Perception & Psychophysics
1997, 59 (5), 675-692

The encoding of rate and talker information during phonetic perception

KERRY P. GREEN
University of Arizona, Tucson, Arizona

and

GAIL R. TOMIAK and PATRICIA K. KUHL
University of Washington, Seattle, Washington

The acoustic structure of the speech signal is extremely variable due to a variety of contextual factors, including talker characteristics and speaking rate. To account for the listener's ability to adjust to this variability, speech researchers have posited the existence of talker and rate normalization processes. The current study examined how the perceptual system encoded information about talker and speaking rate during phonetic perception. Experiments 1–3 examined this question, using a speeded classification paradigm developed by Garner (1974). The results of these experiments indicated that decisions about phonemic identity were affected by both talker and rate information: irrelevant variation in either dimension interfered with phonemic classification. While rate classification was also affected by phoneme variation, talker classification was not. Experiment 4 examined the impact of talker and rate variation on the voicing boundary under different blocking conditions. The results indicated that talker characteristics influenced the voicing boundary when talker variation occurred within a block of trials only under certain conditions. Rate variation, however, influenced the voicing boundary regardless of whether or not there was rate variation within a block of trials. The findings from these experiments indicate that phoneme and rate information are encoded in an integral manner during speech perception, while talker characteristics are encoded separately.

Research over the past 30 years has revealed numerous aspects of the acoustic signal that play a role in phonetic perception. Theories of speech perception have attempted to explain how these acoustic characteristics are processed, integrated, and mapped onto the underlying phonetic representations. Such explanations have been hampered by the complex relationship between the characteristics of the signal and the underlying phonetic representations.

Two factors that have been shown to have a significant impact on the acoustic characteristics are talker variation and rate variation. For example, the size and shape of the vocal tract varies across male, female, and child talkers, creating considerable variation in the acoustic realization of the vocal tract resonances for consonants and vowels. In the case of vowels, this variability results in formant values that vary substantially across talkers, creating acoustic overlap among the categories (Hillenbrand & Houde, 1995; Peterson & Barney, 1952). Fundamental frequency also varies across talkers, with male talkers having lower fundamental frequencies, on average, than female talkers, which in turn are lower than those of child talkers. Changes in speaking rate create variation in the realization of acoustic cues that are temporal in nature, such as voice-onset time (VOT) and transition duration (Miller & Baer, 1983; Miller, Green, & Reeves, 1986; Summerfield, 1981). Across different speaking rates, there is a large amount of overlap in the realization of these cues for consonantal categories contrasting in voicing (e.g., /b/ vs. /p/) or manner of articulation (e.g., /b/ vs. /w/) (Miller & Baer, 1983; Miller et al., 1986; Volaitis & Miller, 1992).

Perception of the speech signal remains accurate despite the variation created by different talkers and speaking rates. Thus, speech perception, like other aspects of perception, exhibits the phenomenon of perceptual constancy. This raises the question of how the perceptual system achieves perceptual constancy, and at what costs. The usual answer is that the perceptual system compensates or normalizes for the variations in the acoustic signal created by the different factors. Developmental data suggest that young infants are very good at contending with variation in both talker and speaking rate (e.g., Eimas & Miller, 1980; Kuhl, 1979, 1983), suggesting that such normalization processes are automatic and present at birth.

There are two ways of conceptualizing such normalization processes. One way is to factor out the nonphonetic variation produced by the context. The result would be a residual signal with acoustic characteristics more closely

This research was supported in part by Research and Training Grant 1 P60 DC-01409 from the National Institute of Deafness and Other Communication Disorders, National Institutes of Health Grant NS-26475 to K.P.G., and National Institutes of Health Grant HD-18286 to P.K.K. We would like to thank Kathryn Fohr, Lisa Kupnis, and Erica Stevens for their help in collecting and analyzing the data. We would also like to thank Joanne Miller and two reviewers for their helpful comments on an earlier version of the manuscript. Please address correspondence to K. P. Green, Psychology Department, Psychology Building, Rm. 312, University of Arizona, Tucson, AZ 85721 (e-mail: kgreen@u.arizona.edu).

aligned to the values of underlying phonetic representations. This approach is sometimes referred to as a "direct" theory of normalization (see Johnson, 1990c). Alternatively, the perceptual system may alter its criteria for analyzing the acoustic signal in accord with the variation produced by different talkers or speaking rates. That is, the contextual variation may produce a modification of the perceptual criteria used to evaluate the acoustic characteristics before they are mapped onto the underlying representations. This approach is referred to as an "indirect" theory of normalization (Johnson, 1990c).

Evidence that talker and rate normalization occurs during speech perception is obtained from a number of studies. For example, with respect to talker variation, listeners' recognition of a particular vowel token varies as a function of the preceding talker context (Johnson, 1990c) as well as acoustic characteristics of the syllable itself, such as fundamental frequency ($F0$) (Johnson, 1990b; Nearey, 1989), which is roughly correlated with vocal tract size. Talker variation is also taken into account in the perception of certain consonants such as fricatives. Johnson (1990a) found a significant shift in the /s/–/ʃ/ boundary as a function of whether the following vowel was produced by a male or a female talker. Finally, some of the acoustic cues used to distinguish the voicing quality of consonants such as $F0$ and first-formant ($F1$) onset frequency vary systematically across talkers. As noted above, $F0$ is roughly correlated with vocal tract size. $F1$ onset frequency tends to be correlated with vocal tract size since small vocal tracts have higher first formant resonances than do large vocal tracts. Jongman and Miller (1990) characterized the perceptual space for stop bursts across different places of articulation with respect to a reference frequency reflecting the speaker's average $F0$ and $F0$ modulations. Such a model may reflect intraspeaker variation, but it is also well suited for reflecting interspeaker variation in the representation of the acoustic information used in the perception of stop consonants. Therefore, it is possible that some of the influence of these two cues on the perception of voicing may reflect the operation of talker normalization processes.[1]

With respect to speaking rate, studies have shown that changes in speaking rate produce consistent changes in overall syllable duration (see Miller, 1981a, for review). Slower speaking rates result in longer syllable durations, and faster speaking rates produce shorter syllable durations. More importantly, changes in syllable duration affect not only the perceived speaking rate of syllables (Green & Miller, 1985; Miller, Aibel, & Green, 1984), but also the phonetic interpretation of such cues as VOT and transition duration (Diehl & Walsh, 1989; Green & Miller, 1985; Green, Stevens, & Kuhl, 1994; Miller et al., 1984; Miller & Liberman, 1979; Summerfield, 1981). For example, changes in syllable duration have been shown to influence not only the perception of boundary tokens along a /b/–/p/ continuum (Green & Miller, 1985; Summerfield, 1981), but also tokens judged as most representative or "prototypical" of a phonetic category (Miller & Volaitis, 1989). Finally, *increasing* overall syllable duration by adding a final consonant to a syllable shifts the VOT boundary toward *shorter* VOT values (Summerfield, 1981; see also Miller & Liberman, 1979, for a similar finding with /b/–/w/). This finding demonstrates that the perceptual system adjusts for changes in speaking rate rather than just the overall duration of a syllable.

Thus, there is evidence for two normalization mechanisms operating during phonetic perception. One adjusts for talker variation and the second adjusts for rate variation (Darwin, McKeown, & Kirby, 1989; Green et al., 1994; Johnson, 1990b, 1900c; Ladefoged, 1967; Miller & Liberman, 1979; Remez, Rubin, Nygaard, & Howell, 1987). One question that arises is whether the information used in either talker or rate normalization is encoded or processed separately from the information used in making phonetic classifications. This question was addressed with respect to talker normalization in a study by Mullennix and Pisoni (1990), using a speeded-classification paradigm developed by Garner (1974).[2] The characteristics of the Garner task enable one to determine whether different dimensions of a stimulus are processed in an integral or separable manner. Mullennix and Pisoni (1990) found that in natural speech tokens, the classification of tokens with respect to their initial consonant (/b/ vs. /p/) took longer when the tokens also varied in talker characteristics (male vs. female talkers), than they did in a control condition in which no such orthogonal variation occurred (only male or only female talkers). In addition, Mullennix and Pisoni found that classification of the tokens with respect to talker characteristics (male vs. female) was longer when there was orthogonal phonetic variation in the tokens (the tokens began with both /b/ and /p/) over a control condition in which there was no such variation (only /b/ tokens or only /p/ tokens). However, the amount of increase in the classification times for the talker dimension was significantly less than the increase in classification times for the phonetic dimension.

On the basis of their findings of mutual, albeit asymmetric, interference between the phoneme and talker dimensions, Mullennix and Pisoni (1990) proposed that the two sources of information were processed in parallel, but that the output of phonetic processing was contingent upon the output of the talker normalization process (see Eimas, Tartten, Miller, & Keuthen, 1978, for a similar notion). Since additional variation along the two dimensions (created by adding either additional word candidates or talkers) produced substantially different patterns of interference, Mullennix and Pisoni proposed that qualitatively different mechanisms were used to encode talker and phoneme information.

The findings of Mullennix and Pisoni (1990) raise the question of whether a similar situation exists for the encoding of speaking-rate information. The studies demonstrating that phonetic segments are normalized with respect to both rate information and talker information suggest that rate information might also be encoded by a mechanism separate from that used to encode phonetic information. However, speaking rate can be defined as

the number of linguistic constituents (either phonetic segments or syllables) per unit of time and, as such, may be an inherently segmental characteristic of the speech signal. It is certainly true that many acoustic cues used to perceive speech segments such as VOT, transition duration, or vowel duration are time-dependent and therefore can be properly interpreted only with respect to the articulation rate of an utterance. Consistent with this view, some models of rate normalization consider rate information to be an intrinsic component of the speech signal rather than extrinsic information that is extracted and applied in a separate stage of processing, as seems to be the case for talker information (see Summerfield, 1981). It is therefore possible that information about speaking rate might be encoded by the same mechanism used to encode phonetic information rather than by a separate mechanism, as talker information appears to be. In a preliminary study, we investigated this question, using synthetic speech tokens (Tomiak, Green, & Kuhl, 1991).

Tomiak et al. (1991) constructed synthetic /bi/ and /pi/ tokens with overall syllable durations of either 118 or 340 msec and presented them to listeners for speeded classification of either the phonemic quality or the speaking rate of the tokens (the short tokens were perceived as being spoken at a fast rate of speech while the long tokens were perceived as being produced at a slow rate of speech). Tomiak et al. found evidence for mutual and symmetric interference between the phoneme and the rate dimensions, suggesting that the two types of information might be encoded by the same mechanism. In a similar experiment, Tomiak et al. also examined the interactions of talker and phonemic characteristics using /bi/ and /pi/ tokens synthesized to correspond to a male and a female talker. In contrast to Mullennix and Pisoni (1990), Tomiak et al. found that phonemic variation had no influence on talker classification (see Eimas Tartter, & Miller, 1981, and Wood, 1974, for similar findings involving the classification of synthetic speech tokens with respect to pitch or phonetic characteristics).

The results of Tomiak et al. (1991) suggest that talker and rate information may be encoded by different kinds of mechanisms during phonetic classification. However, a comparison between their results and those of Mullennix and Pisoni (1990) also indicates that different patterns of interference can occur for natural and synthetic speech tokens. These different patterns may be due to the kind of information available in the two types of tokens with respect to the phoneme and talker dimensions. Due to the basic limitations of speech synthesizers, synthetic speech tokens typically have a smaller variety of acoustic cues than does natural speech, which may account for why synthetic speech is more difficult to process than natural speech (Luce, Feustel, & Pisoni, 1983). Moreover, the absence of certain acoustic cues in synthetic speech may alter the attentional resources used to recognize the speech tokens (see Gordon, Eberhardt, & Rueckl, 1993) and leave fewer resources available for processing the information about talker and rate characteristics. Experiments using natural speech are therefore important for obtaining a more complete understanding of how talker and rate information are processed during phonetic perception.

The purpose of the current study was to investigate the encoding of both talker and rate information using the same set of natural speech tokens which varied with respect to their phonemic characteristics (/b/ vs. /p/) as well as along the two orthogonal dimensions of speaking rate and talker characteristics. As in the Tomiak et al. (1991) and Mullennix and Pisoni (1990) studies, the speeded-classification task was used to investigate whether information for talker or rate was processed separately or interactively with the phonemic information. In Experiment 1, the processing interactions between the voicing and talker dimensions were examined in an attempt to replicate the Mullennix and Pisoni (1990) study. To make the two studies comparable, talker characteristics were varied with respect to the gender (male vs. female) of the talker. The main question posed by this experiment was whether an asymmetry existed in the pattern of interference between the phoneme and talker dimensions. A related question was whether variation along the phoneme dimension would produce a reliable interference along the talker dimension when natural speech tokens were used rather than synthetic speech tokens. In Experiment 2, the processing interactions between the phoneme dimension and the rate dimension were examined. Of primary interest was whether the pattern of phoneme and rate interference would mirror that of the phoneme and talker dimensions. In Experiment 3, the results of Experiment 2 were extended to a different set of tokens that varied along the dimension of place of articulation: /b/ versus /d/. In Experiment 4, how variation in both talker characteristics and speaking rate influenced identification of speech tokens varying in VOT was examined.

EXPERIMENT 1

The purpose of the first experiment was to replicate findings of earlier studies on the interference between phoneme and talker variation (e.g., Mullennix & Pisoni, 1990; Tomiak et al., 1991), using natural speech tokens produced at two different speaking rates. There were two parts to the experiment. The first was done using syllables spoken at a fast rate of speech; the second was identical to the first except that tokens were spoken at a slow rate of speech. Given the amount of variability in speech tokens produced by different talkers at different speaking rates, the second part of the experiment provides a replication of the findings using different stimuli and subjects.

Method

Subjects. The subjects, who received course credit for their participation, were 40 undergraduates at the University of Arizona. All subjects were native speakers of English with no known history of a speech or hearing disorder.

Stimuli. The stimuli were the syllables /bi/ and /pi/ spoken by a male and a female talker at fast and slow speaking rates. Both talkers spoke a Midwestern dialect with no strong regional accent. The

male talker was judged (by the experimenters) to have a clearly "male" voice; the female talker was judged to have a "female"-sounding voice. The talkers were recorded saying the syllables /bi/ and /pi/ at several different speaking rates in a soundproof room using an audio tape recorder (Nagra III) and microphone (Electro-voice 635A). These tokens were low-pass filtered at 9.9 kHz and digitized at a sampling rate of 20 kHz with 12-bit quantization into a lab computer (NEC 386-20) for measuring and editing. Single fast and single slow /bi/ and /pi/ tokens, which sounded like clear instances of their respective phonemic categories, were selected for each talker. The tokens were matched as closely as possible with respect to their overall durations within a particular speaking rate. The durations of these eight tokens are presented in Table 1. The root mean square amplitude levels of the different tokens were digitally equated, and each token was isolated and stored in its own file for presentation purposes.

Procedure. Each subject was randomly assigned to one of two groups of subjects. The first group was presented with the tokens spoken at a fast rate of speech; the second group was presented with the tokens spoken at a slow rate of speech. Each group of 20 subjects was further divided into two additional groups, with 10 of the subjects assigned to a phoneme target group (TG) and 10 assigned to a talker TG. Subjects in the two phoneme TGs were told that they would be listening to versions of the syllables /bi/ and /pi/ produced by a male and a female speaker. For these subjects, the phonemic identity of the tokens ("B" vs. "P") was specified as the target dimension. Subjects in the two talker TGs were told that they would be listening to male and female versions of the syllables /bi/ and /pi/. As in the Mullennix and Pisoni (1990) study, these subjects were instructed to classify the "gender" of the talker as the target dimension.

All subjects classified six randomized sets of stimuli. These sets reflected the three conditions designed to assess the presence/absence of integral processing: control, orthogonal, and correlated. Two of these sets were single-dimension sets in which the target dimension varied while the nontarget dimension was held constant. As an example, for subjects in the phoneme TG (with talker as the irrelevant orthogonal dimension), the first single-dimension set contained the male /bi/ and /pi/ tokens and the second set contained the female /bi/ and /pi/ tokens. Another two sets were correlated sets in which the presentation of each target dimension was paired with a unique instance of the nontarget dimension (male /bi/ and female /pi/ or female /bi/ and male /pi/). Finally, the two orthogonal sets consisted of all four stimuli (male /bi/, female /bi/, male /pi/, female /pi/). Each stimulus was presented 20 times in random order within each test set. The presentation order of the stimulus sets was pseudocounterbalanced across subjects within each TG, with each subject being presented a different order of the six sets of stimuli.

The subjects were given a practice set immediately prior to the presentation of a test set. Within a practice set, each stimulus was presented 10 times in random order. The subjects were required to meet a 90% correct criterion on the practice set before being run in the actual test set. If they failed to meet the criterion, they were presented with the practice set again. Any subjects failing to meet criterion on the second presentation were dropped from the experiment. A total of 4 subjects in the phoneme TG and 5 in the talker TG required a second presentation of a practice set. All 9 of these subjects met criterion on the second presentation. Practice sets served to familiarize the subjects with the experimental task and test sets. In an attempt to reduce the influence of practice/experience on the reaction time (RT) measure (subjects simply getting faster with increased exposure to the task), the first practice–test pairing in each subject's counterbalanced order was repeated at the end of the experiment. The data from the first test set were discarded and excluded from further consideration.

Table 1
Syllable Durations (in Milliseconds) for the Natural Speech Tokens Used in Experiments 1–3, Along With the Ratio of the Overall Durations of the Fast-to-Slow Tokens

Talker	Segment	Rate		Fast-to-Slow Ratio
		Fast	Slow	
		Experiments 1 and 2		
Male	B	185	399	.46
	P	197	404	.49
Female	B	175	380	.46
	P	179	406	.44
		Experiment 3		
Female	B	220	489	.45
	D	211	479	.44

The stimuli were stored on disk at 12-bit quantization in a lab computer (NEC 386-20), reconstructed at a sampling rate of 20 kHz, low-pass filtered at 9.9 kHz, amplified (Yamaha AX-630), and then presented to listeners over headphones (Sennheiser HD 450) at a comfortable listening level of approximately 74 dB SPL.

The subjects were instructed to classify the target segments as quickly and as accurately as possible by pushing one of two appropriately labeled keys on a four-button response box. The stimulus to response-button assignment was switched for half of the subjects in each experimental group. If subjects failed to respond within 1,500 msec of stimulus onset, the response was scored as an error. All responses and latencies were recorded by the computer with RT measured from stimulus onset.

An entire experimental session lasted about 45 min. Prior to the beginning of the experimental session, the subjects were familiarized with the general procedure, the use of the response box, and the stimuli. Stimulus familiarization consisted of having each subject listen to five repetitions of each of the four stimulus tokens before proceeding to the first RT practice set. In an attempt to consistently emphasize the relevant (target) dimension, the presentation order of the stimuli was dependent upon group assignment. For example, the subjects in the phoneme group listened to the male and female /bi/ stimuli, followed by the male and female /pi/ stimuli, while subjects in the talker group listened to the male stimuli (/bi/, /pi/) followed by the female stimuli (/bi/, /pi/). Each stimulus repetition was separated by a 2-sec silent interval. Each stimulus-repetition block was separated from the succeeding block by a 5-sec silent interval.

Results

The mean RT and accuracy for the different TGs across the three different experimental conditions are presented in Table 2 for the fast and the slow tokens. Overall accuracy in classifying the tokens along the two dimensions was quite high, averaging 97% correct or higher in each of the different conditions. A four-way analysis of variance (ANOVA) was used to analyze the mean RTs, with rate (fast vs. slow) and TG (phoneme vs. talker) as between-subjects factors and experimental condition (control, correlated, and orthogonal) and stimulus token (either /b/ and /p/ or male and female) as within-subjects factors. As can be seen in Table 2, the overall pattern of responses for the fast and slow syllables was fairly similar across the different experimental conditions for both TGs. Although the main effect of rate was not significant, there was a significant interaction of experimental condition × rate [$F(2,72) = 3.20, p < .05$]. The nature of this interaction can be seen by comparing the condition means for the fast and slow tokens in Table 2. The fast tokens produced a small 2-msec increase in RT between the control (473 msec) and correlated (475 msec) conditions while

Table 2
Experiment 1: Mean Reaction Times (in Milliseconds) and Percent Accuracy for the Phoneme and Talker Target Groups in All Three Conditions for the Fast and Slow Utterances

Target Dimension	Control		Orthogonal			Correlated			Group Mean
	M	Accuracy	*M*	Accuracy	Difference From Control	*M*	Accuracy	Difference From Control	
				Fast Tokens					
Phoneme	463	98.5%	536	98.5%	+73	492	97.5%	+29	497
Talker	482	97.5%	480	97.3%	−2	457	98.0%	−25	473
Condition means	473		508			475			
				Slow Tokens					
Phoneme	463	97.5%	540	95.8%	+77	448	97.3%	−15	484
Talker	438	97.0%	444	97.5%	+6	395	98.6%	−43	426
Condition means	451		492			422			

the slow tokens produced a 29-msec decrease in RT between the same two conditions (451 and 422 msec for the control and correlated conditions, respectively). The ANOVA also revealed a significant main effect of experimental condition [$F(2,72) = 22.8$, $p < .001$] and a significant condition × TG interaction [$F(2,72) = 10.43$, $p < .0001$]. None of the other main effects or interactions was significant.

The interaction of experimental condition with TG is shown in Figure 1. There is a large increase in RT between the control and the orthogonal conditions for the phoneme TG and a small increase in RT between these two conditions for the talker TG. To examine this interaction further, separate three-way ANOVAs, with rate (fast vs. slow) as a between-subjects factor and experimental condition (control, correlated, and orthogonal) and stimulus token (either /b/ and /p/ or male and female) as within-subjects factors, were conducted on the results from the two TGs. Since this interaction involved a between-subjects variable (TG) and a within-subjects variable (condition), neither post hoc analysis nor planned comparisons could be performed on the cell means. The ANOVA for the phoneme TG showed a significant effect of condition [$F(2,36) = 35.9$, $p < .0001$] and a significant rate × condition interaction [$F(2,36) = 3.82$, $p < .05$]. Planned comparison revealed the nature of this interaction. First, there was a reliable increase in RT between the control and the orthogonal conditions for both speaking rates ($p < .0001$). Second, although there was no reliable difference in RT between the control and correlated conditions for the slow tokens ($p > .3$), there was a significant increase in RT between these two conditions for the fast tokens ($p < .05$). Finally, there was a significant effect of stimulus due to the fact that the /b/ tokens were responded to with significantly slower RTs than the /p/ tokens [$F(1,18) = 15.2$, $p < .001$].

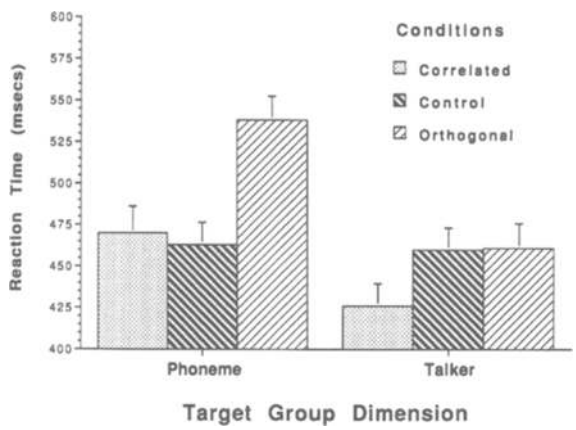

Figure 1. Mean reaction time across the three experimental conditions in Experiment 1 for the phoneme and talker target groups. The tokens classified by the subjects were male and female versions of /bi/ and /pi/. The averages are combined across the fast and slow versions of these tokens.

The ANOVA for the talker TG also produced a significant effect of condition [$F(2,36) = 5.11$, $p < .02$] but no significant rate × condition interaction [$F(2,36) = .58$, $p > .56$]. Planned comparisons revealed that the effect of condition was due entirely to a significant decrease in RT between the control and correlated conditions ($p < .05$). There was no reliable difference in RT between the control and the orthogonal conditions ($p > .85$).

The overall discriminability of the phoneme and talker dimensions was compared by examining the mean RTs in just the control conditions. A three-way ANOVA with rate and TG as between-subjects factors and stimulus token as a within-subjects factor showed no difference between the two TGs [$F(1,36) = .02$, $p > .89$]. This result indicates that the phoneme and talker dimensions were comparable in their overall discriminability. Moreover, neither the effect of rate nor its interaction with TG was significant (both Fs $< .7$, $p > .4$), indicating that this pattern occurred for both the fast and the slow speech tokens. It is possible that the similarity in RTs for the two dimensions was due to a floor effect on subjects' response times rather than an equivalence in discrimination between the two dimensions. However, this is unlikely, because subjects were able to classify the tokens more rapidly in at least one of the correlated conditions than in the control condition.

Finally, the data were examined to determine whether the increase in RT in the orthogonal condition was the result of a speed–accuracy tradeoff. Correlations were calculated on the difference in RT between the control and the orthogonal conditions and the overall accuracy in the control and orthogonal conditions for both the phoneme and the talker TGs. Separate correlations were calculated on the data for the fast and slow tokens. However, because of the high accuracy across all the subjects (*SD*s < 2.5%), none of the correlations was significant (all *p*s > .10).

In summary, the results show orthogonal interference between the two stimulus dimensions for the phoneme TG but no such interference for the talker TG. This pattern of responses occurred for both the fast and the slow speech tokens. These results closely parallel the findings of Tomiak et al. (1991), using synthetic speech tokens and a moderate speaking rate, even though different groups of subjects and very different stimuli were used in the two experiments. Comparison of the mean RTs for the two experiments showed very little difference in either the pattern of RTs for the same two TGs or the actual mean RTs.

Discussion

The results of this first experiment demonstrate that when subjects were instructed to classify speech tokens along a phonemic dimension, RT reliably increased with orthogonal variation along the talker dimension over a control condition in which no such orthogonal variation occurred. However, when subjects classified the same tokens with respect to talker characteristics, RT did not increase with additional orthogonal variation along the phoneme dimension. The asymmetry in the results also demonstrates that the increase in RT between the control and the orthogonal groups was not simply due to the fact that there were two different stimuli in the former condition and four in the latter. Both dimensions were equal in their overall discriminability. If the increase in RT in the orthogonal condition were due simply to the number of different stimuli presented in the orthogonal and control conditions, there should have been an increase in RT in the orthogonal conditions for both the phoneme and the talker dimensions (see also Melara & Marks, 1990).

The results comparing the control and the correlated conditions were not as consistent. When two dimensions are processed integrally, they often produce a decrease in RTs in the correlated condition known as a redundancy gain. We found no evidence of a redundancy gain for either of our two phoneme TGs even though both TGs showed evidence of orthogonal interference (see Mullennix & Pisoni, 1990, for a similar finding). The subjects in the phoneme TG presented with the fast tokens actually produced a reliable increase in RT from the control to the correlated conditions. Such an increase is referred to as a redundancy loss. Why it occurred for this TG is currently unknown. It may have been due to the particular pairing of the different stimulus dimensions. Evidence of a redundancy loss has been obtained in other experiments investigating the processing interactions of integrally related stimulus dimensions, depending upon whether the stimulus dimensions were positively or negatively correlated (Melara, 1989; Pomerantz & Garner, 1973). Alternatively, it may reflect some subjects' reliance on the talker dimension to make the discrimination, even though they were instructed to classify the phoneme dimension. Given the positive correlation between the two dimensions, either dimension could be used to correctly classify the tokens in the correlated condition. Even though the RTs in the control conditions were not reliably different for the phoneme and the talker TGs, the classification along the talker dimension may have required slightly more time than the phoneme classification.

Both talker TGs did show evidence of a significant redundancy gain, even though neither TG showed any evidence of orthogonal interference. Exactly why this pattern occurs is unclear, although Tomiak et al. (1991), using synthetic speech tokens, found a similar pattern of results for their talker TG: no orthogonal interference and a significant redundancy gain. It is not clear, however, whether this represents a true redundancy gain. It is possible that subjects were adopting a selective serial processing strategy in which they used the more discriminable dimension to classify the tokens. The design of these experiments does not enable us to rule out such a strategy, because each subject classified the tokens along just one of the two dimensions. Finally, a redundancy gain can occur with separate processing of the perceptual dimensions (Biederman & Checkosky, 1970). The presence or absence of a redundancy gain should therefore not be considered a strong indication of the way in which the two stimulus dimensions are processed (see Eimas et al., 1978, for discussion).[3]

The results of Experiment 1 demonstrate that (1) interference exists between the phoneme and talker dimensions in a speeded-classification task, and (2) interference between these two dimensions is asymmetric: talker variation interferes with phonetic classification but phonetic variation does not interfere with talker classification. These results are similar to the findings of earlier studies using synthetic speech tokens (Eimas et al., 1981; Tomiak et al., 1991; Wood, 1974). The similarity in the pattern of responses indicates that such an asymmetry is not a property of the quality or discriminability of the stimuli. The results contrast with those of Mullennix and Pisoni (1990), who found that phonetic variation could interfere with speeded classification along a talker dimension for natural speech tokens. The reason for this difference is still unknown. It may reflect differences in the speech tokens used in the experiment. Alternatively, it may reflect differences in experimental procedures. Mullennix and Pisoni had each subject classify the tokens along both stimulus dimensions, whereas in the current experiment different groups of subjects participated in the phoneme and talker TGs.

The pattern of results in the present study can be accounted for by either a serial or a parallel-contingent processing model (see Eimas et al., 1981; Turvey, 1973). For example, in a serial model, the talker characteristics would be encoded first, followed by the encoding of the pho-

netic information by a different processing mechanism. On this view, the talker characteristics are encoded while phonetic variation remains unprocessed. Thus, phonetic variation would not interfere with the processing of talker information. However, since talker characteristics are processed before phonetic information, talker variation would be available to the system when phonetic information was encoded and might be expected to interfere with phonetic processing. Alternatively, a parallel-contingent process model would hold that both talker and phoneme characteristics were processed in parallel, with a decision regarding phoneme characteristics dependent upon an evaluation of talker properties.

The results of this study and of the earlier research of Mullennix and Pisoni (1990) and Tomiak et al. (1991) suggest that classification of talkers with respect to gender is part of a more general mechanism of talker normalization. This assumption underlies other research on talker normalization (see Johnson, 1990c; Jongman & Miller, 1990). However, the results of current work can only specify what happens in gender normalization. Whether such results will generalize to the more typical situation in which there are talker differences within the same gender remains to be determined.

EXPERIMENT 2

This next experiment investigated whether the encoding of speaking rate was accomplished in a manner similar to that of talker information. If so, the pattern of interference between the phonetic dimension of voicing and the dimension of speaking rate should be similar to that found in Experiment 1 between the phoneme and talker dimensions. For two reasons, we predicted that rate variation would interfere with the classification of the tokens along the phoneme dimension. First, as shown in Experiment 1, such other characteristics of the speech signal as talker variation can interfere with phonetic processing. Second, speaking rate is an intrinsic characteristic of many important acoustic cues used in phonetic perception. As such, the encoding of phonetic information may be dependent upon the encoding of rate information. Of primary interest in this experiment was the question of whether phonetic variation would also interfere with classification along the rate dimension. If speaking rate is encoded in the same manner as talker information, phonetic variation should have little influence on classification along the rate dimension.

Method

Subjects. The subjects, who received course credit for their participation, were a different group of 40 undergraduates at the University of Arizona. All subjects were native speakers of English with no known history of a speech or hearing disorder.

Stimuli. The stimuli consisted of the same eight natural speech syllables spoken by the male and female talkers used in Experiment 1.

Procedure. The equipment and procedures used to present the syllables and record response times were identical to those used in Experiment 1. Half of the subjects were presented with the male tokens. Ten of these subjects were assigned to a phoneme TG; the other 10 were assigned to a rate TG. The other half of the subjects were presented with the female tokens, with 10 subjects assigned to the phoneme TG and the remaining 10 to the rate TG. Subjects in the two phoneme TGs were told that they would be listening to versions of the syllables /bi/ and /pi/ produced at fast and slow speaking rates. For these 20 subjects, the phonemic identity of the tokens ("B" vs. "P") was specified as the target dimension. Subjects in the two rate TGs were told that they would be listening to fast and slow versions of the syllables /bi/ and /pi/. For these 20 subjects, the "speaking rate" was specified as the target dimension. A total of 3 subjects in the phoneme TG and 2 in the rate TG failed to meet criterion on the first presentation of a practice set. All subjects met criterion on the second presentation.

Results

The results for the phoneme and rate TGs are presented in Table 3 for the male and female tokens. As in the first experiment, overall accuracy in the two TGs across the different conditions was quite high, averaging better than 95%. A four-way ANOVA was used to analyze the mean RTs, with talker (male vs. female) and TG (phoneme vs. rate) as between-subjects factors and experimental condition (control, correlated, and orthogonal) and stimulus token (either /b/ and /p/ or fast and slow) as within-subjects factors. As can be seen in Table 3, the overall pattern of

Table 3
Experiment 2: Mean Reaction Times (in Milliseconds) and Percent Accuracy for the Phoneme and Rate Target Groups in All Three Conditions for the Male and the Female Utterances

	Condition								
			Orthogonal			Correlated			
Target Dimension	Control				Difference From Control			Difference From Control	Group Mean
	M	Accuracy	*M*	Accuracy		*M*	Accuracy		
				Male Tokens					
Phoneme	452	95.0%	485	96.3%	+33	457	96.5%	+5	465
Rate	525	97.8%	585	96.4%	+60	471	98.0%	−54	527
Condition means	489		535			464			
				Female Tokens					
Phoneme	460	97.5%	501	95.1%	+41	499	98.3%	+39	487
Rate	560	96.5%	608	95.6%	+48	539	98.3%	−21	569
Condition means	510		555			519			

responses for the male and female syllables was fairly similar. This was confirmed by the fact that the effect of talker was not significant and showed no interaction with any other main effect or interaction. The effect of TG was significant due to the fact that the rate TGs produced slower RTs overall [$F(1,36) = 7.6, p < .01$]. This will be discussed more later on. There was also a significant effect of experimental condition [$F(2,72) = 20.7, p < .0001$] and a significant interaction between TG and experimental condition [$F(2,72) = 10.3, p < .0001$]. The nature of this interaction is shown in Figure 2. As can be seen in the figure, both TGs had increases in their overall mean RTs between the control and the orthogonal dimensions, indicating mutual interference between the phoneme and rate dimensions. There were differences between the two TGs with respect to the correlated conditions. For the phoneme TG, there was a small increase in RT between the control and correlated conditions, while for the rate TG there was a decline in RT between these two conditions. To examine this interaction in greater depth, separate three-way ANOVAs, with talker as a between-subjects factor and experimental condition and stimulus token as within-subjects factors, were conducted on the data from the phoneme and rate TGs.

The ANOVA for the phoneme TG showed a significant effect of condition [$F(2,36) = 4.8, p < .02$]. Planned comparisons revealed a significant increase in RT between the control and the orthogonal condition ($p < .005$) and no significant difference between the control and correlated conditions ($p > .07$). The ANOVA for the rate TG also produced a significant effect of condition [$F(2,36) = 24.3$, $p < .0001$]. Planned comparisons revealed a significant difference between the control and orthogonal ($p < .0002$) and the control and correlated conditions ($p < .01$). Thus, the orthogonal interference between the two dimensions was mutual.

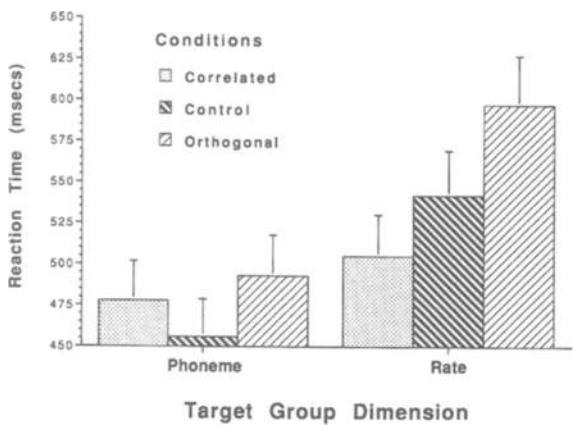

Figure 2. Mean reaction time across the three experimental conditions in Experiment 2 for the phoneme and rate target groups. The tokens classified by the subjects were fast and slow versions of /bi/ and /pi/. The averages are combined across the male and female versions of these tokens.

The effect of stimulus was significant for the rate TG [$F(1,18) = 66.8, p < .0001$]. This was due to the fact that the fast tokens were responded to much faster than the slow tokens (506 and 590 msec, respectively). There was also a significant stimulus × condition interaction [$F(2,36) = 3.71, p < .05$]. Planned comparisons showed that the fast tokens produced reliably shorter RTs than the slow tokens. More importantly, they also revealed a similar pattern across all three conditions for *both* types of tokens. The correlated condition produced reliably shorter RTs than the control condition, which produced RTs that were reliably shorter than those of the orthogonal condition. Thus, the orthogonal interference from the phoneme dimension occurred regardless of the overall speaking rate of the token.

The finding that the two dimensions produced mutual orthogonal interference raised the question of whether the interference between the two dimensions was symmetrical. This was tested by using a two-way ANOVA with talker and TG as between-subjects conditions, to examine the difference scores between the control and orthogonal conditions. The ANOVA showed no significant effect of either TG or talker [both Fs$(1,36) < 2.36, p > .2$]. The interference between the two dimensions was therefore both mutual and symmetrical. This finding also showed that the interaction of TG and experimental condition was due to the differences in the correlated conditions between the two TGs. Currently, it is not clear why such differences occur, although they may be due to overall differences in discriminability between the two dimensions.

The overall discriminability of the two dimensions was assessed with a three-way ANOVA of the control conditions (talker and TG as between-subject factors and stimulus token as a within-subject factor). This ANOVA was significant, suggesting that the phoneme and rate dimensions were not similar in their overall discriminability [$F(1,36) = 11.4, p < .002$]. However, this finding was solely the result of longer RTs to the slow tokens in the rate TG. An examination of the mean RTs for the tokens in each TG indicated that fast tokens in the rate TG (497 msec) were similar to both of the tokens in the phoneme TG (459 and 452 msec for the /b/ and /p/ tokens, respectively), while the slow tokens in the rate TG produced significantly longer RTs (587 msec).[4] It is not clear why the slow tokens produced longer RTs in the rate TG. One possibility is that the perceptual system waits until some significant part of the syllable has been presented before a decision is made about speaking rate (see Diehl, Kluender, Foss, Parker, & Gernsbacher, 1987; although see Miller & Dexter, 1988).

More relevant is an examination of the RTs in the two control sets for the phoneme TG. The subjects in this TG were presented with the fast /bi/ and /pi/ tokens in one control set and the slow tokens in the other. The mean RTs from these two sets (441 and 470 msec for the fast and slow sets, respectively) were then combined to obtain the mean control RT and compared with the orthogonal condition

in the analysis above. One possibility is that the presence of the slow tokens in the orthogonal set slowed the responding primarily to the fast tokens, with little impact on the classification of the slow tokens. Thus, averaging across the fast and slow tokens in the orthogonal condition might result in a mean RT that is significantly different only from the fast control set and not the slow control set. However, true orthogonal interference from variation in speaking rate should produce an increase in RT along the phoneme dimension for both the fast and the slow tokens.

To address this issue, the RTs for the phoneme TG in the orthogonal condition were compared with the RTs in the two different control sets using a two-way ANOVA. This ANOVA had talker (male vs. female) as a between-subjects factor and condition (fast control, slow control, and orthogonal) as a within-subjects factor. The results of the ANOVA indicated a significant effect only of condition [$F(2,36) = 11.7, p < .0001$]. Planned comparisons revealed that the RTs in the fast and slow control sets were reliably different ($p < .01$). More importantly, the mean RT in the orthogonal condition (493 msec) was significantly different from the RTs in both the fast ($p < .0001$) and the slow control sets ($p < .05$). Thus, there was a reliable increase in RT in the orthogonal condition over both control sets, even though the slow tokens produced RTs that were significantly longer than those of the fast tokens. This analysis confirms the finding that orthogonal variation in speaking rate produces interference in the speeded classification of consonants.

Finally, the data were examined for evidence of speed–accuracy tradeoffs. As in Experiment 1, all the subjects responded with a high degree of accuracy indicating no speed–accuracy tradeoffs.

In summary, the results show orthogonal interference between the phoneme and rate dimensions for both the phoneme TG and the rate TG. Moreover, the amount of interference for the two dimensions did not differ statistically. The results also show evidence of a redundancy gain, but only for the rate TG. Whether this was a true redundancy gain or simply the result of selective serial processing of the two correlated dimensions could not be determined.

Discussion

The results of this experiment indicate that when subjects are asked to classify speech tokens along a phonemic dimension in which there is orthogonal variation with respect to speaking rate, their RTs are longer than those of a control condition in which no such orthogonal variation occurs. This finding is similar to that found in Experiment 1, where orthogonal variation along a talker dimension also interfered with phonemic classification of the speech tokens. Moreover, orthogonal variation along the phoneme dimension also interfered with the subjects' classification of the tokens with respect to speaking rate, and the amount of the interference was similar to that of the rate variation on the phonemic classification. This finding indicates that the processing dependency between the phoneme and rate dimensions is both mutual and symmetric. This contrasts with the results of Experiment 1, which showed an asymmetric dependency in the processing of the phoneme and talker dimensions. Taken together, the results from these first two experiments suggest that talker and rate information are encoded in different ways by different mechanisms.

This pattern of responses occurred for both the male and the female talker's tokens and has been obtained using synthetic speech tokens in which speaking rate was manipulated simply by deleting the final portion of the vowel (Tomiak et al., 1991). Thus, the pattern of interference between the rate and phoneme dimensions appears to be the result of the way in which the rate and phoneme dimensions are processed rather than characteristics of the specific speech tokens used in the experiment.

One question that arises is whether integral processing of the phoneme and rate dimensions is dependent upon the particular phonetic category of the tokens. In both Experiment 1 and Experiment 2, the tokens contrasted with respect to their voicing characteristics. The subjects were, in effect, making a voicing decision about the speech tokens.[5] As reported earlier, several studies have shown that the perception of voicing is affected by the overall rate at which a syllable is spoken (e.g., Green & Miller, 1985; Green et al., 1994; Summerfield, 1981). It is therefore possible that the mutual dependency between the phoneme and rate dimensions was a result of the listeners' attending to a phonetic cue that was time or rate dependent.

There are acoustic cues that are not time dependent but are used in the perception of phonetic dimensions other than voicing or manner. Usually, speaking rate has little impact on judgments along those dimensions. One such dimension is place of articulation in which the important acoustic information for categorizing a token as bilabial (e.g., /b/) or alveolar (e.g., /d/) is determined by the frequency characteristics of its second and third formants. Miller (1981b) showed that a change in overall syllable duration had no impact on the phonetic boundary along a continuum created by varying the second formant transition from /b/ to /d/. What is not clear is whether variation in speaking rate will interfere with the speeded classification of /b/ and /d/ tokens. Even though information about speaking rate is not necessary for making a place decision per se, it is necessary for determining other phonetic characteristics of the speech tokens such as their voicing and manner characteristics. It is therefore possible that rate information is encoded regardless of whether attention is focused on a phonetic contrast that is rate dependent. If so, then variation in speaking rate should also interfere with classification of speech tokens that contrast only in their place characteristics. The purpose of Experiment 3 was to address this question.

EXPERIMENT 3

This experiment was similar to Experiment 2 except that a new female talker was recorded saying the syllables /bi/ and /di/ at fast and slow speaking rates. The subjects

in the phoneme TG were asked to classify the tokens as "b" and "d" while the subjects in the rate TG were asked to classify the tokens as either "fast" or "slow."

Method

Subjects. A new set of 20 undergraduates at the University of Arizona participated as subjects in this experiment. All subjects received course credit for their participation and were native speakers of English with no known history of a speech or hearing disorder.

Stimuli. The tokens used in this experiment consisted of natural /bi/ and /di/ tokens produced by a new female talker at fast and slow speaking rates. The talker was recorded and the tokens digitized using the procedures described in Experiment 1. A single fast and slow /bi/ and /di/ token were selected based on the basis of their overall durations. Each token was judged by the experimenters to be a clear representative of its respective phonemic category. The durations of these tokens are presented in Table 1. They were selected such that the overall ratios of the fast and slow durations for each syllable type were similar to those of the tokens used in the previous experiment. The root mean square amplitude levels of the different tokens were digitally equated, and each token was isolated and saved in its own file for presentation purposes.

Procedure. The subjects were run using the same equipment and procedures as in the previous two experiments. However, since tokens from only one talker were used in this experiment, only a single group of subjects was examined. The subjects were divided into the same TGs used in the previous experiment, and the same procedures and criteria were used. Ten of the subjects were assigned to the phoneme TG and 10 were assigned to the rate TG. The subjects in the phoneme TG were told that they would be listening to the syllables /bi/ and /di/ spoken at a fast and a slow speaking rate. For these subjects, the phonemic identity of the tokens ("B" vs. "D") was specified as the target dimension. The subjects in the rate TG were told that they would be listening to fast and slow versions of the syllables /bi/ and /di/. For these subjects, the speaking rate (fast vs. slow) was specified as the target dimension. None of the subjects in the rate TG and only 1 subject in the phoneme TG failed to meet criterion on a practice set. The 1 subject did meet criterion on the second presentation of the practice set.

Results

The results for the phoneme and rate TGs across the different experimental conditions are presented in Table 4. As can be seen in the table, accuracy in the two TGs across the different experimental conditions was high, averaging 97% or better. Given the high accuracy of the subjects, there was no evidence of a speed–accuracy tradeoff in the subjects' responses. Since tokens from only a single talker were used in this experiment, the mean RTs were analyzed using a three-way ANOVA with TG (phoneme vs. rate) as a between-subjects factor and experimental condition (control, correlated, and orthogonal) and stimulus token (either /b/ and /d/ or fast and slow) as within-subjects factors. The ANOVA revealed no significant effect of TG, although the effect of experimental condition was significant [$F(2,36) = 6.6, p < .005$]. As can be seen in the table, the orthogonal condition produced the slowest RTs and the correlated condition the fastest RTs, with the control condition occurring in the middle for both TGs. A planned comparison of the control and orthogonal dimensions indicated that the increase in RT was significant ($p < .05$). A planned comparison of the control and correlated conditions revealed no significant decline in RT for the correlated condition ($p > .14$). With no TG × experimental condition interaction, this pattern of RTs occurred for both target dimensions. As in the previous experiment, the slow tokens produced longer RTs in the rate TG, resulting in a significant stimulus × TG interaction [$F(1,18) = 8.3, p < .01$].

As in Experiment 2, the results of this experiment show that the phoneme and rate dimensions produced mutual orthogonal interference in the classification of tokens along the two dimensions. The magnitude of the interference was tested to determine whether it was symmetrical. A one-way ANOVA of the difference scores between the control and orthogonal conditions for the two TGs showed no reliable difference [$F(1,18) = 1.01, p > .32$]. Thus, as in Experiment 2, the amount of interference between these two dimensions was equal.

As in Experiment 2, the RTs were substantially faster for the phoneme control condition than for the rate control condition, indicating some difference in the overall discriminability between these two dimensions. The ANOVA of just the control RTs indicated a marginal effect of TG ($p > .09$) although, as in the previous experiment, there was also a significant interaction between TG and stimulus token [$F(1,18) = 16.4, p < .001$]. As in Experiment 2, the RTs for the rate TG to the slow tokens were longer than the RTs to the fast tokens. In the phoneme TG, the RTs to the /bi/ and /di/ tokens were nearly identical. Thus, the difference in overall discriminability between the two dimensions is probably the result of longer RTs to the slow tokens.

The mean RT in the orthogonal condition was compared with the mean RTs in the fast and slow control sets for the phoneme TG. This ANOVA revealed a significant effect of condition [$F(1,36) = 7.6, p < .01$]. Planned comparisons revealed no significant difference between the fast and the slow control sets (410 and 435 msec, respectively, $p > .24$). This finding is similar to that of Miller (1981b), who also found that speaking rate had no reli-

Table 4
Experiment 3: Mean Reaction Times (in Milliseconds) and Percent Accuracy for the Phoneme (/b/ vs. /d/) and Rate Target Groups in All Three Conditions for a New Female Talker

	Condition								
			Orthogonal			Correlated			
Target	Control				Difference			Difference	Group
Dimension	*M*	Accuracy	*M*	Accuracy	From Control	*M*	Accuracy	From Control	Mean
Phoneme	424	98.5%	465	97.5%	+41	399	98.5%	−25	429
Rate	512	98.5%	537	96.5%	+25	489	97.0%	−23	513
Condition means	468		501			444			

able effect on the identification times of /ba/ and /da/. Interestingly, while the orthogonal condition (465 msec) was significantly different from the fast control condition ($p < .01$), it was not quite significantly different from the slow control condition ($p = .11$). This difference from Experiment 2 may be due to the fact that half as many subjects were run in this experiment. However, it is also possible that speaking rate had a weaker interference on phonetic classification due to the nature of the phonetic distinction between /b/ and /d/.

In summary, the results in this experiment also indicate orthogonal interference between the phoneme and rate dimensions. This interference occurred for both the phoneme TG and the rate TG and was statistically equal for the two dimensions.

Discussion

The results of this experiment demonstrate that variation in speaking rate interferes with the phonemic classification of syllables even when the syllables differ only with respect to their place of articulation. Thus, rate information is encoded regardless of whether the information is relevant to the particular phonetic contrast that listeners are attending to. This finding is consistent with Miller's notion that the processing of rate information is obligatory (Miller, 1987a, 1987b). Together with the findings of Experiment 2, these results show that listeners are unable to selectively attend to either speaking rate or the phonetic characteristics of speech. Thus, unlike talker characteristics, the processing of speaking rate is an integral part of the processing of the phonetic characteristics of the speech signal.

EXPERIMENT 4

The results of the first three experiments provide evidence that talker and rate information are encoded by different kinds of mechanisms. Talker information is encoded separately from phonetic information either by a separate, parallel process (see Mullennix & Pisoni, 1990) or by an earlier, serial process. Rate information is encoded with phonetic information, perhaps because it is an intrinsic part of many of the acoustic cues used to establish phonetic identity. Thus, even though speaking rate and talker characteristics are both qualities of a talker's voice, they appear to be encoded differently during phonetic perception. Such a finding suggests that talker and rate normalization are accomplished by using different kinds of mechanisms.

The purpose of Experiment 4 was to examine how talker and rate normalization mechanisms jointly influenced the identification of speech tokens varying in both talker characteristics and speaking rate. To address this issue, we examined how speaking rate and talker characteristics influenced the perception of voicing in an initial stop consonant. To investigate the impact of rate normalization, synthetic speech tokens were manipulated with respect to their overall vowel duration. Cutting back on the overall vowel duration causes speech tokens to sound as if they were produced at a faster rate of speech, shifting the voicing boundary toward shorter VOT values. To investigate the impact of talker normalization, we chose to manipulate two acoustic cues that are used to distinguish the voicing quality of consonants and that vary as a function of the gender of the talker: $F0$ and $F1$ onset frequency. Changes in $F0$ have been shown to affect the VOT boundary with an increase in $F0$ resulting in a shorter VOT boundary (Haggard, Ambler, & Callow, 1970; Haggard, Summerfield, & Roberts, 1981). Typically, the effect of $F0$ on the voicing boundary is considered to reflect within-talker variation in the production of voiced and voiceless stops. However, $F0$ does vary as a function of the gender of the talker, with male talkers having lower $F0$s than female talkers (see Peterson & Barney, 1952). Raising $F1$ onset frequency causes listeners to hear more voiceless tokens along a /g/–/k/ continuum (Lisker, 1975). As a result of their smaller vocal tracts, female talkers tend to have higher $F1$ onset frequencies than male talkers.

Previous experiments investigating the influence of speaking rate or talker characteristics (such as $F1$ onset frequency and $F0$) on the voicing boundary have always kept one type of characteristic constant (either rate or talker) and varied the other. The current experiment investigated how these two types of characteristics jointly influenced the perception of voicing by varying both types simultaneously. Four different /bi/–/pi/ continua were synthesized to characterize the variation in the acoustic signal with respect to speaking rate and talker characteristics that had existed in the previous experiments. These tokens consisted of a fast and a slow continuum corresponding to that of a male talker and a fast and a slow continuum corresponding to that of a female talker. These tokens were presented to listeners in three different blocking conditions. The first was a blocked-by-rate condition in which the fast male and female continua were randomized together and presented in one block of trials and the slow male and female continua were presented in a separate block of trials. This condition is similar to the orthogonal condition in Experiment 1. The second was a blocked-by-talker condition. The fast and slow male continua were presented in one block while the fast and slow female continua were presented in the other block. This condition is similar to the orthogonal condition in Experiment 2. Finally, there was a mixed condition in which all the tokens were presented together.

Previous research by Shinn, Blumstein, and Jongman (1985) has shown that presenting speech tokens under blocked conditions can affect how listeners process the acoustic information. Shinn et al. examined the effects of overall syllable duration on the perception of /b/ and /w/. In one of their conditions, a speech continuum was constructed that varied between /ba/ and /wa/ and was patterned after natural speech utterances. By editing these tokens, several additional continua were created with shorter overall syllable durations. The tokens were presented to listeners under two different blocking conditions: a mixed condition in which all the tokens were randomized together and a blocked condition in which the tokens were random-

ized within a particular syllable duration. In the mixed condition, Shinn et al. observed the usual effect of syllable duration on the perception of /b/ vs. /w/ (see Miller & Liberman, 1979). However, in the blocked-by-duration condition, the effect of syllable duration disappeared. These results show that under certain conditions, blocking tokens with respect to overall syllable duration can eliminate the influence of syllable duration on the perception of the initial consonant, although why is not clear.

On the basis of Shinn et al.'s (1985) results, it was predicted that the mixed condition would show an influence of both speaking rate and talker characteristics on the VOT boundary. With regard to the two blocking conditions, we predicted that the dimension that varied within a block would continue to influence the VOT boundary. Of particular interest was whether the dimension held constant would also influence the voicing boundary. For example, when the stimuli were blocked by rate, would the rate dimension continue to influence the voicing boundary? If the results replicated those of Shinn et al., the blocked dimension would have little or no impact on the VOT boundary. However, the results of our previous experiments raised the possibility of alternative outcomes. For example, the results of Experiment 2 had indicated that the speaking rate was encoded along with the phonetic information required to make a voicing decision. Therefore, it seemed quite possible that speaking rate would continue to influence the VOT boundary even when the tokens were blocked by rate. The situation for the condition in which the tokens were blocked by the talker was less clear, because the results of Experiment 1 had indicated that talker characteristics were encoded by a mechanism separate from that involved in phonetic processing. Therefore, talker characteristics might not influence the VOT boundary when the tokens were blocked by the talker.

Method

Subjects. The subjects, who received course credit for their participation, were 30 undergraduates at the University of Washington. All subjects were native speakers of English with no known history of a speech or hearing disorder.

Stimuli. The stimuli consisted of four synthetic /bi/–/pi/ speech continua varying in VOT from 10 to 55 msec. Two of the continua were created by using the Klatt (1980) software synthesizer implemented on an LSI-11/73 computer. The first continuum was synthesized with pitch and formant values appropriate for a male talker. The second continuum was synthesized with pitch and formant values appropriate for a female talker. Each token was 340 msec in duration and contained a 5-msec release burst (AF), followed by a period of formant transitions (30 msec for $F1$ and 45 msec for $F2$ and $F3$) and a 290-msec period of steady-state information. The synthesizer was set in cascade mode, and five formants were synthesized for the male tokens while only four formants were synthesized for the female tokens in order to provide a better approximation of a female voice (see Klatt, 1980). The steady-state values (in hertz) of the formants and their bandwidths were as follows. For the male tokens, $F1$ = 330 Hz, $B1$ = 50 Hz; $F2$ = 2200 Hz, $B2$ = 100 Hz; $F3$ = 3000 Hz, $B3$ = 130 Hz; $F4$ = 3300 Hz, $B4$ = 250 Hz; $F5$ = 3850 Hz, $B5$ = 200 Hz. For the female tokens, $F1$ = 376 Hz, $B1$ = 50 Hz; $F2$ = 2508 Hz, $B2$ = 100 Hz; $F3$ = 3100 Hz, $B3$ = 130 Hz; $F4$ = 3300 Hz, $B4$ = 250 Hz. For the male tokens, the /bi/ endpoints of the continua had starting formant frequencies of 200, 1100, and 2150 Hz; for the female tokens, the starting frequencies were 228, 1254, and 2300 Hz. $F4$ and $F5$ remained constant across the utterances. The release burst was created by stimulating the formants with a noise excitation source (aspiration) and widening the $F1$, $F2$, and $F3$ formant bandwidths to 450, 170, and 200 Hz, respectively. Voicing began 10 msec after the onset of the release burst. Once initiated, voicing amplitude (AV) was held constant over the transition and/or steady-state durations of the tokens. The pitch contour for the male tokens rose from 100 to 120 Hz during the first 55 msec and then fell linearly to 90 Hz at syllable offset; the pitch contour for the female tokens rose from 180 to 206 Hz during the first 55 msec and then fell linearly to 160 Hz at syllable offset. Stimulus offsets were tapered by linearly interpolating the bandwidth of the first formant ($B1$) from 50 to 100 Hz over the last 45 msec. To make the remainder of the continua, the time from the onset of the release burst to the onset of voicing was increased by increasing the duration of aspiration in 5-msec increments. This had the effect of increasing $F0$ and $F1$ onset frequency at the time of voicing across the continua. Each continuum contained a total of 10 tokens, with VOTs ranging from 10 to 55 msec. These tokens are referred to as the "slow" continua.

Two additional male and female continua were created by shortening the duration of the original tokens to 118 msec. This was accomplished by deleting 232 msec of the final steady-state portion of the vowel without tapering the offset of either the amplitude or $F0$. The editing was performed using a commercial waveform analysis/editing package. All editing was performed at a zero-crossing in the signal. These tokens are referred to as the "fast" tokens. Thus, four /bi/–/pi/ continua were created corresponding to fast (118 msec) and slow (340 msec) speaking rates produced by a male and a female talker. Although not explicitly tested, the tokens sounded like good synthetic approximations of male and female talkers to the experimenters. More importantly, they were clearly different on the basis of their talker characteristics.

Procedure. Subjects were assigned randomly to one of three different experimental groups. Each group received the same speech continua. The only difference across the three groups was the way in which the stimuli were blocked during presentation. The first group of subjects was presented with the fast and slow male tokens in one block of trials and the fast and slow female tokens in the other block of trials. This condition is referred to as the blocked-by-talker group. The second group of subjects was presented with the fast male and female tokens in one block of trials and the slow male and female tokens in a second block of trials. This condition is referred to as the blocked-by-rate group. The third group of subjects was presented with tokens from all four continua randomized across the two blocks of trials. This condition is referred to as the mixed group. The order of presentation of the different blocks of trials was counterbalanced across the subjects in each group.

For the blocked-by-rate and the blocked-by-talker groups of subjects, each block contained 21 different randomized orders of two speech continua. The first order of each block was considered practice and was excluded from the data analysis. For the mixed group, the first block consisted of 11 different randomized orders of each of the four speech continua. The first order was considered practice and was excluded from the data analysis. The second block consisted of 10 randomized orders of the four continua. By the end of the experiment, each subject was presented with 21 repetitions of each of the four continua, for a total of 840 trials.

An entire experimental session lasted approximately 45 min. The subjects were tested individually in a small, sound-attenuated room. The stimuli were presented on-line, using a lab computer (DEC LSI-11/73). The stimuli were output at a 10-kHz sampling rate, low-pass filtered at 4.9 kHz, amplified (Yamaha A-420 stereo amplifier), and presented to subjects over headphones (Telephonics TDH-39P) at a comfortable listening level (approximately 78 dB SPL). The subjects indicated their responses by pressing the appropriately labeled

button on a two-button Microsoft mouse attached to a computer terminal (NDS GP-29). The lab computer recorded all responses and controlled the rate of presentation. The stimuli were presented every 2 sec.

Results

Two types of analyses were conducted on the identification data from this experiment. In the first type of analysis, the VOT boundaries for each of the four different continua were calculated for each subject by fitting a linear regression line to the boundary region of each subject's identification function, taking as the category boundary the VOT value that corresponded to 50% of the /b/ responses. Boundary regions were defined as that part of the identification function ranging between the last VOT value at which the percentage of /b/ responses was 90% or greater and the first VOT value at which the percentage of /b/ responses was 10% or less. The mean VOT boundaries for the four continua in the three experimental conditions are presented in Table 5. These boundaries for each condition were then analyzed using a two-factor ANOVA with rate (fast vs. slow) and talker (male vs. female) as within-subjects factors. A similar analysis was performed on the total percent of /b/ responses for each of the four continua for each subject. These means are presented in Table 6. Since the ANOVAs for these two different analyses revealed the same patterns of effects, the results of only the phoneme boundaries are described.

Consider first the results for the blocked-by-talker group. The identification function for these conditions are shown in Figure 3. As can be seen in the figure, blocking the tokens by talker influenced the identification of the tokens in the boundary region of the identification functions. There was no indication of an effect at the end points of the continua. Blocking by talker resulted in a shift in the VOT boundary of 4.94 msec between the fast and the slow tokens and only a 1.10-msec shift in the VOT boundary between the female and male tokens. The ANOVA revealed a significant effect of rate [$F(1,9) = 20.93, p < .002$] but no effect of talker [$F(1,9) = .88, p > .37$]. The interaction between these two factors was also not significant [$F(1,9) < .01, p > .9$].

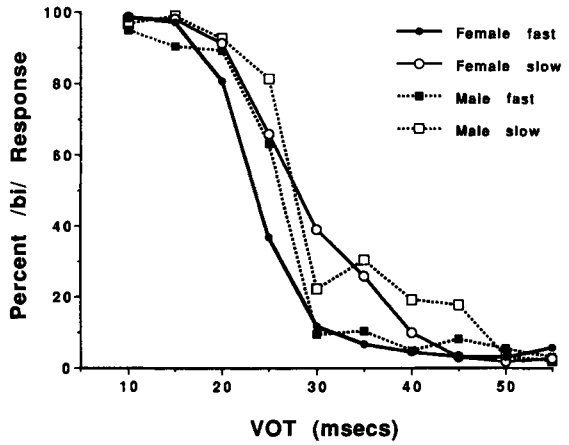

Figure 3. Group identification functions for the four /bi/–/pi/ continua in the "blocked-by-talker" condition in Experiment 4.

The results of the blocked-by-talker condition for different rates are consistent with those of previous studies showing that changes in overall syllable duration produce a significant shift in the VOT boundary of a syllable initial consonant (Green & Miller, 1985; Summerfield, 1981). Of greater interest is the finding that differences in talker, which included changes in $F0$ and $F1$ onset frequency, produced no significant effect on the VOT boundary. This result shows that when there is no variation in talker characteristics among the tokens presented in a block of stimuli, there is no impact on the processing of VOT. Subjects were able to ignore the talker characteristics and focus only on the rate variability when identifying the tokens.

Next, consider the results from the blocked-by-rate conditions (Figure 4). Again, blocking influenced the boundary region of the continua and had no influence on the end points. In this condition, there was a 4.92-msec shift in the VOT boundary between the male and female tokens and a smaller, 2.16-msec shift between the fast and slow tokens. The two-way ANOVA revealed a significant effect of talker [$F(1,9) = 10.61, p < .01$] and of

Table 5
Experiment 4: Mean Voice Onset Time Boundaries (in Milliseconds) Collapsed Across the Talker and Rate Dimensions for the Different Blocking Conditions

		Blocked by:					
		Talker		Rate		Randomized	
		M	*SE*	*M*	*SE*	*M*	*SE*
Female tokens	fast	24.96	.78	21.21	.81	22.39	1.75
	slow	29.88	1.22	23.11	1.22	26.81	2.04
Male tokens	fast	26.05	.83	25.87	.82	26.97	1.57
	slow	30.99	1.55	28.28	.81	28.60	1.03
Average across rates	female	27.42		22.16		24.60	
	male	28.52		27.08		27.78	
Differences		−1.10		−4.92		−3.18	
Average across talkers	fast	25.50		23.54		24.68	
	slow	30.44		25.70		27.71	
Differences		−4.94		−2.16		−3.03	

Table 6
Experiment 4: Mean Total /b/ Responses for Each Subject Collapsed Across the Talker and Rate Dimensions for the Different Blocking Conditions

		Blocked by:					
		Talker		Rate		Randomized	
		M	*SE*	*M*	*SE*	*M*	*SE*
Female tokens	fast	35.35	1.52	29.99	2.86	30.49	3.47
	slow	45.10	2.38	32.45	2.84	37.96	4.17
Male tokens	fast	38.73	1.78	39.19	1.48	39.75	2.78
	slow	47.57	2.65	43.32	1.41	44.29	1.83
Average across rates	female	40.22		31.22		34.22	
	male	43.15		41.26		42.02	
Differences		−2.93		−10.04		−7.80	
Average across talkers	fast	37.04		34.59		35.12	
	slow	46.33		37.88		41.13	
Differences		−9.29		−3.29		−6.01	

rate [$F(1,9) = 15.43, p < .005$]. The interaction between these two variables was not significant [$F(1,9) = .3, p > .6$]. The results of this condition are important for two reasons. First, they show that the differences in talker characteristics in these synthetic tokens can produce a significant shift in the VOT boundary. Second, they show that even when there is no rate variation across the tokens within a block of trials, speaking rate continues to influence the VOT boundary for synthetic speech tokens, although the size of that shift is smaller than when there is rate variation within a block of trials.[6]

We investigated whether there was a reliable difference between the two blocking conditions with respect to the effect of speaking rate on the VOT boundary. The difference in the VOT boundaries between the fast and slow continua (for both talkers) was calculated for the blocked-by-talker condition and the blocked-by-rate condition. The mean shift in VOT between the fast and slow tokens was 4.94 msec when the tokens were blocked by talker and only 2.16 msec when the same tokens were blocked by speaking rate. These difference scores were submitted to a two-way ANOVA with condition (talker vs. rate blocking) and talker (male vs. female) as the main factors. The ANOVA indicated a significant effect of condition [$F(1,18) = 5.32, p < .05$] but no effect of talker [$F(1,18) = .1, p > .75$]. The interaction of these two variables was also not significant [$F(1,18) = 11, p > .74$]. This analysis indicates that speaking rate had a reliably larger impact on the VOT boundary when the tokens varied in speaking rate within a block of trials (blocked-by-talker condition) than when they did not (blocked-by-rate condition).

Finally, consider the results from the mixed condition. The identification functions for these conditions are presented in Figure 5. As can be seen in Table 5, the shift in the VOT boundary was approximately the same for the two types of tokens: a 3.18-msec shift between the male and female tokens and a 3.03-msec shift between the fast and slow tokens. However, a two-factor ANOVA indicated a significant effect only of rate [$F(1,9) = 9.89, p < .02$] and not of talker [$F(1,9) = 2.49, p > .15$].[7] The interaction between these variables was not significant [$F(1,9) = 2.45, p > .15$].

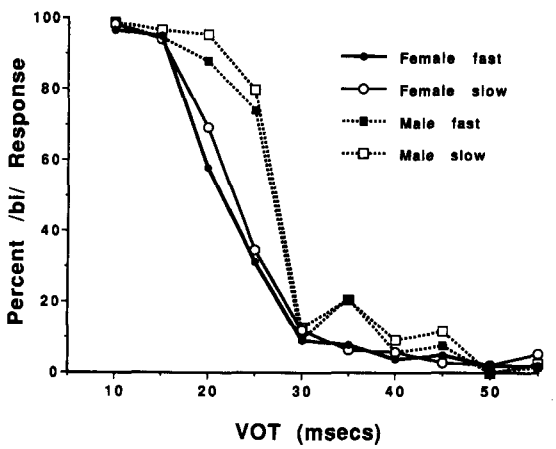

Figure 4. Group identification functions for the four /bi/–/pi/ continua in the "blocked-by-rate" condition in Experiment 4.

Discussion

The results of the blocked-by-talker condition indicate that when talker characteristics are kept constant across a block of trials, the relevant acoustic information that specifies a change in talker has little influence on the phonetic classification of voicing. The results from the blocked-by-rate condition demonstrate that when the tokens within a block of trials vary with respect to talker, talker characteristics can influence the perception of voicing when there is no other variation across the speech tokens. The results of the mixed condition indicate some limitations on that influence. Specifically, when another dimension such as speaking rate is also varying across trials, the influence of talker characteristics was eliminated in the analysis of the VOT boundaries but marginal in the total number of /b/ analyses.

The dimension of speaking rate shows a different pattern of influence. Speaking rate influenced the perception of voicing regardless of whether or not there was rate variation within a block of trials. It also influenced the perception of voicing even when there was additional variation in talker characteristics, as in the mixed condi-

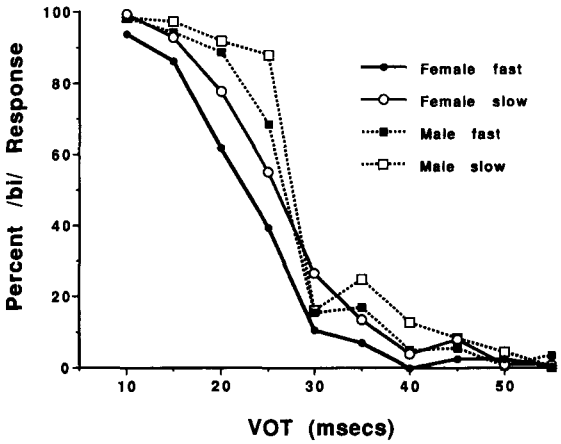

Figure 5. Group identification functions for the four /bi/–/pi/ continua in the "mixed" condition in Experiment 4.

tion. However, the amount of influence was reliably greater when the tokens were blocked by talker than when they were blocked by rate. This result is similar to that of Shinn et al. (1985), who also found a greater influence of speaking rate on the perception of /ba/–/wa/ when tokens varied across speaking rates within a block of trials than when the same tokens were blocked by speaking rate. The finding demonstrates that blocking does have an effect on the way rate variation influences the perception of voicing.

How does one account for the effects of blocking on the processing of either rate or talker characteristics? One possibility is a perceptual "tuning" account proposed by Nusbaum and Morin (1992). Nusbaum and Morin found that monitoring times for various phonemic targets were influenced by the amount of talker variation in the target and the distractor items. Monitoring times were faster when a single talker produced all the tokens in a given block of trials than when the items were produced by two or more talkers. According to Nusbaum and Morin, talker characteristics are computed and held in working memory to aid in the acoustic-to-phonetic mapping during speech perception. When the talker remains the same, the phonetic system can simply access the previously computed talker characteristics held in working memory. However, when the talkers vary randomly from trial to trial, talker characteristics need to be recomputed, and this takes resources away from phonetic processing. Thus, the perceptual system "tunes" to the characteristics of a particular talker during phonetic perception.

Some of the results of this last experiment are consistent with this view. For example, the perceptual system does appear to tune to rate characteristics. This is indicated by the fact that a reliable difference in the VOT boundary between the fast and slow tokens remained even when the tokens were blocked by speaking rate. However, the results with respect to talker characteristics are not consistent with this notion. When the tokens were blocked by talker, there was no difference in the VOT boundary between the male and the female tokens. If the perceptual system had tuned or normalized for a particular talker, then there should have been a difference in the VOT boundary when the tokens were blocked by talker.

An alternative possibility is that variation along a particular dimension (either rate or talker) within a block of trials draws attention to that dimension. This increased attention may result in that dimension's being given more "perceptual weight" than other acoustic dimensions in the recognition of the segment. This is similar to Nosofsky's (1986; Nosofsky, Clark, & Shin,1989) notion that attention can serve to warp the underlying psychological space onto which stimulus information is mapped during perceptual classification. There is evidence to suggest that this kind of warping of the perceptual space begins very early in life and is a function of linguistic experience (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992). According to Nosofsky, attention to a particular dimension will stretch the psychological space and result in better discriminability along that dimension. Taking attention away from a dimension results in a shrinkage of the psychological space, which results in much poorer discriminability along the dimension.

Previous studies have shown that listening conditions can influence the relative importance of different acoustic cues during phonetic perception. For example, Gordon et al. (1993) showed that directing attention to a concurrent task altered the relative importance of different acoustic cues during speech perception. Similarly, Miller and Wayland (1993) showed that adding noise to the speech signal increased the importance of overall syllable duration in the perception of /b/ versus /w/. Perhaps variation in speech tokens along a particular dimension within a block of trials works in a similar fashion by directing attention to that dimension, increasing its importance during speech perception. Whether this is accomplished via a "stretching" of the underlying psychological space, as suggested by Nosofsky, or through a reweighting of the information along a particular dimension when it is mapped onto underlying phonetic representations remains to be determined.[8]

Interestingly, varying both dimensions within a block of trials in the mixed condition resulted in intermediate shifts in the VOT boundary as compared with the two blocking conditions. The effect of speaking rate on the VOT boundary was significant in this condition, while the effect of talker was nearly significant. One possibility is that the attentional resources were split over the two dimensions, resulting in only moderate influence of both dimensions on the VOT boundary.

Although we currently favor the attentional explanation as an account for the results in this experiment, it should be pointed out that there are other possible accounts for the effect of blocking on the VOT boundary, including a reduction in stimulus uncertainty. The results of this experiment do not favor one explanation over the other and additional studies will be necessary to determine which account provides the best explanation for the blocking effects obtained in this experiment. The important point about the findings from this experiment is that

blocking did have an effect on the way in which the talker and rate dimensions influenced the VOT boundary, and that influence was asymmetric: speaking rate influenced the VOT boundary regardless of the blocking condition, whereas talker characteristics influenced the VOT boundary only in the situation in which there was just talker variation and no other variation among the stimuli.

GENERAL DISCUSSION

The results of these experiments point out differences in the way in which talker characteristics and speaking rate affect phonetic perception. The processing of speaking rate and phonetic perception appear to be tightly coupled, while the processing of talker characteristics seems to be more independent. Phonetic variation produced no interference in the classification of the tokens with respect to talker. Although the processing of talker characteristics can interfere with processing along the phoneme dimension, it does not do so consistently. If the talker differences are not made perceptually prominent in some way, the acoustic characteristics that distinguish talkers have little impact on phonetic perception.

What type of processing model can account for the pattern of results obtained in these experiments? A serial model in which the talker characteristics are encoded at an early level of analysis followed by phonetic encoding processes accounts for the speeded-classification results but has difficulty in accounting for the effects of blocking conditions on the VOT boundary. A strictly serial model would predict that talker characteristics would always influence the phonetic encoding processes, since they occur first and provide input to the phonetic encoding system. However, as the results from Experiment 4 demonstrate, the influence of talker characteristics on the VOT boundary varies with respect to blocking conditions. The blocking manipulations may have produced a change in the attention focused on the talker characteristics. Thus, for a serial model to account for the results of these experiments, there must be a mechanism by which the processing system can ignore the talker characteristics and yet still process the rest of the speech signal.

Mullennix and Pisoni (1990) proposed that talker and phonetic characteristics were encoded in a parallel-contingent fashion (Turvey, 1973). The output of acoustic-to-phonetic processes is dependent upon an output with respect to talker characteristics, while the encoding of talker characteristics can be carried out somewhat independently of the acoustic-to-phonetic encoding. The results of the current study are consistent with this view and emphasize the relative independence of the processing of talker characteristics with respect to phonetic encoding processes. However, the results also demonstrate that although phonetic encoding processes *can* depend upon the analysis of talker characteristics, the use of such information is not obligatory (Mullennix & Pisoni, 1990). It is possible to have phonetic decisions which are not affected by the characteristics of the talker's voice unless attention is directed toward that information. When additional variation is present in the signal from another dimension, say, speaking rate, attention to the talker dimension may be reduced, resulting in a diminished influence on the voicing boundary. This possibility would predict that variation along the talker dimension should produce less interference on speeded classification along the phoneme dimension in the presence of additional variation in speaking rate. To our knowledge, a three-way orthogonal variation has not been investigated in a speeded-classification task. However, the results of such an experiment could prove to be informative.

The encoding of speaking rate shows a different pattern of results and therefore needs to be accounted for in a different manner in such a model. The processing of the phoneme and rate dimensions is integral and symmetrical. It is not possible to process one dimension without also processing the other. Thus, the output of phonetic encoding processes is dependent upon the analysis of speaking rate. Even though the two dimensions are processed in an integral manner, it is possible that attention might affect the degree to which speaking rate influences encoding along the phoneme dimension. However, unlike talker characteristics, the perceptual system cannot completely ignore rate variation in the tasks used in this study. Thus, the influence of speaking rate on phonetic processing is obligatory, as Miller (1987a) has argued. One way of accounting for these characteristics is to argue that the encoding of phoneme and rate characteristics are part of the same process. This process would be linked to a separate, parallel process which extracts information about talker characteristics.

The link between these two processes allows for talker normalization during phonetic perception. However, the interaction between the two processes is not mandatory. The results from the current study suggest that the output from the process responsible for encoding talker information can be ignored depending upon attentional resources. Consistent with this notion are the findings of Logan and Pastore (1990), who found that talker normalization occurs only when the perceptual system can detect that a change in talker had been made. If there was no change in talker or if for some reason the perceptual system fails to notice the change in talker, talker normalization does not occur.

In summary, the results of this study indicate that information about a talker's voice and speaking rate are encoded by different kinds of mechanisms during phonetic processing. This finding is of interest in light of recent experiments investigating the impact of talker and rate variation on both the recognition of speech and the memory for words. These studies have shown an advantage for single-talker conditions over multitalker conditions in the recognition of speech presented in noise (Mullennix, Pisoni, & Martin, 1988; Nygaard, Sommers, & Pisoni, 1994) or for the monitoring of phonetic segments (Nusbaum & Morin, 1992). Similar advantages have been obtained for single-rate conditions over multirate conditions (Sommers, Nygaard, & Pisoni, 1994). Although talker and rate variation have similar impacts on the processing of

speech, the results of the current study indicate that these two dimensions are encoded by different kinds of mechanisms during phonetic perception.

REFERENCES

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

DARWIN, C. J., MCKEOWN, J. D., & KIRBY, D. (1989). Perceptual compensation for transmission channel and speaker effects on vowel quality. *Speech Communication*, **8**, 221-234.

DIEHL, R. L., KLUENDER, K. R., FOSS, D. J., PARKER, E. M., & GERNSBACHER, M. A. (1987). Vowels as islands of reliability. *Journal of Memory & Language*, **26**, 564-573.

DIEHL, R. L., & WALSH, M. A. (1989). An auditory basis for the stimulus-length effect in the perception of stops and glides. *Journal of the Acoustical Society of America*, **85**, 2154-2164.

EIMAS, P. D., & MILLER, J. L. (1980). Contextual effects in infant speech perception. *Science*, **209**, 1140-1141.

EIMAS, P. D., TARTTER, V. C., & MILLER, J. L. (1981). Dependency relations during the processing of speech. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech* (pp. 283-309). Hillsdale, NJ: Erlbaum.

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

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

GORDON, P. C., EBERHARDT, J. L., & RUECKL, J. G. (1993). Attentional modulation of the phonetic significance of acoustic cues. *Cognitive Psychology*, **25**, 1-42.

GREEN, K. P., & MILLER, J. L. (1985). On the role of visual rate information in phonetic perception. *Perception & Psychophysics*, **38**, 269-276.

GREEN, K. P., STEVENS, E. B., & KUHL, P. K. (1994). Talker continuity and the use of rate information during phonetic perception. *Perception & Psychophysics*, **55**, 249-260.

HAGGARD, M. P., AMBLER, S., & CALLOW, M. (1970). Pitch as a voicing cue. *Journal of the Acoustical Society of America*, **31**, 613-617.

HAGGARD, M. P., SUMMERFIELD, A. Q., & ROBERTS, M. (1981). Psychoacoustical and cultural determinants of phoneme boundaries: Evidence from trading *F*0 cues in the voiced-voiceless distinction. *Journal of Phonetics*, **9**, 49-62.

HILLENBRAND, J., & HOUDE, R. A. (1995). Vowel recognition: Formants, spectral peaks, and spectral shape. *Journal of the Acoustical Society of America*, **98**, 2949.

JOHNSON, K. (1990a). Compensation for talker variability and vowel variability in the perception of fricatives. *Journal of the Acoustical Society of America*, **87**, S118.

JOHNSON, K. (1990b). Contrast and normalization in vowel perception. *Journal of Phonetics*, **18**, 229-254.

JOHNSON, K. (1990c). The role of perceived speaker identity in *F*0 normalization of vowels. *Journal of the Acoustical Society of America*, **88**, 642-654.

JONGMAN, A., & MILLER, J. D. (1990). Method of location of burst-onset spectra in the auditory-perceptual space: A study of place of articulation in voiceless stop consonants. *Journal of the Acoustical Society of America*, **89**, 867-873.

KLATT, D. H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, **67**, 971-995.

KUHL, P. K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustical Society of America*, **66**, 1668-1679.

KUHL, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior & Development*, **6**, 263-285.

KUHL, P. K., WILLIAMS, K. A., LACERDA, F., STEVENS, K. N., & LINDBLOM, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, **255**, 606-608.

LADEFOGED, P. (1967). *Three areas of experimental phonetics*. London: Oxford University Press.

LISKER, L. (1975). Is it VOT or a first-formant transition detector? *Journal of the Acoustical Society of America*, **57**, 1547-1551.

LOGAN, R. J., & PASTORE, R. E. (1990). Talker normalization and speaker recognition by humans: One mechanism or two? *Journal of the Acoustical Society of America*, **87**, S70.

LUCE, P. A., FEUSTEL, T. C., & PISONI, D. B. (1983). Capacity demands in short-term memory for synthetic and natural speech. *Human Factors*, **25**, 17-32.

MELARA, R. D. (1989). Dimensional interaction between color and pitch. *Journal of Experimental Psychology: Human Perception & Performance*, **15**, 69-79.

MELARA, R. D., & MARKS, L. E. (1990). Dimensional interactions in language processing: Investigating directions and levels of crosstalk. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 539-554.

MILLER, J. L. (1981a). Effects of speaking rate on segmental distinctions. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech* (pp. 39-74). Hillsdale, NJ: Erlbaum.

MILLER, J. L. (1981b). Some effects of speaking rate on phonetic perception. *Phonetica*, **38**, 159-180.

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

MILLER, J. L. (1987b). Rate-dependent processing in speech perception. In A. W. Ellis (Ed.), *Progress in the psychology of language* (pp. 119-157). Hillsdale, NJ: Erlbaum.

MILLER, J. L., AIBEL, I. L., & GREEN, K. (1984). On the nature of rate-dependent processing during phonetic perception. *Perception & Psychophysics*, **35**, 5-15.

MILLER, J. L. & BAER, T. (1983). Some effects of speaking rate on the production of /b/ & /w/. *Journal of the Acoustical Society of America*, **73**, 1751-1755.

MILLER, J. L., & DEXTER, E. R. (1988). Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology: Human Perception & Performance*, **14**, 369-378.

MILLER, J. L., GREEN, K.P., & REEVES, A. (1986). Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast. *Phonetica*, **43**, 106-115.

MILLER, J. L., & LIBERMAN, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semi-vowel. *Perception & Psychophysics*, **25**, 457-465.

MILLER, J. L., & VOLAITIS, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*, **45**, 506-512.

MILLER, J. L., & WAYLAND, S. C. (1993). Limits on the limitations of context conditioned effects in the perception of [b] and [w]. *Perception & Psychophysics*, **54**, 205-210.

MULLENNIX, J. W., & PISONI, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics*, **47**, 379-390.

MULLENNIX, J. W., PISONI, D. B., & MARTIN, C. S. (1988). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America*, **85**, 365-378.

NEAREY, T. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America*, **85**, 2088-2113.

NOSOFSKY, R. M. (1986). Attention, similarity and the identification-categorization relationship. *Journal of Experimental Psychology: General*, **115**, 39-57.

NOSOFSKY, R. M., CLARK, S. E., & SHIN, H. J. (1989). Rules and exemplars in categorization, identification, and recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 282-304.

NUSBAUM, H. C., & MORIN, T. M. (1992). Paying attention to differences among talkers. In Y. Tohkura, Y. Sagisaka, & E. Vatikiotis-Bateson (Eds.), *Speech perception, speech production, and linguistic structure* (pp. 113-134). Tokyo: OHM.

NYGAARD, L. C., SOMMERS, M. S., & PISONI, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, **5**, 42-46.

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

POMERANTZ, J. R., & GARNER, W. E. (1973). Stimulus configuration in selective attention tasks. *Perception & Psychophysics*, **14**, 565-569.

REMEZ, R., RUBIN, P., NYGAARD, L., & HOWELL, W. (1987). Perceptual

normalization of vowels produced by sinusoidal voices. *Journal of Experimental Psychology: Human Perception & Performance*, **13**, 40-61.

SHINN, P. C., BLUMSTEIN, S. E., & JONGMAN, A. (1985). Limitations of context conditioned effects in the perception of [b] and [w]. *Perception & Psychophysics*, **38**, 397-407.

SOMMERS, M. S., NYGAARD, L. C., & PISONI, D. B. (1994). The effects of speaking rate and amplitude variability on perceptual identification. *Journal of the Acoustical Society of America*, **96**, 1314-1324.

SUMMERFIELD, Q. (1981). On articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 1074, 1095.

TOMIAK, G. R., GREEN, K. P., & KUHL, P. K. (1991). Phonetic coding and its relationship to talker and rate normalization. *Journal of the Acoustical Society of America*, **90**, S2363.

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.

VOLAITIS, L. E.., & MILLER, J. L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America*, **92**, 723-735.

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

NOTES

1. The effects of *F*0 and *F*1 onset frequency are usually considered to reflect intratalker variation in the production of voiced and voiceless stop consonants. However, since *F*0 and *F*1 are presumably evaluated with respect to talker characteristics during vowel perception, it seems possible that such an evaluation would also be used to normalize for the initial consonant. At issue here is when, during the processing of the initial consonant, the information about *F*0 and *F*1 is made available to the perceptual system and what mechanism is responsible for providing such information. Clearly, more research is needed to determine whether the role of *F*0 in the perception of voicing reflects only intraspeaker adjustment or also some more global normalization for talker differences.

2. A typical Garner task involves the selection of items that vary along two different perceptual dimensions (e.g., shape and color: a black circle, black square, red circle, and red square). Various combinations of these four objects are then presented to subjects under three different experimental conditions: control, orthogonal, and correlated. In each condition, the subjects are required to make a two-choice speeded classification of the items. In the control condition, the items are selected such that they vary only along a single dimension (e.g., color: black circle and red circle), with the other dimension (shape) held constant. The subjects are asked to classify the tokens along the dimension of color, and the condition provides a base level of performance for classifying the items along one of the dimensions.

In the orthogonal condition, all four items are presented; however, the subjects again classify the items only along the relevant dimension. This condition examines the impact of the irrelevant dimension on the classification of the items. An increase in RT to classify the tokens in the orthogonal condition over the control condition indicates that variation along the irrelevant dimension influences the processing of the attended dimension. This pattern of responses, termed "orthogonal interference," is indicative of "integral" processing of the two dimensions. No increase in RT in the orthogonal condition over the control condition demonstrates that the two dimensions can be selectively attended to and is taken as evidence that the two dimensions are processed separately.

In the correlated condition, two items are selected such that they have different specifications along both dimensions (e.g., red circle and black square) and again presented for classification along just one of the dimensions. In this condition, the correlated dimension provides redundant information for classifying along the target dimension. If the two dimensions are processed integrally, the redundant information can result in a decrease in RT from the control or single-dimension condition to the correlated condition. This decrease in RT is typically called a "redundancy gain."

3. We included the correlated conditions in the design of these experiments for two reasons. First, their inclusion made the design comparable to previous studies that had investigated the processing interactions between segmental and nonsegmental speech dimensions (e.g., Eimas et al., 1981; Mullennix & Pisoni, 1990; Tomiak et al., 1991; Wood, 1974). Second, even though redundancy gains are not conclusive evidence for integral processing of two stimulus dimensions, their presence can provide supportive evidence.

4. It is possible that the overall difference in discriminability between the two target dimensions may be responsible for the pattern of responses obtained in this experiment. However, there are two reasons why this is unlikely. First, the pattern of responses closely resembles the findings found in a previous study using synthetic tokens (Tomiak, Mullennix, & Sawusch, 1987) in which the overall discriminability between the phoneme and rate dimensions was not significantly different. Second, the RTs to the fast tokens were comparable to the RTs in the phoneme target group. As reported earlier, both the fast and the slow tokens showed evidence of orthogonal interference from the phoneme dimension. Thus, we consider it unlikely that the orthogonal interference between the phoneme and rate dimensions is due only to an overall discriminability difference between the two dimensions.

5. Eimas et al. (1981) have shown that orthogonal interference occurs among various segmental dimensions regardless of whether subjects classify the tokens along those dimensions (e.g., voiced vs. voiceless) or just categorize the speech token (e.g., /b/ vs. /p/).

6. It should be pointed out that for both groups of subjects, talker characteristics had a much more variable influence on the VOT boundaries than did speaking rate, resulting in much smaller mean square errors for the rate variable than for the talker variable in the ANOVAs. Thus, even though the difference in the VOT boundaries in the two blocking conditions are nearly mirror images of one another, speaking rate had a reliable influence in both conditions, while talker characteristics influenced the phoneme boundary only when tokens varied in talker within a block of trials.

7. In the analysis of the total /b/ responses, the variable of talker was marginally significant [$F(1,9) = 3.89, p = .08$].

8. It is difficult to compare the results from the current experiment and those of Gordon et al. (1993). Although the task conditions served to reallocate attention in both studies, the nature of those task conditions was quite different. In the Gordon et al. study, attention was actually diverted to another, concurrent task. According to Gordon et al., this resulted in less competition between a strong cue for voicing (VOT) and a weak cue (*F*0) and allowed the weak cue to have a greater influence on the voicing decision. In the current experiment, attention may have been diverted from one speech dimension to another by the blocking condition. Thus, variation with respect to speaking rate and no variation with respect to talker may result in a shift of attention from the talker dimension to the rate dimension. This may result in a shrinking of the perceptual space along the talker dimension and therefore reduced discriminability.

(Manuscript received February 1, 1995;
revision accepted for publication July 9, 1996.)