

## Processing Two Dimensions of Nonspeech Stimuli: The Auditory-Phonetic Distinction Reconsidered

Mark J. Blechner and Ruth S. Day  
*Yale University and Haskins Laboratories*

James E. Cutting  
*Wesleyan University and  
Haskins Laboratories*

Nonspeech stimuli were varied along two dimensions—intensity and rise time. In a series of speeded classification tasks, subjects were asked to identify the stimuli in terms of one of these dimensions. Identification time for the dimension of rise time increased when there was irrelevant variation in intensity; however, identification of intensity was unaffected by irrelevant variation in rise time. When the two dimensions varied redundantly, identification time decreased. This pattern of results is virtually identical to that obtained previously for stimuli that vary along a linguistic and a nonlinguistic dimension. The present data, taken together with those from other studies using the same stimuli, suggest that the mechanisms underlying the auditory-phonetic distinction should be reconsidered. The results are also discussed in terms of general models of multidimensional information processing.

Several contemporary accounts of speech perception have emphasized the organization of processing into a hierarchy of levels, including auditory, phonetic, phonological, lexical, syntactic, and semantic (Fry, 1956; Stevens & House, 1972; Studdert-Kennedy, in press). The distinction between phonetic and higher levels has been commonly accepted by linguists and psychologists for some time. Recently, however, much attention has been directed toward the auditory-phonetic distinction (e.g., Fant, 1967; Stevens & Halle, 1967; Studdert-Kennedy, Shankweiler, & Pisoni, 1972). Fry (1956), in an early discussion of the levels-of-processing view of speech perception, emphasized the role of the "physical-psychological transformation" that occurs in the recognition of phonemes

from the acoustical signal. The important characteristic of this transformation is that there is no one-to-one relationship between "the number and arrangement of physical clues and the sound which is recognized" (Fry, 1956, p. 170). Fry did not state that the physical-psychological transformations characteristic of speech are exclusive to speech. However, this possibility was emphasized by later work which viewed speech perception as mediated by articulatory mechanisms (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Stevens & House, 1972). Such a view made it desirable and perhaps necessary to partition all sounds into two general classes: those which are speech and those which are not.<sup>1</sup>

Definitions concerning which criteria must be met in order for sounds to be classified as speech have varied in the literature. The present paper assumes a two-part definition of speech: Sounds that (a) can be articulated by the human vocal apparatus; and (b) can be recoded into higher order linguistic units. According to this definition, phonetic processing may

---

This research was supported by National Institute of Mental Health Training Grant PHS5T01MH05276-27 to Yale University, and by National Institute of Child Health and Human Development Grant HD-01994 to the Haskins Laboratories. The authors thank Michael Studdert-Kennedy for a discussion of the issues raised in this paper, Robert L. Plotz for assistance in running the experiment, and Richard M. Levich for statistical advice.

Requests for reprints should be sent to Mark J. Blechner at Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06510.

---

<sup>1</sup> Some authors use the terms *linguistic and nonlinguistic* or *verbal and nonverbal* to indicate the same distinction.

refer to articulatory processes either directly (Liberman et al., 1967) or implicitly (Stevens & House, 1972), and follows some system of linguistic organization such as a distinctive feature system (Chomsky & Halle, 1968; Jakobson, Fant, & Halle, 1963). Furthermore, while all sounds undergo auditory processing, only speech sounds undergo phonetic processing.

To assess the psychological reality of the auditory-phonetic distinction, many experiments have been conducted. Results from several paradigms suggest that even though speech sounds differ along a wide variety of acoustic dimensions, they are perceived in ways that are qualitatively distinct from the way nonspeech sounds are perceived. For example, the main difference between the phonemes /ba/ and /da/ is the direction and extent of the second-formant transition (Liberman, Delattre, Cooper, & Gerstman, 1954), while the distinction between /ba/ and /pa/ lies in the latency between the initial plosive burst and the onset of voicing (Lisker & Abramson, 1964). Yet, for both distinctions, there is no one-to-one relationship between changes in the acoustic patterns and probabilities of identification. Instead, item identifications remain at or near 100% as one phoneme or the other, with an abrupt crossover at the "phoneme boundary." More importantly, whereas most stimulus dimensions in the environment, such as frequency, intensity, and brightness, can be discriminated from one another much more accurately than they can be identified (Miller, 1956), this is not the case for several linguistic dimensions. Instead, two acoustically different stimuli that lie within the same phoneme category are discriminated at near-chance level; two stimuli that lie in separate categories but differ by the same acoustic increment are discriminated with very few errors. This nonlinear mode of perception is called *categorical perception* (Liberman, Harris, Hoffman, & Griffith, 1957; for a review, see Studdert-Kennedy, in press).

Other experimental operations have shown processing differences for speech and nonspeech stimuli. Recent work using

a selective adaptation paradigm has shown that repeated presentation of a consonant-vowel (CV) syllable produces systematic shifts in the phoneme boundary (Eimas, Cooper, & Corbit, 1973; Eimas & Corbit, 1973). Some experiments suggest that the basis of this adaptation is phonetic rather than auditory (Cooper, 1975), although this evidence is not conclusive (Studdert-Kennedy, in press). The auditory-phonetic distinction seems to be further supported by dichotic identification tasks that often reveal right-ear advantages for speech stimuli (e.g., Kimura, 1961; Shankweiler & Studdert-Kennedy, 1967) and left-ear advantages for nonspeech stimuli (e.g., Chaney & Webster, 1966; Curry, 1967; Kimura, 1964).

In addition, several experiments have investigated the relationship of auditory and phonetic processes in selective attention tasks using stimuli that vary along two dimensions. When both dimensions are linguistic (e.g., initial stop consonant and vowel in CV syllables), selective attention for either dimension is impaired by irrelevant variation in the other dimension (Wood & Day, 1975). This pattern of symmetric interference also occurs when both dimensions are nonlinguistic, such as pitch and intensity (Wood, 1975a). However, when one dimension is linguistic and the other is not, a pattern of asymmetric interference appears; reaction time (RT) for identification of stop consonants is impaired by irrelevant variation in pitch, but RT for pitch identification is not increased significantly by irrelevant variation in stop consonant (Day & Wood, 1972; Wood, 1974, 1975a). These results have been interpreted to support the auditory-phonetic distinction.

The dichotic listening, categorical perception, selective adaptation, and speeded classification experiments appear to comprise a set of converging operations (Garner, Hake, & Eriksen, 1956) on the psychological reality of the auditory-phonetic distinction. Recently, however, this view has been brought into question by experiments in which the stimuli are sawtooth-wave tones differing in rise time

(see Cutting, in press). While rise time can cue a speech distinction, such as the difference between the syllables /ʃa/ and /tʃa/, sawtooth waves are not perceived as speech. Instead, they sound comparable to a plucked or bowed violin string. Although these "plucks" and "bows" are not processed phonetically, they are processed similarly to speech in several ways. They are perceived categorically (Cutting & Rosner, 1974), and their identification functions shift in the same manner as for speech following selective adaptation (Cutting, Rosner, & Foard, in press). Ear-advantage data for plucks and bows are not decisive.

The present experiment seeks to investigate further the processes by which the plucks and bows are perceived and their relationship to the auditory-phonetic distinction. The two-choice speeded classification procedure developed by Garner and Felfoldy (1970) and modified for use in auditory experiments by Day and Wood (1972; Wood, 1974, 1975a; Wood & Day, 1975) was used to determine how the dimensions of rise time and intensity interact. If symmetric interference necessarily results when stimulus dimensions are of the same general class, that is, both linguistic or nonlinguistic, then selective attention to either rise time or intensity should suffer from irrelevant variation in the other dimension. However, if a pattern of asymmetric interference occurs, it would be clear that such a pattern need not be based on separate auditory and phonetic levels of processing. This pattern of results, along with those of other experiments using pluck and bow stimuli, would lead one to question the mechanisms underlying the auditory-phonetic distinction as currently conceived.

The present experiment is also concerned with the processing of multidimensional stimuli in general. The pattern of asymmetric interference suggests a *serial* model; only the dimension processed first interferes with the other. However, this view has been challenged by the finding that when the two dimensions of the same stimulus vary redundantly, subjects can

identify them more quickly than when only one dimension varies (Wood, 1974). This *redundancy gain* (Garner & Felfoldy, 1970) seems to argue against a serial model. If pitch processing were completed before processing of the stop consonant began, a subject should not be able to use redundant information about the stop consonant to speed up identification of pitch. Instead, a *parallel* model, which posits that the processing of the two dimensions overlap or are simultaneous, would better account for the redundancy gain. Information processing models that propose strict serial or parallel processing do not seem to offer an adequate explanation of Wood's (1974) finding of asymmetric interference with redundancy gain. It seems likely, instead, that subjects have some degree of freedom about the kinds of processing that they use in different task conditions. The present experiment, by also including a task that varies the two dimensions redundantly, seeks to distinguish the conditions in which processing strategies are optional from those in which they are mandatory.

## METHOD

### *Stimuli*

Stimuli varied along two dimensions—intensity and rise time. They were derived from the sawtooth waves used by Cutting and Rosner (1974), generated on the Moog synthesizer at the Presser Electronic Studio at the University of Pennsylvania. The original stimuli differed in rise time and resembled the sound of a plucked or bowed violin string. The pluck and bow stimuli reached maximum intensity in 10 and 80 msec, respectively. Two more stimuli of lower amplitude were created by attenuating the original stimuli 7 db., using the Pulse Code Modulation (PCM) system at Haskins Laboratories (Cooper & Mattingly, 1969). Thus the final four stimuli were loud pluck, soft pluck, loud bow, and soft bow. The absolute level of the loud and soft stimuli were 75 and 68 db. re  $20 \mu\text{N}/\text{m}^2$ , respectively. All stimuli were truncated to 800 msec in duration (the original stimuli were approximately 1,050 msec), and then digitized and stored on disc file using the PCM system. Items were reconverted to analog form at the time of tape recording.

### *Tapes*

All tapes were prepared on the PCM system. Test stimuli were recorded on one channel of the audio tape. On the other channel, brief pulses were

TABLE 1  
STIMULUS SETS FOR EACH DIMENSION AND  
CONDITION

Dimension	Condition		
	Control	Correlated	Orthogonal
Rise time	Loud pluck Loud bow or Soft pluck Soft bow	Loud pluck Soft bow	Loud pluck Soft pluck Loud bow Soft bow
Intensity	Loud pluck Soft pluck or Loud bow Soft bow	Loud pluck Soft bow	Loud pluck Soft pluck Loud bow Soft bow

synchronized with the onset of each test stimulus; these pulses triggered the reaction time counter during the experimental session.

A display tape was prepared to introduce the listeners to the stimuli. The four stimuli were played in the same order several times, beginning with three tokens of each item, then two of each, and finally one of each. Practice tapes were also prepared, consisting of a randomized order of 20 items, five of each stimulus. There were two practice tapes, each with a different random order.

The eight test tapes each contained 64 stimuli with a 2-sec interstimulus interval. Each tape was composed of different subsets of the four stimuli, depending on the condition of the experiment. In the *control* condition, the stimuli varied along only one dimension, while the other dimension was held constant. Thus, for example, one intensity control tape consisted of loud and soft bows only, while the other consisted of loud and soft plucks. For half the subjects, the nontarget dimension (in this case, rise time) was held constant at one value (pluck), whereas for the other half, it was held constant at the other (bow). Rise-time control tapes were constructed in an analogous fashion. In the *orthogonal* condition, both dimensions varied independently. Hence, the two tapes for this condition contained all four kinds of stimuli, in different random orders. In the *correlated* condition, the dimensions varied in a completely redundant manner: All of the pluck stimuli were loud and all of the bow stimuli were soft. See Table 1 for a complete outline of the stimuli in each condition.

### Subjects and Apparatus

The six subjects (five males and one female, from 19 to 27 years of age) participated in all six tasks. All reported no history of hearing trouble.

The tapes were played on an Ampex AG-500 tape recorder and the stimuli were presented through calibrated Telephonics headphones (Model TDH39-300Z). Subjects sat in a sound-insulated room and responded with their dominant hand on the two telegraph keys mounted on a wooden board. Throughout the experiment, the left key was used for pluck and loud responses, while the right key was used for bow and soft responses. The pulse

on one channel of the tape triggered a Hewlett-Packard 522B Electronic Counter. When a response on either telegraph key stopped the counter, the reaction time was registered onto paper tape by a Hewlett-Packard 560A digital recorder for subsequent analysis. The listener's response choice was recorded manually by the experimenter.

### Procedure

At the start of the session, subjects were informed of the general nature of the experiment and of the dimensions which they would be asked to identify. They were told that the difference in rise time could be compared to the difference in sound between a plucked and a bowed violin string.

For preliminary training, subjects heard the display sequence twice. Next, they listened to the two sets of 20 practice items and responded verbally, first attending to rise time and then to intensity. This insured that the subject could perceive the differences along both dimensions. They then repeated the same practice trials, responding with a key press rather than verbally, in order to become familiar with the mode of response. Subjects were instructed to respond as quickly as possible without making errors.

The order of presentation for the six test tapes was determined by a balanced Latin square design. Before the test trials of each condition, appropriate instructions were read. Subjects were then given eight practice trials to help stabilize reaction time performance and to familiarize them with the identification task and stimulus set for that particular condition. In the control conditions subjects were told which dimension to attend to and the value at which the other dimension would be held. In the orthogonal conditions, they were instructed to attend to one dimension and to ignore variation in the other dimension. In the correlated conditions, they were instructed to attend to one dimension, but were encouraged to use the additional information from the other dimension.

### RESULTS AND DISCUSSION

Both dimensions were easy to identify. In the practice trials, no subject made more than 2% errors. During test trials, the highest mean error rate for any condition was 1.8%. A three-way analysis of variance of the error data (Subjects  $\times$  Conditions  $\times$  Dimensions) revealed no significant main effects nor interactions. Therefore, the error data are not considered in detail in this discussion.

### *Asymmetric Interference with Redundancy Gain*

For the reaction time data, median RT was calculated for each individual block of trials for each subject, and means of

TABLE 2

MEDIAN REACTION TIME IN MILLISECONDS AND PERCENT ERRORS FOR EACH DIMENSION AND CONDITION

Dimension	Condition		
	Control	Correlated	Orthogonal
Rise time	426.5 (1.32)	393.7 (.06)	480.0 (1.66)
Intensity	442.3 (1.34)	406.4 (.06)	442.5 (1.83)

Note. Percent errors are shown in parentheses.

three alternative explanations of the redundancy gain were ruled out, as in Wood (1974). First, the possibility of a different speed-accuracy trade-off in the two correlated conditions could be eliminated by the lack of significant differences in the error data, as noted above.

Second, the possibility that the redundancy gain could be due to selective serial processing (see Felfoldy & Garner, 1971) was considered. If subjects use this strategy in the correlated condition, they merely attend to the more discriminable of the two dimensions, regardless of the instructions. Thus, their RT data would show that neither correlated condition is faster than the faster control condition. To test for the occurrence of the selective serial processing strategy, the RT data for each correlated condition was tested against each subject's faster control condition with an analysis of variance and a subsequent comparison of means using the Newman-Keuls method. The correlated conditions were still found to be faster than the faster control condition,  $F(2, 63) = 13.5$ ,  $p < .001$ . Therefore, the redundancy gain cannot be attributed to selective serial processing.

<sup>2</sup> Wood and Day, in all of their reaction time experiments, transformed their data, so that any RT longer than 1 sec was set equal to 1 sec. This was done to correct for possible malfunctioning of the equipment, such as failure of the response key to make electrical contact, or temporary inattention of the subject. While unusually long reaction times due to equipment trouble or lapsing of the subject's attention should be transformed, the continual resetting of long RT values to 1 sec is arbitrary and could distort the data. If arithmetic means are to be used, very long data points can be more equitably adjusted by using reciprocal RT values. Alternatively, medians or trimeans can be used, with similar effect.

medians for each condition across subjects were computed.<sup>2</sup> In addition, the untransformed RT data were subjected to a complete four-way factorial analysis of variance (Subjects  $\times$  Conditions  $\times$  Dimensions  $\times$  Within Cell).

Median RT data are presented in Table 2. For the dimension of rise time, there was an increase of 53.5 msec from the control to the orthogonal condition, while for the dimension of intensity, there was an increase of only .2 msec. In the analysis of variance, the effect of dimensions (intensity versus rise time) was not significant, while the effect of conditions (control, orthogonal, and correlated) was significant,  $F(2, 126) = 33.23$ ,  $p < .001$ . In addition, the Dimensions  $\times$  Conditions interaction was significant,  $F(2, 126) = 6.43$ ,  $p < .01$ . In order to differentiate interference effects from redundancy gains, a contrast of the interactions between the two dimensions and only the control and orthogonal conditions was performed, omitting correlated conditions. This contrast was significant,  $F(1, 63) = 16.58$ ,  $p < .001$ . Thus there was an asymmetric pattern of interference; intensity variation interfered with the processing of rise time while rise time had virtually no effect on the processing of intensity. This finding is especially interesting given that intensity was somewhat more discriminable than rise time, although the difference of 14.2 msec between the control conditions was not statistically reliable.

The effect of redundancy gains was assessed by two methods. A contrast of the conditions effect showed the correlated conditions to be significantly different from the control conditions,  $F(1, 63) = 35.1$ ,  $p < .001$ . A subsequent comparison of the individual means using the Newman-Keuls procedure showed the correlated conditions in both dimensions to differ significantly from the respective control condition. The different correlated conditions, like the control conditions, did not significantly differ from each other.

In order to determine whether the redundancy gain could rightfully be considered as evidence of parallel processing of the two dimensions in this experiment,

Finally, because the stimulus sets in the two correlated conditions were the same, whereas in the control conditions they were different, it is possible that the redundancy gain could be based on differential transfer between control and correlated conditions (Biederman & Checkosky, 1970). To test this explanation, an analysis of variance of the control and correlated conditions (Subjects  $\times$  Conditions  $\times$  Order of Presentation  $\times$  Within Cell) was performed. The control condition presented first was 16 msec slower than the second, suggesting a possible practice effect, although this difference was not reliable. The correlated conditions presented first and second differed by only 2 msec—not significantly. Thus the transfer between the correlated conditions was less than or equal to the transfer between the control conditions, so that differential transfer does not account for the redundancy gain.

The pattern of results in this experiment is remarkably similar to those of Wood (1974). The relationship of intensity to rise time matches that of initial stop consonant both in the asymmetric pattern of interference and in the significant redundancy gain. In Garner's (1974) terminology, the dimensions of rise time and intensity are therefore *asymmetrically integral*. In the following discussion we will first consider the implications of our results in terms of general information-processing models and then reconsider the auditory-phonetic distinction.

#### *Relevance to Theories of Multidimensional Processing*

The present results pose problems for information-processing models that try to account for perception only in terms of the serial or parallel handling of information. The data suggest that both stimulus and task characteristics may affect the mode of processing, so that neither a strict serial nor a strict parallel model can account for the whole picture. This view agrees with the recent suggestions of several authors (Garner, 1974; Nickerson, 1971; Townsend, 1971). One alternative to the strictly serial and parallel models, for

example, is to view the present results as one point on a performance-resource function (Norman & Bobrow, 1975). Perhaps the processing of rise time is resource limited (relatively low in priority and competing with other processes for limited psychological resources) but the processing of intensity is data limited (relatively high in such priority and limited only by the clarity of the signal). An alternative view is that the two processes overlap temporally and that one is contingent on the other, as suggested by Turvey (1973). Perhaps the processing of rise time and intensity (as well as that of place of articulation and pitch) begin simultaneously. Both kinds of information can be combined to produce a redundancy gain in the correlated condition. However, it may be that only the orthogonal condition, which requires information *gating* (Posner, 1964), can reveal the contingency of one kind of information on the other. Current theories, however, do not account for why this contingency relationship might affect one task and not the other. Further research is needed on this point.

One useful approach might be to vary the discriminability of the two dimensions and to note the changes that occur in both the orthogonal and correlated conditions. In a pilot study of speeded classification, we used rise time and pitch where the latter dimension was considerably more discriminable than the former. The result was that the subjects always processed the more discriminable dimension first. Thus, in the orthogonal condition, there was asymmetric interference, and in the correlated condition, RT performance was equal to the faster control condition (the Selective Serial Processing pattern). However, it would be more interesting to determine whether, by manipulating discriminability in the reverse direction, rise time can be made to interfere with intensity. How much more discriminable than intensity would rise time have to be for a pattern of mutual interference to appear? If RT performance in the correlated condition were only as fast as the faster control condition, one might conclude that re-

dundancy gains are impervious to the effects of contingency processing relationships between dimensions. Such a finding would be congruent with the results of the experiment reported here.

#### *Decision-Combination Model of Redundancy Gain*

Many different processing models have been proposed to account for redundancy gains (Morton, 1969), but there is no explicit way to distinguish among them for reaction time data. However, Wood (1975b) has provided a normative model, based on the theory of decision combination, that predicts both the size and the shape of the reaction time distribution for redundant dimensions. This model assumes that completion times for processing each stimulus dimension are independent random variables, and that the response choice on any given trial is initiated as soon as a decision is reached about either stimulus dimension. If the processing-time distributions for the stimulus dimensions overlap, a performance gain for redundant information is predicted.

Wood applied the decision-combination model to his own data (Wood, 1974) and to simulated distributions based on the data of Biederman and Checkosky (1970). In all cases, he found that the model predicted redundancy gains that were slightly but consistently larger than those obtained.

The data of the present experiment were also analyzed in terms of the decision-combination model.<sup>3</sup> Following Wood (1975b) reaction time distributions for each subject in each control condition were calculated (20-msec cell widths), and then were transformed into  $z$  units.<sup>4</sup> The single-dimension distributions were then combined across subjects and entered into Equation 1:

$$h(t_i) = f(t_i) \sum_{j=1}^n g(t_j) + g(t_i) \sum_{j=1}^n f(t_j) - f(t_i)g(t_i), \quad (1)$$

where  $f(t)$  and  $g(t)$  represent the distributions for the individual dimensions, and

TABLE 3

MEANS AND STANDARD DEVIATIONS OF OBTAINED AND PREDICTED REACTION TIME DISTRIBUTIONS

Condition	<i>M</i>	<i>SD</i>
Rise-time control condition	419.9	87.0
Intensity control condition	435.6	95.5
Correlated condition: Obtained*	400.7	84.2
Correlated condition: Predicted	395.2	72.3

\* Pooled data from the two correlated conditions, which, as noted in the text, did not differ significantly from one another.

$h(t)$  represents the predicted distribution when dimensions vary redundantly.

The results of this analysis are displayed in Table 3. The model predicts a redundancy gain 5.5 msec larger than that actually obtained. This result parallels Wood's finding that the model predicts redundancy gains that are slightly but consistently

<sup>3</sup> Wood (1974) set all reaction times greater than 1 sec equal to 1 sec. In the subsequent analysis of the same data, Wood (1975b) does not mention this transformation, although its use is inappropriate when applying the decision combination model, since a "lump" in the reaction time distribution at 1,000 msec would result. Some form of preliminary transformation is necessary, however, because extreme outlying scores would result in a large number of empty cells when calculating reaction time distributions. In the present data, such outliers were confined to very high values; therefore, for the purpose of applying the decision-combination model, it was decided not to consider the 5% highest reaction times for each subject in each condition. This method is justified when, as in the present experiment, error rates are very low, so that there are outliers at high values but not at low values. The method would not be justified if speed were emphasized much more such that high error rates and extremely low reaction times occurred.

<sup>4</sup> The normalization of individual subjects' reaction time distributions is quite important, especially when there are large differences in the overall reaction times for each subject, as in the present data. Wood (1975b) found that his model predicted redundancy gains adequately when he simulated Biederman and Checkosky's (1970) data, with normal distributions based on their reported means and standard deviations. In the present data, however, the simulation of a normal distribution from the mean and standard deviation results in a predicted redundancy gain considerably larger than that actually obtained. This is mainly because the large across-subject variance in the data produces greater overlap between the two control condition distributions than that which actually occurred for any individual subject.

larger than those obtained. It also demonstrates another way in which there is similar processing of place of articulation and of rise time.

It is important to note, however, that the decision-combination model does not account for asymmetric effects. The model is sensitive to differences in discriminability between stimulus dimensions, but it is indifferent to the nature of the individual dimensions. For example, it would predict the same redundancy gain whether intensity were perceived  $n$  msec faster than rise time, or vice versa. Thus the model provides a potent norm against which future research may test the hypothesis that certain pairs of stimulus dimensions produce asymmetric redundancy gains.

#### *Relevance to the Auditory-Phonetic Distinction*

The present results bear not only on the way that different levels of processing interact, but also on the broader question of which levels of processing are important in human auditory perception. In light of the present data, which show asymmetric interference between two "auditory" dimensions, it seems unwise to lump together all acoustic properties that do not provide linguistic cues. At the very least, a heterogeneous conception of auditory analyzing systems, with processing levels of increasing complexity, seems preferable.

Assuming that there are several levels of auditory processing, what are the specific parameters that best describe these levels? The research on auditory and phonetic levels of processing suggested that acoustic factors could not adequately account for the data from experiments using speech stimuli, and took recourse instead to models based on the linguistic-nonlinguistic distinction, invoking reference to articulatory processes.

The present data, along with those of other pluck and bow studies, suggest that the importance of the linguistic-nonlinguistic dimension may have been overrated and that the role of acoustic factors may

have been underrated. Thus, rise time seems to be perceived in many of the same ways, regardless of whether it characterizes a speech sound, as in /ʃa/ and /tʃa/, or a nonspeech sound, as in plucks and bows (see Cutting & Rosner, 1974, for experiments that make this comparison directly). Similarly, Miller, Pastore, Wier, Kelly, and Dooling (in press) found that noise-buzz sequences with varying relative onset times were also perceived categorically. The stimuli varied in a manner analogous to the voice-onset time continuum; thus there appears to be a comparable mode of perception for such stimuli regardless of whether they are speech or nonspeech.

The results of research with phonemes and with plucks and bows are analogous, but they are not identical. The lack of significant laterality data is one notable exception. Therefore, the linguistic-nonlinguistic distinction, which has considerable heuristic value in much research, cannot be blankly discarded. However, in the future, as acoustic factors that underlie speech sounds are understood in greater detail, less weight may be given to "special" modes of processing accorded to speech (see, e.g., Kuhn, 1975).

It may well be that nonacoustic factors do determine certain perceptual processes, but that the linguistic-nonlinguistic dimension is not the most accurate way of describing these factors. Speech, after all, is the most important hierarchically coded system of sound, but it is not the only one. Consider rise time: By the definition of the term *speech* established above, it is clear that plucks and bows are not phonetic, since they cannot be articulated by the human vocal tract. However, their status with respect to the second part of the definition, that the sounds can be recoded into higher order linguistic units, is less clear. Certainly, plucks and bows are not part of spoken language; but they do comprise lower level components in the "language" of music, which, like human spoken language, can be divided into hierarchical levels of organization (e.g., pitch, timbre, and harmony; see



Nattiez, 1975, for a more complete discussion of this problem). Perhaps the type of processing results from the interaction of the acoustic nature of a sound and its codability within a hierarchically organized system of sound, whether linguistic or not.

In conclusion, we do not question that levels of processing separate certain linguistic and nonlinguistic dimensions of the same stimuli. We suggest, rather, that the crux of the auditory-phonetic distinction is, as Fry suggested, a "physical-psychological transformation." The nature of this transformation probably cannot be accounted for solely in terms of the linguistic-nonlinguistic distinction. Instead it may be based on acoustic properties alone, on the coding of sounds within a hierarchically organized system, or on the interaction of acoustic properties with such a system.

## REFERENCES

- Biederman, I., & Checkosky, S. F. Processing redundant information. *Journal of Experimental Psychology*, 1970, 83, 486-490.
- Chaney, R. B., & Webster, J. C. Information in certain multidimensional sounds. *Journal of the Acoustical Society of America*, 1966, 40, 447-455.
- Chomsky, N., & Halle, M. *The sound pattern of English*. New York: Harper & Row, 1968.
- Cooper, F. S., & Mattingly, I. G. Computer controlled PCM system for investigation of dichotic speech perception. *Journal of the Acoustical Society of America*, 1969, 46, 115. (Abstract)
- Cooper, W. E. Selective adaptation to speech. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. Lindman, & D. B. Pisoni (Eds.), *Cognitive theory*. Hillsdale, N. J.: Erlbaum, 1975.
- Curry, F. K. W. A comparison of left-handed and right-handed subjects on verbal and non-verbal dichotic listening tasks. *Cortex*, 1967, 3, 343-352.
- 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 musiclike sounds: Implications for theories of speech perception. *Quarterly Journal of Experimental Psychology*, in press.
- Day, R. S., & Wood, C. C. Interactions between linguistic and nonlinguistic processing. *Journal of the Acoustical Society of America*, 1972, 51, 79. (Abstract)
- Eimas, P. D., Cooper, W. E., & Corbit, J. D. Some properties of linguistic feature detectors. *Perception & Psychophysics*, 1973, 13, 247-252.
- Eimas, P. D., & Corbit, J. D. Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, 1973, 4, 99-109.
- Fant, G. Auditory patterns of speech. In W. Wathen-Dunn (Ed.), *Models for the perception of speech and visual form*. Cambridge, Mass.: M.I.T. Press, 1967.
- Felfoldy, G. L., & Garner, W. R. The effects on speeded classification of implicit and explicit instructions regarding redundant dimensions. *Perception & Psychophysics*, 1971, 9, 289-292.
- Fry, D. B. Perception and recognition in speech. In M. Halle, H. G. Lunt, & C. H. van Schoonveld (Eds.), *For Roman Jakobson*, The Hague: Mouton, 1956.
- Garner, W. R. *The processing of information and structure*. Potomac, Md.: Erlbaum, 1974.
- Garner, W. R., & Felfoldy, G. L. Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, 1970, 1, 225-241.
- Garner, W. R., Hake, H. W., & Eriksen, C. W. Operationism and the concept of perception. *Psychological Review*, 1956, 63, 149-159.
- Jakobson, R., Fant, C. G. M., & Halle, M. *Preliminaries to speech analysis*. Cambridge, Mass.: M.I.T. Press, 1963.
- Kimura, D. Cerebral dominance and the perception of verbal stimuli. *Canadian Journal of Psychology*, 1961, 15, 166-171.
- Kimura, D. Left-right differences in the perception of melodies. *Quarterly Journal of Experimental Psychology*, 1964, 16, 355-358.
- Kuhn, G. M. On the front cavity resonance and its possible role in speech perception. *Journal of the Acoustical Society of America*, 1975, 58, 428-433.
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. Perception of the speech code. *Psychological Review*, 1967, 74, 431-461.
- Liberman, A. M., Delattre, P. C., Cooper, F. S., & Gerstman, L. J. The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs*, 1954, 68, 1-13.
- 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.
- Lisker, L., & Abramson, A. S. A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 1964, 20, 384-422.
- Miller, G. A. The magical number seven plus or minus two, or, some limits on our capacity for processing information. *Psychological Review*, 1956, 63, 81-97.
- 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: An example of categorical perception.

- Journal of the Acoustical Society of America*, in press.
- Morton, J. The use of correlated stimulus information in card sorting. *Perception & Psychophysics*, 1969, 5, 374-376.
- Nattiez, J. J. Sémiologie musicale: L'état de la question. *Acta Musicologica*, 1975, 42, 153-171.
- Nickerson, R. S. Binary-classification reaction time: A review of some studies of human information-processing capabilities. *Psychonomic Monograph Supplements*, 1971, 4 (17, Whole No. 65).
- Norman, D. A., & Bobrow, D. G. On data-limited and resource-limited processes. *Cognitive Psychology*, 1975, 7, 44-64.
- Posner, M. I. Information reduction in the analysis of sequential tasks. *Psychological Review*, 1964, 71, 491-504.
- Shankweiler, D. P., & Studdert-Kennedy, M. Identification of consonants and vowels presented to the left and right ears. *Quarterly Journal of Experimental Psychology*, 1967, 19, 59-63.
- Stevens, K. N., & Halle, M. Remarks on analysis by synthesis and distinctive features. In W. Wathen-Dunn (Ed.), *Models for the perception of speech and visual form*. Cambridge, Mass.: M.I.T. Press, 1967.
- Stevens, K. N., & House, A. S. Speech perception. In J. V. Tobias (Ed.), *Foundations of modern auditory theory* (Vol. 2). New York: Academic Press, 1972.
- Studdert-Kennedy, M. Speech perception. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics*. New York: Academic Press, in press.
- Studdert-Kennedy, M., Shankweiler, D. P., & Pisoni, D. B. Auditory and phonetic processes in speech perception: Evidence from a dichotic study. *Cognitive Psychology*, 1972, 3, 455-466.
- Townsend, J. T. A note on the identifiability of parallel and serial processes. *Perception & Psychophysics*, 1971, 10, 161-163.
- Turvey, M. On peripheral and central processes in vision: Inferences from an information-processing analysis of masking with patterned stimuli. *Psychological Review*, 1973, 80, 1-52.
- Wood, C. C. Parallel processing of auditory and phonetic information in speech perception. *Perception & Psychophysics*, 1974, 15, 501-508.
- 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. (a)
- Wood, C. C. A normative model for redundancy gain in speech perception. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. Lindman, & D. B. Pisoni (Eds.), *Cognitive theory*, Hillsdale, N. J.: Erlbaum, 1975. (b)
- Wood, C. C., & Day, R. S. Failure of selective attention to phonetic segments in consonant-vowel syllables. *Perception & Psychophysics*, 1975, 17, 346-350.

(Received June 2, 1975)