

Effects of word frequency, contextual diversity, and semantic distinctiveness on spoken word recognition

Brendan T. Johns, Thomas M. Gruenenfelder,^{a)} David B. Pisoni, and Michael N. Jones

Department of Psychological and Brain Sciences, Indiana University, 1101 East 10th Street, Bloomington, Indiana 47405

johns4@indiana.edu, tgruener@indiana.edu, pisoni@indiana.edu, jonesmn@indiana.edu

Abstract: The relative abilities of word frequency, contextual diversity, and semantic distinctiveness to predict accuracy of spoken word recognition in noise were compared using two data sets. Word frequency is the number of times a word appears in a corpus of text. Contextual diversity is the number of different documents in which the word appears in that corpus. Semantic distinctiveness takes into account the number of different semantic contexts in which the word appears. Semantic distinctiveness and contextual diversity were both able to explain variance above and beyond that explained by word frequency, which by itself explained little unique variance.

© 2012 Acoustical Society of America

PACS numbers: 43.71.Gv, 43.71.An [SGS]

Date Received: May 2, 2012 **Date Accepted:** June 11, 2012

1. Introduction

One of the oldest and most robust findings in the spoken word recognition literature is the word frequency effect. Words that occur more frequently in the language are recognized in noise more accurately than words that occur less frequently in the language (see, e.g., [Howes, 1957](#); [Pollack *et al.*, 1959](#); [Savin, 1963](#)). The word frequency effect is extremely robust and ubiquitous. When comparing recognition accuracy for different groups of words, researchers are careful to equate the different word groups on word frequency. All viable theories of spoken word recognition have one mechanism or another to account for the word frequency effect ([Dahan *et al.*, 2001](#); [Goldinger, 1998](#); [Morton, 1969, 1979](#); [Norris, 1994, 2006](#)).

A robust word frequency effect also occurs in the visual word recognition literature: Words presented visually are processed more quickly the more frequently that they occur in the language (see, e.g., [Balota *et al.*, 2004](#); [Broadbent, 1967](#); [Forster and Chambers, 1973](#); [Krueger, 1975](#)). Recently, however, [Adelman *et al.* \(2006\)](#) reported that, for printed words, contextual diversity (the number of different documents in which a word occurs) accounted for more variance in word-naming (reading a word aloud) and lexical decision tasks (deciding if a string of letters is a word) than did word frequency. Word frequency, in fact, explained little unique variance beyond that accounted for by contextual diversity ([Adelman *et al.*, 2006](#)).

In a more recent study, [Jones *et al.* \(2012\)](#) further examined the effects of word frequency and contextual diversity on naming and lexical decision times using visually presented words. They also assessed the effects of a new variable, semantic distinctiveness. The semantic distinctiveness of a word takes into account how semantically dissimilar from one another the documents in which the word appears are. The more dissimilar those documents, the higher the word's semantic distinctiveness ([Jones *et al.*, 2012](#)).

^{a)}Author to whom correspondence should be addressed.

The original definition of semantic distinctiveness developed by Jones *et al.* (2012) involved first computing the semantic dissimilarity of all pairs of documents in a corpus and then, for a given word, examining the distribution of dissimilarity values across all documents pairs in which that word occurred. A less computationally intensive approach, also described in Jones *et al.*, that is intended to approximate the original definition is summarized and used here. The computation begins with a word-by-document matrix that simply records the documents in which a word occurs. A word's meaning is represented by the vector in that matrix corresponding to that word. (This approach is quite common in computational studies of lexical semantic memory; see, e.g., Landauer and Dumais, 1997.) When a new document is encountered, a new column is added to the matrix, corresponding to that document. A word's value in that new document vector is determined by its current context as follows. First, the current context is computed as

$$\text{Context} = \sum_{i=1}^n T_i, \quad (1)$$

where n is the number of words in the new document and T_i is the memory vector of a particular word in the document. The strength with which the word is then encoded into the new column is determined by the similarity of the current context to the earlier contexts in which the word occurred—the higher the similarity, the less strongly the word is encoded. Similarity is computed using the vector cosine between the word's existing memory vector and the current context, as computed previously. The cosine is passed through an exponential transformation (see Shepard, 1987) such that high similarity of context is transformed into low distinctiveness, and low similarity of context is transformed into high distinctiveness. The magnitude of the transformation is controlled by the λ parameter, set to 5.5 in this study, based on the optimal value found by Jones *et al.* (2012). This transformed value is the semantic distinctiveness, SD,

$$\text{SD} = e^{-\lambda \cos(\text{context}, \text{word}_i)}, \quad \lambda > 0. \quad (2)$$

This value of SD is then encoded into this word's row in the new column in the word by document memory matrix. A word's overall semantic distinctiveness is then simply the sum of the word's vector elements, i.e., its magnitude. Words that occur in more semantically unique contexts will have a higher magnitude than words that appear in redundant or highly predictable contexts, given equal word frequencies.

The present study extended the earlier work of Jones *et al.* (2012) to an analysis of the accuracy of open-set spoken word recognition in noise. Word frequency (WF), contextual diversity (CD), and semantic distinctiveness (SD) were computed for words from three different corpora. Multiple regression analyses were then performed in order to determine the relative ability of each of these independent variables to predict unique variance in the accuracy of spoken word recognition in noise in two different data sets. Note that to a large extent, all three of these variables are simply different ways of counting frequency, ranging from counting raw frequencies (WF) through elimination of double counting words that occur multiple times in the same document (CD) to weighting each count by the dissimilarity of the context a word appears in relative to its other, earlier occurrences.

2. Method

2.1 Stimuli

Two different sets of proportion correct word recognition scores were analyzed. In both sets, spoken words were presented in isolation and the listener's task was open-set word identification. Data set 1 (DS1) consisted of proportion correct scores to 910 CVC spoken words. Words were presented over headphones in white, band-limited (low-passed filtered at 4.8 kHz) Gaussian noise at one of three signal-to-noise

ratios (SNR): -5 , $+5$, and $+15$ dB SPL. Each word was presented to ten different listeners at each of the three SNR. Hence, each word was tested a total of 30 times. These data were originally collected as part of an earlier study carried out by [Luce and Pisoni \(1998\)](#). Additional details may be found in that paper.

The second data set (DS2) consisted of proportion correct scores to 1428 spoken words selected to be a representative sample of American English across five different variables: Word frequency, initial phoneme, syllabic structure, number of phonemes, and number of syllables using a computer-readable version of Webster's Pocket Dictionary (see [Luce and Pisoni, 1998](#); [Nusbaum *et al.*, 1984](#)). The words were also presented over headphones in isolation to listeners in six-talker babble at one of three different SNR: 0 , $+5$, and $+10$ dB SPL. Each word was spoken by two different talkers, one female and one male. In the present analysis, results were collapsed across the two talkers. A total of 192 listeners participated in the study, with each hearing one-quarter of the entire stimulus set. Across all listeners, each word was presented a total of 48 times, 16 times at each of the three different SN ratios. Additional details may be found in [Felty *et al.* \(2009\)](#).

Several covariates of the words in each stimulus set were used in the present analyses. These covariates included lexical or phonological density (or neighborhood size), frequency-weighted density, and, for DS2, length in number of phonemes, and length in number of syllables. Values for these covariates were taken from the Hoosier Mental Lexicon ([Nusbaum *et al.*, 1984](#)). Lexical or phonological density is the number of words that the target could be changed into through the deletion, addition, or substitution of a single phoneme ([Landauer and Streeter, 1973](#)). Each such word is called a lexical neighbor of the target word. Frequency-weighted density is the sum of the log frequency of each of those neighbors.

2.2 Participants

For both DS1 and DS2, listeners were undergraduate students at Indiana University. None reported a history of hearing or speech disorders. All reported to be native speakers of American English.

2.3 Corpora

Three different corpora were used to compute WF, CD, and SD values: (1) the Touchstone Applied Science Associates (TASA) corpus ([Landauer and Dumais, 1997](#)), (2) a Wikipedia corpus ([Recchia and Jones, 2009](#)), and (3) a Usenet corpus ([Shaoul and Westbury, 2011](#)). The TASA corpus consists of 37 600 documents, whereas the Wikipedia and Usenet corpora consisted of 40 000 documents. These corpora were chosen to represent a diverse range of the English language.

2.4 Analyses

The analysis methods employed in this paper emulated those used by [Adelman *et al.* \(2006\)](#) and [Jones *et al.* \(2012\)](#). As in these studies, all WF, CD, and SD values were transformed to a log scale. The amount of unique variance contributed by one of the three variables of interest, call it X, independent of the other two, Y and Z, was determined as follows. First, a multiple regression analysis was performed that included Y and Z, as well as other covariates (see the following) as independent variables. That analysis was then repeated, but now with X also included as an independent variable. The proportionate increase in R^2 was then interpreted as the unique variance accounted for by X.

3. Results

Regression analyses were conducted in which the phonological density, number of phonemes, and number of syllables, in addition to WF, CD, and SD, were included as

Table 1. Unique variance predicted by phonological density, WF, DC, and SD from data set 1.

Corpus	SNR ^a	Effect [ΔR^2 (%) ^b			
		Density	WF ^c	CD ^d	SD ^e
TASA	Overall	17.02***	0.0	0.0	3.19 [†]
	-5	11.11*	1.51	1.26	0.11
	+5	29.62***	0.0	0.98	5.06*
	+15	8.16*	0.0	1.63	4.77
WIKI	Overall	23.33***	1.66	0.0	5.0
	-5	14.28*	7.14 [†]	0.0	7.38 [†]
	+5	37.5***	1.78	1.33	5.35 [†]
	+15	12.14 [†]	0.0	0.0	1.43
Usenet	Overall	24.35***	1.28	6.41*	20.51***
	-5	17.91**	0.0	4.18	37.31***
	+5	32.85***	4.28	8.57*	14.28**
	+15	15.62*	3.12	6.25	12.5 [†]

^aSNR is signal-to-noise ratio.

^{b†} $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

^cWF is word frequency.

^dCD is contextual diversity.

^eSD is semantic distinctiveness.

covariates. Table 1 contains the results for DS1, and Table 2 contains the results for DS2. For the DS1 analysis, number of syllables and number of phonemes were not included because all CVCs are, by definition, monosyllabic and contain three phonemes. Tables 1 and 2 show that even when other conventional properties of a word

Table 2. Unique variance predicted by phonological density, number of phonemes, and number of syllables WF, DC, and SD from data set 2.

Corpus	SNR ^a	Effect [ΔR^2 (%) ^b					
		Density	# Phon ^c	# Syll ^d	WF ^e	CD ^f	SD ^g
TASA	Overall	6.48***	2.31**	0.0	0.0	2.77**	6.94***
	0	4.32**	1.23 [†]	0.43	0.0	3.7**	8.64***
	+5	7.91***	1.69*	0.0	0.56	4.51***	8.47***
	+10	8.64***	2.46**	0.0	0.0	0.62	3.71**
WIKI	Overall	5.11***	4.54***	0.56	0.56	0.0	3.97**
	0	1.05 [†]	3.25*	0.57	0.81	0.0	6.5**
	+5	7.23***	3.94**	0.65	0.0	1.04	5.92***
	+10	7.24***	3.95**	0.66	0.0	1.31	5.72***
Usenet	Overall	11.29***	9.6***	1.69 [†]	1.11	5.02**	11.73***
	0	8.8***	8.0***	1.6	0.0	2.8 [†]	20.85***
	+5	14.86***	5.4**	0.67	0.0	1.35	10.14***
	+10	10.07***	13.95***	3.1*	0.75	2.79 [†]	7.75**

^aSNR is signal-to-noise ratio.

^{b†} $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

^cPhon is number of phonemes in the target word.

^dSyll is the number of syllables in the target word.

^eWF is word frequency.

^fCD is contextual diversity.

^gSD is semantic distinctiveness.

were included in the regression analysis, the SD variable was still able to account for an additional significant amount of unique variance not predicted by the other contributing variables. The effects in DS2 were more consistent than those for DS1 (which contained only monosyllabic CVC words), suggesting that SD has a greater effect when tested with a more lexically diverse data set. CD accounted for some unique variance, but these effects were much smaller than the effects of SD, suggesting that SD is the more important variable. Additionally, WF accounted for very little unique variance across all of the tests. This overall pattern of results demonstrates that these structural contextual variables, especially semantic distinctiveness, account for a significant amount of variance over and above other important variables in word recognition. This analysis was repeated with lexical density replaced with frequency-weighted density (Luce and Pisoni, 1998). The results were the same, with the use of frequency-weighted density slightly reducing the overall fit. Both of these analyses were repeated without including the covariates of density, number of phonemes, or number of syllables. The results for WF, CD, and SD were essentially the same as in the larger analysis.

Table 3 shows all possible pairwise correlations between WF, CD, and SD for each of the two data sets and each of the three corpora. As can be seen, all these correlations were highly significant. In particular, there is a strong correlation between WF and SD, as would be expected if, in fact, SD is a major contributor to word frequency effects.

4. Discussion

The present study examined the ability of word frequency, contextual diversity, and semantic distinctiveness to explain unique variance in the accuracy of spoken word recognition in noise. For the data set consisting of monosyllabic CVCs, although the effects did not always reach conventional levels of statistical significance, semantic distinctiveness was the most successful of the three variables in explaining unique variance in the accuracy of spoken word recognition. Contextual diversity did explain some unique variance, but for fewer combinations of corpus and noise level than did semantic distinctiveness. When contextual diversity did explain unique variance, it explained less than semantic distinctiveness. Word frequency explained little unique variance.

For the larger data set consisting of a representative sample of American English words (DS 2), word frequency again failed to explain any unique variance. In contrast, semantic distinctiveness consistently explained unique variance using all three corpora. Contextual diversity also explained some unique variance, although the results were not as consistent across corpora and noise levels. In addition, the amount of unique variance explained by contextual diversity was consistently less than that explained by semantic distinctiveness. Overall, then, the results indicate that a complete

Table 3. Correlations (R) among word frequency (WF), contextual diversity (CD), and semantic distinctiveness (SD).

Data set	Corpus	R^a		
		WF-CD	WF-SD	CD-SD
DS1	TASA	0.98	0.922	0.945
	Wiki	0.971	0.787	0.847
	Usenet	0.884	0.456	0.707
DS2	TASA	0.986	0.947	0.929
	Wiki	0.969	0.867	0.897
	Usenet	0.987	0.799	0.765

^aAll correlations are significant at the $p < 0.001$ level.

understanding of the effects of word frequency on spoken word recognition require taking into account not only the raw number of occurrences of a word, but also the number of different contexts in which it occurs and the semantic dissimilarity of those various contexts.

The present results are correlational and do not prove a causal relation between semantic distinctiveness and spoken word recognition accuracy, just as earlier demonstrations of the word frequency effect were correlational, and did not prove a causal relation. Just as another variable (semantic distinctiveness) seems to underlie the word frequency effect, it may well turn out that some other variable, potentially with no connection to semantics, underlies the effect of semantic distinctiveness observed here.

Hence, the results do not prove, but they do strongly suggest, that explanations of what have hitherto been called word frequency effects in word recognition studies need to take semantics into account. Equation (2) suggests ways of doing so. In logogen-based models of spoken word recognition (Morton, 1969, 1979), each repetition of a word is assumed to permanently lower the threshold of the logogen corresponding to that word. Thus, less evidence is required in order to recognize the word. If the amount of that decrease is proportional to the distinctiveness of the context on that repetition, as it is in Eq. (2), then semantic distinctiveness effects may result. Similarly, one way to account for word frequency effects in connectionist models (McClelland and Elman, 1986; Norris, 1994) is to assume that each repetition of a word increases the weight of the connection from phoneme units to the corresponding word unit (Dahan *et al.*, 2001). If the amount of that increase is proportional to the result of Eq. (2), then semantic distinctiveness effects may also result. In Bayesian models of spoken word recognition (Norris, 2006) one way to explain word frequency effects is to assume that a word's *a priori* probability is determined by its frequency of occurrence in the language. Replacing word frequency with semantic distinctiveness as the prior should, in such models, also result in semantic distinctiveness effects.

The above-mentioned approaches emphasize the effects of semantic distinctiveness on a word's memorial representation. Multiple-trace or exemplar theories (Hintzman, 1986, 1988) of spoken word recognition (Goldinger, 1998) stress, in addition, the semantic processing done at the time of recognition. These models assume that each individual trace of a spoken word contains not only acoustic-phonetic information, but also highly detailed contextual information. Presumably, that contextual information could include semantic information. In that case, the greater the variety of contexts in which a word appears, the greater the probability that, whatever semantic context happens to be active on a given trial in an isolated word recognition task, at least one stored exemplar would be activated sufficiently for correctly recognizing the word. Indeed, Goldinger (1998), especially in his discussion of distributed models of word recognition, seems to have foreshadowed such an approach.

Exhaustively delineating and discriminating these various explanations of the effects of semantic distinctiveness is beyond the scope of the present report. The main point of these new analyses is to document that any account of word frequency effects in spoken word recognition would seem to need to include a role for semantics. It is the interaction of repetition (frequency of occurrence) with semantic context, not just repetition alone, that appears to be responsible for what has, up to now, been described as word frequency effects. Our expectation is that further explorations of these contextual effects will lead to a more detailed understanding of the close connections between perception and semantics.

Acknowledgments

This work was supported by National Institutes of Health, Grant No. DC00111-34 to D.B.P., and NSF, Grant No. BCS-1056744 to M.N.J. B.T.J. and T.M.G. contributed equally to this work.

References and links

- Adelman, J. S., Brown, G. D. A., and Quesada, J. F. (2006). "Contextual diversity, not word frequency, determines word-naming and lexical decision time," *Psychol. Sci.* **17**, 814–823.
- Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., and Yap, M. J. (2004). "Visual word recognition of single-syllable words," *J. Exp. Psychol. Gen.* **133**, 283–316.
- Broadbent, D. E. (1967). "Word-frequency effect and response bias," *Psychol. Rev.* **74**, 1–15.
- Dahan, D., Magnuson, J. S., and Tanenhaus, M. K. (2001). "Time course of frequency effects in spoken-word recognition: Evidence from eye movements," *Cognitive Psychol.* **42**, 317–367.
- Felty, R. A., Buchwald, A., and Pisoni, D. B. (2009). "Error analysis of spoken word recognition," *Research on Spoken Language Processing: Progress Report No. 29* (Department of Psychological and Brain Sciences, Indiana University Speech Research Laboratory, Bloomington, IN), pp. 183–196.
- Forster, K. I., and Chambers, S. M. (1973). "Lexical access and naming time," *J. Verbal Learn. Verbal Behav.* **12**, 627–635.
- Goldinger, S. D. (1998). "Echoes of echoes? An episodic trace theory of lexical access," *Psychol. Rev.* **105**, 251–279.
- Hintzman, D. L. (1986). "'Schema abstraction' in a multiple-trace memory model," *Psychol. Rev.* **93**, 411–428.
- Hintzman, D. L. (1988). "Judgments of frequency and recognition memory in a multiple-trace memory model," *Psychol. Rev.* **95**, 528–551.
- Howes, D. H. (1957). "On the relation between the intelligibility and frequency of occurrence of English words," *J. Acoust. Soc. Am.* **29**, 296–305.
- Jones, M. N., Johns, B. T., and Recchia, G. (2012). "The role of semantic diversity in lexical organization," *Can. J. Exp. Psychol.* **66**, 115–124.
- Krueger, L. E. (1975). "Familiarity effects in visual information processing," *Psychol. Bull.* **82**, 949–974.
- Landauer, T. K., and Dumais, S. T. (1997). "A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge," *Psychol. Rev.* **104**, 211–240.
- Landauer, T. K., and Streeter, L. A. (1973). "Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition," *J. Verbal Learn. Verbal Behav.* **12**, 119–131.
- Luce, P. A., and Pisoni, D. B. (1998). "Recognizing spoken words: The neighborhood activation model," *Ear Hear.* **19**, 1–36.
- McClelland, J. L., and Elman, J. L. (1986). "The TRACE model of speech perception," *Cogn. Psychol.* **18**, 1–86.
- Morton, J. (1969). "The interaction of information in word recognition," *Psychol. Rev.* **76**, 165–178.
- Morton, J. (1979). "Word recognition," in *Structures and Processes*, edited by J. Morton and J. C. Marshall (MIT Press, Cambridge, MA), pp. 108–156.
- Norris, D. (1994). "Shortlist: A connectionist model of continuous speech recognition," *Cognition* **52**, 189–234.
- Norris, D. (2006). "The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process," *Psychol. Rev.* **113**, 327–357.
- Nusbaum, H. C., Pisoni, D. B., and 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* (Psychology Department, Indiana University, Speech Research Laboratory, Bloomington, IN), pp. 357–376.
- Pollack, I., Rubenstein, H., and Decker, L. (1959). "Intelligibility of known and unknown message sets," *J. Acoust. Soc. Am.* **31**, 273–279.
- Recchia, G., and Jones, M. N. (2009). "More data trumps smarter algorithms: Comparing pointwise mutual information with latent semantic analysis," *Behav. Res. Methods* **41**, 647–656.
- Savin, H. B. (1963). "Word-frequency effect and errors in the perception of speech," *J. Acoust. Soc. Am.* **35**, 200–206.
- Shaoul, C., and Westbury, C. (2011). "A USENET corpus (2005-2010)," <http://www.psych.ualberta.ca/~westburylab/downloads/usenetcorpus.download.html> (Last viewed 01/29/2012).
- Shepard, R. N. (1987). "Toward a universal law of generalization for psychological science," *Science* **237**, 1317–1323.