Perception & Psychophysics
1990, 47 (4), 379-390

Stimulus variability and processing dependencies in speech perception

JOHN W. MULLENNIX and DAVID B. PISONI
Indiana University, Bloomington, Indiana

Processing dependencies in speech perception between voice and phoneme were investigated using the Garner (1974) speeded classification procedure. Variability in the voice of the talker and in the cues to word-initial consonants were manipulated. The results showed that the processing of a talker's voice and the perception of voicing are asymmetrically dependent. In addition, when stimulus variability was increased in each dimension, the amount of orthogonal interference obtained for each dimension became significantly larger. The processing asymmetry between voice and phoneme was interpreted in terms of a parallel-contingent relationship of talker normalization processes to auditory-to-phonetic coding processes. The processing of voice information appears to be qualitatively different from the encoding of segmental phonetic information, although they are not independent. Implications of these results for current theories of speech perception are discussed.

The production of human speech is characterized by a large number of individual differences between talkers. Such factors as structural differences in vocal tract size and shape (Fant, 1973; Joos, 1948; Peterson & Barney, 1952), glottal characteristics (Carr & Trill, 1964; Carrell, 1984; Monsen & Engebretson, 1977), and dynamic articulatory control (Ladefoged, 1980), and so forth, manifest themselves in the speech waveform in terms of a wide variety of acoustic differences between talkers. One of the major issues in speech perception concerns the manner in which the acoustic differences between talkers are processed in perceiving spoken language. It is likely that several processes and/or mechanisms are involved in perceptual compensation for voice information. Some researchers have characterized such likely processes as "normalizing" or "adjusting" the acoustic differences between talkers (e.g., Summerfield, 1975; Summerfield & Haggard, 1973). However, the manner in which these processes operate has not been clearly described, nor has a precise characterization of such processes been developed. Although some research has been devoted to this problem, for the most part, the perceptual consequences of these compensatory processes have received little attention. In most studies in speech perception over the last 40 years, researchers have used speech produced by only one talker. Indeed, only one token of each utterance is often used. Such severe limits on stimulus material prevent any systematic assessment of the role of stimulus variability in speech perception.

There has been some research examining the perceptual consequences of processing differences between talkers in studies of vowel and consonant perception (Assmann, Nearey, & Hogan, 1982; Fourcin, 1968; Rand, 1971; Verbrugge, Strange, Shankweiler, & Edman, 1976; Weenink, 1986), word recognition (Creelman, 1957; Mullennix, Pisoni, & Martin, 1989), and memory (Martin, Mullennix, Pisoni, & Summers, 1987). These studies have shown that changes in the voice of the talker from trial to trial within an experiment produce reliable decrements in overall task performance. The presence of these effects can be interpreted in terms of a "processing cost" to the perceptual system that is induced by the stimulus variability in the talker's voice. For instance, in one recent study, Mullennix et al. (1989) examined the effects of talker variability on spoken word recognition. In several experiments, using perceptual identification and word naming tasks, we found that recognition was significantly worse for words produced by different talkers than for the same words produced by a single talker. Furthermore, the decrement in performance caused by talker variability increased when the acoustic information in the speech signal was degraded by using a special distortion technique. Because perceptual performance was consistently worse when the words were produced by different talkers, we argued that a resource-demanding perceptual mechanism was used to compensate for acoustic differences between talkers. In addition, because these effects were greater when the early acoustic information in the signal was disrupted, we suggested that the processing of voice information appears to be closely related to processes involved in the early perceptual encoding of the input sig-

The research reported here was supported, in part, by NIH Research Grant NS-12179-11 and, in part, by NIH Training Grant NS07134-09, to Indiana University in Bloomington, IN. The authors would like to thank our colleague Linda B. Smith and two anonymous reviewers for comments on an earlier version of the paper and Luis Hernandez for programming assistance. John Mullennix is now at the Department of Psychology, Wayne State University, Detroit, MI 48202. Reprints can be obtained from either author; David B. Pisoni's address is: Speech Research Laboratory, Department of Psychology, Indiana University, Bloomington, IN 47405.

nal into an initial phonetic representation. These results provide a first step at characterizing the nature of talker-related perceptual processes. However, the relationship of these processes to other phonetic coding processes and to the higher level processes involved in spoken word recognition and lexical access are largely unknown and remain a topic for additional investigation. This paper is intended as a first step in that direction.

One important issue concerning "talker normalization" processes is their relation to the auditory-to-phonetic coding processes of speech. Do the perceptual processes used to encode voice information function independently of processes that are used to encode phonetic information in the speech signal? Or, are talker normalization processes and phonetic coding processes interrelated? A major objective of the present study was to investigate the relationship of talker normalization processes and auditory-to-phonetic coding processes and assess their interactions. One way to determine whether perceptual processes are related to one another is to assess whether stimulus dimensions relevant to both types of processes are perceived independently of one another or whether there is some dependency relation between them. In the present study, we examined the processing relations between talker normalization and auditory-to-phonetic coding processes. We used the speeded classification technique, which was specifically designed for the study of processing interactions between stimulus dimensions (see Garner, 1974).

One hypothesis that has been proposed to account for talker variability effects is that a resource-limited talker normalization process is involved in encoding (Mullennix et al., 1989; see also Nusbaum & Morin, 1989). Mullennix et al. (1989) suggested that perceptual deficits due to changes in a talker's voice occur because of competition for processing resources used by talker normalization processes and other perceptual processes involved in speech perception. It is conceivable that each time a different voice is encountered, resources must be allocated to talker normalization processes until speaker-dependent perceptual operations are completed. If this is the case, perceptual deficits may arise from the additional processing load induced by changes in the voice of the talker from trial to trial. If selective attention to phonetic coding processes is interfered with by processes involved in talker normalization, then the effects of talker variability may be intimately dependent on the role of selective attention in speech perception. By examining the processing interactions between phonetic and speaker-related stimulus dimensions, we hoped to obtain further information about the role of selective attention in speech perception and spoken word recognition and to assess the interactions of these two stimulus dimensions.

In previous studies with perceptual identification and naming tasks, it has been found that trial-to-trial variability in the voice of the talker produces significant decrements in word recognition performance (Creelman, 1957; Mullennix et al., 1989). In the present study, the voice of the talker and the acoustic-phonetic composition of word-initial consonants were manipulated in a speeded classification task so that latency measures could be obtained. If trial-to-trial changes in variability have detrimental effects on performance on this task, the results would provide additional evidence that stimulus variability from trial to trial produces significant perceptual effects on spoken word recognition. By manipulating the variability in both stimulus dimensions, we hoped to obtain further information about the potential interactions of these two variables.

In order to examine the nature of any processing dependencies between talker normalization and auditory-to-phonetic coding processes, and to assess the extent to which talker normalization processes are related to selective attention, a modified version of the selective attention procedure described by Garner (1974) was used. Over the years, this procedure has been adopted by a number of researchers to examine processing dependencies between auditory and phonetic dimensions (Blechner, Day, & Cutting, 1976; Carrell, Smith, & Pisoni, 1981; Eimas, Tartter, Miller, & Keuthen, 1978; Miller, 1978; Pastore et al., 1976; Tomiak, Mullennix, & Sawusch, 1987; Wood, 1974; Wood & Day, 1975). These studies have shown that certain stimulus dimensions relevant to speech are processed as integral dimensions, often displaying a mutual dependence on each other.

The experimental procedure developed by Garner (1974) involves the use of a two-choice speeded classification task. Subjects are required to attend selectively to one stimulus dimension while simultaneously ignoring another stimulus dimension. Two stimulus dimensions are combined in various ways to form three types of stimulus sets: a control set, an orthogonal set, and a correlated set. In the control set, the unattended dimension is held constant while the attended dimension varies randomly. The control set for each dimension provides a baseline measure for classifying each dimension and permits one to assess whether both dimensions are, a priori, equally discriminable. In the orthogonal set, both the attended and the unattended dimensions vary randomly. The degree to which response latencies increase from the control set to the orthogonal set for each dimension indicates the extent to which the stimulus dimensions are processed separably or in an integral fashion. If stimulus dimensions are classified as quickly in the orthogonal conditions as they are in the control conditions, then the stimulus dimensions are said to be processed independently. That is, decisions about the relevant dimension are unaffected by the irrelevant dimension. However, if there is a significant increase in response latencies from the control conditions to the orthogonal conditions, the stimulus dimensions are said to be processed in a dependent manner. Apparently, subjects cannot ignore or "filter out" variation in the irrelevant dimension. This result, which is termed *orthogonal interference*, indicates that a failure of selective attention to the attended dimension has occurred. Finally, in the correlated condition, one particular value of one dimension is always paired with another particular value

of the other dimension. The presence of decreased response latencies in this condition as opposed to the control condition is called a *redundancy gain*. A redundancy gain indicates that the information in the nonattended stimulus dimension can be used to facilitate perceptual classification of the attended dimension. Although the presence of a redundancy gain can be interpreted as further evidence for integrality of dimensions (see Garner, 1974; Garner & Felfoldy, 1970), it is best thought of as additional evidence, and it is not absolutely crucial for making assertions about integral processing. However, under certain circumstances, the presence of redundancy gains can provide important diagnostic evidence regarding the serial/parallel nature of the processes involved (Wood, 1974, 1975) or the presence of a selective serial processing strategy (Biederman & Checkosky, 1970; Felfoldy & Garner, 1971).

In the present study, the processing relationships between talker normalization and phonetic coding were examined by manipulating the talker's voice and the cues to phonetic categorization. To avoid confusion, the two stimulus factors selected were called the *voice* factor and the *word* factor. The voice factor involved variations in the gender of the talker (i.e., male vs. female). The word factor involved variations in the phonetic feature of voicing (/b/ vs. /p/) in the initial position in English words. When subjects were required to attend to voice, the required response was "male voice" or "female voice"; when the subjects were required to attend to the word, the required response was "b" or "p." By examining subjects' performance in classifying the stimuli during the selective attention procedure, we hoped to assess the degree to which the two stimulus dimensions are processed independently. In this way, we hoped to gain some insight into the nature of talker normalization and its relation to the phonetic coding of speech.

In this study, we also investigated the effects of stimulus variability in speech perception. Word variability and talker variability were manipulated, by changing the composition of the orthogonal stimulus set: Word variability was manipulated by increasing the number of different words used within the orthogonal set; talker variability was manipulated by increasing the number of male and female talkers who produced the words used within the orthogonal set. Through comparison of the amount of orthogonal interference obtained across such conditions, the effects of stimulus variability on speeded classification performance could be assessed for both stimulus dimensions.

A number of predictions about the outcome of the first experiment can be made. First, we consider the processing of word and voice dimensions. If there is no increase in response latencies from the control condition to the orthogonal condition for attending to either the word or the voice dimensions, this pattern of results would suggest that word and voice dimensions are processed independently of one another. However, if there are significant increases in response latencies from control to orthogonal conditions for both stimulus dimensions, this would suggest that the voice and word dimensions are processed in a mutually dependent manner. These results would also imply that auditory-to-phonetic coding processes and talker normalization processes are highly interrelated. If redundancy gains are obtained for either dimension, this would provide further evidence of a processing dependency and would permit one to conclude that the two processes may operate in parallel. The presence of processing dependencies in these conditions would be consistent with the idea that the processing of voice information in speech is mandatory and requires selective attention.

Second, we consider the effects of increasing the amount of stimulus variability within each dimension. If the difference in response latencies between control and orthogonal conditions becomes greater as stimulus variability on that dimension increases, this result would suggest that increases in stimulus variability produce greater demands on selective attention and/or processing time. This outcome would provide further support for the proposal that the effects of talker variability observed in our previous studies are related to changes in selective attention to phonetically relevant information in the speech signal.

EXPERIMENT 1

Method

Subjects. Seventy-two undergraduate students enrolled in introductory psychology courses at Indiana University volunteered to be subjects. Each subject took part in one 1-h session and received partial course credit for participating in the experiment. All subjects were native speakers of English who reported no history of a speech or hearing disorder at the time of testing.

Stimulus Materials. The stimuli consisted of 16 naturally spoken English words obtained from eight male and eight female talkers, all of whom spoke with a midwestern dialect. The stimuli were English monosyllabic words selected from the corpus of words used in the modified rhyme test (House, Williams, Hecker, & Kryter, 1965). One half of the words began with the consonant *b*, and one half of the words began with the consonant *p*. Each talker's utterances were recorded on audiotape in a sound-attenuated booth (IAC Model 401A), using an Electro-Voice Model D054 microphone and a Crown 800 series tape recorder. Each stimulus item was pronounced in citation format in unique randomized lists for each talker. The words were subsequently converted to digital form via a 12-bit analog-to-digital converter at a 10-kHz sampling rate and then stored as digital files. The target words were digitally edited to produce the final experimental materials used in the study. RMS amplitude levels among words were digitally equated, using a software package designed to modify digital waveforms. All of the words in the experiment had been previously tested for intelligibility in a separate experiment, using a different group of listeners. All items received identification scores of 95% or above when presented in isolation. In the present study, items were carefully selected so that the stimuli used across different sets were equated in terms of mean intelligibility scores. This was done in order to avoid any possible confounds that could arise from uncontrolled intelligibility differences across sets.

Procedure. Three experimental factors were manipulated: stimulus dimension, stimulus set condition, and stimulus variability. Stimulus dimension was manipulated within subjects, by requiring subjects to attend either to the word or to the voice when they classified each stimulus item. Stimulus set condition was manipulated within subjects, by presenting the stimuli in a control set, an or-

thogonal set, or a correlated set. Stimulus variability was manipulated between subjects, by varying the composition of the orthogonal stimulus sets to create four experimental conditions. In the 2W×2T condition, the orthogonal set contained 2 words spoken by 2 talkers. In the 4W×4T condition, the orthogonal set contained 4 words (2 *b* words, 2 *p* words) spoken by 4 talkers (2 male, 2 female). In the 8W×8T condition, the orthogonal set contained 8 words (4 *b* words, 4 *p* words) spoken by 8 talkers (4 male, 4 female). And, in the 16W×16T condition, the orthogonal set contained 16 words (8 *b* words, 8 *p* words) spoken by 16 talkers (8 male, 8 female). In the first three conditions, all words spoken by all talkers were presented in the experiment. However, in the 16W×16T condition, a subset of words spoken by different talkers was used, in order to keep the number of trials the same as in the other three conditions. Thus, in the 16W×16T condition, all 16 words appeared and all 16 talkers appeared, but any 1 word was only spoken by 4 talkers, and all of the 16 talkers only produced 4 different words (see Table 1). With the assignment of talkers to words in this manner, the increase in variability from the 8W×8T condition to the 16W×16T condition was not directly analogous to the increases in variability observed from condition to condition for the 2W×2T, 4W×4T, and 8W×8T conditions.[1]

The subjects were divided equally into groups and randomly assigned to the four experimental conditions. Depending on the particular condition, subjects were required to attend to either the word or the voice in order to make a response. For the word condition, the subjects classified the stimulus as beginning with either an initial *b* or *p* consonant. For the voice condition, the subjects classified the stimulus as to whether it was spoken by a male or a female talker.

Table 1
List of Words Used in the Orthogonal Stimulus Sets for Each Stimulus Variability Condition as a Function of Talker

Condition	Word	Male Talker	Female Talker
2W×2T	bad	1	1
	pad	1	1
4W×4T	bad	1,2	1,2
	buff	1,2	1,2
	pad	1,2	1,2
	puff	1,2	1,2
8W×8T	bad	1,2,3,4	1,2,3,4
	buff	1,2,3,4	1,2,3,4
	beach	1,2,3,4	1,2,3,4
	bill	1,2,3,4	1,2,3,4
	pad	1,2,3,4	1,2,3,4
	puff	1,2,3,4	1,2,3,4
	peach	1,2,3,4	1,2,3,4
	pill	1,2,3,4	1,2,3,4
16W×16T	bad	1,2	3,4
	buff	2,3	4,5
	beach	3,4	5,6
	bill	4,5	6,7
	back	5,6	7,8
	beak	6,7	8,1
	bit	7,8	1,2
	buck	8,1	2,3
	pad	3,4	1,2
	puff	4,5	2,3
	peach	5,6	3,4
	pill	6,7	4,5
	pack	7,8	5,6
	peak	8,1	6,7
	pit	1,2	7,8
	pun	2,3	8,1

Note—The particular talkers are denoted by a talker number corresponding to one of the eight male talkers or one of the eight female talkers under their respective categories.

For each of the four stimulus variability conditions, the subjects received three sets of trials: control trials, correlated trials, and orthogonal trials. Thus, all subjects received three sets of trials in which they classified stimuli by word and three sets of trials in which they classified stimuli by voice. In all of the control conditions, the target stimulus dimension was varied while the irrelevant dimension was held constant. For example, one control set for the word dimension consisted of the words *bad* and *pad* spoken in a male voice, while the other control set for the word dimension consisted of the words *bad* and *pad* spoken in a female voice. Each control set always contained two stimuli only. In the correlated conditions, one value along the target dimension was always correlated with a unique value along the irrelevant dimension. For example, one correlated set consisted of *bad* in the male voice and *pad* in the female voice, while the other correlated set consisted of *bad* in a female voice and *pad* in a male voice. The correlated conditions also contained only two stimuli. In the orthogonal conditions, the stimulus dimensions varied independently. In these sets, all *b* and *p* words were presented in both male and female voices. The composition of the orthogonal sets varied across the four stimulus variability conditions.

The stimuli used in the control and correlated sets across all stimulus variability conditions were identical. These stimulus sets were formed by selecting the appropriate stimuli for each set from the words *bad* and *pad* spoken by one male talker and one female talker. However, the stimuli used in the orthogonal sets differed for the variability conditions (see Table 1).

Subjects received a total of six stimulus sets per session. The control, correlated, and orthogonal conditions were presented once for classification by the voice dimension and once again for classification by the word dimension. The subjects classified the first three sets in each session for one stimulus dimension and then classified the last three sets for the other stimulus dimension. The order of dimensions was counterbalanced across subjects and the order of stimulus sets was counterbalanced by means of a Latin square design. Half of the subjects received a word dimension control condition consisting of the words *bad* and *pad* spoken in a male voice and half of the subjects received a word dimension control condition consisting of the words *bad* and *pad* spoken in a female voice. In addition, half of the subjects received a voice dimension control condition consisting of the word *bad* spoken in male and female voices and half of the subjects received a voice dimension control condition consisting of the word *pad* spoken in male and female voices.

Within each stimulus set, 64 randomized test trials occurred. For the control and correlated sets, 32 repetitions of 2 stimuli were used. For the orthogonal sets, 16 repetitions of each stimulus occurred in the 2W×2T condition, 4 repetitions of each stimulus in the 4W×4T condition, and 1 repetition of each stimulus in the 8W×8T and 16W×16T conditions. Before each set of test trials, a set of 12 practice trials was presented to familiarize subjects with the experimental procedures and the specific condition. The 12 practice trials consisted of 12 stimulus items randomly selected from the set of test trials subsequently presented. Six items were drawn from each response category.

The stimuli were presented binaurally over matched and calibrated TDH-39 headphones to the subject at a listening level of 80 dB SPL. The subjects were run in small groups, in sound-treated booths containing headphones and two-button response boxes. The subjects were instructed to respond as quickly and as accurately as possible by pushing one of two buttons on a computer-controlled response box in front of them. A warning light was illuminated before the presentation of each stimulus. For the practice trials, after all subjects made a response, feedback was provided about the correct alternative for that trial, with the illumination of a light located above the response button corresponding to the correct choice. The subjects did not receive feedback during any of the test trials. Presentation of each stimulus occurred 3 sec after all subjects had made

a response or 3 sec after a 2-sec response interval had elapsed. A 15-sec interval occurred between each practice set and the appropriate test set. A 1-min rest period was inserted after each test set. Stimulus-to-response button assignment was counterbalanced across subjects. Identification accuracy and response latencies were recorded for all trials. Responses over 2,000 msec were scored as incorrect and eliminated from subsequent analysis. Response latencies were measured from stimulus onset. Stimulus presentation and data collection were controlled on-line by a PDP-11/34A computer.

Results

The data were analyzed in terms of overall percent correct identification and response latencies. For each subject, mean percent correct and mean response latencies were calculated over each of the stimulus set conditions for each stimulus dimension. Response latencies were analyzed for correct responses only.

Response latencies. Table 2 displays the mean response latencies and standard deviations collapsed over subjects for the control, orthogonal, and correlated conditions for the word and voice dimensions for each of the four stimulus variability conditions. The individual response latencies were plotted using an analysis program that estimated the normality of the response time (RT) distribution for each condition. The data indicated that the RT distributions in all conditions were approximately normal. Thus, the following data analyses are based on subjects' mean response latencies.[2]

A four-way ANOVA was conducted on the latency data for the factors of stimulus dimension, stimulus set, stimulus variability, and set order. A significant main effect of stimulus dimension was obtained [$F(1,48) = 12.8$, $p < .001$]. Response latencies were faster for classifying the voice dimension than the word dimension (493.3 msec for the voice and 521.4 msec for the word dimensions). A significant main effect of stimulus set was also obtained [$F(2,96) = 185.5$, $p < .001$]. Latencies were fastest in the correlated condition, slower in the control condition, and slowest in the orthogonal condition. This is the general pattern observed when there are processing dependencies between stimulus dimensions.

Newman-Keuls post hoc tests revealed that performance in the orthogonal condition differed significantly from performance in the control and correlated conditions. A significant interaction of stimulus dimension with stimulus set was obtained [$F(2,96) = 15.5$, $p < .001$]. Post hoc tests of this interaction revealed that performance in the orthogonal condition differed as a function of stimulus dimension, while performance in the control and correlated conditions did not. That is, performance in the orthogonal condition was much slower when the relevant dimension was word and the irrelevant dimension was voice than vice versa. A significant interaction of stimulus set with stimulus variability condition was observed [$F(6,96) = 7.0$, $p < .001$]. Post hoc tests revealed that performance in the orthogonal condition in the 2W×2T condition differed significantly from performance in the orthogonal conditions of the 4W×4T, 8W×8T, and 16W×16T conditions. Finally, a significant effect of order was found [$F(5,48) = 4.2$, $p < .01$]. The ordering of the stimulus sets within each session had a substantial effect on overall performance, with mean RT performance as a function of order ranging from 433.4 to 592.2 msec. No other significant differences between conditions were observed.

These analyses indicate that response latencies varied reliably as a function of the relevant stimulus dimension attended to and as a function of the stimulus set condition. In order to examine the effects of stimulus set condition on response latencies more closely, a series of one-way ANOVAs was conducted between the control conditions and the orthogonal and correlated conditions for each dimension in all four stimulus variability conditions.

First, we consider the response latencies for the 2W×2T condition. For both the word and the voice dimensions, the increase in latencies from the control condition to the orthogonal condition was significant [$F(1,17) = 8.5$, $p < .01$, and $F(1,17) = 6.9$, $p < .02$, respectively]. This result indicates that when subjects attend to either dimension, they cannot selectively ignore irrelevant variation in the other dimension. Significant differences in latencies between the control condition and the correlated condition for each dimension were not observed, indicating the absence of any redundancy gains.

For the 4W×4T condition, the increase in latencies from control condition to orthogonal condition was also significant for both word and voice dimensions [$F(1,17) = 53.1$, $p < .0001$, and $F(1,17) = 19.6$, $p < .001$, respectively]. A significant decrease in latencies from the con-

Table 2
Mean Response Latencies (in milliseconds) and Mean Standard Deviations Collapsed over Subjects for Stimulus Variability Condition and Word and Voice Dimensions as a Function of Stimulus Set Condition

		Control		Orthogonal		Correlated		
Condition	Dimension	*M*	*SD*	*M*	*SD*	*M*	*SD*	Interference
2W×2T	Word	501.7	159.6	560.1	127.4	478.4	144.2	+58.4
	Voice	470.7	115.5	494.2	130.5	463.1	116.0	+23.5
4W×4T	Word	493.2	118.8	587.2	108.1	482.4	131.7	+94.0
	Voice	484.8	105.1	561.8	147.6	487.5	130.3	+77.0
8W×8T	Word	513.9	96.7	630.5	91.2	466.7	113.3	+116.6
	Voice	473.4	77.5	544.6	85.9	480.2	98.9	+71.2
16W×16T	Word	469.5	106.1	629.0	102.5	444.0	104.6	+159.5
	Voice	460.5	130.3	552.5	141.5	446.0	94.3	+92.0

trol condition to the correlated condition was not observed for either dimension. In the 8W×8T condition, latencies again increased from control to orthogonal conditions for both word and voice dimensions [$F(1,17) = 55.6$, $p < .0001$, and $F(1,17) = 22.7$, $p < .0001$, respectively]. A significant decrease in latencies from control condition to correlated condition was observed, but only for the word dimension [$F(1,17) = 11.2$, $p < .01$]. Finally, for the 16W×16T condition, orthogonal interference was also present for word and voice dimensions [$F(1,17) = 68.5$, $p < .001$, and $F(1,17) = 26.8$, $p < .001$, respectively]. As in the 8W×8T condition, a redundancy gain was found only for the word dimension [$F(1,17) = 10.2$, $p < .006$]. Overall, in all four conditions, orthogonal interference for both dimensions was found, and, in two of the four conditions, redundancy gains for the word dimension were present. Taken together, these results are consistent with the hypothesis that the processing of word and voice dimensions is mutually dependent.[3]

Figure 1 shows the amount of orthogonal interference (in milliseconds) for the word and voice dimensions for each of the four stimulus variability conditions. For all four conditions, significant increases in orthogonal interference were obtained when subjects were required to attend to either the word or the voice dimension. The pattern of results shows clearly that the processing of each dimension affects classification of the other dimension. Moreover, this effect increases as stimulus variability increases. Thus, each dimension affects decisions on the other dimension and does so to a greater degree as stimulus variability increases.

A closer examination of the amount of orthogonal interference present for each dimension across all four conditions shows, however, that the amount of interference was greater for the word dimension than for the voice dimension. Thus, perception of the word dimension appears to be subject to more interference by irrelevant variation in the voice dimension than vice versa. Although the two stimulus dimensions are processed in a mutually dependent manner, a reliable processing asymmetry is also present in these data.

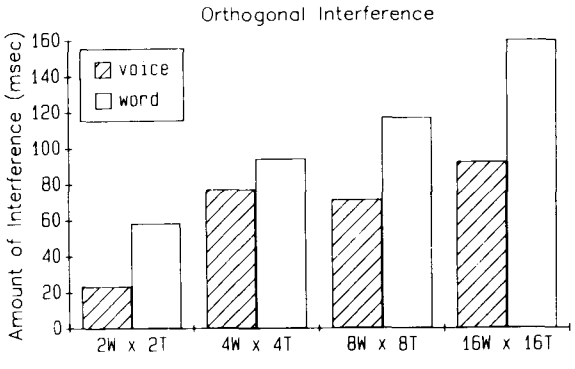

Figure 1. The amount of orthogonal interference (in milliseconds) for all stimulus variability conditions in Experiment 1. Interference is shown as a function of word and voice dimensions.

Upon further inspection of Figure 1, it is clear that stimulus variability reliably affects performance across all conditions. The amount of orthogonal interference obtained for the word and voice dimensions increases as stimulus variability increases. In order to quantify these observations, a two-way ANOVA was carried out to assess the amount of orthogonal interference obtained for the factors of stimulus dimension and stimulus variability condition. A significant main effect of stimulus variability was obtained [$F(3,68) = 9.2$, $p < .0001$], indicating that as stimulus variability increased for a given dimension, the amount of orthogonal interference also increased for that dimension. Post hoc tests revealed that only the 2W×2T condition and the 16W×16T condition differed significantly from one another. A significant main effect of stimulus dimension was also observed [$F(1,68) = 12.8$, $p < .001$]. Overall, the amount of orthogonal interference obtained for the word dimension was significantly larger than the amount of interference obtained for the voice dimension. This result supports the processing asymmetry observed earlier. Irrelevant variation in the voice dimension interfered more with processing of the word dimension than vice versa.

One account of the asymmetrical pattern of interference observed here may be related to discriminability of the two dimensions. Under some circumstances, an asymmetrical pattern of interference may be present because of differences in the relative discriminability of the target dimensions (see Eimas et al., 1978; Garner, 1974). If one dimension is inherently more discriminable than the other dimension, the more discriminable dimension may be easier to process when it is the relevant dimension but harder to ignore when it is the irrelevant dimension. In the present study, the asymmetrical pattern of interference could have been due to the greater discriminability of the voice dimension as compared with the word dimension. One method of assessing whether stimulus dimensions in this task differ in discriminability is to compare the response latencies obtained in the control conditions for each dimension. If response latencies are significantly faster in the control condition for one dimension rather than the other, this would provide support for the idea that the faster dimension is more discriminable. Applying this criteria to the present results, if the latencies for the voice dimension control condition were faster than those obtained for the word dimension control condition, then the asymmetrical pattern of interference could be explained simply on the basis of the inherent discriminability of the individual target dimensions.

In order to test this hypothesis, we carried out separate one-way ANOVAs on the latency data for the two control conditions. The results of these analyses indicated that performance for the word and voice dimension control conditions did not differ significantly within any of the

stimulus variability conditions. Thus, the asymmetry we observed does not appear to be due to inherent differences in discriminability between the two dimensions. Instead, the asymmetry reflects a real difference in processing time between word and voice dimensions.[4]

Identification data analyses. Table 3 shows the mean percent correct identification data collapsed over subjects for the control, orthogonal, and correlated conditions for word and voice dimensions for all stimulus variability conditions. A three-way ANOVA was conducted on the identification data. A significant main effect of stimulus set condition was obtained [$F(2,136) = 41.1$, $p < .001$]. Identification was most accurate in the correlated condition, less accurate in the control condition, and least accurate in the orthogonal condition. Post hoc tests revealed that identification performance in the orthogonal condition differed significantly from performance in both the control and correlated conditions. No other significant main effects or interactions were obtained.

In considering the identification and the latency data together, the pattern of results suggests that speed-accuracy tradeoffs did not occur in the present experiment. Post hoc tests showed that identification performance did not differ between the control and correlated conditions, while identification performance was worse in the orthogonal condition as compared with the other two conditions. Since the increase in latencies from control to orthogonal conditions was not accompanied by a parallel increase in accuracy, and since the decrease in latencies from control to correlated conditions was not accompanied by a parallel decrease in accuracy, further analyses on the data to test for speed-accuracy tradeoffs were not carried out.

Discussion

The results of Experiment 1 are important in two respects. First, we found that in all four stimulus variability conditions, subjects were unable to attend selectively to either word or voice while performing a speeded classification task. Information about word-initial consonants and information about the talker's voice appear to be processed together in a mutually dependent manner. Furthermore, the nature of this processing interaction appears to be asymmetrical. The processing of the voice dimension affected phonetic classification more than vice versa.

The second important result concerns the effects of stimulus variability. When stimulus variability was increased by increasing both the number of words and the number of talkers, more interference was observed for both word and voice dimensions. The increase in response latencies as a function of stimulus variability is consistent with earlier research showing that variability in the voice of the talker produces detrimental effects on spoken word recognition (Creelman, 1957; Mullennix et al., 1989). Thus, the effects of stimulus variability not only are present in perceptual identification and naming tasks, but apparently generalize to two-choice speeded classification tasks as well.

We should note here that, in Experiment 1, two sources of variability were manipulated together. It is possible that variability from trial to trial in the acoustic characteristics of the initial consonants may have resulted in greater demands on the perceptual system in encoding phonetic information relevant to the identification of the initial consonant. On the other hand, talker variability may have affected performance because of perceptual adjustments that are required to compensate for the acoustic differences due to changes in the talker's voice. Since both word variability and talker variability were manipulated together, it is impossible to assess whether the increase in orthogonal interference produced by the increase in stimulus variability was due to one or both sources of variability. In order to examine the separate contributions of word and talker variability, Experiment 2 was conducted.

Table 3
Mean Percent Correct Identification Collapsed over Subjects for All Conditions as a Function of Stimulus Dimension and Stimulus Set Condition

Condition	Dimension	Control	Orthogonal	Correlated
2W×2T	Word	98.3	97.8	98.9
	Voice	99.0	97.2	98.4
4W×4T	Word	98.8	97.2	99.5
	Voice	97.7	97.7	99.1
8W×8T	Word	98.2	96.3	98.9
	Voice	97.7	96.7	98.2
16W×16T	Word	98.9	97.2	98.9
	Voice	98.7	96.8	99.1

EXPERIMENT 2

In Experiment 1, talker variability and word variability were covaried across conditions by simultaneously increasing the number of words and talkers across orthogonal stimulus sets. In Experiment 2, the effects of talker variability and word variability were examined by manipulating each source of variability independently. Instead of increasing both the number of words and the number of talkers together to form orthogonal stimulus sets, only the number of words or only the number of talkers was increased to create orthogonal stimulus sets to be compared against control sets. By arranging the sets in this manner, variability on one dimension at a time can be manipulated and its effects on the pattern of orthogonal interference examined. Thus, the separate contributions of talker variability and acoustic-phonetic variability underlying the effects found in Experiment 1 could be assessed.

Method

Subjects. Subjects were 80 volunteers drawn from the Indiana University community. Each subject took part in one 1-h session and was paid $5 for participating in the experiment. All subjects

Table 4
Mean Response Latencies (in milliseconds) and Mean Standard Deviations Collapsed over Subjects for Stimulus Variability Condition and Stimulus Dimension as a Function of Stimulus Set Condition

Variability Condition	Dimension	Stimulus Set Condition: Control *M*	Control *SD*	2W×2T *M*	2W×2T *SD*	4W×2T *M*	4W×2T *SD*	8W×2T *M*	8W×2T *SD*	16W×2T *M*	16W×2T *SD*
Word	Word	530.0	135.0	563.2	146.2	613.1	161.5	642.0	197.3	657.2	203.7
Word	Voice	455.6	102.7	461.9	97.7	485.8	116.9	474.3	114.1	507.6	149.3
		Control		2W×2T		2W×4T		2W×8T		2W×16T	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Talker	Word	528.0	113.2	548.6	103.6	528.0	113.6	535.3	111.1	558.9	160.8
Talker	Voice	513.0	126.7	563.9	160.0	611.3	154.7	601.3	199.1	605.4	184.5

were native speakers of English who reported no history of a speech or hearing disorder at the time of testing.

Stimulus Materials. The stimuli were drawn from the same corpus used in Experiment 1.

Procedure. Three experimental factors were manipulated: stimulus dimension, stimulus set condition, and stimulus variability condition (talker variability or word variability). As in Experiment 1, stimulus dimension was manipulated by requiring subjects to attend to either word or voice. Stimulus set condition was manipulated by presenting the stimuli in a control set and in four different orthogonal sets. The control sets for each stimulus dimension were identical to those used in Experiment 1, and the orthogonal sets varied in composition. Stimulus variability was manipulated by increasing either talker variability or word variability while holding the other dimension constant. In the talker variability condition, the number of words contained in the four orthogonal sets remained at two while the number of talkers varied across the orthogonal sets from 2 to 16. These orthogonal sets will be referred to as the 2W×2T set, the 2W×4T set, the 2W×8T set, and the 2W×16T set. In the word variability condition, the number of talkers remained at 2, whereas the number of words contained in the four orthogonal sets varied across orthogonal sets from 2 to 16. These orthogonal sets will be referred to as the 2W×2T set, the 4W×2T set, the 8W×2T set, and the 16W×2T set. Stimulus dimension and stimulus variability were manipulated between subjects, and stimulus set was manipulated within subjects.

The subjects were divided equally into four groups and randomly assigned to the experimental conditions. The subjects received a total of five stimulus sets per session. Half of the subjects classified stimuli by word and half of the subjects classified stimuli by voice. The subjects were presented with the appropriate control condition for the stimulus dimension they classified. The stimuli used in the control sets for each dimension were identical across subjects. For each of the stimulus dimensions classified, half the subjects received orthogonal stimulus sets varying in talker variability, and half the subjects received orthogonal stimulus sets varying in word variability.

The order of stimulus sets was counterbalanced by means of a Latin square design. The number of stimulus repetitions was adjusted for each stimulus set to produce 64 test trials. All other aspects of the experiment were identical to the procedures used in Experiment 1.

Results

Response latencies. Table 4 displays the mean response latencies and standard deviations collapsed over subjects for the stimulus set conditions as a function of stimulus dimension and stimulus variability condition. The individual response latencies were plotted to estimate normality of the distribution for each condition. As in Experiment 1, the distributions were approximately normal. A four-way ANOVA was then conducted on the latency data for the factors of stimulus dimension, stimulus set, stimulus variability, and order. A significant main effect of stimulus set was obtained [$F(4,304) = 15.8, p < .001$]. Response latencies were fastest in the control condition and increased as a function of increasing stimulus variability (506.7, 534.4, 559.6, 563.3, and 582.3 msec, re-

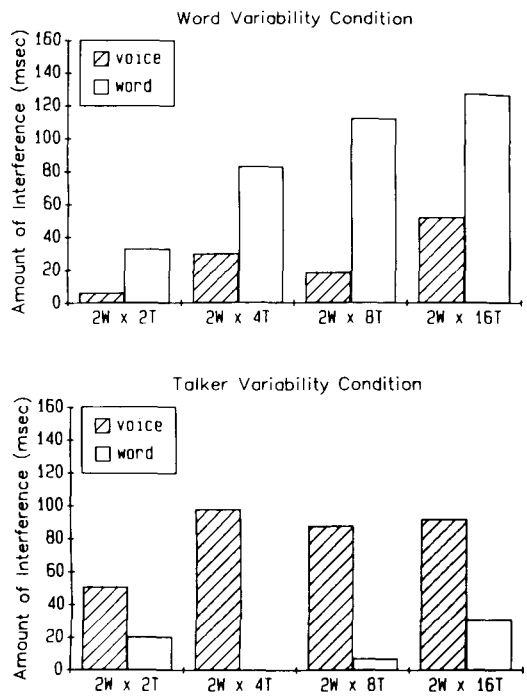

Figure 2. The amount of orthogonal interference (in milliseconds) for variability conditions in Experiment 2. The top panel displays the data for the word variability condition and the bottom panel displays the data for the talker variability condition. Interference is shown as a function of word and voice dimensions.

spectively, for the five stimulus set conditions). A significant interaction of stimulus variability with stimulus dimension was also obtained [$F(1,76) = 7.4, p < .01$]. Post hoc tests indicated that latencies in the word variability condition were slower for attending to the word than to the voice, while latencies in the talker variability condition were slower for attending to the voice as opposed to the word (see Figure 2).

A significant three-way interaction between the three experimental variables was also observed [$F(4,304) = 6.1$, $p < .001$]. When word variability increased, latencies steadily increased across stimulus sets when the word was attended to, but latencies changed relatively little when the voice was attended to. On the other hand, when talker variability increased, latencies increased and leveled off across stimulus sets when the voice was attended to, but they changed relatively little when the word was attended to (see Figure 2). Examination of the overall pattern of interference reveals an interesting and potentially important difference between the two conditions. As shown in the top panel, the amount of interference increases for the word variability condition only when the word is attended and variability in the words is increased. However, the same pattern is not observed for comparable changes in talker variability in the lower panel. In order to determine whether the pattern of orthogonal interference across stimulus sets differed for these two conditions, separate linear trend analyses were performed. The results of the analyses indicated that the pattern of interference across orthogonal sets for the word variability condition (word attended to) fit a linear model only ($F = 6.0$, $p < .02$), while the pattern of interference across orthogonal sets for the talker variability condition (voice attended to) did not fit a linear, quadratic, or cubic model. This difference in the linearity of interference gains across orthogonal sets suggests that differences in processing under the two experimental conditions are present and that they lead to differential patterns of interference in both experiments.

Identification data analyses. Table 5 shows the mean percent correct identification data collapsed over subjects for stimulus sets and word and voice dimensions for both stimulus variability conditions. A three-way ANOVA was conducted on the identification data. A significant main effect of stimulus set condition was obtained [$F(1,76) = 4.0, p < .01$]. However, identification actually differed very little across stimulus sets (98.9%, 98.4%, 98.2%, 97.8%, and 98.2%, respectively). Post hoc tests revealed no significant differences between stimulus set conditions. Thus, the pattern of identification and latency results again suggests that speed-accuracy tradeoffs did not occur.

Discussion

The results of Experiment 2 provide further information about the stimulus variability effects found in Experiment 1. The primary finding in Experiment 2 is that when variability along a dimension is increased, selective attention to that dimension is impaired relative to appropriate control conditions. For example, when word variability increased across the orthogonal sets, the amount of interference obtained was much greater when subjects were required to classify the word dimension than when they were required to classify the voice dimension, a finding that would be anticipated on the basis of earlier results. What is interesting about the present results is that the effects of variability are quite different for the two dimensions under examination. Attention to the word dimension is linearly related to the number of words in the stimulus set, whereas attention to voice is not. When word variability increases, the amount of orthogonal interference when words are attended to increases fairly steadily and in a linear fashion. However, when talker variability increases, an initial increase in interference is present, but then the amount of interference levels off and appears to asymptote.[5] These particular effects are remarkably similar to patterns of performance obtained in visual sorting tasks when set size is increased. Smith and Kemler (1978) found that increasing the number of items in a stimulus set had an initial detrimental effect on performance that quickly leveled off when subjects were told to process the stimuli on the basis of an integral dimension. However, when subjects processed the stimuli on the basis of a separable dimension, performance steadily became worse as the number of items in the set increased. Smith and Kemler interpreted the first pattern of performance as evidence that the subjects classified the stimulus relations in a holistic fashion. They interpreted the second pattern of performance as evidence that subjects classified/processed the stimulus relations in terms of dimensional structure. If Smith and Kemler's interpretations are extended to the present results, it would appear that voice information is processed in a more "holistic" manner,

Table 5
Mean Percent Correct Identification Collapsed over Subjects for Variability Condition and Stimulus Dimension as a Function of Stimulus Set Condition

Variability Condition	Dimension	Stimulus Set Condition				
		Control	2W×2T	4W×2T	8W×2T	16W×2T
Word	Word	99.8	98.4	97.2	98.1	97.8
Word	Voice	98.7	98.1	98.9	97.4	98.2
		Control	2W×2T	2W×4T	2W×8T	2W×16T
Talker	Word	98.5	98.5	98.9	98.3	98.6
Talker	Voice	98.7	98.4	97.9	97.4	98.1

whereas the acoustic-phonetic information required for phoneme identification is processed in a more "dimensionally analyzable" manner.

GENERAL DISCUSSION

Taken together with other recent findings from our laboratory, the present set of results shows that the perceptual processes used to encode information about a talker's voice are closely related to the processes involved in the encoding of the signal into a phonetic representation. A phonetically related stimulus dimension and a voice-related stimulus dimension were found to be processed as in a mutually dependent manner in a speeded classification task. Because neither talker information nor phonetic information can be selectively ignored when subjects are required to attend to specific aspects of a spoken word, we conclude that the processes involved in phonetic coding and the processes involved in encoding characteristics of a talker's voice do not operate independently of one another.

The presence of interference effects in the speeded classification task also demonstrates that the processing of voice information is a mandatory encoding operation in speech perception (Fodor, 1983; Miller, 1987). Because information about a talker's voice cannot be selectively ignored, selective attention to phonetic information is interfered with by irrelevant variation in voice.[6] Given the present findings, with respect to previous research on talker variability, it seems reasonable to conclude that decreases in spoken word recognition performance produced by changes in the voice of the talker (Creelman, 1957; Mullennix et al., 1989) may be due to changes in selective attention caused by the mandatory processing of the talker's voice.

The processing dependencies observed in the present study provide further insight into the relationship between auditory-to-phonetic coding processes and talker normalization processes. The asymmetric pattern of interference observed, with greater interference caused by the irrelevant variation in the voice dimension, suggests that the analysis of phonetic information contained in word-initial consonants is more dependent on the prior or concurrent analysis of voice information than vice versa. Asymmetries of this kind in speech perception have been interpreted in terms of serial and parallel models of processing (see Eimas et al., 1978; Wood, 1974, 1975). As stated by Eimas et al. (1978), the mechanisms of analysis involved in the processing of place and manner phonetic dimensions, "while functioning in temporally overlapping and interactive fashion, are, to some extent, hierarchically arranged, in that some processes of analysis require the outputs from other analyzers before their own analyses can be completed" (p. 18). Thus, Eimas et al. suggested that the phonetic dimensions were processed in what is called a parallel-contingent manner (see Turvey, 1973). The processes used to extract each phonetic dimension operate in parallel, while information from manner-of-articulation analyzers is used by the place-of-articulation analyzers in a hierarchically driven manner. A similar idea was also proposed by Wood (1974), on the basis of his finding of an asymmetric processing relation between the phonetic dimensions of place of articulation and fundamental frequency. Wood (1974) argued that a hybrid serial/parallel model of processing was appropriate to account for the pattern of his results.

With regard to the present study, we obtained significant interference for both the word and the voice dimensions. The magnitude of interference was greater for the word dimension than for the voice dimension. However, we also observed redundancy gains in some conditions for the word dimension. Because interference was obtained on both dimensions, it is likely that talker normalization processes and auditory-to-phonetic processes operate in parallel. However, because the interference was asymmetric and because the redundancy gains indicated that only the redundant voice information was used to assist classification of the word dimension, auditory-to-phonetic coding processes may be partially contingent on the prior output of the talker normalization processes. On the basis of the present findings, it appears that processing of these dimensions occurs in a manner best described as parallel contingent (Turvey, 1973). If there exist multiple information-processing components or modules in speech perception, it is possible that a subprocess or set of subprocesses operates to encode the talker's voice and another set of subprocesses operates to encode phonetically related auditory information, and that they operate in parallel. As these subprocesses are carried out, auditory-to-phonetic processes must wait for at least part of the output from talker-related analysis routines before any further phonetic analysis of the input signal can proceed. Thus, in effect, a hierarchically driven contingency of processing exists between talker normalization processes and auditory-to-phonetic coding processes, so that talker normalization processes can be carried out at a somewhat earlier level in the perceptual system.

With regard to the effects of word and talker variability, the present findings show clearly that an increase in stimulus variability produces increases in response latencies. This result provides additional evidence supporting the findings obtained in previous studies on spoken word recognition (Creelman, 1957; Mullennix et al., 1989) and vowel and consonant perception (Assmann et al., 1982; Fourcin, 1968; Rand, 1971; Verbrugge et al., 1976; Weenink, 1986) demonstrating that talker variability affects speech perception and spoken word recognition. Apparently, the perceptual system compensates for the acoustic differences due to talker variability, and this form of compensation produces reliable and robust effects on the processing system.

One additional result obtained in the present study concerns the pattern of orthogonal interference found for word and voice dimensions. Increases in stimulus variability affect the processing of words and the voice quite differently. Perceptual performance in classifying voice

is initially affected by increases in talker variability, but further increases in variability have little effect. However, perceptual performance declines in a linear fashion when word variability is increased. The difference in the patterns of interference due to the different sources of variability is consistent with the view that two qualitatively different types of processes are utilized in the two situations (Smith & Kemler, 1978). It is possible that the perceptual processing of voice information utilizes "holistic" analyzers, whereas the encoding of acoustic-phonetic information requires "dimensional" analyzers of some sort. Thus, the encoding of information about a talker's voice may be carried out by perceptual mechanisms that are qualitatively quite different from those used to encode acoustic-phonetic information about a word.

In summary, the results of the present investigation suggest that talker normalization processes and acoustic-phonetic perceptual processes are closely interrelated. Selective attention to information in the speech signal relevant to either type of process appears to be affected by the mandatory processing of the information relevant to the other process. Despite the findings that these two processes are closely related, the encoding of voice information differs in two ways from the encoding of acoustic-phonetic information about spoken words. First, decisions about a talker's voice show less interference from irrelevant variation of words than vice versa. Second, decisions about a talker's voice do not show set size effects due to increases in stimulus variability. Subjects apparently can attend to dimensions of voice and selectively ignore irrelevant variation in the words. However, they have much more difficulty attending to words when there is simultaneously irrelevant variation in the voice of the talker. Taken together, the present results suggest that perceptual normalization processes used to encode information about a talker's voice appear to be fundamentally quite different from the early auditory-to-phonetic coding processes involved in phonetic perception and spoken word recognition.

REFERENCES

ASSMANN, P. F., NEAREY, T. M., & HOGAN, J. T. (1982). Vowel identification: Orthographic, perceptual, and acoustic aspects. *Journal of the Acoustical Society of America*, **71**, 975-989.

BIEDERMAN, I. J., & CHECKOSKY, S. F. (1970). Processing redundant information. *Journal of Experimental Psychology*, **83**, 486-490.

BLECHNER, M. J., DAY, R. S., & CUTTING, J. E. (1976). Processing two dimensions of nonspeech stimuli: The auditory-phonetic distinction reconsidered. *Journal of Experimental Psychology: Human Perception & Performance*, **2**, 257-266.

CARR, P. B., & TRILL, D. (1964). Long-term larynx-excitation spectra. *Journal of the Acoustical Society of America*, **36**, 2033-2040.

CARRELL, T. D. (1984). Contributions of fundamental frequency, formant spacing, and glottal waveform to talker identification. *Research on Speech Perception* (Tech. Rep. No. 5). Bloomington: Indiana University, Department of Psychology.

CARRELL, T. D., SMITH, L. B., & PISONI, D. B. (1981). Some perceptual dependencies in speeded classification of vowel color and pitch. *Perception & Psychophysics*, **29**, 1-10.

CREELMAN, C. D. (1957). Case of the unknown talker. *Journal of the Acoustical Society of America*, **29**, 655.

EIMAS, P. D., TARTTER, V. C., MILLER, J. L., & KEUTHEN, N. J. (1978). Asymmetric dependencies in processing phonetic features. *Perception & Psychophysics*, **23**, 12-20.

FANT, G. (1973). *Speech sounds and features*. Cambridge, MA: MIT Press.

FELFOLDY, G. L., & GARNER, W. R. (1971). The effects on speeded classification of implicit and explicit instructions regarding redundant dimensions. *Perception & Psychophysics*, **9**, 289-292.

FODOR, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

FOURCIN, A. J. (1968). Speech-source interference. *IEEE Transactions on Audio & Electroacoustics*, **ACC-16**, 65-67.

GARNER, W. R. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum.

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

HOUSE, A. S., WILLIAMS, C. E., HECKER, M. H. L., & KRYTER, K. D. (1965). Articulation-testing methods: Consonantal differentiation with a closed-response set. *Journal of the Acoustical Society of America*, **37**, 158-166.

JOOS, M. A. (1948). Acoustic phonetics. *Language*, **24**(Suppl. 2), 1-136.

LADEFOGED, P. (1980). What are linguistic sounds made of? *Language*, **56**, 485-502.

MARTIN, C. S., MULLENNIX, J. W., PISONI, D. B., & SUMMERS, W. V. (1987). Effects of talker variability on recall of spoken words. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 676-684.

MILLER, J. L. (1978). Interactions in processing segmental and suprasegmental features of speech. *Perception & Psychophysics*, **24**, 175-180.

MILLER, J. L. (1987). Mandatory processing in speech perception. In J. L. Garfield (Ed.), *Modularity in knowledge representation and natural-language understanding* (pp. 309-322). Cambridge, MA: MIT Press.

MONSEN, R. B., & ENGEBRETSON, A. M. (1977). Study of variations in the male and female glottal wave. *Journal of the Acoustical Society of America*, **62**, 981-993.

MULLENNIX, J. W., PISONI, D. B., & MARTIN, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America*, **85**, 365-378.

NUSBAUM, H. C., & MORIN, T. (1989, May). *Perceptual normalization of talker differences*. Paper presented at the 117th meeting of the Acoustical Society of America, Syracuse, NY.

PASTORE, R. E., AHROON, W. A., PULEO, J. S., CRIMMINS, D. B., GOLOWNER, L., & BERGER, R. S. (1976). Processing interaction between two dimensions of nonphonetic auditory signals. *Journal of Experimental Psychology: Human Perception & Performance*, **2**, 267-276.

PETERSON, G. E., & BARNEY, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, **24**, 175-184.

RAND, T. C. (1971). Vocal tract size normalization in the perception of stop consonants. *Haskins Laboratories Status Reports on Speech Research*, **SR-25/26**, 141-146.

SMITH, L. B., & KEMLER, D. G. (1978). Levels of experienced dimensionality in children and adults. *Cognitive Psychology*, **10**, 502-532.

SUMMERFIELD, Q. (1975). *Acoustic and phonetic components of the influence of voice changes and identification times for CVC syllables* (Report of Speech Research in Progress, Vol. 2, No. 4, pp.73-98). The Queen's University of Belfast, Belfast, Ireland.

SUMMERFIELD, Q., & HAGGARD, M. P. (1973). *Vocal tract normalisation as demonstrated by reaction times* (Report on Research in Progress in Speech Perception Vol. 2, No. 2, pp. 1-12). The Queen's University of Belfast, Belfast, Ireland.

TOMIAK, G. R., MULLENNIX, J. W., & SAWUSCH, J. R. (1987). Integral processing of phonemes: Evidence for a phonetic mode of perception. *Journal of the Acoustical Society of America*, **81**, 755-764.

TURVEY, M. T. (1973). On peripheral and central processes in vision: Inferences from an information-processing analysis of masking with patterned stimuli. *Psychological Review*, **80**, 1-52.

VERBRUGGE, R. R., STRANGE, W., SHANKWEILER, D. P., & EDMAN, T. R. (1976). What information enables a listener to map a talker's vowel space? *Journal of the Acoustical Society of America*, **60**, 198-212.

WEENINK, D. J. M. (1986). The identification of vowel stimuli from men, women, and children. *Proceedings 10 from the Institute of Phonetic Sciences of the University of Amsterdam*, 41-54.

WOOD, C. C. (1974). Parallel processing of auditory and phonetic information in speech discrimination. *Perception & Psychophysics*, **15**, 501-508.

WOOD, C. C. (1975). A normative model for redundancy gains in speeded classification: Application to auditory and phonetic dimensions in speech discrimination. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. Lindman, & D. B. Pisoni (Eds.), *Cognitive Theory* (Vol. 1. pp. 55-77). Hillsdale, NJ: Erlbaum.

WOOD, C. C., & DAY, R. S. (1975). Failure of selective attention to phonetic segments in consonant-vowel syllables. *Perception & Psychophysics*, **17**, 346-350.

NOTES

1. However, note that arranging the orthogonal set in this manner for this condition still preserved an increase in stimulus variability as reflected by the greater variation in the range of words and talkers used in the orthogonal set as compared with the other three conditions.

2. The latency data were also analyzed in two other ways: by eliminating the data of "outlier" subjects whose mean response latencies fell outside of two standard deviations around the mean for any condition, and by using median response latencies instead of mean response latencies. The results from these two alternative analyses did not substantially differ from the present results; hence we report only the analyses based on mean response latencies.

3. In order to assess whether the redundancy gains observed for the word dimension may have been due to a selective serial processing strategy adopted by the subjects (Biederman & Checkosky, 1970), a one-way ANOVA was conducted on the latency data over all groups for the fastest control condition and the fastest correlated condition for each subject. A significant main effect of condition was observed [$F(1,71) = 7.1$, $p < .01$], with mean latencies faster for correlated versus control conditions (438.0 and 452.6 msec, respectively). Thus, it appears that the redundancy gains observed were not due to a selective serial processing strategy.

4. An argument could also be made that even though performance in the control conditions did not differ, there may have been discriminability differences not exhibited, because of the presence of floor effects. We have no evidence from the present experiment to support or refute such a possibility.

5. Of additional interest is that, going back to Experiment 1, the increases in interference for word and voice dimensions also suggest a similar pattern (see Figure 1).

6. It may be pointed out that this interpretation of the mandatory nature of talker normalization is based on data in the present study, which was obtained with relatively unpracticed subjects. It is possible that with extensive practice, listeners could be trained to process word and voice dimensions in a separable manner. This result would suggest that, under certain conditions, the processing of voice information is not obligatory. However, we feel that the experimental conditions of the present study provide a more "ecologically valid" test of what human listeners actually do during speech perception rather than what they are capable of doing under conditions of extensive training. Thus, we believe our results reflect the perceptual processing occurring under normal listening conditions more closely than they do a highly artificial training environment.

(Manuscript received August 19, 1988;
revision accepted for publication August 17, 1989.)