

(appears in BLS 28)

Effects of signal-independent factors in speech perception

José R. Benkí*

University of Michigan, Ann Arbor

Benkí, J. R. (2002) Effects of signal-independent factors in speech perception. *Proceedings of the 28th annual meeting of the Berkeley Linguistics Society*, edited by J. Larson and M. Paster (Berkeley Linguistics Society, Berkeley), pp. 63-74.

1. Introduction and background

In human speech recognition, listeners use sensory information from the speech signal to match a stimulus with an internal representation. The accuracy of that process is affected by many factors, including, but not limited to, the acoustic-phonetic properties of the stimulus, whether the stimulus is a familiar lexical item, its frequency of usage, and whether the stimulus is confusable with other words.

Boothroyd and Nittrouer (1988) quantified the advantage of lexical status afforded to listeners by comparing the recognition of familiar CVC words with CVC nonsense (though phonotactically English) syllables using the j-factor model. While they did not evaluate contextual effects due to usage frequency or neighborhood density, the design is well suited to quantification of effects as well. The present study is investigation of how lexical status, frequency, and neighborhood density affect the speech recognition in noise, through a replication and an extended analysis of the first experiment in Boothroyd and Nittrouer. The j-factor model, proposed as a metric of context effects insensitive to overall performance level, is used to quantify the effects of these factors. All of these factors are measured with the j-factor model. The effect of neighborhood density is particularly interesting because it is primarily due to the first two segments.

Boothroyd and Nittrouer propose two measures of context effects that are relatively insensitive to the degree of signal degradation or overall performance level. The present study uses the second of these measures, the j-factor, which quantifies the recognition of a whole as a function of the recognition of its parts. From probability theory, the recognition probability of a whole is the product of the marginal recognition probabilities of its parts. For CVC syllables, $p(\text{syll}) = p(C_1) p(V) p(C_2)$. Assuming the recognition probabilities of individual segments or phonemes in CVC syllables are statistically independent and

* Special thanks to Katherine Cassady and Sherene Flemmings for running the experiment. A preliminary version of this work was presented a Cognitive Science Cognitive Neuroscience workshop in August 2001 at the University of Michigan, organized by Julie Boland and Rick Lewis. I am grateful to Pam Beddor, John Kingston, Terry Nearey, and other audience members for helpful comments. All errors are my own.

approximately equal (Fletcher, 1953), $p(\text{syll}) = p(\text{seg})^j$, where j represents the number of independently perceived segments in the syllable. The j -factor can be empirically determined by calculating the logarithms of recognition probabilities of whole syllables and segments in an identification task, yielding $j = \log(p(\text{syll})) / \log(p(\text{seg}))$. A finding of $j = n$ (where $n = 3$ for CVC stimuli) is consistent with independent recognition of the segments and implies that listeners are not exploiting contextual information. The reduction of j below n is a measure of the effect of context. At the limit of $j = 1$, the recognition of any one segment is all that is needed to recognize the whole.

In the first experiment of their study, Boothroyd and Nittrouer measured j -factors for CVC words ($j = 2.46 \pm 0.08$) nonsense syllables ($j = 3.07 \pm 0.14$), concluding that $j = 3.07$ for nonsense targets is consistent with perception of three independent units. The finding of $j = 2.46$ for word targets is interpreted by Boothroyd and Nittrouer as a measure of the contextual advantage for words.

The j -factor reduction indicates that the higher recognition probabilities of meaningful syllables are due in part at least to the higher predictability of words relative to nonwords (cf. Allen, 1994). The j -factor quantifies this lessening of statistical independence among the segments of meaningful syllables.

On the basis of a computational simulation of Boothroyd and Nittrouer's experiment, Nearey (1998) suggests that the j -factor effects could be reproduced in a Luce choice model of word recognition as a bias that favors words over nonsense syllables. If the j -factor measures bias, then manipulations of bias in a word recognition task should affect the j -factor. If facilitation for high frequency words is the result of a bias (Broadbent, 1967; Norris 1986), then high frequency words are predicted to have lower j -factors than low frequency words.

Potential confusors to a given stimulus can affect recognition (Savin, 1963). The neighborhood activation model (NAM; Luce & Pisoni, 1998), quantified in (1), proposes that phonetic neighbors compete with the actual target for activation in a Luce choice model. Degree of phonetic overlap between the neighbor and the target stimulus determines the degree of competition. The log usage frequency is used as a weight for both the target and its neighbors.

$$(1) \quad p(ID_S) = \frac{p(S|S) \log(freq_S)}{p(S|S) \log(freq_S) + \sum_j p(N_j|S) \log(freq_j)}$$

The probability of identifying a stimulus S is $p(ID_S)$; $p(S|S) \log(freq_S)$ is the frequency-weighted stimulus word probability of S given S (FWSWP), and $\sum_j p(N_j|S) \log(freq_j)$ is the sum of the frequency-weighted probabilities of each neighbor N_j of S given S (FWNP).

For empirical evaluation of the model, Luce and Pisoni use the Kucera-Francis (Kucera & Francis, 1967) usage frequencies, and their own confusion matrices of nonsense syllables in noise. The conditional probability of an item is

estimated by multiplying the conditional marginal probabilities of the constituent segments obtained from the confusion matrices.

Accuracy should be positively correlated with FWSWP, the stimulus probability based on acoustic-phonetic salience weighted by frequency, but negatively correlated with FWNP, the frequency-weighted probability of competitors. These qualitative predictions as well as the quantitative predictions of (1) are borne out in experiments reported by Luce and Pisoni.

The j-factor model can be applied to the parts of (1) to measure the contextual advantages of words with high and low values of FWSWP and FWNP. In the case of the FWSWP, only variation from usage frequency and not stimulus probability should be measurable with the j-factor. It may be that the stimulus probability factor dominates the frequency weight, in which case the FWSWP should not have any context effect, as measured by the j-factor.

On the other hand, neighborhood density, quantified by the FWNP, should be correlated with the j-factor if the j-factor is inversely related to bias, as suggested by Nearey. Consider the case of a listener perceiving partial phonetic information of a target word in a dense phonetic neighborhood. Given the partial phonetic information, the probabilities of non-target potential responses are large, so any bias in favor of the target will be reduced. If the partial phonetic information delimits a sparse phonetic neighborhood, the probabilities of the non-target competitors are low, and bias for the target should be high. Under this account, words with high values of FWNP (low bias) should have high j-factors, while words with low values of FWNP (high bias) should have low j-factors.

Boothroyd and Nittrouer's design offers an opportunity to test these predictions of frequency and neighborhood density, since the words span a range of usage frequencies, and are phonetically balanced with the nonsense syllables.

2. Method

The procedure for Boothroyd and Nittrouer's Experiment 1, in which participants identified CVC nonsense and word syllables at different noise levels, was followed as closely as possible, except that stimulus presentation and response collection was done online. Proportion correct of phonemes and whole syllables of different subsets of the test items were subsequently used in j-factor analyses.

Forty-three young adults were recruited from an undergraduate linguistics course at the University of Michigan and participated for course extra credit. All were native speakers of English and reported no known hearing problems.

The same lists of CVC syllables developed by Boothroyd and Nittrouer, consisting of 120 words and 120 nonsense items were used for this study. Both the word and nonsense syllable lists were phonetically balanced such that the phonemes in the sets of 10 initial consonants /b p d t k s h m l r/, 10 vowels /i ɪ eɪ ε u ou ɔ æ ɑ aɪ/, and 10 final consonants /p d t g k s z m n l/ were evenly distributed in the word and nonsense syllable lists.

Each item was read by the author, a native speaker of American Midwest English, in the carrier phrase “You will write ... please” in a sound-treated room and was recorded to DAT with a Realistic Highball microphone and a Tascam DA-30 digital tape deck at a sampling rate of 48 kHz. The recording of each item embedded in the carrier phrase was converted to a WAV file at the same sampling rate and stored on computer disk. The overall level of each stimulus was adjusted so that the peak amplitudes of all stimuli were matched.

The experiment was run using software running in the Matlab (version 6.1) environment on four Windows NT laptop computers in an anechoic chamber. Signal-dependent (though uncorrelated) noise (Schroeder, 1968) was added online at one of four S/N ratios (–14 dB, –11 dB, –8 dB, –5 dB). The resulting stimuli in their carrier phrases were presented for identification binaurally via AKG headphones with the volume set to a comfortable listening level, presented in 24 random blocks of 10 random targets, each block containing all words or nonsense syllables.

Thirty-seven participants were randomly assigned to one of the four S/N ratios (11 participants at –14 dB, 9 at –11 dB, 9 at –8 dB, and 8 at –5 dB). Six participants at the beginning of the study were assigned to other S/N ratios (1 at –10 dB, 2 at –9 dB, 2 at –7.5 dB, and 1 at –4 dB) in order to determine a range of S/N ratios for performance levels approximate of those in Boothroyd and Nittrouer. All participants were instructed in writing that they would be listening to real and nonsense consonant-vowel-consonant syllables of English presented in a carrier phrase with noise, and were to type what they heard using standard English orthography for both the words and nonsense items. A brief list of examples of English orthography for spelling nonsense items was provided.

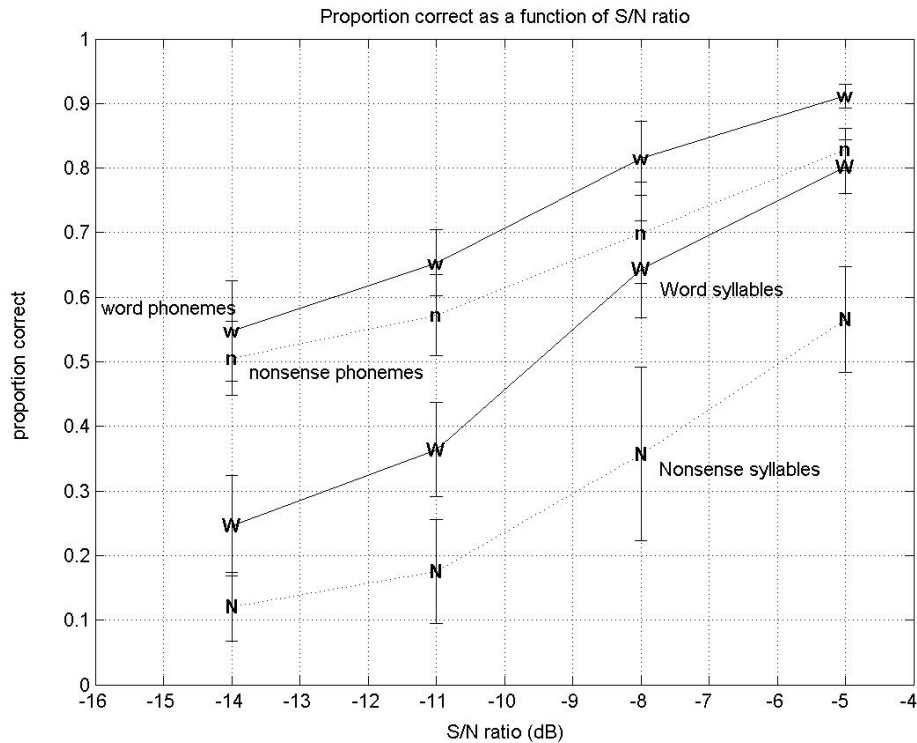
3. Results and analysis

Each phoneme response was scored as correct if it matched the corresponding stimulus phoneme, and incorrect otherwise, with the following adjustments. First, /a/ and /ɔ/ were counted as matching vowels. In their stimulus list preparation and response analysis, Boothroyd and Nittrouer regarded the vowels /a/ and /ɔ/ as distinct phonemes, and these distinctions were maintained in the preparation of the stimuli for the present study. However, these vowels are merged in the English spoken by many of the participants, and were therefore counted as the same vowel for scoring purposes. Missing consonants were scored as “null” responses and incorrect. If a cluster was reported for one of the consonants, it was scored as “other” and incorrect, unless one of the elements was a correct response and the epenthetic consonant occurred between the vowel and the correct consonant. In those cases, half were scored as vowel errors (“other”) and half as consonant errors (“other”).

Using the above criteria, the observed probabilities of correct recognition of nonsense phonemes, word phonemes, nonsense syllables, and word syllables for each participant were calculated and are plotted in (2).

Effects of signal independent factors

(2)



The range and pattern of performance is quite similar to those reported by Boothroyd and Nittrouer. The S/N ratios are about 11 dB lower in the present study, which may be due to differences in the quality of the stimuli, type of noise (they used spectrally-shaped white noise that was the same level for all of the stimuli of a given S/N ratio, instead of the signal-correlated noise used here), or the experimental procedure.

Proportions correct of phonemes and syllables for each condition were converted into j-factors for each participant and averaged for an estimate of the j-factor for each condition. Because measurement errors for probabilities near zero or unity have a large effect on the estimate of the j-factor, if either the phoneme or syllable probability was less than 0.05 or greater than 0.95, the resulting j-factor was not included in the calculation of average j-factor or subsequent statistical tests, following Boothroyd and Nittrouer.

4.1. Lexical status

The j-factors averaged across participants for the high and low context condition for each comparison are shown in (3). There is no significant difference between the nonsense syllable j-factor $j=3.07$ and $n=3$ ($t(40)=1.69$, $p=0.0991$), as predicted by independent perception of phonemes in nonsense syllables and consistent with Boothroyd and Nittrouer. A paired comparison of words and nonsense syllable j-factors shows the word j-factor mean, $j=2.35$, to be significantly less than the

nonsense syllable *j*-factor mean ($t(40)=11.196$, $p<0.00001$), diagnostic of phonemes in words not being perceived independently of each other, or of a bias in favor of words.

(3)

The *j*-factors for high and low context conditions of lexical status, frequency, FWNP, and FSWP averaged over participants.

Comparison	<i>j</i> -factors with 95% confidence intervals	
	High context	Low context
Lexical status: word/nonsense syllable	2.34 ± 0.08 (B&N 2.46 ± 0.08)	3.07 ± 0.08 (B&N 3.07 ± 0.14)
High freq./Low freq.	2.25 ± 0.10	2.46 ± 0.10
Low FWNP/High FWNP	2.11 ± 0.09	2.61 ± 0.11
High FSWP/Low FSWP	2.46 ± 0.13	2.39 ± 0.09

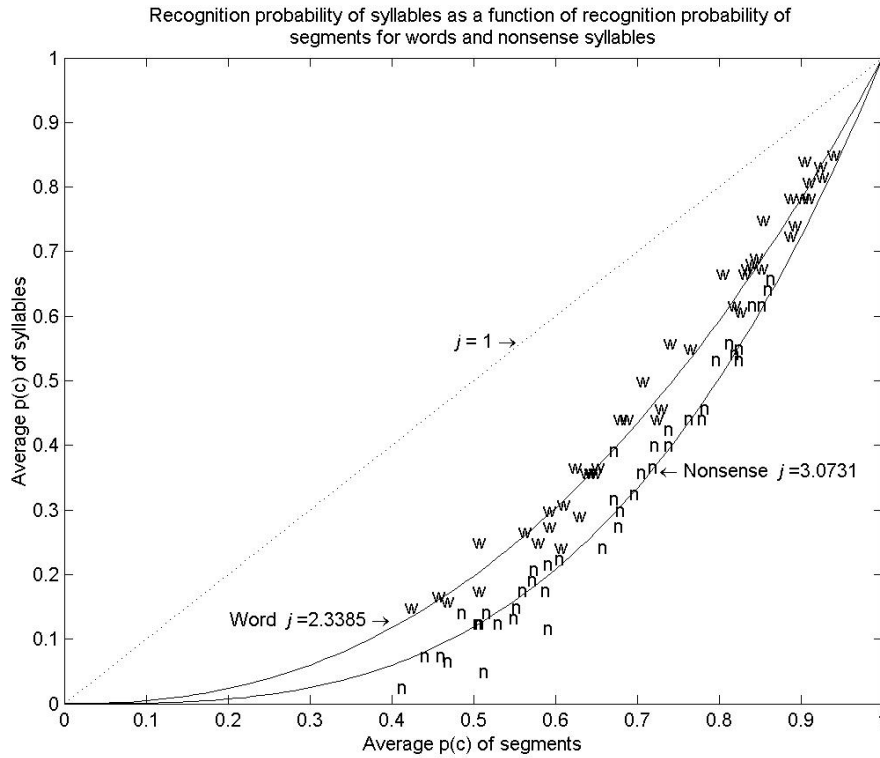
Values for each participant are plotted in (4), with best-fitting curves for words and nonsense syllables. Each point represents average syllable recognition probability as a function of average phoneme recognition probability for words or nonsense syllables of a single individual. The *j*-factors are not significantly correlated with phoneme recognition probability for either nonsense syllables ($R^2=0.0650$, $F(1,40)=2.7104$, $p=0.1077$) or for words ($R^2<0.00001$, $F(1,42)=0.0002$, $p=0.9998$). The lack of correlation with phoneme recognition probability and the good fit across the range of recognition probability supports the use of the *j*-factor as an index of context effects independent of recognition probability.

4.2. Word frequency

Word frequency effects were measured by dividing word trials into high and low frequency groups using the median log Kucera-Francis frequency of all the words (3.29) as a cutoff. The high frequency words have a mean log Kucera-Francis frequency of 4.90, while the low frequency words have a mean log Kucera-Francis frequency of 2.46. Average phoneme and syllable recognition probabilities were calculated for high and low frequency words for each participant, and converted to *j*-factors as shown in (3). As expected, the high frequency words have a lower *j*-factor ($j=2.25$) than the low frequency words ($j=2.46$), consistent with the prediction of a bias in favor of high frequency words, with the magnitude being about a third of the size of the lexical status effect. A paired comparison indicates that the difference between the mean high and low frequency *j*-factors is significant ($t(42)=3.809$, $p=0.00045$). The difference is also significant if a familywise $\alpha=0.05$ error rate is maintained using a Bonferroni criterion for four tests (the cutoff for familywise $\alpha=0.05$ for four tests is $t=2.4949$; Hays 1994, p.1007).

Effects of signal independent factors

(4)



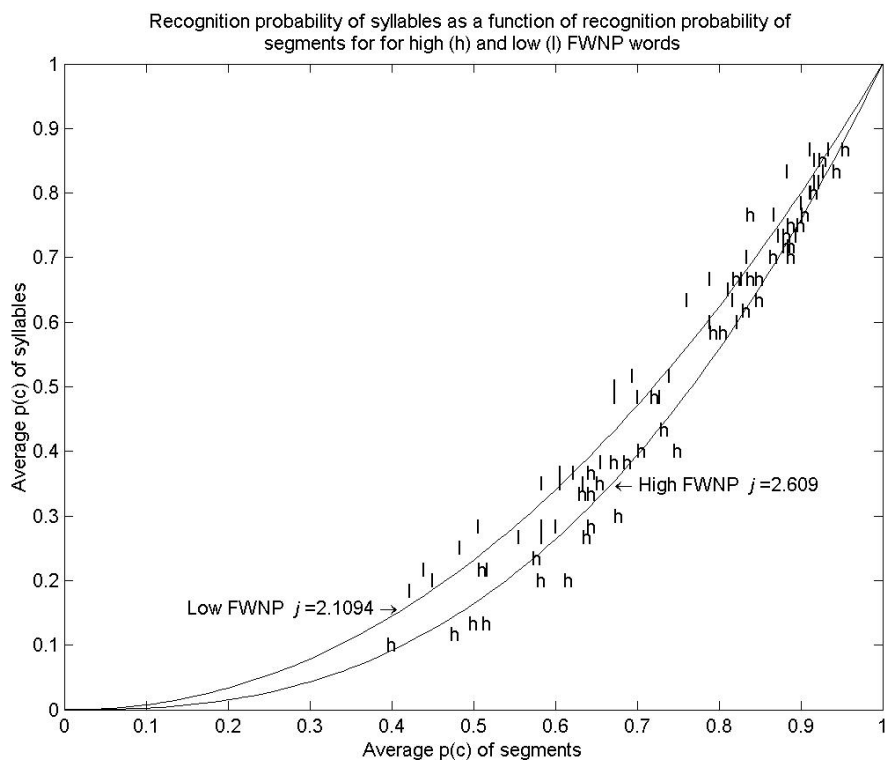
4.3. Neighborhood density and stimulus word probability

An online version of Webster’s Pocket Dictionary (*Webster’s Seventh Collegiate Dictionary*, 1967; Nusbaum, Pisoni, & Davis, 1984) was used to determine the neighbors for each target word. All neighbors differed with the target by one segment, with a substitution or a deletion for the third non-matching segment, but no insertions. In order to compute conditional phoneme probabilities for computing the FWNP and FWSWP, confusion matrices were calculated for C_1 , V , and C_2 by collapsing the nonsense syllable responses across all participants. The cells of each confusion matrix were used to calculate the conditional probabilities needed to compute FWNP and FWSWP for each target word.

The median FWNP and FWSWP values were used as cutoffs to divide the target words into high and low FWNP groups and high and low FWSWP groups. A j -factor was calculated for each participant for the high and low groups of both FWNP and FWSWP; the average j -factors are in (3).

A paired comparison of the mean j -factors low and high FWNP words indicates that the difference is significant ($t(41)=8.691$, $p<0.00001$). The magnitude of the effect is nearly as large as that between words and nonsense syllables, confirming the expectation that targets with sparse phonetic neighborhoods ($j=2.11$) have a large contextual advantage over targets with dense phonetic neighborhoods ($j=2.61$). Data for individuals for words with high and low values of FWNP, and average j -factors are plotted in (5).

(5)



The effect of FWSWP is nonsignificant ($t(37)=1.050$, $p=0.3004$), with $j=2.46$ for high FWSWP words and $j=2.40$ for low FWSWP words. This nonresult is consistent with the phoneme probability overwhelming the frequency weighting, resulting in accuracy differences but not in j -factor differences.

The contextual advantage afforded to words with sparse phonetic neighborhoods can be further investigated by subdividing the neighborhood of a given target into those neighbors sharing C_1V with the target, those sharing VC_2 , and those sharing C_1C_2 . The FWNP can be calculated for these three different neighborhoods to divide the target words into high and low FWNP groups for j -factor analyses. The results of these analyses are in (6). The C_1V neighborhood shows a significant difference ($t(39)=6.507$, $p<0.00001$) between low ($j=2.15$) and high ($j=2.52$) FWNP targets in favor of words with low density neighborhoods. The difference in j -factor for the C_1C_2 neighborhood is slight and in the opposite direction as might be expected (low density, $j=2.41$; high density $j=2.29$) but significant ($t(39)=2.386$, $p=0.0220$). However, the difference for the C_1C_2 neighborhood is *not* significant if a familywise $\alpha=0.05$ error rate is maintained using a Bonferroni criterion for three tests (the cutoff for familywise $\alpha=0.05$ for three tests is $t=2.3954$; Hays 1994, p.1007). The difference for VC_2 neighborhoods is not significant ($t(41)=0.004$, $p=0.9966$). The vast majority of the difference in the j -factors for low and high FWNP words seems to arise from C_1V neighborhood structure.

Effects of signal independent factors

(6)

Neighborhood *j*-factor analysis. The following values are reported for each type of neighborhood: the average number of neighbors (with S.D.), the average *j*-factors with 95% C.I. for low FWNP (low density neighborhood), high FWNP words (high density neighborhood) words.

Neighborhood type	Mean number of neighbors	<i>j</i> -factor for low FWNP words	<i>j</i> -factor for high FWNP words
All neighbors	20.8 (4.8)	2.11 ± 0.09	2.61 ± 0.11
C ₁ V neighbors	7.1 (2.4)	2.15 ± 0.11	2.52 ± 0.10
VC ₂ neighbors	5.7 (2.2)	2.35 ± 0.08	2.34 ± 0.08
C ₁ C ₂ neighbors	8.0 (3.5)	2.41 ± 0.08	2.29 ± 0.11

It is unlikely that the neighborhood analysis results are because of an excessive number of C₁V neighbors relative to the other two types of neighbors. Column 1 of (6) shows the mean number of neighbors per target. The average 20.8 neighbors per target are roughly equally divided between the three types of neighbors.

5. Discussion

The interpretation of the result $j=n$, as was found for nonsense CVC syllables with $n=3$ segments, is consistent with the hypothesis that the constituent segments of syllables are perceived independently. But what does the result $j<n$ mean? Boothroyd and Nittrouer suggest it is a measure of the reduction of independent perceptual units. Words are perceived as consisting of $j=2.35$ independent units, with each phoneme consisting of about 0.78 units.

Nearey (1998), proceeding from a computational simulation of Boothroyd and Nittrouer's results, suggests that small reductions (around 1 or less for $n=3$) in the *j*-factor could arise from a bias in favor of particular items in a Luce choice model. Results of $2<j\leq n$ are consistent with independent perception of n segments, and reduction of *j* below n quantifies the amount of bias involved for those items. The present result for word frequency, that high frequency words have lower *j*-factors than low frequency words, are entirely consistent with this interpretation and would support bias accounts of word frequency effects in word recognition. However, it is important to note that the *j*-factors reported here are averages over groups of words. Under certain situations of high context, it is possible that a gestalt model of word recognition, in which words are perceived as wholes, would be a more appropriate interpretation for results of $j=1$.

The neighborhood density results suggest that the implementation of bias must be understood with reference to temporal distribution of the acoustic-phonetic information. Recall that as predicted, words with sparse neighborhoods had a lower *j*-factor than words with dense neighborhoods, consistent with a bias favoring words from sparse neighborhoods. Importantly, this result largely holds true for neighborhoods defined by CV neighbors but not VC neighbors.

Correct perception of the beginning of a word in a sparse phonetic neighborhood delimits a small set of potential candidates. A listener can then focus attention on just those phonetic features that distinguish the members of this small set to achieve correct recognition despite reduced acoustic-phonetic information present at the end of the syllable. This account of how listeners use bias, based on a dynamic analysis of the effects of neighborhood density, offers support for the dynamic aspects (but perhaps not the strict autonomy) of the cohort theory of word recognition (Marslen-Wilson, 1989).

The lack of any significant *j*-factor effect for VC neighborhood indicates that in open-response identification, contextual information from correct recognition of syllable-final material is not used in order to reevaluate or sharpen the perception of earlier-occurring syllable-initial material in the same way that contextual knowledge guides perception of upcoming material (but cf. Salasoo & Pisoni, 1985). What is measured by the *j*-factor for words in sparse C_1V neighborhoods does not just narrow the set of possible word candidates so that guesses can be more effective, but seems to have genuine perceptual effects. If the role of bias were merely to narrow such a set, along the lines of what Broadbent (1967) calls the sophisticated guessing model, then one would expect a significant reduction of the *j*-factor for words in sparse VC_2 neighborhoods as well.

This asymmetric effect is consistent with claims for more robust acoustic-phonetic information in the speech signal for onsets than for codas (Wright, 2001), which suggests some active compensatory strategy on the part of listeners, since syllable-final consonants must be correctly identified in languages that have them, such as English. If the strategy takes place according to the account outlined above, with listeners focusing attention whenever expected neighborhood density permits, then recognition rates for codas should be low when expected neighborhood is dense, and high when the expected neighborhood is sparse.

A preliminary comparison of the average recognition rates for C_1 and C_2 is consistent with these predictions. Nonsense syllables, whose segments are highly unpredictable, show $p(C_1)-p(C_2)=0.2124$, providing a baseline for the advantage of onsets over codas. Words, with $p(C_1)-p(C_2)=0.0798$, show lower differences than the nonsense syllable difference. The recognition rates for C_1 and C_2 of low C_1V FWNP words—sparse expected neighborhood—are nearly equal at $p(C_1)-p(C_2)=0.0226$, while the comparable rate for high FWNP words—dense expected neighborhood—is $p(C_1)-p(C_2)=0.1704$.

Frequency of usage and neighborhood density effects could be explained here in terms of a criterion bias shift, supporting feedforward models of top-down effects in word recognition such as Merge (Norris, McQueen, and Cutler, 2000) or FLMP (Massaro, 1998) over feedback models such as TRACE (McClelland and Elman, 1986). However, given the dynamic nature of the neighborhood density effect as reported here, the operationalization of bias appears to narrow the difference between feedforward and feedback models. The current findings indicate that bias appears to improve the efficiency of perception of phonological structure in the coda when acoustic-phonetic information is impoverished. This

interpretation of bias as measured by the j-factor may represent attentional priming effects (Grossberg & Stone, 1986; see also Grossberg, 2000; Luce, Vitevitch and Goldinger, 2000).

6. Conclusion

Support has been provided for Boothroyd and Nittrouer's j-factor as a robust and replicable measure of the effects of context in human speech perception. The j-factor represents the number of perceptually independent parts within a whole, and can be interpreted as a bias in favor of words over nonsense syllables, words with higher usage frequencies over words with lower usage frequencies, or of words from sparse neighborhoods over words from dense neighborhoods. The neighborhood density effect is dynamic, such that the neighborhood is primarily determined by the first two segments of a CVC word. This dynamic effect appears to improve perception of codas of CVC words in sparse CV neighborhoods.

Future work will use the j-factor model to investigate context effects in other types of stimuli besides English CVC monosyllables, such as longer words, and other languages with more different syllable structure. Investigation of the model's assumptions is also planned, such as approximately equal recognition rates of all phonemes, and whether segments (as opposed to syllables or features) are the proper units of analysis.

References

- Allen, J. 1994. How do humans process and recognize speech? *IEEE Transactions on Speech and Audio Processing* **2**: 567-577.
- 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.
- Broadbent, D. E. 1967. Word-frequency effect and response bias. *Psychological Review* **74**: 1-15.
- Fletcher, H. 1953. *Speech and hearing in communication*. New York, NY: Van Nostrand.
- Goldinger, S. D. 1997. Words and voices: Perception and production in an episodic lexicon. In K. Johnson & J. W. Mullennix (eds.) *Talker variability in speech processing*. San Diego, CA: Academic Press.
- Grossberg, S. & Stone, G. 1986. Neural dynamics of word recognition and recall: Attentional priming, learning, and resonance. *Psychological Review* **93**: 46-74.
- Grossberg, S. 2000. Brain feedback and adaptive resonance in speech perception. *Behavioral and Brain Sciences* **23**: 332-333.
- Hays, W. L. 1995. *Statistics*. 5th ed. Fort Worth, TX: Harcourt Brace.
- Johnson, K. 1997. Speech perception without normalization: An exemplar model. In K. Johnson & J. W. Mullennix (eds.), *Talker variability in speech processing*. San Diego, CA: Academic Press.

José R. Benkí

- Kucera, F., & Francis, W. 1967. *Computational analysis of present day American English*. Providence, RI: Brown University Press.
- Luce, P. A., & Pisoni, D. B. 1998. Recognizing spoken words: The neighborhood activation model. *Ear and Hearing* **19**: 1-36.
- Luce, P. A., Goldinger, S. D., & Vitevitch, M. S. 2000. It's good...but is it ART? *Behavioral and Brain Sciences* **23**: 336.
- Marslen-Wilson, W. 1989. Access and integration: projecting sound onto meaning. In W. Marslen-Wilson (ed.) *Lexical representation and process* Cambridge, MA: MIT Press.
- Massaro, D. W. 1998. *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge, MA: MIT Press.
- McClelland, J. L., & Elman, J. L. 1986. The TRACE model of speech perception. *Cognitive Psychology* **18**: 1-86.
- Nearey, T. 1998. On the factorability of phonological units in speech perception. Paper presented at the Sixth Conference on Laboratory Phonology, University of York, July 1998.
- Norris, D. G. 1986. Word recognition: Context effects without priming. *Cognition* **22**: 93-136.
- Norris, D. G., McQueen, J. M., & Cutler, A. 2000. Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences* **23**: 299-370.
- 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: Speech Research Laboratory, Psychology Department, Indiana University.
- Salasoo, A., & Pisoni, D. B. 1985. Interaction of knowledge sources in spoken word recognition. *Journal of Memory and Language* **24**: 210-231.
- Savin, H. 1963. Word-frequency effects and errors in the perception of speech. *Journal of the Acoustical Society of America* **35**: 200-206.
- Schroeder, M. R. 1968. Reference signal for signal quality studies. *Journal of the Acoustical Society of America* **44**: 1735-1736.
- Webster's Seventh Collegiate Dictionary*. 1967. Los Angeles, CA: Library Reproduction Service.
- Wright, R. 2001. Perceptual cues in contrast maintenance. In E. Hume & K. Johnson (eds.) *The role of speech perception in phonology* San Diego, CA: Academic Press. pp.251-277.

José R. Benkí
Department of Linguistics
1076 Frieze Bldg.
University of Michigan
Ann Arbor, 48109-1285

benki@umich.edu