CONFUSION PATTERNS AND RESPONSE BIAS IN SPOKEN WORD RECOGNITION OF GERMAN DISYLLABIC WORDS AND NONWORDS

Robert Felty

University of Michigan
robfelty@umich.edu

ABSTRACT

The abundant research on lexical access in the last 30 years has shown that context effects such as lexical status, morphological complexity, and neighborhood density can affect word recognition. Very little research has investigated interactions between perceptual distinctiveness and context effects. This study used a spoken word recognition in noise experiment with German words and nonwords to research this interaction. Results showed a processing advantage for monomorphemic words over bimorphemic words, and that listeners are particularly sensitive to morphological information when presented with highly confusable stimuli.

Keywords: Detection theory, noise, confusion, German, bias

1. INTRODUCTION

Spoken word recognition is known to be a complex process of integrating acoustic, lexical, and grammatical information. Lexical information can be used to recover degraded acoustic information, as in the phonemic restoration effect [14]. Morphological information has also been shown to affect lexical access [12, 7]. Few studies to date have investigated the effects of lexical and grammatical information in degraded spoken word recognition. This experiment is designed to discover how morphological and frequency information might be used by listeners hearing a degraded acoustic signal.

2. METHOD

2.1. Participants

Thirty-two paid participants were recruited via flyer from the University of Konstanz. All participants reported being native speakers of German and having no known hearing impairments.

2.2. Materials

The stimuli consisted of CVCCVC trochees, including 150 nonwords and 150 German words (75 monomorphemic and 75 bimorphemic). The words included nouns and adjectives with and without inflectional suffixes, selected from the CELEX database [1]. The monomorphemic and bimorphemic words were matched for lexical frequency and neighborhood density. The nonword stimuli were generated from the distribution of phonemes in the word stimuli, which ensured that the word and nonword lists were largely phonotactically balanced. A subset of the materials excluding stimuli which contained post-vocalic ʀ are reported in the analysis here, including 94 nonwords, 36 mono- and 43 bimorphemic words.

The stimuli were recorded at the University of Michigan in an anechoic chamber directly into .wav format with a sampling rate of 44.1 kHz. Each item was read by a male speaker of German embedded in the carrier phrase "Sagen Sie ___ einmal". Each stimulus was extracted from the carrier phrase, padded with 100 ms of silence on both sides, and amplitude-normalized.

2.3. Procedure

The stimuli were presented to the participants in isolation over closed headphones in a quiet room. The stimulus presentation and response collection was controlled by software developed in Matlab, which mixes signal-dependent noise [11] with the recorded stimuli, and allows for the collection of open response data typed in via the keyboard. Two different signal-to-noise-ratios (S/Ns) — 2 dB and 7 dB — were determined from pilot results. Half of the participants heard the stimuli presented at S/N = 2 dB and half at S/N = 7 dB. Listeners were instructed that they would hear disyllabic words and nonwords mixed with noise, and that they should type what they hear, using standard German orthography.

The experiment began with two practice blocks (one word, one nonword) of 10 stimuli each, in order to familiarize the participants with the task. The main experiment consisted of 20 blocks of 15 stimuli each; to make the task less demanding for participants, stimuli were blocked according to lexical status. Participants only heard each stimulus once, but had no time limit to type in their answer.

2.4. Analysis

The data were analyzed using the j-factor model [4], which provides a measure of the number of independent units in a stimulus. A finding of $j = n$ (where n is the number of phonemes in the stimulus) is interpreted as evidence that phonemes are perceived independently of one another. Previous studies [4, 10, 2] have consistently found $j = 3$ for CVC nonwords, and $j \approx 2.5$ for CVC words, which has been interpreted as a response bias for words [8].

3. PREDICTIONS

Although no studies to date have used the j-factor model to analyze disyllabic stimuli, several predictions can be made based on previous studies using the j-factor model with CVC stimuli [4, 10, 2].

1. $j_{nonword} \approx 6$. This result would provide evidence that phonemes in nonwords are perceived independently of one another.
2. $j_{word} \approx 5$. Given that previous studies using CVC stimuli have found $j_{word} \approx 2.5$ [4, 10, 2], it is logical to hypothesize that j_{word} will be twice as large for CVCCVC stimuli.
3. $j_{bi} > j_{mono}$. This prediction follows from the hypothesis that morphological units are stored in the lexicon, and that increasing the number of morphemes in a word should add to the number of independent units.
4. $j_{word} \propto$ density: Neighborhood density provides an inhibitory effect, such that words in dense neighborhoods are more difficult to process than words in sparse neighborhoods [2].
5. Listeners rely more heavily on lexical and grammatical information in the absence of clear acoustic information. Thus, effects of morphology should be greatest for highly confusable stimuli.

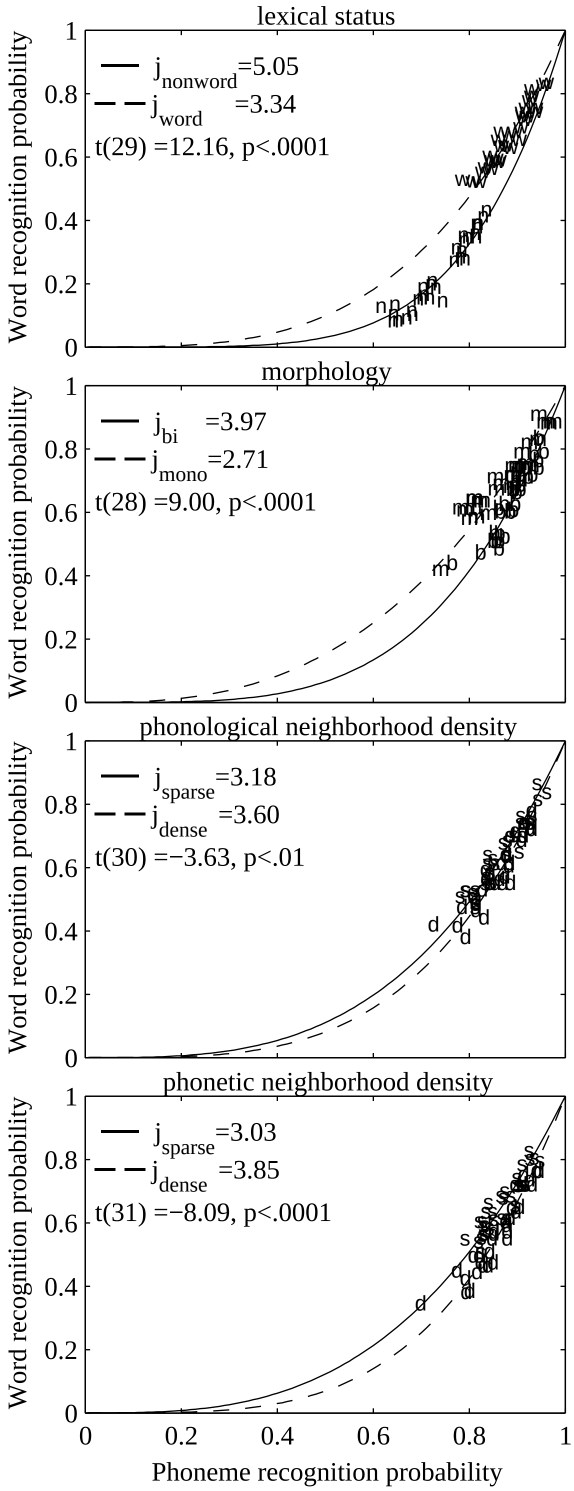
4. RESULTS AND DISCUSSION

Results of the j-factor analysis are shown in Figure 1.

4.1. Lexical Status

As predicted, nonwords had significantly higher j-scores than words, which is consistent with previous studies using the j-factor model [4, 10, 2], and can be interpreted as a bias towards words. The actual values for j are somewhat unexpected. The result of $j_{nonword} \approx 5.05$ is substantially lower than the predicted value of 6. There are several explanations for this result. It is possible that the nonwords chosen in this experiment had a particularly high phonotactic probability, resulting in a lesser degree of perceptual independence than expected. Stimuli with

Figure 1: j-factor results — Each point represents the average results for one subject. Curves represent $y = x^j$, where j is the mean from each group. Statistics shown are from paired t-tests (one-tailed); before computing the statistics, all points lying in the floor or ceiling ranges (> .95 or < .05) were removed, but are still shown on the plot.



high phonotactic probability should have a lesser degree of perceptual independence than those with low phonotactic probability, which should be reflected in lower j-scores. However, results of a separate analysis of phonotactic probability analysis do not support this hypothesis. Another explanation is that the stimuli were all trochees (i.e. disyllables with initial stress), and therefore the set of 4 possible vowels for the second vowel were highly restricted (/ʊ ɪ ə ɔ/ compared to the set of 18 possible vowels for the first vowel (/i ɪ y ʏ e ɛ ø œ u ʊ o ɔ a aː ə ay ɔy au/).

In addition, the result of $j_{word} \approx 3.34$ is also lower than predicted. This is likely partially due to the trochaic syllable structure of the words, as for the nonwords, but it is also possible that bias towards words does not scale linearly with word length. Additional research using stimuli of a variety of lengths can address these questions.

4.2. Morphology

As predicted, j of bimorphemic words was significantly higher than that of monomorphemic words ($j_{bi} = 3.97$, $j_{mono} = 2.71$, $p < .0001$). This is evidence that monomorphemes have a greater amount of lexical context than bimorphemes. Each phoneme in a monomorphemic word contributes to the meaning of the word, while phonemes part of an inflectional suffix in bimorphemic words do not contribute to the meaning of the word, but serve only a grammatical function.

4.3. Neighborhood Density

The effect of neighborhood density was also investigated. Two different metrics of neighborhood density were computed: a phonological and a phonetic measure. The phonological measure simply counted the number of words differing by only phoneme (see [9]), while the phonetic measure weighted each neighbor according to perceptual confusability based on confusion matrices from the nonword data, (see [6, 2]). Sparse and dense groups were created from the word list using a median split. As shown in Figure 1, the effect of neighborhood density was significant using both the phonological and the phonetic measure, though the phonetic measure shows a much larger effect, as measured by the difference in j. Subsequent linear regression analyses showed that the phonetic measure accounted for 14.5% of the variation in j ($F(1,182) = 13.78, p < .001$), while the regression analysis using the phonological measure was not significant. These results are consistent with previous results using the j-factor model [2, 3], and underscore the importance of including detailed perceptual information in models of spoken word recognition.

4.4. Perceptual distinctiveness, morphology, and response Bias

As predicted, the mean j of monomorphemic words was significantly lower than that of bimorphemic words. This can be interpreted in several non-mutually exclusive ways. One interpretation is that morphemes add to the overall number of independent units of a word. Another interpretation is that bimorphemic words are less predictable than monomorphemic words, and therefore the phones are less independent of one another than in monomorphemic words. Consider two words, one monomorphemic and one bimorphemic, with an equal number of neighbors. The bimorphemic neighbor will likely (and in the case of the German certainly) include neighbors which share the same lemma, whereas the monomorphemic words should not include such neighbors. A listener presented with a bimorphemic word whose neighbors share the same root must rely on wordform frequency rather than lemma frequency as a predictor of which response is more probable. This is problematic, as several studies have shown that listeners are more attuned to lemma frequency than wordform frequency [5, 13]. If listeners are primarily depending on lemma frequency to make educated guesses, then they must use a strategy based on something other than lemma frequency when choosing between bimorphemic neighbors differing only in their final consonant. Such a strategy could include raw acoustics and knowledge about the distribution of affixes.

These strategies can be tested by investigating the degree of acoustic salience and response bias in the data. The final consonants in the bimorphemic stimuli were restricted to the phonemes /ʀ s m n/, which, along with /ə/ constitute all of the possible inflectional endings for nouns and adjectives in German. Two of these, /m/ and /n/ are known to be highly confusable with one another. In addition, /n/ occurs as an inflectional ending much more frequently than /m/. Thus it is highly probable that both acoustic factors as well as response bias could be playing a role in the perception of these two final consonants. In order to investigate this further, a Signal Detection Theory (SDT) analysis was carried out.

To carry out the SDT analysis, the original confusion matrices for each S/N were transformed into 2x2 submatrices. An SDT analysis was then applied to each submatrix. From the results shown in Table 1, several conclusions can be drawn — (1) in the absence of lexical context effects (i.e. in the nonword condition), /m/ and /n/ are highly confusable, with a small bias towards /n/, (2) /m/ and /n/ are perceived as most distinct in the monomorphemic con-

Table 1: Signal Detection Theory analysis of /m/ and /n/ submatrix in final position. For this analysis /m/ is considered to be the target stimulus. Positive values of c indicate a bias towards /n/. The final two columns list the total number of presentations of /m/ and /n/ which were used to compute the SDT analysis

	d'	c	/m/	/n/
Nonwords				
lower S/N (2 dB)	-0.182	0.555	240	240
higher S/N (7 dB)	0.664	0.743	240	240
Bimorphemes				
lower S/N (2 dB)	1.616	0.984	128	352
higher S/N (7 dB)	1.913	0.556	128	352
Monomorphemes				
lower S/N (2 dB)	3.514	0.239	48	192
higher S/N (7 dB)	4.733	-0.060	48	192

dition, and (3) bias towards /n/ is greatest in the bimorphemic case. The increase in distinction in the monomorphemic case can be interpreted as a result of the greater ability to distinguish between neighbors based on lexical frequency information. The bias towards /n/ in the bimorphemic case can be interpreted as evidence that listeners are exploiting the fact that the /n/ ending occurs most frequently among all possible inflectional endings in German, and they are therefore choosing /n/ more frequently. The results of the SDT analysis show that listeners seem to be depending on a combination of acoustics, lemma frequency, and morphological distribution to make their decisions.

5. CONCLUSION

This experiment has addressed several context effects in spoken word recognition. The j-factor analysis showed that phonemes are perceived roughly independently of one another in nonwords, and that there is a strong bias towards words over nonwords. Morphology also can effect spoken word recognition, in that j was significantly higher for bimorphemic words than for monomorphemic words. Neighborhood density had a robust effect on word recognition, such that words in sparse neighborhoods showed a strong bias over words in dense neighborhoods. Moreover, a phonetically based measure of neighborhood density accounted for a much larger portion of the variation in the data than a phonologically based measure. Finally, an SDT analysis showed that morphological information can interact with acoustic information, both in terms of perceptual distinctiveness and response bias.

ACKNOWLEDGMENTS

I wish to thank the Linguistics department at the University of Konstanz for their hospitality and assistance, particularly Aditi Lahiri and Mathias Scharinger. I am also grateful for the support and insights of my advisors at the University of Michigan, José Benkí and Pam Beddor.

6. REFERENCES

[1] Baayen, R., H, Rijn, H. 1993. The CELEX lexical database (cd-rom). Philadelphia: Linguistics Data Consortium, University of Pennsylvania.

[2] Benkí, J. 2003. Quantitative evaluation of lexical status, word frequency and neighborhood density as context effects in spoken word recognition. *Journal of the Acoustical Society of America* 113(3), 1689–1705.

[3] Benkí, J., Felty, R. October 2006. Relative contributions of initial and final similarity to neighborhood density effects on english spoken word recognition. Poster presented at the Fifth International Conference on the Mental Lexicon: Montreal, Quebec.

[4] Boothroyd, A., Nittrouer, S. 1988. Mathematical treatment of context effects in phoneme and word recognition. *Journal of the Acoustical Society of America* 84, 101–114.

[5] Clahsen, H., Isenbeiss, S., Hadler, M., Sonnenstuhl, I. 2001. The mental representations of inflected words: an experimental study of adjectives and verbs in German. *Language* 77(3), 510–543.

[6] Luce, P., Pisoni, D. 1998. Recognizing spoken words: The neighborhood activation model. *Ear and Hearing* 19, 1–36.

[7] Marslen-Wilson, W. 2001. Access to lexical representations: Cross-linguistic issues. *Language and Cognitive Processes* 16(5/6), 699–708.

[8] Nearey, T. M. 2001. Phoneme-like units and speech perception. *Language and Cognitive Processes* 16, 673–681.

[9] Newman, R. S., Sawusch, J. R., Luce, P. A. 1997. Lexical neighborhood effects in phonetic processing. *Journal of Experimental Psychology* 23(1), 873– 889.

[10] Olsen, W., Tasell, D. V., Speaks, C. 1997. Phoneme and word recognition for words in isolation and sentences. *Ear and Hearing* 18(3), 175–188.

[11] Schroeder, M. 1968. Reference signal for signal quality studies. *Journal of the Acoustical Society of America* 44, 1735–1736.

[12] Taft, M. 1979. Recognition of affixed words and the word frequency effect. *Memory & Cognition* 7, 263–272.

[13] Vannest, J., Newport, E. L., Bavelier, D. October 2006. How frequent is a word? reexamining base and surface frequencies. Poster presented at the Fifth International Conference on the Mental Lexicon: Montreal, Quebec.

[14] Warren, R. 1970. Perceptual restoration of missing speech sounds. *Science* 167, 392–393.