

Discrimination of intensity differences carried on formant transitions varying in extent and duration

JAMES E. CUTTING

*Wesleyan University, Middletown, Connecticut 06437
and Haskins Laboratories, New Haven, Connecticut 06511*

and

MICHAEL F. DORMAN

*Arizona State University, Tempe, Arizona 85281
and Haskins Laboratories, New Haven, Connecticut 06511*

Dorman (1974) found that small-intensity differences carried on the initial portions of consonant-vowel syllables were not discriminable. Similar differences carried on steady-state vowels and on isolated formant transitions, however, were readily discriminable. He interpreted the difference between the first and latter conditions as a phonetic effect. Yet Pastore, Ahroon, Wolz, Puleo, and Berger (1975) found similar results using sine-wave analogs to Dorman's stimuli. They concluded that the effect is not phonetic, and that it is attributable to simple backward masking. The present studies observed the discriminability of intensity differences carried on formant transitions varying in extent and duration. Results support the conclusion of Pastore et al. to the extent that the effect is clearly not phonetic. However, these results and others suggest that simple peripheral backward masking is not a likely cause; instead, recognition masking may be involved. Moreover, the finding that phonetic-like processes occur elsewhere in audition does not necessarily impugner the existence of a speech processor; phonemic and phonological processes remain, as yet, unmatched.

Perhaps the most impressive characteristic of speech perception is the efficiency of information reduction (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). The speed and ease of phonemic segmentation is reflected in the rapid transformation of a 40,000-bit/sec acoustic signal into a 40-bit/sec phonetic string (Liberman, Mattingly, & Turvey, 1972), suitable for considerably further savings by conversion into higher-order, meaningful linguistic elements. One empirical manifestation of this process is *categorical perception*, a phenomenon in which phonetic properties of a syllable are rapidly extracted and separated from the acoustic waveform. In a discrimination task, acoustically different stop consonants that are labeled the same are typically perceived to be identical. Stops labeled as different, on the other hand, even though they may differ physically by the same amount, are readily perceived to be dissimilar (Liberman, Harris, Hoffman, & Griffith, 1957; Mattingly, Liberman, Syrdal, & Halwes, 1971; Pisoni, 1973). Thus, for example, acoustic

information about trajectories of formant transitions—information that contributes directly to the phonemic percept—cannot be retrieved readily from sensory memory.

Dorman (1974) found that phonemically irrelevant acoustic information suffers a similar fate. He found that intensity differences carried on formant transitions of consonant-vowel (CV) syllables were largely undetectable. However, the same differences were eminently detectable when carried on steady-state vowels or on formant transitions isolated outside the syllable context. It appears that information-reduction mechanisms relevant for speech do not distinguish between phonemically relevant and irrelevant information at this level. This is as it should be. Liberman et al. (1972, p. 323), for example, suggest that "the distinction between speech and nonspeech is not made at some early stage on the basis of general acoustic characteristics," but rather after many speech-relevant processors have been polled for proper speech-like features. In other words, both phonemically relevant and irrelevant auditory signals share some, probably many, early processing stages. This view is supported by the results of a recent study (Pastore, Ahroon, Wolz, Puleo, & Berger, 1975) which show that intensity differences carried on frequency ramps before steady-state sine waves are as difficult to discriminate as intensity

Supported by a National Institute of Child Health and Human Development grant (HD-01994) to the Haskins Laboratories and a seed grant from Wesleyan University to the first author. We thank R. E. Pastore for his helpful comments. Reprint requests should be sent to J. E. Cutting, Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511.

differences carried on formant transitions of CV syllables.

Dorman's (1974) earlier account of the inability to discriminate intensity differences on formant transitions is incorrect. He noted the similarity between the poor discriminability of intensity differences on formant transitions and poor discriminability of formant frequency within a phoneme category. Both effects were attributed to the uniquely categorical, linguistic processing accorded stop consonants: "After the acoustic cues for stop consonants have been recorded into a *phonetic* [categorical] *representation*, all of the acoustic information is stored in a relatively inaccessible short-term auditory memory" (p. 86, italics added). The effect, however, is not necessarily the result of linguistic coding, since categorical perception occurs in several nonlinguistic domains (Cutting, in press; Cutting & Rosner, 1974; Cutting, Rosner, & Foard, 1976; Miller, Wier, Pastore, Kelly, & Dooling, 1976; see also Lane, 1965, 1967; Locke & Kellar, 1973). Moreover, it does not appear contingent on categorical perception or phonemic processing at all, since the stimuli of Pastore et al. are likely neither to be perceived categorically (see Pisoni, 1971, Experiment 2) nor phonemically (see Cutting, 1974, Experiment 3).

Pastore et al. noted another problem with Dorman's account of his results. They suggested that to change the carrier waveform from a CV syllable to a steady-state vowel syllable, as Dorman did, is to change the task at the same time from one of simple backward-masking detection to one of pedestal detection (see Tanner, 1958; Tanner & Sorkin, 1972). We concur that formal parallels are unmistakable between pedestal detection and the detection of intensity differences carried at the beginning of the vowels. Thus, Dorman's steady-state vowel control does not appear to eliminate simple backward masking as a cause for poor discriminability of intensity differences carried on CV syllables: Pedestal detection experiments appear to be a special kind of masking experiment.

Several important questions about masking arise. First, how might backward masking function in speech perception? For example, if phonemically irrelevant information can be masked at an auditory level, why is it that phonemically relevant information is not masked as well, rendering speech incomprehensible? Second, Pastore et al. do not suggest a particular relationship between backward-masking detection and pedestal detection tasks. For example, do the two tasks differ in degree or in kind: Should we expect intermediate detectability for speech syllables whose transitions are midway between those of a CV and a steady-state vowel? Or should we expect that all syllables with transitions, regardless of their extent or duration, would inhibit detection of intensity differences since only

the steady-state vowel stimulus meets the requisite of having a true pedestal? Experiment I explores the detectability of intensity differences carried on the formant transitions of these intermediate stimuli. The discussion and Experiment II, which follows thereafter, explore the plausibility of simple backward masking vs. backward recognition masking as a cause of our results.

EXPERIMENT I

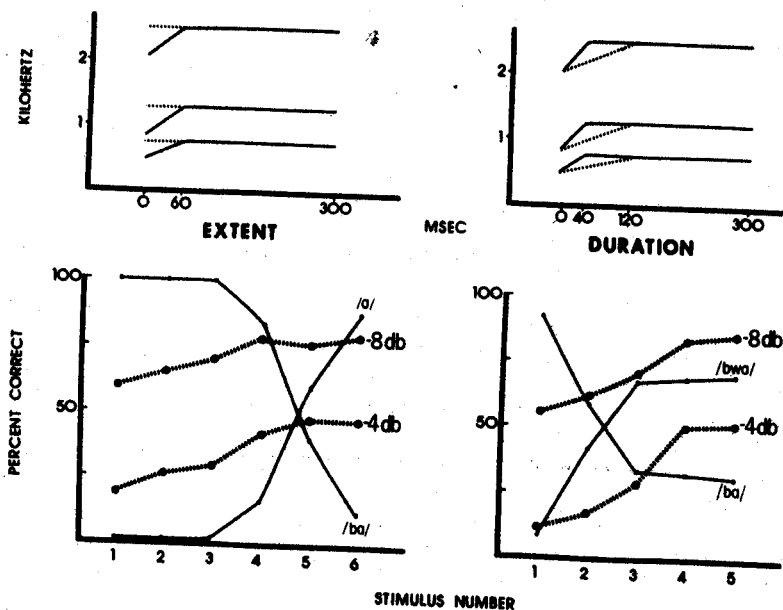
Method

Two arrays of three-formant speech stimuli were generated on the Haskins Laboratories parallel-resonance synthesizer. One array consisted of six items differing in the *extent* of formant transitions, with all items identifiable as /ba/ or /a/; the other array consisted of five items differing in *duration* of formant transitions, with all items identifiable as /ba/ or /bwa/. All stimuli were 300 msec in duration and had a flat pitch contour of 100 Hz. Steady-state /a/ resonances for both arrays centered on 769, 1,232, and 2,525 Hz for first, second, and third formants, respectively. The six-item /ba/-to-/a/ array contained stimuli whose formant transitions were 60 msec in duration. Transitions decreased in extent by equal increments over this array, in corresponding fashion for all three formants. Stimulus 1 (the prototype /ba/) transitions began at 513, 846, and 2,180 Hz for the three formants, respectively; and Stimulus 6 (the steady-state vowel /a/) began with formants of 769, 1,232, and 2,525 Hz. Intermediate stimuli had intermediate starting frequencies for each formant. The five-item /ba/-to-/bwa/ array contained stimuli whose formant transitions always began at 513, 856, and 2,180 Hz, but whose transition durations lasted 40, 60, 80, 100, and 120 msec for Stimuli 1 through 5, respectively. The end-point stimuli of both arrays are shown schematically in the top panels of Figure 1. Stimuli were digitized and stored on disk file using the pulse code modulation system at Haskins. Further stimulus alteration consisted of decreasing the initial portions of all stimuli by 0, 4, and 8 dB. For the /ba/-to-/a/ array, the decreased portion was always 60 msec in duration (like that used by Dorman, 1974), and for the /ba/-to-/bwa/ array, it conformed to the duration of the formant transitions: 40, 60, 80, 100, or 120 msec. In this manner, each of the 11 stimuli was synthesized in three renditions. For an indication of overall amplitude envelope shape of these stimuli, see Figure 2 of Dorman (1974).

Four diotic stimulus sequences were recorded on audio tape, one identification sequence consisted of random orders of the standard (0-dB) stimuli, with 48 and 40 items, respectively, for the extent and duration stimuli; each item in each array appeared eight times. The interval between each item in both sequences was 3 sec. Listeners wrote down BAH or AH and BAH or BWAH to identify members of the arrays. Discrimination sequences consisted of 90 and 75 AX trials for the /ba/-to-/a/ and /ba/-to-/bwa/ arrays: (6 and 5 stimuli in the arrays, respectively) \times (3 intensities to be discriminated: 0-, 4-, and 8-dB differences between members of the AX pair) \times (5 observations per pair). Each discrimination trial began with a 100-msec, 1,000-Hz warning tone, followed by 500 msec of silence, followed by Stimulus A, another 500-msec silent interval, and Stimulus X. Stimulus A was always the standard stimulus, whereas Stimulus X had formant structures identical to Stimulus A but with its initial portions attenuated by 0, 4, or 8 dB. There was a 3.5-sec interval between the offset of Stimulus X and the onset of the warning tone for the subsequent trial. Listeners wrote down S for *same* if they thought the AX items were identical, and D for *different* if they were not.

Thirteen Wesleyan University students listened en masse to the four sequences as part of a course project. All were native American English speakers with little experience at listening to synthetic speech. They listened to the audio tapes played on a

Figure 1. In upper panels, schematic spectrograms indicating the arrays of stimuli whose formant transitions differ in extent or in duration. Actual bandwidth of the formants is 60, 90, and 120 Hz for the first, second, and third formants, respectively; very narrow bands are shown only for purposes of clarity in revealing acoustic variation. Respective identification and discrimination results are superimposed on one another in the lower panels.



Crown CX-822 tape recorder, broadcast in a quiet room over an Ampex AA-620 loudspeaker. All listeners sat between 8 and 18 ft from the loudspeaker, which for the standard items delivered approximately 75 dB SPL.

Results

All results are shown in the lower panels of Figure 1. In the left-hand panel, identification functions for /ba/ and /a/ are superimposed on two discrimination functions, those for judgments of 4- and 8-dB differences. Stimuli 1 through 4 were consistently identified as /ba/, and only Stimulus 6 was identified consistently as /a/. The identification "boundary" appears to be located near Stimulus 5, where the two complementary identification functions cross. Discrimination functions (percent correct discrimination of intensity differences at each comparison) show that 8-dB judgments were consistently more successful than the 4-dB judgments [$F(1,44) = 65.1, p < .001$]. There was no interaction of intensity with stimulus location along the array; therefore, collapsing across the two intensity differences, there was a significant increase in discriminability as the formant transitions decreased in extent [$F(5,144) = 3.15, p < .025$]. Moreover, a trend test (Winer, 1962, p. 132) proved this increase to be linear [$F(1,64) = 49.3, p < .001$], with no significant quadratic, cubic, or other higher-order components. The D responses on AA trials (those with 0-dB difference) were scored as false alarms, and the detectability of the intensity differences was then assessed independent of possible response bias. A generally linear increase was obtained: The d' scores for 4-dB judgments were .44, .60, .84, 1.10, 1.08, and 1.15; and those for 8-dB judgments were 1.61, 1.56, 1.89, 2.20, 1.92, and 2.09, respectively, for the six different transition extents.

Results for the /ba/-to-/bwa/ array are shown in the lower right-hand panel of the same figure. Identification functions are somewhat unimpressive: Only Stimulus 1 was consistently identified as /ba/ and, whereas Stimuli 3 and 5 were primarily identified as /bwa/, none was so identified with a consistency exceeding 72%. The identification "boundary," if one can be said to exist, appears to be near Stimulus 2. The pattern of discrimination results followed very closely that for the previous set of stimuli. Again 8-dB judgments were superior to 4-dB judgments [$F(1,120) = 58.9, p < .001$], discriminability increased across the stimulus array [$F(4,120) = 7.9, p < .001$], and that increase was linear [$F(1,51) = 73.0, p < .001$] without significant higher-order components. This linear pattern was repeated in terms of detectability: 4-dB d' scores were .24, .45, .84, 1.26, and 1.28; and 8-dB scores were 1.60, 1.70, 1.96, 2.12, and 2.22, respectively, for the five different transition durations.¹

Discussion

Two aspects of our results support the primary conclusions of Pastore et al. (1975): The inability to detect intensity differences carried on the formant transitions of stop consonants is a *psychoacoustic* rather than *phonetic* effect. First, there is no abrupt increase in detectability of intensity differences as the stimulus arrays change from /ba/ to /a/ for those stimuli differing in extent of transitions, and from /ba/ to /bwa/ for those differing in duration of transitions. If the availability of acoustic information were somehow inhibited by the processing of the highly encoded stop consonant in particular, one would have expected a quantal increase in discriminability in the /ba/-to-/a/ array at about

Stimulus 5. Clearly none exists, and thus the effect cannot be directly related to categorical perception. Studdert-Kennedy, Liderman, Harris, and Cooper (1970), among others, would predict discontinuities in the discrimination functions at this point if the phenomenon were related to categorical perception. Second, the increase in discriminability is linear for both arrays. Such linear increases are also at variance with the nonlinear, categorical-like processes associated with phonetic perception.

Our results demonstrate interaction between rate of frequency change and the discrimination of intensity change on formant transitions. That is, for the /ba/-to-/a/ array in particular, the less frequency change that occurs, the more perceptible the intensity differences become. Thus, frequency and intensity appear to be yoked in the percept and contribute in an interactive manner to the traces available to short-term auditory memory. Of course, as Pastore et al. admit, finding a psychoacoustic basis for the inability to detect such intensity differences here does not rule out the possibility that a similar outcome could result from processes occurring at other levels. In visual masking, for example, Turvey (1973) demonstrated that when viewers were unable to report a target, the contour information may have been masked peripherally or centrally. At both levels the effect is similar—viewers are unable to identify the target.

A second look at simple backward masking. The secondary conclusion of Pastore et al.—that these results are caused by simple backward masking—is more suspect. While they do not mention these issues, the type of phenomenon they refer to appears to be threshold masking rather than recognition masking (Massaro, 1973, 1975), the locus of the backward masking appears to be peripheral, not central, and it appears to result from target-mask integration, not interruption [see Kahneman (1968) and Turvey (1973) for arguments with respect to vision]. From this view of masking, one might not expect to find evidence in any experimental paradigm of the ability to detect 4- to 9-dB intensity differences carried on formant transitions. That is, this information would be buried in background noise considerably prior to the decision-making process. There are several reasons to suspect, however, that the intensity information in the Dorman (1974) and present studies persists throughout the system.

First, there appears to be a strong effect on detectability of mixing vs. blocking stimuli within a test. The d' scores of Dorman's study and the present study are compared in Table 1 for intensity differences carried on CV syllables as opposed to those on steady-state vowels. Direct comparisons are difficult (a) since Dorman used attenuations of 7.5 and 9.0 dB, whereas we used attenuations of 4 and 8 dB, (b) since Dorman used the carrier stimuli /bae/ and

Table 1
Differential Detectability (d') of Intensity Differences Carried on CV and V Syllables

| | Intensity (dB) | CV | V |
|---------------|-------------------|------|------|
| Dorman (1974) | 7.5 | .32 | 3.28 |
| | 9.0 | .62 | 3.98 |
| Experiment I | 4.0 | .44 | 1.15 |
| | 8.0 | 1.61 | 2.09 |

/æ/, whereas we used /ba/ and /a/, and (c) since Dorman's listeners heard his stimuli through ear-phones, whereas we played them over a loudspeaker in a reverberant room. Nevertheless, a striking trend can be seen when d' scores for his stimuli are compared with those for Stimuli 1 and 6 from the /ba/-to-/a/ array.

In the present study, by mixing the CV and V stimuli together with several intermediate items, the detectability of the intensity differences carried on the CV syllables increased considerably. It decreased, on the other hand, for those differences carried on steady-state vowels. It would appear, then, that a large proportion of the effect is attributable to context, not to masking. That is, detectability varies according to previous experience and expectations within the experiment. The difference in detectability for intensity differences in CV and V syllable changed from a standard score of more than 3.2 (for Dorman's 9-dB discriminations) to one of less than .5 (for our 8-dB discriminations). Such a finding appears to be at variance with the hypothesized effect of simple peripheral masking, and suggests (a) that the intensity information is available at some level of perceptual analysis, and (b) that recognition masking rather than threshold masking may be involved in the Dorman (1974) and Pastore et al. (1975) results.

A second avenue of reasoning comes from the many studies of categorical perception of stop consonants, and the fate of within-phoneme-category formant frequency information. In ABX (Liberman et al., 1957), odd-ball (Mattingly et al., 1971), and AX (Pisoni, 1971, 1973) paradigms, the discrimination of frequency differences carried on formant transitions has been found to be categorical—that is, the frequency difference in formant transitions within the same phonemic category is discriminated at about chance, while that across categories is discriminated very easily. Despite essentially chance within-category performance, frequency information is not *masked* in auditory processing nor *lost* in the auditory-to-phonetic transformation (see Barclay, 1972; Pisoni & Lazarus, 1974). Pisoni & Tash (1974), for example, have shown that "same" reaction times (RTs) to physically different but phonemically identical stop consonants are slower than "same" RTs to physically identical stop consonants. Thus, even though the discrimination response implies that

the two signals were perceived identically, and by inference that there was no distinguishing information left about formant trajectories, the RTs indicate that at some level in the nervous system the information was present. We would expect a similar outcome in an RT analysis with the signals used in the present study. That is, we suspect that the "same" RTs to the physically different (4 dB) signals would be slower than the "same" RTs in the physically identical (0 dB) condition. Experiment II was conducted to test this hypothesis.

EXPERIMENT II

Method

Two stimuli were selected from Experiment I: Stimulus 1 (/ba/) and Stimulus 6 (/a/) from the array whose transitions differed in extent. Both were generated in three renditions: The initial 60 msec was attenuated by 0, 4, and 8 dB. One discrimination sequence was assembled exactly as in Experiment I. It contained 120 AX trials: (2 Stimuli) × (3 intensities to be discriminated) × (20 observations per item). Listeners pressed, as rapidly as possible, one of two telegraph keys to indicate whether the two items within a trial were the same or different. Reaction times were fed on-line into a PDP-11 computer for analysis. They were measured from the onset of the second item to the onset of the keypress.

Four students and staff members at Haskins volunteered for the experiment. All were naive to the purposes of the task. They listened, in groups of two, to stimuli reproduced on an Ampex AG-500 tape recorder and transmitted binaurally through a listening station to Telephonics headphones (TDH-39).

Results and Discussion

The most important reaction-time results are shown in Table 2—mean RTs for "same" responses for the 0- and 4-dB discriminations. Few "same" responses were made for 8-dB trials, so they are not included. The difference in RTs for the two conditions ranged from 85 and 336 msec for the four listeners; the results for three listeners were statistically robust by a Mann-Whitney U test on individual reaction times, while those for the other listener approached significance. (U scores were converted into standard z units, as shown in Table 2). These results indicate that intensity information not discriminated on a particular trial is *not* masked in absolute terms, but is represented in some form

throughout the information-processing system.² The representations for the two stimuli within a trial that differ in the amplitude of their onsets of 4 dB are more difficult to match than are those pairs with the same onset amplitude. These results are congruent with those of Pisoni and Tash (1974) using speech syllables, and with prior results of Emmerich, Gray, Watson, and Tanis (1972) using nonspeech stimuli.

CONCLUDING DISCUSSION

The results of the present studies suggest, first, that the relationship between pedestal-detection and recognition-masking experiments is one of degree rather than kind. There is no discontinuity between the two. Second, the results support the primary conclusion of Pastore et al. (1975): The relative inability to discriminate intensity differences carried on the formant transitions of CV syllables, as compared to those carried on the initial portions of steady-state vowel syllables, is an effect that is psychoacoustic rather than phonetic.

Third, our results demonstrate differences between types of masking. Pastore et al. appear to attribute the inability to discriminate differences carried on formant transitions to simple backward masking. Simple masking, according to Licklider (1951), is the opposite of analysis: Information is simply not processed, and the implication is that masked information is irretrievably buried in background noise. However, the results of Experiment II shows that phonemically irrelevant acoustic information remains accessible to the listener in some form. This suggests that recognition masking is the phenomenon involved in the Dorman (1974) and Pastore et al. (1975) experiments and Experiment I of the present investigation. Moreover, recognition masking is selective in its effect on auditory vs. phonetic memory codes. Fourth, a comparison of the detectability scores from Dorman's (1974) study and those from Experiment I also suggest that this information is not masked absolutely even in recognition terms, but may be used or unused as a function of context in an experimental session.

Table 2
Mean Reaction Time (and Number) of "Same" Responses to Intensity Differences Carried on the Initial 60 msec of CV Syllables

| Listener | Intensity Difference | | |
|---------------|----------------------|-----------|-----------------------|
| | 0 dB | 4 dB | |
| T.B. | 659 (19) | 942 (9) | $z = 3.02 p < .002^*$ |
| P.B. | 609 (16) | 694 (11) | $z = 1.19 p < .12$ |
| W.F. | 612 (17) | 818 (8) | $z = 1.92 p < .03$ |
| H.S. | 669 (18) | 1005 (4) | $z = 2.89 p < .002$ |
| Mean of Means | 637 | 865 | |

Note—Maximum number of trials per cell is 20.
*One-tailed

On the "Speech Processor"

Pastore et al. suggest that a speech processor is an unneeded construct to account for results in the AX discrimination task. We agree. Nevertheless, whereas our results support this position, we must not ignore the necessity for some such device at some level. The level at which any device is specific to speech is currently a crucial question. Several effects thought to demonstrate the psychological reality of phonetic processing (Wood, 1975, p. 16) have been found to occur in purely auditory domains (Blechner, Day, & Cutting, 1976; Cutting, in press; Cutting & Rosner, 1974; Miller et al., in press; Pastore, Ahroon, Puleo, Crimmins, Golowner, & Berger, 1976). Thus, the mechanism that extracts phonetic information from the speech signal may be the same device that is used elsewhere, for example in the processing of music-like sounds. In other words, phonetic-like processing may not be speech-specific processing. Yet these recent findings cut only into the lowest tier of the speech-language hierarchy—that of phonetic processing. The perception of different allophones of the same phoneme as being the same—such as the /p/s in *pit*, *spit*, and *tip*—and the parsing of syllables from a continuous speech stream seem to be processes without nonspeech analogs. Unless (or until) analogs are found, the notion of a speech processor is not impugned by the existence of phonetic-like processes elsewhere in audition.

REFERENCES

- BARCLAY, J. R. Noncategorical perception of a voiced stop: A replication. *Perception & Psychophysics*, 1972, 11, 269-273.
- BLECHNER, M. J., DAY, R. S., & CUTTING, J. E. Processing two dimensions of nonspeech stimuli: The auditory-phonetic distinction reconsidered. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, 2, 257-266.
- CUTTING, J. E. Two left-hemisphere mechanisms in speech perception. *Perception & Psychophysics*, 1974, 16, 601-602.
- CUTTING, J. E. The magical number two and the natural categories of speech and music. In N. S. Sutherland (Ed.), *Tutorial essays in psychology*. Hillsdale, N. J: Erlbaum, in press.
- CUTTING, J. E., & ROSNER, B. S. Categories and boundaries in speech and music. *Perception & Psychophysics*, 1974, 16, 564-570.
- CUTTING, J. E., ROSNER, B. S., & FOARD, C. F. Perceptual categories for music-like sounds: Implications for theories of speech perception. *Quarterly Journal of Experimental Psychology*, 1976, 28, in press.
- DORMAN, M. F. Discrimination of intensity differences in formant transitions in and out of syllable context. *Perception & Psychophysics*, 1974, 16, 84-86.
- EMMERICH, D. S., GRAY, J. L., WATSON, C. S., & TANIS, D. C. Response latency, confidence, and ROCs in auditory signal detection. *Perception & Psychophysics*, 1972, 11, 65-72.
- KAHNEMAN, D. Method, findings, and theory in studies of visual masking. *Psychological Bulletin*, 1968, 70, 404-426.
- LANE, H. Motor theory of speech perception: A critical review. *Psychological Review*, 1965, 72, 275-309.
- LANE, H. A behavioral basis for the polarity principle in linguistics. *Language*, 1967, 43, 494-511.
- LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. P., & STUDDERT-KENNEDY, M. Perception of the speech code. *Psychological Review*, 1967, 74, 631-661.
- LIBERMAN, A. M., HARRIS, K. S., HOFFMAN, H. S., & GRIFFITH, B. C. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 1957, 54, 358-368.
- LIBERMAN, A. M., MATTINGLY, I. G., & TURVEY, M. T. Language codes and memory codes. In A. W. Melton & E. Martin (Eds.) *Coding processes in human memory*, Washington, D. C: Winston, 1972.
- LICKLIDER, J. C. R. Basic correlates of the auditory stimulus. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.
- LOCKE, S., & KELLAR, L. Categorical perception in a nonlinguistic mode. *Cortex*, 1973, 9, 355-369.
- MASSARO, D. W. A comparison of forward versus backward recognition masking. *Journal of Experimental Psychology*, 1973, 100, 434-436.
- MASSARO, D. W. Backward recognition masking. *Journal of the Acoustical Society of America*, 1975, 58, 1059-1065.
- MATTINGLY, I. G., LIBERMAN, A. M., SYRDAL, A., & HALWES, T. G. Discrimination in speech and nonspeech modes. *Cognitive Psychology*, 1971, 2, 131-157.
- MILLER, J. D., WIER, C. C., PASTORE, R. E., KELLY, W. M., & DOOLING, R. M. Discrimination and labeling of noise-buzz sequences with varying noise lead times: An example of categorical perception. *Journal of the Acoustical Society of America*, 1976, in press.
- PASTORE, R. E., AHROON, W. A., PULEO, J. S., CRIMMINS, D. B., GOLOWNER, L., & BERGER, R. S. Processing interaction between dimensions of nonphonetic auditory signals. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, 2, 267-276.
- PASTORE, R. E., AHROON, W. A., WOLZ, J. P., PULEO, J. S., & BERGER, R. S. Discrimination of intensity differences on formant-like transitions. *Perception & Psychophysics*, 1975, 18, 224-226.
- PISONI, D. B. *On the nature of categorical perception of speech sounds* (PhD thesis, University of Michigan, 1971). *Dissertation Abstracts International*, 1972, 32, 6693B (University Microfilms No. 72-14,964).
- PISONI, D. B. Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 1973, 13, 253-260.
- PISONI, D. B., & LAZARUS, J. H. Categorical and noncategorical* modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America*, 1974, 55, 328-333.
- PISONI, D. B., & TASH, J. Reaction times to comparisons within and across phonetic boundaries. *Perception & Psychophysics*, 1974, 15, 285-290.
- 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. What is masking? *Journal of the Acoustical Society of America*, 1958, 30, 919-921.
- TANNER, W. P., & SORKIN, R. D. The theory of signal detectability. In J. V. Tobias (Ed.), *Foundations of modern auditory theory* (Vol. 2). New York: Academic Press, 1972.
- TURVEY, M. T. On peripheral and central processes in vision: Inferences from an information-processing analysis of masking with patterned stimuli. *Psychological Review*, 1973, 80, 1-52.
- WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.
- WOOD, C. C. Auditory and phonetic levels of processing in speech perception: Neurophysiological and information-processing analyses. *Journal of Experimental Psychology: Human Perception and Performance*, 1975, 1, 3-20.

NOTES

1. We assume that the variance of the distributions of noise and signal plus noise are the same for all stimuli, and also that the criteria are the same for all stimuli.

2. One may complain that we are confounding stimulus information (0- vs. 4-dB change) with correctness of responses, and that one normally expects longer RTs for incorrect responses since different processes are thought to be involved. Our response

is that until this study was made, the results of Dorman (1974) and Pastore et al. (1975) would suggest that a "same" response to a 4- and 9-dB discrimination was the correct response. The same logic prompted the study of Pisoni and Tash (1974).

(Received for publication February 26, 1976;
revision accepted June 21, 1976.)