Perception & Psychophysics
1975, Vol. 17 (4), 346-350

Failure of selective attention to phonetic segments in consonant-vowel syllables

CHARLES C. WOOD and RUTH S. DAY
Yale University, New Haven, Connecticut 06520

Perception & Psychophysics
1975, Vol. 17 (4), 346-350

Failure of selective attention to phonetic segments in consonant-vowel syllables

CHARLES C. WOOD and RUTH S. DAY
Yale University, New Haven, Connecticut 06520

Subjects performed a two-choice speeded classification task that required selective attention to either the consonant or the vowel in synthetic consonant-vowel (CV) syllables. When required to attend selectively to the consonant, subjects could not ignore irrelevant variation in the vowel. Similarly, when required to attend selectively to the vowel, they could not ignore irrelevant variation in the consonant. These results suggest that information about an initial stop consonant and the following vowel is processed as an integral unit.

Experiments using both synthetic and natural speech stimuli have suggested that the acoustic cues for an initial stop consonant and the following vowel are interdependent (for reviews see Cole & Scott, 1974; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Stevens & House, 1972). Such interdependence of consonant and vowel information at the acoustic level may be associated with interactions between consonants and vowels at various stages of perceptual processing. For example, Strange, Verbrugge, and Shankweiler (1974) reported that the presence or absence of consonant context affected subjects' ability to identify vowel targets. Identification of vowels presented in either fixed or variable consonant context was significantly superior to identification of the same vowels presented in isolation. Pisoni and Tash (1974) compared "same-different" response times (RTs) for consonant targets and vowel targets in pairs of successive consonant-vowel (CV) syllables. For both types of target, RTs for correct "same" responses were significantly shorter when the stimulus pair shared both the consonant and vowel than when only the target segment was shared. Similarly, RTs for correct "different" responses were shorter when the stimulus pair differed in both the consonant and vowel than when they differed in only the target segment. One well-known interpretation of these and related results is that adjacent phonemes within a syllable are not processed as discrete and independent segments, but are processed as syllabic units. While there is continuing disagreement concerning the primary *psychological* units in speech perception (e.g., Cole & Scott, 1974; Liberman et al., 1967; Massaro, 1972; McNeill & Lindig, 1973; Savin & Bever, 1970; Wickelgren, in press), there is general agreement that phonetic information is transmitted by *acoustic* units of roughly syllable length (for a recent review, see Studdert-Kennedy, in press).

The present experiment attempted to provide a stringent test of the interdependence of consonant and vowel processing, and differs from previous experiments in two respects. First, previous experiments have typically used tasks that allowed or in some cases even encouraged subjects to capitalize on possible interactions between consonant and vowel information (cf. Strange et al., 1974). In contrast, the present experiment explicitly attempted to dissociate consonant and vowel processing by using a selective attention task. Second, the stimulus set employed in the present experiment was specifically restricted relative to those of previous experiments in order to optimize conditions for independent consonant and vowel processing. If interactions between consonant and vowel processing are obtained under these conditions, one may be reasonably confident that consonant-vowel interactions reflect a basic characteristic of speech processing, not simply an optional processing strategy available to the listener.

Subjects were required to attend selectively to either the consonant or vowel target dimension in synthetic CV syllables and to ignore irrelevant variation in the other (nontarget) dimension. An ability to attend selectively to the target dimension would imply some degree of dissociation or independence of consonant and vowel processing. In contrast, an inability to attend selectively to the target dimension would provide strong support for the hypothesis that information about an initial stop consonant and the following vowel is processed as an integral unit.

This research was supported in part by National Institute of Child Health and Human Development Grant HD-01994 to the Haskins Laboratories. A preliminary version of this paper was presented at the 83rd Meeting of the Acoustical Society of America (Day & Wood, 1972b). C. C. Wood's present address is Division of Neuropsychiatry, Walter Reed Army Institute of Research, Washington, D.C. 20012. R. S. Day is also associated with Haskins Laboratories, New Haven, Connecticut 06510.

Table I

Stimulus Sets for Each Target Dimension and Condition

Target Dimension	Condition: Control	Condition: Orthogonal
Stop Consonant	/ba/ /da/ or /bae/ /dae/	/ba/ /da/ /bae/ /dae/
Vowel	/ba/ /bae/ or /da/ /dae/	/ba/ /da/ /bae/ /dae/

METHOD

A two-choice speeded classification task similar to that used by Day and Wood (1972a) and Garner and Felfoldy (1970) was employed. Subjects received blocks of 64 trials in which either the consonant or the vowel was specified as target dimension. Each target dimension had two possible values, /b/ vs. /d/ for the consonant, and /a/ vs. /ae/ for the vowel. On each trial, subjects were required to identify as rapidly as possible which value of the target dimension had been presented by pressing one of two response buttons.

RTs for identification of the consonant and vowel targets were obtained in two different conditions: (a) a single-dimension *control* condition, in which the target dimension varied randomly with the nontarget dimension held constant; and (b) a two-dimension *orthogonal* condition, in which both the target dimension and the nontarget dimension varied randomly and independently. For both target dimensions, the only difference between the control and orthogonal conditions was the presence or absence of irrelevant variation in the nontarget dimension. Thus, a comparison of the RTs from these two conditions indicates the degree to which the consonant and vowel dimensions could be processed independently of irrelevant variation in the other dimension.

Eight Yale University undergraduates served as paid volunteers. Each received two blocks of 64 trials in each of the four conditions, with presentation order counterbalanced across subjects. The stimulus sets for each condition are shown in Table 1. In each orthogonal condition, Table 1 shows that the stimulus sets for both blocks of 64 trials were identical. In the control conditions, however, there were two possible sets for each condition. Each subject received each of the control stimulus sets shown in Table 1 in separate blocks of trials. By constructing the stimulus sets in this manner, each of the four stimuli was presented equally often both within and between conditions. Before beginning the actual experiment, subjects received practice in all four conditions. In addition, they were given at least eight practice trials immediately preceding each block of trials in order to familiarize them with the condition they would receive on that block.

The stimuli were three-formant synthetic syllables generated by the Haskins Laboratories parallel resonance synthesizer. The syllables containing /b/ vs. /d/ differed only in the initial segments of the second and third formant transitions, while those containing /a/ vs. /ae/ differed in the steady-state formant frequencies (/a/: 718, 1,075, and 2,525 Hz; /ae/: 666, 1,695, and 2,348 Hz). All stimuli were prepared to be equal in acoustic parameters other than those explicitly varied for experimental purposes; they were 300 msec in duration, with a fundamental frequency of 104 Hz and falling fundamental frequency and amplitude contours. The stimuli were presented binaurally at 5-sec interstimulus intervals from an Ampex AG-500 tape recorder through Grason-Stadler TDH-39 earphones. Signals synchronized with stimulus onset and subjects' responses were recorded on a Harmon-Kardon CAD-4 cassette recorder for off-line RT computation by a Hewlett-Packard counter-timer and digital printer.

RESULTS AND DISCUSSION

Mean RTs for identification of the consonant and vowel target dimensions are shown for the control and orthogonal conditions in Table 2, together with the error rate for each condition. The RTs in the orthogonal condition were substantially longer than those in the control condition for both dimensions: 50 msec longer for consonants and 66 msec for vowels. The results of a factorial analysis of variance showed that the main effects of condition and target dimension were highly significant ($F = 215.60$, $df = 1,127$, $p < .001$; and $F = 122.37$, $df = 1,127$, $p < .001$, respectively). In addition, the Condition by Target Dimension interaction was also significant ($F = 4.18$, $df = 1,127$, $p < .05$). The error rates did not differ significantly among conditions.

The large RT differences between control and orthogonal conditions indicate that subjects were unable to attend selectively to either the consonant or the vowel and ignore the other dimension. These results suggest that information about an initial stop consonant and the following vowel is processed as an integral unit. The RT differences between consonant and vowel dimensions and the Condition by Dimension interaction are discussed below.

Integral and Separable Dimensions

Failures of selective attention similar to those obtained with speech stimuli in the present experiment have been obtained for a variety of visual and auditory dimensions (Garner, 1974; Garner & Felfoldy, 1970; Wood, 1975). Garner and Felfoldy (1970) used the term "integral" to refer to pairs of dimensions yielding this pattern of results. Operationally, integral dimensions "...produce a redundancy gain when the dimensions are correlated and some measure of speed or accuracy of discrimination is used, and produce interference in speed of classification when selective attention is

Table 2
Mean Response Time and Error Rate for Each Target Dimension and Condition

Target Dimension	Condition: Control	Condition: Orthogonal
Stop Consonant	400 (3.8)	450 (3.6)
Vowel	348 (3.4)	414 (3.9)

Note—Response times are given in milliseconds and error rates in percent.

required with orthogonal stimulus dimensions" (Garner & Felfoldy, 1970; p. 328). The present results support those of Wood (1975) in suggesting that integral dimensions are limited neither to the visual modality studied by Garner and Felfoldy (1970) nor to pairs of relatively "elementary" stimulus continua such as brightness and hue or pitch and loudness. Rather, the finding that abstract phonetic segments appear to function as integral dimensions lends additional support and generality to the concept of dimensional integrality in information processing.

Linguistic and Nonlinguistic Dimensions

Not all dimensions have produced results corresponding to a strict dichotomy between integral and separable dimensions. For example, Day and Wood (1972a) and Wood (1974, 1975) found that irrelevant variation in a nonlinguistic dimension of a CV syllable (fundamental frequency) produced substantial interference with identification of a linguistic dimension (the initial stop consonant). In contrast, when subjects were required to identify fundamental frequency, there was only slight interference from irrelevant variation in the stop consonant. This unidirectional or asymmetric interference between linguistic and nonlinguistic dimensions is consistent with other data that distinguish between auditory and phonetic levels of processing in speech perception (for a recent review, see Studdert-Kennedy, in press). As is the case in a number of other experimental paradigms (e.g., Day & Bartlett, 1972; Day, Cutting, & Copeland, 1971; Wood, 1975; Wood, Goff, & Day, 1971), the linguistic status of a dimension may be an important factor in speeded classification experiments.

Temporal Factors

The temporal aspects of speech stimuli introduce an important new variable into the concept of dimensional integrality that has not been systematically investigated. The visual stimulus dimensions in the Garner and Felfoldy (1970) experiments were always presented simultaneously at stimulus onset and did not vary over time. However, with speech stimuli, the identity of each phonetic segment is often cued by multiple acoustic events and the cues for different phonetic segments can be both overlapping (simultaneous) and nonoverlapping (successive) in time. Therefore, the dimension of time becomes an important independent variable to be investigated in such experiments. In the present experiment, acoustic information about both the consonant and vowel segments was simultaneously available from the onset of the syllable. Liberman (1970; Liberman et al., 1967) has suggested that such "parallel transmission" of information about an initial stop consonant and the following vowel provides an opportunity for parallel processing of consonant and vowel information at some stage of perceptual analysis. The results of the present experiment and those of Pisoni and Tash (1974) strongly support this interpretation.[1]

However, it is also of interest to consider stimuli in which the acoustic cues for the target phonemes are not simultaneously available from stimulus onset. For example, consider the case of syllables identical to those used in the present experiment but reversed in time to form VC syllables. Since the formant transition cues for the stop consonant would occupy approximately the final 50 msec, 250 msec of "pure" vowel information would be available for vowel identification before the onset of any consonant information. Under these circumstances, vowels might well be identified with minimal interference from the irrelevant variation in the consonant. As the duration of the vowel is decreased, however, greater interference from irrelevant variation in the consonant might be expected to occur. Thus, the present results should not be generalized beyond the particular phoneme targets and temporal relationships employed in this experiment. By systematic variation of the syllable structure and phoneme targets, this selective attention paradigm may be used to investigate interactions between phonetic segments during perceptual processing.

Stimulus Discriminability

In addition to the temporal relationship between the acoustic cues for each target segment, the role of stimulus discriminability in this and related experiments must also be considered. In the present experiment, the significant main effect of target dimension indicates that vowels were identified faster than the consonants regardless of experimental condition (Table 1). In the control condition, vowels were identified 52 msec faster and in the orthogonal condition 36 msec faster. This result is consistent with the backward masking data reported by Pisoni (1972), in which identification of the vowel in a CV syllable was impaired by a masking stimulus over much shorter target-mask intervals than the identification of the consonant of the same syllable. One possible interpretation of these results is that vowel information is extracted from a syllable "before" information about the initial consonant, possibly due to additional stages of processing for the consonant. For example, based on faster RTs for detection of syllable targets than phoneme targets, Savin and Bever (1970) concluded that "syllables are perceived before their constituent phonemes" (p. 299). However, such interpretations overlook the potentially important role of stimulus discriminability in RT and related experiments. Since RT decreases with increasing stimulus discriminability (e.g., Imai & Garner, 1965; Thurmond & Alluisi, 1963), and since

vowels are typically more discriminable than stop consonants according to a signal-to-noise ratio criterion (cf. Miller & Nicely, 1955; Pickett, 1957), it would appear more parsimonious to interpret the faster vowel RTs in terms of their greater discriminability.[2] Similar considerations concerning stimulus discriminability apply to the backward masking data, as well as a variety of other paradigms currently used in speech perception experiments. For example, Darwin and Baddeley (1974) have emphasized the importance of relative discriminability in investigations of short-term memory for speech stimuli. Additional experiments are required to determine whether processing times for consonants and vowels differ for reasons other than differences in stimulus discriminability.

In summary, the results of the present experiment indicate that subjects could not attend selectively to either the consonant or the vowel in synthetic CV syllables. These results suggest that information about an initial stop consonant and the following vowel is processed as an integral unit. Together with the "redundancy gain" approach of Wood (1974, in press) and the same-different RT experiments of Pisoni and Tash (1974), the selective attention paradigm appears to be a useful tool for investigating interactions between phonetic segments during perceptual processing.

REFERENCES

Cole, R. A., & Scott, B. Toward a theory of speech perception. *Psychological Review*, 1974, **81**, 348-394.

Darwin, C. J., & Baddeley, A. D. Acoustic memory and the perception of speech. *Cognitive Psychology*, 1974, **6**, 41-60.

Day, R. S., & Bartlett, J. C. Separate speech and nonspeech processing in dichotic listening? *Journal of the Acoustical Society of America*, 1972, **51**, 79A.

Day, R. S., Cutting, J. C., & Copeland, P. M. Perception of linguistic and nonlinguistic dimensions of dichotic stimuli. *Haskins Laboratories Status Report on Speech Research*, 1971, **SR-27**, 193-197.

Day, R. S., & Wood, C. C. Interactions between linguistic and nonlinguistic processing. *Journal of the Acoustical Society of America*, 1972, **51**, 79A. (a)

Day, R. S., & Wood, C. C. Mutual interference between two linguistic dimensions of the same stimuli. *Journal of the Acoustical Society of America*, 1972, **52**, 175A. (b)

Garner, W. R. The stimulus in information processing. *American Psychologist*, 1970, **25**, 350-358.

Garner, W. R. Information integration and form of encoding. In A. W. Melton and E. Martin (Eds.), *Coding processes in human memory*. Washington: Winston, 1972.

Garner, W. R. *The processing of information and structure*. Potomac, Md: Lawrence Erlbaum, 1974.

Garner, W. R., & Felfoldy, G. L. Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, 1970, **1**, 225-241.

Imai, S., & Garner, W. R. Discriminability and preference for attributes in free and constrained classification. *Journal of Experimental Psychology*, 1965, **69**, 596-608.

Liberman, A. M. The grammars of speech and language. *Cognitive Psychology*, 1970, **1**, 301-323.

Liberman, A. M., Cooper, F. S., Shankweiler, D., & Studdert-Kennedy, M. Perception of the speech code. *Psychological Review*, 1967, **74**, 431-461.

Lockhead, G. R. Processing dimensional stimuli: A note. *Psychological Review*, 1972, **79**, 410-419.

Massaro, D. W. Preperceptual images, processing time, and perceptual units in auditory perception. *Psychological Review*, 1972, **79**, 124-155.

McNeill, D., & Lindig, K. The perceptual reality of phonemes, syllables, words, and sentences. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 419-430.

Miller, G. A., & Nicely, P. An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, 1955, **27**, 338-352.

Pickett, J. M. Perception of vowels heard in noises of various spectra. *Journal of the Acoustical Society of America*, 1957, **29**, 613-620.

Pisoni, D. B. Perceptual processing time for consonants and vowels. *Haskins Laboratories Status Report on Speech Research*, 1972, **SR-31/32**, 83-92.

Pisoni, D. B., & Tash, J. "Same-different" reaction times to consonants, vowels, and syllables. *Journal of the Acoustical Society of America*, 1974, **55**, 436(A).

Savin, H. B., & Bever, T. G. The nonperceptual reality of the phoneme. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 295-302.

Stevens, K. N., & House, A. S. Speech perception. In J. V. Tobias (Ed.), *Foundations of modern auditory theory*. New York: Academic Press, 1972. Pp. 1-62.

Strange, W., Verbrugge, R., & Shankweiler, D. Consonantal environment specifies vowel identity. *Journal of the Acoustical Society of America*, 1974, **55**, S54.

Studdert-Kennedy, M. Speech perception. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics*. Springfield, Ill: Thomas (in press).

Thurmond, J. B., & Alluisi, E. A. Choice time as a function of stimulus dissimilarity and discriminability. *Canadian Journal of Psychology*, 1963, **17**, 326-337.

Wickelgren, W. A. Phonetic coding and serial order. In E. C. Carterette and M. P. Friedman (Eds.), *Handbook of perception*. (Vol. 7) *Language and speech*. New York: Academic Press, in press.

Wood, C. C. Parallel processing of auditory and phonetic information in speech discrimination. *Perception & Psychophysics*, 1974, **55**, 501-508.

Wood, C. C. A normative model for redundancy gains in speech discrimination. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. Lindman, and D. B. Pisoni (Eds.), *Cognitive theory* (Vol. I). Potomac, Md: Lawrence Erlbaum, in press.

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, **104**, 3-20.

Wood, C. C., Goff, W. R., & Day, R. S. Auditory evoked potentials during speech perception. *Science*, 1971, **171**, 1248-1251.

NOTES

1. There is some degree of disagreement concerning whether results such as those of Pisoni and Tash (1974) and the present experiment are best interpreted in terms of dimensional integrality or parallel processing (see discussions by Garner, 1970, 1972, 1974; Garner & Felfoldy, 1970; Lockhead, 1972). A distinction between the integrality and parallel processing interpretations of such interdependence may be feasible based on comparison of empirical RT distributions with those predicted by each interpretation. For a more detailed discussion of this point, see Wood (in press).

2. The significant Condition by Target Dimension interaction indicates that irrelevant variation in the stop consonant produced slightly greater interference with vowel identification than the reverse. The same considerations about stimulus discriminability and temporal relationship of acoustic cues discussed above also apply to this result.

(Received for publication October 14, 1974; accepted November 15, 1974.)