 1998), is a robust predictor of the similarity effects observed in open-set word recognition and word repetition tasks. These findings suggest that a complex systems approach may add little to our earlier understanding of similarity effects in spoken word recognition. This conclusion does need to be qualified by two additional observations. First, we analyzed behavioral data only for monosyllabic words, which included approximately 20% of the original corpus. Hence, our observations may not generalize across the entire mental lexicon. Second, we constructed our network using the DAS rule which admittedly is at best a rather crude measure of the perceptual similarity of two words. On the other hand, even under these constraints, we identified a new global variable, the clustering coefficient, that does appear to affect spoken word recognition nearly as robustly as does local neighborhood size. This second finding, if verified and replicated, suggests that non-local, global properties of a word's position in the mental lexicon can have important consequences for how listeners recognize that word in isolation and in context.

## References

- Abler, W.L. (1989). On the particulate principle of self-diversifying systems. *Journal of Social & Biological Structures*, 12, 1–13.
- Achacosa, T.B. & Yamamoto, W.S. (1992). *AY's Neuroanatomy of C. elegans for Computation*. CRC Press: Boca Raton, FL.
- Albert, R. & Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74, 47–97.
- Albert, R., Jeong, H., & Barabási, A.-L. (1999). Diameter of the world-wide web. *Nature*, 401, 130 - 131.
- Anderson, R.B. (2001). The power law as an emergent property. *Memory & Cognition*, 29, 1061-1068.
- Batagelj, V. & Mrvar, A. (1998). Pajek—A program for large network analysis. *Connections*, 21, 47–57.
- Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286, 509–512.
- Barabási, A.-L., Albert, R., & Jeong, H. (1999). Mean-field theory for scale-free random networks. *Physica A*, 272, 173–187.
- Brown, S. & Heathcote, A. (2003a). Bias in exponential and power function fits due to noise: Comment on Myung, Kim and Pitt. *Memory and Cognition*, 31, 656-661.
- Brown, S. & Heathcote, A. (2003b). Averaging learning curves across and within participants. *Behaviour Research Methods, Instruments & Computers*, 35, 11-21.
- Dorogovtsev, S. N. & Mendes, J.F.F. (2001). Language as an evolving word web. *Proceedings of the Royal Society of London B*, 268, 2603 – 2606.
- Elman, J.L. & McClelland, J.L. (1986). Exploiting lawful in variability in the speech waveform. In J.S. Perkell & D.H. Klatt (Eds.), *Invariance and variability in speech processing*, Hillsdale, NJ: Erlbaum, 360–385.
- Faloutsos, M., Faloutsos, P, & Faloutsos, C. (1999). On power-law relationships of the Internet topology. *Computing and Communications Review*, 29, 251–262.

- Ferrer, R. & Solé, R.V. (2001). The small world of human language. *Proceedings of the Royal Society of London B*, 268, 2261–2265.
- 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.
- Goldinger, S.D., Luce, P.A., Pisoni, D.B., & Marcario, J.K. (1992). Form-based priming in spoken word recognition: The roles of competition and bias. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 18, 1211–1238.
- Greenberg, J.H. & Jenkins, J.J. (1964). Studies in the psychological correlates of the sound system of American English. *Word*, 20, 157–177.
- Huberman, B.A. & Adamic, L.A. (1999). Growth dynamics of the World-Wide Web. *Nature*, 401, 131.
- Huberman, B.A., Pirollo, P.L.T., Pitkow, J.E. & Lukose, R.M. (1998). Strong regularities in World-Wide Web surfing. *Science*, 280, 95–97.
- Jeong, H., Tombor, B., Albert, R., Oltavi, Z.N. & Barabási, A.-L. (2000). The large-scale organization of metabolic networks. *Nature*, 407, 651–654.
- Kirk, K.I., Pisoni, D.B., & Osberger, M.J. (1995). Lexical effects on spoken word recognition by pediatric cochlear implant users. *Ear & Hearing*, 16, 470 – 481.
- 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 theories of word recognition. *Journal of Verbal Learning and Verbal Behavior*, 12, 119–131.
- Lawrence, S. & Giles, C.L. (1998). Searching the World-Wide Web. *Science*, 280, 98–100.
- Lawrence, S. & Giles, C.L. (1999). Accessibility of information on the web. *Nature*, 400, 107–109.
- Liljeros, F., Edling, C.R., Amaral, L.A.N., Stanley, H.E., & Aberg, Y. (2001). *Nature*, 411, 907.
- Luce, P.A. & Pisoni, D.B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19, 1–36.
- Marslen-Wilson, W.D. (1987). Parallel processing in spoken word recognition. *Cognition*, 25, 71–102.
- Marslen-Wilson, W.D. (1989). Access and integration: Projecting sound onto meaning. In W.D. Marslen-Wilson (Ed.), *Lexical access and representation*. Cambridge, MA: Bradford, 3–24.
- McClelland, J.L. & Elman, J.L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86.
- Milgram, S. The small world problem. *Psychology Today*, 2, 60–67.
- Newell, A. & Rosenbloom, P.S. (1981). Mechanisms of skill acquisition and the law of practice. In J.R. Anderson (Ed.), *Cognitive skills and their acquisition*. Hillsdale, NJ: Lawrence Erlbaum.
- Newman, M.E.J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*. 98, 404–409.
- Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52, 189–234.
- 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.
- Storkel, H.L. (2004). Do children acquire neighborhoods? An investigation of similarity neighborhoods in lexical acquisition. *Applied Psycholinguistics*, 25, 201–221.
- Triesman, M. (1978a). A theory of the identification of complex stimuli with an application to word recognition. *Psychological Review*, 85, 525–570.
- 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.
- Vitevitch, M.S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 28, 735–747.

- Vitevitch, M.S. (2004). Phonological neighbors in a small world: What can graph theory tell us about word learning? Unpublished manuscript, Department of Psychology, University of Kansas, Lawrence, KS.
- 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.
- Watts, D.J. & Strogatz, S.H. (1998). Collective dynamics of ‘small-world’ networks. *Nature*, 393, 440-442.
- Webster’s Seventh Collegiate Dictionary. (1967). Los Angeles: Library Reproduction Service.
- Wuchty, S. (2001). Scale-free behavior in protein domain networks. *Molecular & Biological Evolution*, 18, 1694-1702.

