650

Detectability of words and nonwords in two kinds of noise

Bruno H. Repp and Ram Frost
Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511-6695

(Received 1 March 1988; accepted for publication 11 July 1988)

Recent models of speech perception emphasize the possibility of interactions among different processing levels. There is evidence that the lexical status of an utterance (i.e., whether it is a meaningful word or not) may influence earlier stages of perceptual analysis. To test how far down such "top-down" influences might penetrate, an investigation was conducted to determine whether there is a difference in detectability of words and nonwords masked by amplitude-modulated or unmodulated broadband noise. The results were negative, suggesting either that the stages of perceptual analysis engaged in the detection task are impermeable to lexical top-down effects, or that the lexical level was not sufficiently activated to have any facilitative effect on perception.

PACS numbers: 43.71.Es, 43.71.An

INTRODUCTION

In experimental psychology, it is now common to conceptualize perceptual and cognitive processes in terms of parallel interactive networks. These models have also been applied to speech perception (e.g., McClelland and Elman, 1986; Samuel, 1986). An unconstrained version of such a model predicts interactions among all processing levels in the system. The task of experimental psychologists thus is to determine whether specific kinds of interactions do occur, and if they do not, to introduce constraints into the model. Thus, for example, there is evidence that influences of semantic context do not penetrate to earlier levels of analysis in speech perception (Samuel, 1981; Connine and Clifton, 1987). On the other hand, studies of phenomena such as phoneme restoration (Samuel, 1981) and the category boundary shift along word–nonword continua (Connine and Clifton, 1987) have suggested that the lexical status of an utterance (i.e., whether it is a meaningful word or not) may affect the extraction of phoneme information from an impoverished or ambiguous stimulus. It is not known, however, whether even earlier levels of perceptual analysis, such as the extraction of phonetic features from speech, can be affected by lexical status.

In this study, therefore, we asked whether lexical status can affect the *detectability* of speech in noise.[1] A positive finding would provide the strongest possible evidence for lexical "top-down" effects on speech perception, because a detection task taps into the earliest levels of perceptual processing. On the other hand, a demonstration that these early levels are immune to lexical influences would place a further constraint on interactive models of speech perception, and thereby would help reveal the internal architecture of the speech perception system.

The level of perceptual analysis employed in a speech detection task may depend on the kind of masking noise employed. In unmodulated (UM) white noise, any kind of auditory evidence can be used for detection of speech, so that very little information enters the perceptual system from near-threshold stimuli. In amplitude-modulated (AM) noise, on the other hand, if it has the same amplitude envelope as the speech to be detected (cf. Horii *et al.*, 1971), subjects must detect spectral evidence of speech against a background of appropriate speech envelope features (cf. Van Tasell *et al.*, 1987; Frost *et al.*, 1988). The spectral features in combination with the envelope features may provide sufficient information for activation of a narrow range of lexical candidates, perhaps not enough for accurate identification, but enough to generate a facilitative top-down flow to earlier perceptual stages. In the present experiment, therefore, we compared the detectability of structurally similar words and nonwords in either AM or UM noise.

I. METHODS

A. Subjects

Twelve paid undergraduate subjects participated. All were native speakers of English and reported having normal hearing.

B. Stimuli

We used the same words and nonwords as Frost *et al.* (1988). Each set comprised 24 disyllabic stimuli beginning with a stop consonant, each containing from four to six phonemes and being stressed on the first syllable. The word frequencies ranged from 0–438, with a median of 60, according to Kucera and Francis (1967). The nonwords were genera-

 0001-4966/88/111929-04$00.80

ted from a different set of disyllabic words by changing one or two phonemes, without violating the phonotactic rules of English. All stimuli were spoken by a woman and were recorded in a sound-insulated booth. The utterances were digitized at a 20-kHz sampling rate and low-pass filtered at 9.6 kHz.

An AM noise masker was generated for each individual utterance by reversing the polarity of sampling points with a probability of 0.5 (see Schroeder, 1968). Each masking noise thus had exactly the same amplitude envelope and overall level as its speech mate, but it had no spectral structure.[2] A UM masking noise was obtained for each individual stimulus by excerpting a segment of exactly the same duration from a longer file of white noise (sampled from a General Radio 1390-A random noise generator), applying 5-ms amplitude ramps to avoid abrupt onsets and offsets, and attenuating the noise until its average level matched that of the speech.

Speech-plus-noise stimuli were obtained by digitally adding the waveforms of a speech stimulus and each of its two matched noise maskers. Each masking noise thus began and ended with the speech. In the waveform-adding procedure, weights were applied to both digitized files to vary the signal-to-noise (S/N) ratio while keeping the overall level of the added stimulus constant. Thus the level of the speech decreased as that of the noise increased. Five S/N ratios, spaced 2 dB apart, were chosen on the basis of pilot data for each noise condition. They ranged from -18 to -10 dB for the modulated noise, and from -32 to -24 dB for the UM noise.[3] Since these ratios were all negative, changes in S/N ratio entailed primarily changes in the absolute level of the speech signal, and only negligible changes in noise level.

Each of the two stimulus sets, words and nonwords, was divided randomly into two sets of 12, which were then combined to form two parallel stimulus sets, each containing 12 words and 12 nonwords. Each of these two sets was presented in each of the two noise conditions. The corresponding four stimulus sequences were recorded on audiotapes. Each of the tapes contained 36 familiarization trials representing S/N ratios of increasing difficulty, followed by five blocks of 24 experimental trials. Each of the blocks contained one instance of each stimulus, at a fixed S/N ratio. Successive experimental blocks repeated the same stimuli in a different random order and simultaneously increased the S/N ratio by 2 dB. The task thus started with the most difficult condition and became progressively easier. The reason for this was to reduce practice effects, since the stimuli were the same in each block; however, there was only a slim chance of actually identifying any word or nonword at even the most favorable S/N ratio.

C. Procedure

Each subject listened to one AM noise tape and one UM noise tape containing different speech stimuli. Half the subjects listened to one pair of tapes and half to the other. Half the subjects received one noise condition first, and half the other. All subjects listened binaurally in a quiet room over TDH-39 earphones. The instructions emphasized that detection, not identification, of the speech was required. The playback level was calibrated on a voltmeter using white noise recorded at the beginning of the tape. The average levels of individual stimuli at the subjects' earphones ranged from 83–88 dB SPL.

II. RESULTS

The results are shown in Fig. 1. Performance (percent correct detections) increased steadily as S/N ratio increased, from near chance to between 85 and 90 percent correct in each noise condition. There was no difference between words and nonwords in either AM or UM noise. The slopes of the detectability functions were also similar in the two noise conditions.

Because the speechlike features of the AM noise may have caused a response bias to say "yes" (cf. Frost *et al.*, 1988), the results were also analyzed in terms of separate indices of discriminability (sensitivity) and bias, following the procedures of Luce (1963). The discriminability index d assumes values in the same general range as the d' of signal detection theory, with zero representing chance performance. The bias index b assumes positive values for a tendency to say "yes" and negative values for a tendency to say "no."[4]

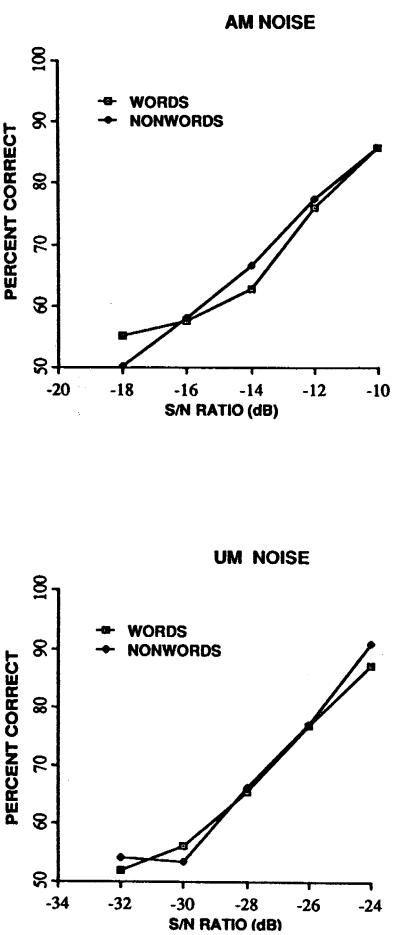

FIG. 1. Percent correct detections as a function of S/N ratio for words and nonwords in two noise conditions.

TABLE I. Discriminability and bias indices as a function of lexical status, noise type, and S/N ratio.

	S/N ratio (dB)									
	AM noise					UM noise				
	−18	−16	−14	−12	−10	−32	−30	−28	−26	−24
Discriminability										
Words	0.26	0.33	0.56	1.46	2.01	0.12	0.28	0.71	1.43	2.05
Nonwords	−0.03	0.38	0.76	1.41	1.96	0.19	0.18	0.78	1.49	2.40
Bias										
Words	−0.40	−0.03	−0.08	−0.15	0.30	−0.52	−0.31	−0.14	−0.13	−0.11
Nonwords	−0.38	0.09	−0.03	0.34	0.20	−0.27	−0.04	−0.27	0.16	−0.04

The indices, averaged over subjects, are shown in Table I. The *d* indices confirm that performance was similar for words and nonwords, in both noise conditions. In a repeated-measures analysis of variance on these indices (with the factors noise type, S/N ratio, and lexical status), no effect except the obvious one of S/N ratio [$F(4,44) = 86.11$, $p < 0.001$] even remotely approached significance.

The *b* indices show that subjects had a tendency to say "no" at the lowest S/N ratio, but as soon as performance rose above chance, there was no clear response bias in either direction. There were no differences in bias between words and nonwords or between noise types. This was confirmed in an analysis of variance on the *b* indices, which revealed only a significant effect of S/N ratio, $F(4,44) = 5.95$, $p = 0.006$.

III. DISCUSSION

Our results show that structurally similar words and nonwords are equally detectable in noise, even when the noise provides a background of appropriate speech envelope features. Thus it appears that lexical influences did not penetrate to the levels of perceptual analysis required for the detection task. We conclude tentatively that the earliest stages of speech analysis are not permeable to lexical top-down effects, and that this needs to be taken into account in interactive models of speech perception.

Our conclusion is tentative because there are two other possible reasons for our negative results. First, stimulus information may have been too limited to lead to lexical activation that was sufficiently constrained to have a facilitative top-down effect. This was especially true in the UM noise condition, although in the AM noise condition, because of the combination of spectral and envelope features, there was a good chance of more focused lexical activation, especially at the higher S/N ratios. Even so, there was considerable uncertainty about the possible lexical choices, which contrasts with other tasks in which lexical top-down effects have been observed (Samuel, 1981; Connine and Clifton, 1987).

The other possible reason for our negative findings is that subjects in the detection task may have employed a purely auditory strategy. Even in AM noise, detection of the voice fundamental, for example, may have been sufficient. The subjects were not forced to detect phonetic features or phonemes as such. It is, of course, a matter of theoretical viewpoint whether or not early auditory analysis is considered to be part of the speech perception system. If the lowest level units are assumed to be phonetic features (as in the TRACE model of McClelland and Elman, 1986), then it may be argued that lexical top-down effects do not occur in a detection task because the speech perception system is not engaged.

ACKNOWLEDGMENT

This research was supported by NICHD Grant HD-01994 to Haskins Laboratories.

[1]It is well known that familiar words are *identified* better than unfamiliar words in noise (e.g., Howes, 1957). Although this "word frequency effect" may reflect an influence of lexical representations on early perceptual processes, an interpretation in terms of response bias has generally been considered sufficient (see, e.g., Broadbent, 1971).

[2]After the random sample-reversal procedure, the noise had a flat spectrum, like white noise. However, because it had to be converted into sound at the same time as the speech, which had been input with high-frequency preemphasis, the noise underwent high-frequency de-emphasis at the output stage. Thus the AM noise, in contrast to the UM noise, actually had a sloping spectrum (about 6 dB/oct above 1000 Hz, and less below).

[3]The S/N ratios for the two types of noise were not directly comparable because the former were exact (i.e., they remained constant with changes in speech level within a stimulus), while the latter related to the average speech level only (i.e., they varied with local changes in speech level). They would have been more similar, had the UM S/N ratios been specified with respect to the peak level of the speech. (See Horii *et al.*, 1971, for discussion of these issues.)

[4]Given a total of 12 responses per subject and condition, values of 0.5 and 11.5 were substituted for response frequencies of 0 and 12, respectively, so as to obtain finite *d* and *b* indices. See Wood (1976) for an earlier application of these indices in a speech perception task. The original names of *d* and *b* are $-1 - \ln \eta$ and $\ln b$, respectively (Luce, 1963).

Broadbent, D. E. (**1971**). *Decision and Stress* (Academic, London).

Connine, C. M., and Clifton, C., Jr. (**1987**). "Interactive use of lexical information in speech perception," J. Exp. Psychol.: HPP **13**, 291–299.

Frost, R., Repp, B. H., and Katz, L. (**1988**). "Can speech perception be influenced by simultaneous presentation of print?," J. Mem. Lang. (in press).

Horii, Y., House, A. S., and Hughes, G. W. (**1971**). "A masking noise with speech-envelope characteristics for studying intelligibility," J. Acoust. Soc. Am. **49**, 1849–1856.

Howes, D. (**1957**). "On the relation between the intelligibility and frequency of occurrence of English words," J. Acoust. Soc. Am. **29**, 296–305.

Kucera, H., and Francis, W. N. (**1967**). *Computational Analysis of Present-Day American English* (Brown U. P., Providence, RI).

Luce, R. D. (**1963**). "Detection and recognition," in *Handbook of Mathematical Psychology*, edited by R. D. Luce, R. R. Bush, and E. Galanter (Wiley, New York).

McClelland, J. L., and Elman, J. L. (**1986**). "The TRACE model of speech perception," Cog. Psychol. **18**, 1–86.

Samuel, A. G. (**1981**). "Phonemic restoration: Insights from a new methodology," J. Exp. Psychol.: General **110**, 474–494.

Samuel, A. G. (**1986**). "The role of the lexicon in speech perception," in *Pattern Recognition by Humans and Machines, Vol. 1*, edited by E. C. Schwab and H. C. Nusbaum (Academic, New York), pp. 89–111.

Schroeder, M. R. (**1968**). "Reference signal for signal quality studies," J. Acoust. Soc. Am. **43**, 1735–1736.

Van Tasell, D. J., Soli, S. D., Kirby, V. M., and Widin, G. P. (**1987**). "Speech waveform envelope cues for consonant recognition," J. Acoust. Soc. Am. **82**, 1152–1161.

Wood, C. C. (**1976**). "Discriminability, response bias, and phoneme categories in discrimination of voice onset time," J. Acoust. Soc. Am. **60**, 1381–1389.