Discrimination of intensity differences on formant-like transitions

RICHARD E. PASTORE, WILLIAM A. AHROON, JR., JAMES P. WOLZ,
JOSEPH S. PULEO, and ROBERT S. BERGER

State University of New York at Binghamton, Binghamton, New York 13901

Dorman (1974) studied the discrimination of intensity differences on formant transitions in and out of syllable context. He interpreted his results as suggesting that the acoustic features of his stop-consonant/vowel syllable were recoded into a phonetic representation, then stored in an inaccessible form of auditory short-term memory. The Dorman results are replicated with analogous pure-tone and FM-glide conditions. The results of both studies are explained in terms of specified acoustic properties of the signals and thus provide no evidence for a special phonetic recoding.

Dorman (1974) examined the discriminability of intensity differences in the formant transitions of a stop-consonant/vowel syllable relative to the discriminability of comparable portions of two control stimuli. This study found essentially chance performance with the stop-consonant and nearly perfect performance with a steady-state vowel and a nonspeech pitch glide. Dorman argued that the acoustic factors of simple discriminability and backward masking are present in one or both control conditions and therefore cannot account for the marked performance decrement with the speech stimuli. He conjectured that the acoustic properties of the stop-consonants are recoded into a phonetic representation which is stored in a relatively inaccessible form of short-term auditory memory.

The Dorman study was based on earlier studies on "categorical perception," an apparently universal (Mattingly, 1972) and unique property of the perception of speech (Mattingly, Liberman, Syrdal, & Halwes, 1971; Studdert-Kennedy, Liberman, Harris, & Cooper, 1970). In a recent study, Miller, Pastore, Wier, Kelley, and Dooling (1974) reported categorical perception of nonspeech noise-buzz sequences. According to these authors, earlier failures to find categorical perception in similar noise-buzz sequences could be attributed to a number of factors, including not asking the questions properly. The purpose of the present study was to examine more closely possible acoustic explanations for the difference noted by Dorman between his stop-consonant and two control stimuli.

The three computer-generated stimuli employed by

Dorman (1974) are represented in Figure 1a. The stimulus labeled D1 is the stop-consonant/vowel syllable /bae/. D2 and D3 are the control stimuli employed by Dorman. The five stimuli represented in Figure 1b are simple, pure-tone analogues to the three stimuli employed by Dorman. The present study used a yes/no task in addition to the same/different task employed by Dorman. In both studies, the task of the observer was to discriminate the presence or absence of a 7.5-dB¹ decrement in the initial 60-msec portion of the comparison stimulus (labeled A in Figure 1). The A portion of the standard stimulus and the B portion of both the standard and comparison stimuli were 72 dB SPL.

There are a number of factors involved in the various tasks. Table 1 is a summary of major factors present in each stimulus. All stimulus conditions except D3 and T3 include some form of backward

Figure 1. (a) Three stimuli employed by Dorman (1974). (b) Five stimuli employed in the present study. The stimuli can be grouped into three conceptual classes: (D1, T1, G1), (D2, T2), and (D3, T3, G3). The observers' task was to discriminate between a 7.5-dB and a 0-dB decrement in the intensity of the A portion of each stimulus relative to the intensity of the B portion.

The authors wish to thank Charlotte S. MacLachy for her helpful comments and criticisms in the preparation of this manuscript. They also thank Ms. MacLachy and Ms. Judith K. Ahroon for their assistance in the conduct of this study. Send requests for reprints to: Richard E. Pastore, Psychoacoustics Laboratory, Department of Psychology, State University of New York at Binghamton, Binghamton, New York 13901.

Table 1
Factors Involved in Tasks

Condition		Back-ward Masking	Reference		Phonetic Properties
Stimulus	Task		External	Internal	
D1	S/D	X	X	.	X
T1	S/D	X	X	.	.
G1	S/D	X	X	.	.
D2	S/D	X	X	X	.
T2	S/D	X	X	X	.
D3	S/D	.	X	.	.
T3	S/D	.	X	.	.
G3	S/D	.	X	.	.
T1	Y/N	X	.	X	.
T2	Y/N	X	.	.	.

masking (Dorman, 1974). However, the A and B portions of the stimulus in conditions D2 and T2 (see Figure 1) are identical in spectral composition. Therefore, conditions D2 and T2 represent pedestal detection experiments in which the observer must detect a change in the intensity of a portion of an otherwise steady-state stimulus. Tanner (1958; also Rabb, Osman, & Rich, 1963) has shown that performance in pedestal detection tasks is typically superior to performance in simple detection tasks. Since the pedestal serves as a reference which is internal to the stimulus waveform, the "standard" in the same/different task, at best, is redundant and performance should be similar in the yes/no and same/different tasks with the T2 stimulus. Only D1 and D2 are phonetic. If the discrimination results with the tonal stimuli described in Figure 1b are equivalent to those analogous stimuli in the Dorman study (Figure 1a), then parsimony would demand that Dorman's results be attributed to the simple acoustic properties of the stimuli.

METHOD

Subjects

Seven adults with clinically normal hearing, six with previous experience as psychoacoustic listeners, served as observers in the study. The observers typically ran in pairs, with each observer in a separate double-wall, double-floor commercial, sound-insulated chamber. All seven observers ran under the five pure-tone conditions. Three of these observers ran under the two FM-glide conditions. Each observer ran three 50-trial blocks per condition. The order in which the conditions were run varied across observers.

Procedure

The five stimulus conditions in the present study are summarized in Figure 1b. Each stimulus was composed of two portions, labeled A and B in Figure 1. As with the stimuli employed by Dorman, the A and B portions were 60 and 240 msec in duration, respectively. In both the present study and the Dorman study, the B portion of the first and second class of stimuli [(D1, T1, G1); (D2, T2)] was 72 dBA, and was absent in the third class (D3, T3, G3). In the present study, the B portion of the stimuli was a 1,500-Hz pure tone, while in the Dorman study it was the first three formants of the synthetic vowel syllable /ae/. For the A portion, we used a 1,400-Hz tone (T1), a

1,500-Hz tone (T2, T3), or a 60-msec uniform glide beginning at 1,300 Hz and ending at 1,500 Hz (G1, G3). All signals were gated on and off with essentially instantaneous rise and fall times at random phase.²

With all five conditions, a same/different procedure was employed in which the observers' task was to detect the presence or absence of a 7.5-dB decrement in A relative to B. With equal a priori probability, A was thus either 72 or 64.5 dBA. The temporal parameters of the trials were identical to those employed by Dorman, and the standard for the task was always the 0-dB decrement (A equal to 72 dBA) condition. The T1 and T2 stimuli also were run with a yes/no procedure which was identical to the same/different procedure without the standard.

RESULTS AND DISCUSSION

The mean value of percent correct for each of the seven conditions in the present study are plotted in Figure 2. Also plotted are percent correct calculated from the reported probabilities of hits and false alarms for the three conditions run by Dorman.³ The data for the conditions within each of the three separate classes of stimuli are equal within the tolerance of measurement error. The detectability of the initial portion of both the second (D2, T2) and third (D3, T3, G3) classes of stimuli is almost perfect, while it is near chance in the first class of stimuli (D1, T1, G1). The differences in performance between the T1 stimulus and the T2 and T3 stimuli is statistically significant ($t = 19.44$; $p \ll .005$).

The nearly perfect performance with the third class of stimuli would indicate that the isolated stimulus decrement is highly detectable. Since the first and second classes of stimuli follow the general stimulus paradigm employed in backward-masking experiments (cf. Elliott, 1971), backward masking should be considered a primary candidate to explain the poor performance with the first class of stimuli. Dorman suggested and attempted to test this hypothesis. He

Figure 2. The mean percent correct for each stimulus in the present study and the Dorman (1974) study. One standard deviation is plotted on the upper side of each mean for the present study. The Dorman data are based on one 32-trial block per observer. The data in the present study are based on three 50-trial blocks per observer.

assumed that the same masking should be present with the D1 and D2 stimuli and rejected this masking hypothesis when he found nearly perfect performance with this second class of stimuli. He then conjectured that the poor performance with the D1 stimulus might be due to the phonetic properties of these stimuli which result in the activation of a special speech processor. The replication of the Dorman results with analogous pure-tone stimuli (see Figures 1 and 2) would indicate that the poor performance with the first class of stimuli probably is due to some common acoustic property rather than any special phonetic property (see Table 1).

We hypothesized that the second class of stimuli changed the task from a simple backward-masking detection task to a pedestal detection task. Tanner (1958) demonstrated that pedestal detection tasks result in higher observer efficiency, which he hypothesized as being due to the pedestal serving as a standard internal to the stimulus waveform. If there is such an internal standard, then the external standard would be redundant. The replication of the results for the T2 stimulus with the yes/no task indicate that the standard in the T2 same/different task indeed is unnecessary, and would tend to support the above hypotheses. Since the internal standard provided by the B segment of the second class of stimuli would counteract any backward-masking effects of this portion of the stimulus, this second class of stimuli cannot be used to evaluate the importance of backward masking in the first class of stimuli. Therefore, based on the data in the present study and in the Dorman study, it is not possible to reject backward masking as the primary cause of the poor performance with stop-consonant/vowel syllables relative to performance with pitch glides or steady-state vowels.

The lack of support for a special speech factor in the Dorman study as evaluated in the context of the present study indicates that the speech properties of the Dorman D1 stimulus were not a significant factor in his study. These results are not a direct test of the

general notion of a special process or series of processes concerned only with the perception of speech. They have a bearing on the validity of theories which propose such a process only in the general context of the many studies which compare the processing of auditory speech and nonspeech signals.

REFERENCES

- DORMAN, M. F. Discrimination of intensity differences on formant transitions in and out of syllable context. *Perception & Psychophysics*, 1974, **16**, 84-86.
- ELLIOTT, L. L. Backward and forward masking. *Audiology*, 1971, **10**, 65-76.
- MATTINGLY, I. G. Speech cues and sign stimuli. *American Scientist*, 1972, **60**, 327-337.
- MATTINGLY, I. G., LIBERMAN, A. M., SYRDAL, A. K., & HALWES, T. Discrimination in speech and nonspeech modes. *Cognitive Psychology*, 1971, **2**, 131-157.
- MILLER, J. D., PASTORE, R. E., WIER, C. C., KELLY, W. J., & DOOLING, R. J. Discrimination and labeling of noise-buzz sequences with varying noise-lead times. *Journal of the Acoustical Society of America*, 1974, **55**, 390(A).
- RABB, D. H., OSMAN, E., & RICH, E. Effects of waveform correlation and signal duration on detection of noise bursts in continuous noise. *Journal of the Acoustical Society of America*, 1963, **35**, 1942-1946.
- STUDDERT-KENNEDY, M., LIBERMAN, A. M., HARRIS, K. S., & COOPER, F. S. Motor theory of speech perception: A reply to Lane's critical review. *Psychological Review*, 1970, **77**, 234-249.
- TANNER, W. P., JR. What is masking? *Journal of the Acoustical Society of America*, 1958, **30**, 919-921.

NOTES

1. Dorman employed both a 7.5 and a 9.0 decrement in the A portion of his stimuli relative to the B portion.
2. Two observers replicated the T1 and T2 same/different experiments, using signals gated through Grason-Stadler 1287B electronic switches with 2.5-, 5-, and 10-msec rise/fall times. The results were equivalent to those reported for the signals with instantaneous rise/fall times.
3. The probabilities of hit and false alarm reported by Dorman yield a d' value of 3.22, which is at variance with his reported value of 3.72 for the D1 stimulus.

(Received for publication December 20, 1974;
revision received March 10, 1975.)