Journal of Experimental Psychology:
Human Perception and Performance
1991, Vol. 17, No. 1, 278–288

0096-1523/91/$3.00

Integral Processing of Visual Place and Auditory Voicing Information During Phonetic Perception

Kerry P. Green and Patricia K. Kuhl
University of Washington

Results of auditory speech experiments show that reaction times (RTs) for place classification in a test condition in which stimuli vary along the dimensions of both place and voicing are longer than RTs in a control condition in which stimuli vary only in place. Similar results are obtained when subjects are asked to classify the stimuli along the voicing dimension. By taking advantage of the "McGurk" effect (McGurk & MacDonald, 1976), the present study investigated whether a similar pattern of interference extends to situations in which variation along the place dimension occurs in the visual modality. The results showed that RTs for classifying phonetic features in the test condition were significantly longer than in the control condition for the place and voicing dimensions. These results indicate a mutual and symmetric interference exists in the classification of the two dimensions, even when the variation along the dimensions occurs in separate modalities.

The acoustic realization of phonetic segments is very complex due to the large number of auditory cues that influence listeners' judgments. Numerous studies have examined phonetic perception by manipulating combinations of cues in an attempt to understand how different cues are combined during speech perception. Studies of trading relations and context effects show that the phonetic response is not a simple combination of different acoustic cue dimensions but rather a complex integration of cues in which one cue's influence often depends on the value of another cue (e.g., Fitch, Halwes, Erickson, & Liberman, 1980; Repp, 1982, for review).

Studies have also shown that visual information from a talker's face is integrated with auditory information during phonetic perception. The "McGurk effect" (McGurk & MacDonald, 1976; MacDonald & McGurk, 1978), the demonstration that an auditory /ba/ presented with a video /ga/ produces the percept /da/, indicates that visual information can influence the perceived place of articulation of well-specified auditory tokens (Summerfield, 1983, for review).

Green and Kuhl (1989) have shown that the presentation of visual place information influences not only the perceived place of articulation, a feature that is known to be highly subject to visual influences, but also the voicing feature. Green and Kuhl paired members of an auditory /ibi–ipi/ continuum with a videodisplay of a talker saying /igi/. When the auditory tokens were presented without the videodisplay, they were all heard as either /ibi/ or /ipi/. However, consistent with the McGurk effect, when both auditory and visual information were presented to observers, the tokens were perceived as ranging from /idi/ to /iti/. Green and Kuhl's experiment examined whether perception of the (nonvisible) voicing feature was also altered in the auditory–visual situation. They found that it was; the results showed that the voicing boundary for the auditory–visual tokens was significantly longer than the voicing boundary for the same auditory tokens presented without the visual information. This suggests that the auditory and visual information is processed as a "whole unit" rather than as separable phonetic features, with vision providing place information and audition providing manner and voicing information (MacDonald & McGurk, 1978).

The purpose of this study's experiments was to examine the integration of auditory and visual information for speech using a different experimental procedure, the selective attention paradigm described by Garner (1974). The selective attention paradigm, as applied by Garner, requires subjects to make a two-choice speeded classification of four objects 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 the objects would be presented to subjects under three different experimental conditions: Control, Orthogonal, and Correlated. In the Control condition, two objects are selected that vary along only a single dimension (e.g., color: black circle and red circle) with the other dimension, shape, held constant. Subjects classify the tokens along that dimension (i.e., black or red). This condition provides a base level of performance for classifying the objects when they differ along only one of the dimensions.

In the Orthogonal condition, all four objects are presented to subjects who again classify the stimuli along that dimension. The Orthogonal condition examines the role of the unattended dimension. An increase in the reaction times (RTs) for the Orthogonal over the Control condition indicates that the variation along the irrelevant dimension influences classification along the relevant dimension. This pattern of results is taken as evidence that the two dimensions are

---

This research was supported by National Institutes of Health Grant NS-26475 to Kerry P. Green and National Institutes of Health Grant HD-18286 to Patricia K. Kuhl.

We thank Jim Cutting, Arthur Samuel, and two anonymous reviewers for their helpful comments on an earlier version of this article. We also thank Rosemarie Hardy and Shari Speigler for their assistance in collecting the data. A portion of these data were presented at the spring meeting of the Acoustical Society of America, Seattle, 1988.

Correspondence concerning this article should be addressed to Kerry P. Green, Cognitive Science, Psychology Building, Room 312, University of Arizona, Tucson, Arizona, 85721.

perceptually processed in some complex, integrated manner. If the RTs between the two conditions are similar, then the variation in the irrelevant dimension does not interfere with the classification of the attended dimension, and the two perceptual dimensions are processed separately.

Finally, in the Correlated condition, two objects that have different specifications along both dimensions (e.g., a red circle and a black square) are presented for classification. The results from this condition are more difficult to interpret, although a decrease in RTs from the Control to the Correlated condition is often considered as evidence for the integration of the two perceptual dimensions because the covarying information along the irrelevant dimension provides redundant information that can be used to help classify the stimuli. However, this decrease in RTs for the correlated conditions (termed a *redundancy gain*) has been shown to be indicative of both integral (Lockhead, 1972) and separate processing of perceptual dimensions (Beiderman & Checkosky, 1970) and therefore a lack of a redundancy gain should not be considered a strong indication of separate processing (see Eimas, Tartter, Miller, & Keuthen, 1978, for discussion).

Garner (1974) describes a series of experiments in which the selective attention paradigm was successfully applied to situations of known integral dimensions with the expected results. In addition, the paradigm has also been applied to various speech and nonspeech stimuli (Blechner, Day, & Cutting, 1976; Eimas et al., 1978; Eimas, Tartter, & Miller, 1981; Tomiak, Mullennix, & Sawusch, 1987; Wood, 1974; Wood & Day, 1975). These studies have demonstrated that many of the acoustic cues of speech show evidence of integral processing. For example, Eimas et al. (1981) used this paradigm to investigate the classification of speech tokens along the dimensions of Place and Voicing. Using as their stimuli the syllables /ba/, /pa/, /da/, and /ta/, they found a reliable increase in RTs for the Orthogonal over the Control conditions for both the Place and Voicing dimensions. Moreover, the amount of increase in RTs was similar for both dimensions. Thus, there was a mutual and symmetric interaction among the acoustic cues underlying the Place and the Voicing dimensions.

The McGurk effect provides an ideal situation for determining whether the integral processing of the Voicing and Place dimensions is related to the modality in which the information is presented. Taking advantage of the McGurk effect, it is possible to create four auditory–visual stimuli that are perceived as unified phonetic percepts, and in which the Voicing dimension of each stimulus varies only in the auditory modality, and the Place dimension varies only in the visual modality. The following set of experiments use the selective attention paradigm to examine the processing interactions between the Voicing and Place dimensions when the variation along the two dimensions occurs in separate modalities. If visual place information does in fact influence classification along the Voicing dimension as indicated by Green and Kuhl (1989), then there ought to be a reliable increase in RT for the Orthogonal condition over the corresponding Control condition when subjects are asked to attend to the Voicing dimension, even though the visual information does not provide any direct information regarding the voicing feature.[1] In addition, if the auditory voicing information influences the classification along the Place dimension, then there ought to be a similar increase in RT when subjects are asked to attend to the Place dimension. Alternatively, because the Voicing and Place dimensions are varied in separate modalities, it is possible that the information for these two dimensions can be processed separately. If so, then subjects ought to be able to selectively attend to the information in one modality or the other, resulting in little or no difference in the mean RTs between the Orthogonal and Control conditions for either dimension.

Experiment 1: Auditory Speech Information Only

The purpose of the first experiment was to establish that the auditory tokens used in this study demonstrate integral processing in a manner similar to stimuli used in previous studies (e.g., Eimas et al., 1981). In addition, this experiment provided an Auditory-Only (AO) condition with which to compare the results of the later auditory–visual (AV) studies.

Method

Subjects

The subjects were 12 undergraduate students who were given course credit as an incentive to participate. Subjects had no reported history of a speech or hearing disorder.

Materials

The stimuli consisted of four auditory tokens (/ibi/, /ipi/, /idi/, and /iti/) naturally produced by a female talker. The tokens were low-pass filtered at 9.89 KHz, digitized using a 20 KHz sampling rate with 12 bits of amplitude quantization, and stored on a computer (LSI 11-73).

Numerous repetitions of the four stimuli were randomized and then output at a 20 KHz sampling rate, low-pass filtered at 9.89 KHz, and recorded onto the second channel of a ¾-in. videotape. A total of eight different randomizations or blocks were created. The first four blocks were used for the Control condition. Each block consisted of only two of the four stimuli varying along a single dimension (see Table 1). The fifth and sixth blocks, used for the Correlated conditions, consisted of two of the four stimuli simultaneously varying along both dimensions in a correlated manner. Each of these first six blocks contained 10 practice trials and 60 test trials, evenly divided between the two test stimuli. The seventh and eighth blocks, used for the Orthogonal conditions, consisted of all four stimuli, varying along both dimensions in a completely orthogonal manner. These two blocks contained 12 practice trials and 60 test trials, evenly divided among all four stimuli.

Procedure

Each subject participated in two 40-min. sessions conducted on different days. Each subject was presented with two of the four control

---

[1] Previous studies have shown that the visual modality provides information about place of articulation but not about voicing (Binnie, Montgomery, & Jackson, 1974; Green & Miller, 1985).

Table 1
Experimental Conditions and the Auditory Stimuli Used in Experiment 1

Experimental condition	Block number	Stimuli presented	Classified as
Control	1	/ibi;ipi/	voiced-voiceless
	2	/idi;iti/	voiced-voiceless
	3	/ibi;idi/	bilabial-alveolar
	4	/ipi;iti/	bilabial-alveolar
Correlated	5	/ibi;iti/	voiced-voiceless
	6	/ipi;idi/	bilabial-alveolar
Orthogonal	7	/ibi;ipi/ /idi;iti/	voiced-voiceless
	8	/ibi;ipi/ /idi;iti/	bilabial-alveolar

blocks (either 1 and 3 or 2 and 4), and both of the correlated and the orthogonal blocks (see Table 1). Three of the six test blocks were presented in each session, with the order of the blocks counterbalanced across subjects. The subjects were asked to listen to each token and to classify the consonant as either voiced or voiceless, or bilabial or alveolar, depending on the experimental block. The subjects indicated their decision by pressing the appropriate button on a response pad located in front of them. This pad was actually a two-button mouse (Microsoft Corporation) hooked up to a computer terminal (NDS GP-29) and could comfortably fit under the palm of the subject's hand. The response buttons were labeled with the feature dimensions appropriate to the experimental condition.

Before the start of each session, the mapping between the feature dimension labels and the four different phonemes was explained to each subject. To provide the subjects with practice in applying the labels to the four phonemes, the experimenter read out loud from a randomized list of the phonemes, and the subjects replied with the correct feature dimension label. If the subject was incorrect, the experimenter informed them of the correct label and continued. This procedure continued until the experimenter had completed the 60-item list, or until the subject had responded correctly 20 consecutive times, whichever took longer. Most of the subjects required only a single reading of the list.

The subjects were tested individually in a small, dimly lit, sound-attenuated room. The videotape was played on a videocassette player (Sony VP 2000) located in an adjoining control room. The audiosignal was presented through the loudspeaker of a color videomonitor (NEC JC-1215MA) located approximately 46 in. in front of the subject. The audiosignal was presented at a comfortable listening level of approximately 65 dB SPL (A scale, fast) measured at the peak intensity of the second vowel on a sound-level meter (Bruel & Kjaer #2203) placed at the appropriate distance and height of the subjects' heads.

Results

For each subject, the overall geometric mean RT and percentage of accurate responses was calculated in each of the three experimental conditions, for both phonetic dimensions.[2] All of the subjects responded at better than 98% accuracy in each of experimental conditions, demonstrating that they had no problems with the task or in assigning the phonetic feature labels to the stimuli.

The individual mean RTs for each subject were analyzed using a two-factor repeated-measures ANOVA with Experimental Condition and Phonetic Dimension as the main effects. The results of this analysis indicated a reliable Condition effect, $F(2,22) = 7.61$, $p < .005$, but no effect of Phonetic Dimension, $F(1,11) = 1.75$, $p > .2$ or the interaction of Condition × Dimension, $F(2,22) = 1.62$, $p > .2$. The overall means for each phonetic dimension across the three experimental conditions are presented in Table 2. Post hoc analysis using Duncan's Multiple Range Test (Kirk, 1968) indicated a reliable increase in the mean RT from the Control to the Orthogonal conditions for both the Place and the Voicing dimensions ($p < .05$); however, there was no significant difference between the Control and the Correlated conditions for either dimension.[3]

The results of this experiment clearly demonstrate a mutual interference between the two dimensions of Place and Voicing, indicating that the auditory information underlying these two dimensions is processed in an integral manner. These results closely mirror the findings of Eimas et al. (1981) for the same two dimensions, using similar stimuli.[4] In addition, there was no evidence of a redundancy gain between the Control and Correlated conditions. However, this finding is also consistent with previous studies by Eimas et al. (1981), and as mentioned earlier, the presence of a redundancy gain can be considered indicative of either integral or independent processing of the stimulus dimensions. In summary, this experiment was successful in providing a set of auditory stimuli that demonstrate integral processing of place and voicing and were therefore appropriate for pairing with the visual stimuli.[5]

---

[2] The geometric mean is the mean of the logarithmic transforms of each Ss RT score. This procedure helps to reduce the amount of positive skewness that tends to occur in RT distributions (see Kirk, 1968).

[3] Because of the possibility that the Correlated condition was masking a significant Condition × Dimension interaction, these data were reanalyzed comparing just the Control and Orthogonal conditions. The results of this analysis again demonstrated no significant Condition × Dimensions interaction, $F(1,11) = 2.45$, $p > .14$, indicating that there was no significant difference in the increase of RT between the Control and Orthogonal conditions for the two phonetic dimensions.

[4] Although our RTs are much longer than Eimas et al.'s, the difference is almost completely attributable to the fact that our stimuli consisted of medial consonants, whereas Eimas et al. used syllable initial consonants. When the duration of the initial vowel and silence (approximately 285 ms) is subtracted out of our RTs, they become comparable to those reported in Eimas et al.

[5] One criticism that might be raised with regard to these findings is that even though the subjects were asked to decompose and classify the stimuli into their underlying dimensions, the subjects may still have used the phoneme labels for the classification of the tokens. In fact, Eimas et al. (1981) have shown that similar results are obtained regardless of whether subjects are asked to classify the stimuli in terms of their underlying feature dimensions or in terms of their phonetic labels. A problem that arises is whether the increase in RTs between the Control and Orthogonal conditions for the attended stimulus dimension is actually the result of interference from the unattended dimension. An alternative possibility is that the subjects are simply sorting four items in the Orthogonal condition and only two items in the Control conditions, and sorts of four items take a longer time to complete than sorts of two items. However, Eimas et

Table 2
Reaction Time (ms) for the Two Phonetic Dimensions as a Function of the Three Experimental Conditions

Phonetic dimension	Experimental condition		
	Correlated	Control	Orthogonal
Auditory Speech Information Only—Experiment 1			
Place	799	800	1006
Voicing	774	821	931
Auditory and Visual Speech Information—Experiment 2			
Place	661	709	799
Voicing	723	896	984
Auditory Nonspeech and Visual Speech Information—Experiment 3			
Place	592	541	542
Pitch	666	787	798

Experiment 2: Auditory and Visual Speech Information

The results of the previous experiment demonstrated a mutual and symmetric influence on the classifications along the auditory Voicing and Place dimensions. The purpose of the next experiment was to determine whether a similar situation occurs between the auditory Voicing and the visual Place dimensions. To address this issue, it was necessary to create four auditory–visual tokens such that the auditory information for these tokens varied only along the Voicing dimension, and the visual information varied only along the Place dimension. This was accomplished by pairing just the auditory /ibi/ and /ipi/ from the previous experiment, which differ only in their voicing characteristic, with a visual /ibi/ and /igi/, which differ only in their place characteristics. Because of the McGurk effect, the four auditory–visual tokens that resulted from these pairings were perceived as either /ibi/, /ipi/, /idi/, or /iti/.

The specific question addressed in this experiment was whether the variation along the visual Place dimension would interfere with the classification along the auditory Voicing dimension, and whether the variation along the auditory Voicing dimension would also interfere with the classification along the visual Place dimension. If so, then the pattern of results between the Control and Orthogonal conditions should be similar to the results in the previous experiment when the variation along both dimensions occurred only in the auditory modality. Alternatively, because the voicing and place information are presented in different modalities, it may be possible for observers to process the visual place information separately from the auditory voicing information, perhaps by selectively attending to one modality or another. Previous research by Roberts and Summerfield (1981) has indicated that, at least at the initial stages of speech processing, the auditory and visual information are processed separately. If this is also the case in the present situation, then there ought to be no difference in the RTs for the Control and the Orthogonal conditions on either of the two stimulus dimensions.

Method

Subjects

The subjects were 12 new undergraduate students who were given course credit as an incentive to participate. Subjects had no reported history of a speech or hearing disorder and all had normal or corrected-to-normal vision.

Materials

Auditory stimuli. The auditory stimuli consisted of the /ibi/ and /ipi/ tokens used in the previous experiment.

Visual stimuli. The visual stimuli were videotaped repetitions of a female talker saying /ibi/ and /igi/. The talker was videotaped in such a way that only the region surrounding the talker's mouth (including the lips, jaw, and oral cavity) was visible on the videotape (see Green & Kuhl, 1989, for specific details). The videotaping procedure produced an excellent close-up of the talker's mouth with good resolution of the tongue position during articulation.

A single /ibi/ and /igi/ were selected on the basis of the quality of the articulation and the lack of any identifying extraneous movements or features. Six blocks of 70 repetitions or trials were created using a video-editing console (JVC VE-92) connected to two ¾ in. videocassette machines (JVC CR8250). Each block consisted of 10 practice and 60 test trials. The practice and test trials were evenly divided between the /ibi/ and /igi/ stimuli. These six blocks were used for the four Control and the two Correlated conditions. An additional two blocks consisting of 12 practice and 60 test trials were created for the two Orthogonal conditions.

Each trial was separated by approximately 1,300 ms of video black and consisted of 1 s of videodisplay before the onset of articulation, the articulation itself, and 1 s of videodisplay following the offset of the articulation. In addition, there was a 1-s fade-up from video black and a 1-s fade-out to video black at the start and end of each trial, respectively. This prevented any abrupt visual onsets or offsets that might have produced masking or interference effects.

Auditory–visual stimuli. The auditory–visual stimuli were created by pairing each auditory /ibi/ with /ipi/ with either the visual /ibi/ or /igi/. In the control sessions, the stimuli varied only along the auditory Voicing dimension, or the visual Place dimension. For example, in Control Condition 1, the auditory /ibi/ and /ipi/ were paired with the visual /ibi/ and were perceived as /ibi/ or /ipi/. For

---

al. (1978, 1981) have demonstrated that the increase in RT between the Control and Orthogonal conditions is not always symmetrical for different phonetic dimensions. For example, the dimensions of Place and Voicing produce a similar increase in RT between the Control and Orthogonal conditions, whereas the dimensions of Place and Manner do not. Specifically, the Place dimension produced a significantly greater increase in RT than the Manner dimension. These findings are difficult to account for with the alternative possibility. If the subjects were simply sorting the four stimuli on the basis of their nominal (phonetic) labels, why should one group of four phonetic stimuli produce asymmetric increases, whereas a different group of phonetic stimuli produce symmetric increases? It is our conclusion that the Eimas et al. findings indicate that at least some of the increase in RT between the Control and Orthogonal conditions reflects the integral processing of the underlying phonetic dimensions and not a difference in the number of items that are sorted in the two conditions.

Control Condition 2, the two same two auditory stimuli were paired with just the visual /igi/ and perceived as /idi/ or /iti/. In both conditions, the variation between the two test stimuli was along the auditory Voicing dimension. The visual dimension stayed the same. For Control Condition 3, the auditory /ibi/ was paired with both the visual /ibi/ and the visual /igi/. These two stimuli were perceived as /ibi/ and /idi/, respectively. In Control Condition 4, the auditory /ipi/ was paired with the same two visual stimuli which were perceived as /ipi/ and /iti/. For these two control conditions, the two test stimuli varied only along the visual Place dimension, whereas the auditory portion stayed the same.

In Correlated Condition 5, the auditory /ibi/ was paired with the visual /ibi/, and the auditory /ipi/ was paired with the visual /igi/. These two stimuli were perceived as /ibi/ and /iti/, respectively. In Correlated Condition 6, the auditory /ipi/ was paired with the visual /ibi/, and the auditory /ibi/ was paired with the visual /igi/. These two stimuli were perceived as /ipi/ and /idi/, respectively. Thus, perceptually, the two test stimuli varied along the auditory Voicing and the visual Place dimensions in a correlated manner in each of the two conditions.

Finally, in each of the two Orthogonal conditions, the two auditory stimuli were paired with each of the two visual stimuli to create four stimuli that were perceived as /ibi/, /ipi/, /idi/, and /iti/. These four stimuli varied simultaneously along both the auditory Voicing and the visual Place dimensions in a completely orthogonal manner. The actual auditory–visual pairings in each of the experimental conditions are depicted in Table 3.

The auditory stimuli were dubbed onto the videotape using a videocassette recorder (JVC CR8250) and a lab computer (LSI 11-73). The syllables were output at a 20 KHz sampling rate in a randomized and predetermined order, low-pass filtered at 9.89 KHz, and then recorded onto the second channel of the test videotape with a high degree of accuracy (see Green & Kuhl, 1989, for specific details of the dubbing procedure).

Procedure

The apparatus and procedure were identical to that of Experiment 1 with the single exception that both the audio- and the videosignal were presented over the videomonitor located in the testing room. At the viewing distance of 46 in., the open mouth of the talker displayed on the videomonitor subtended a visual angle of 3.65 degrees. The contrast and brightness controls on the videomonitor were set at about their midpoint levels. None of the subjects reported having any problems using the response pad while keeping their attention focused on the videomonitor.

Results

As in Experiment 1, the percentage of correct responses and geometric mean RT of each subject were calculated separately for the two stimulus dimensions in each of the three experimental conditions. An examination of the subject's accuracy indicated that these subjects averaged over 97% correct in each of the stimulus conditions.

The fact that our auditory and visual stimuli often conflicted raises the question of whether there was any influence of this conflict on the subjects' response times. To address this issue, we first analyzed the RTs to the four different auditory–visual stimulus pairs for the Control conditions as a function of the auditory–visual pairings (regardless of the dimension attended to). The mean RTs to the four stimuli are presented in Table 4. The data were analyzed using a two-factor, within-subjects ANOVA with Visual Token (b vs. g) and Audio Token (b vs. p) as the two main effects. The results revealed a significant effect of Visual Token, indicating that

Table 3
Experimental Conditions and the Auditory–Visual Stimuli Used in Experiment 2

	Stimuli				
Condition	Aud +	Vis =	Perceived as	Dimension varied	Classified as
Control					
1	/ibi/ +	/ibi/ =	/ibi/	Auditory Voicing	voiced-voiceless
	/ipi/ +	/ibi/ =	/ipi/		
2	/ibi/ +	/igi/ =	/idi/	Auditory Voicing	voiced-voiceless
	/ipi/ +	/igi/ =	/iti/		
3	/ibi/ +	/ibi/ =	/ibi/	Visual Place	bilabial-alveolar
	/ibi/ +	/igi/ =	/idi/		
4	/ipi/ +	/ibi/ =	/ipi/	Visual Place	bilabial-alveolar
	/ipi/ +	/igi/ =	/iti/		
Correlated					
5	/ibi/ +	/ibi/ =	/ibi/	Auditory Voicing & Visual Place	voiced-voiceless
	/ipi/ +	/igi/ =	/iti/		
6	/ipi/ +	/ibi/ =	/ipi/	Auditory Voicing & Visual Place	bilabial-alveolar
	/ibi/ +	/igi/ =	/idi/		
Orthogonal					
7	/ibi/ +	/ibi/ =	/ibi/	Auditory Voicing & Visual Place	voiced-voiceless
	/ipi/ +	/ibi/ =	/ipi/		
	/ibi/ +	/igi/ =	/idi/		
	/ipi/ +	/igi/ =	/iti/		
8	/ibi/ +	/ibi/ =	/ibi/	Auditory Voicing & Visual Place	bilabial-alveolar
	/ipi/ +	/ibi/ =	/ipi/		
	/ibi/ +	/igi/ =	/idi/		
	/ipi/ +	/igi/ =	/iti/		

Table 4
Reaction Time (ms) to the Four Auditory–Visual Tokens in the Control Conditions (Experiment 2)

Auditory token	Visual token	Perceived as	Reaction time
/b/	/b/	"b"	702
/p/	/b/	"p"	760
/b/	/g/	"d"	853
/p/	/g/	"t"	893

those audio stimuli paired with the video /g/ were reliably slower than those same tokens paired with the video /b/, $F(1,11) = 5.8$, $p < .04$. These results indicate that when the auditory and visual stimuli conflicted in terms of place of articulation, the subjects were reliably slower in responding during the Control conditions, than when the stimuli were congruent with respect to place of articulation. Thus, even though the subjects did not report any "discrepancy" between the auditory and visual tokens during the experiment, the pattern of RTs in the Control conditions indicate that they were affected in the amount of time that it took them to respond, even when they were attending to the Voicing dimension.

An examination of the Audio Token effect, and the interaction between the two main effects showed that neither was significant, $F(1,11) = .67$, $p > .4$ and $F(1,11) = .22$, $p > .6$ respectively. These data indicate that when the auditory and visual stimuli conflicted in terms of voicing, there is no effect on RTs in the Control condition. This finding is to be expected given that previous studies have demonstrated that subjects cannot accurately determine the voicing characteristic of visual speech tokens (Binnie, Montgomery, & Jackson, 1974; Green & Miller, 1985).

Because the previous analysis indicated that certain AV stimuli pairs may produce more interference than others, the data from this experiment were analyzed using a three-factor ANOVA with Experimental Condition (three levels: Control, Correlated, Orthogonal), Phonetic Dimension (two levels: Place, Voicing), and Response Choice two levels: Bilabial/ Voiced, Alveolar/Voiceless) as the three main effects. The results of this analysis indicated a reliable effect for experimental condition, $F(2,22) = 14.72$, $p < .0005$, Phonetic Dimension, $F(1,11) = 21.08$, $p < .0001$, and the interaction of Condition × Dimension, $F(2,22) = 4.89$, $p < .05$. All other effects and interactions were not significant ($p > .1$). The group means for each of the phonetic dimensions as a function of experimental condition are presented in Table 2.

Consider first the results from the Control and Orthogonal conditions for the Voicing dimension. As shown in the table, there is a significant increase in the mean RT from the Control to the Orthogonal condition ($p < .05$), indicating that the place information again interfered with the classification along the Voicing dimension. This result is similar to that obtained in Experiment 1 in which only auditory stimuli were used. However, in this experiment, the variation along the Place dimension occurred only in the visual modality. Therefore, the visual place information must have influenced the classification of the auditory voicing information. This result is in line with previous findings by Green and Kuhl (1989) using a different kind of experimental paradigm.

Next, consider the results for the Place dimension. Again, there is a significant increase in the mean RTs from the Control to the Orthogonal conditions ($p < .05$), demonstrating that variation along the voicing information interfered with the classification along the visual Place dimension. Because that variation occurred only in the auditory modality, the increase in RTs must have been due to the influence of the auditory voicing on the classification of the visual Place dimension.

An examination of the mean RTs for the Control and the Orthogonal conditions for the Place dimension indicated they were significantly faster than their corresponding conditions for the Voicing dimension ($p < .01$). This pattern of results is different from that of Experiment 1, in which just the auditory information was presented. This finding may indicate that visual place information can be processed faster than auditory place information, or that there is intersensory facilitation for the analysis of place information (see, e.g., Gielen, Schmidt, & Van Den Heuvel, 1983). However, an alternative explanation is that information about place of articulation from the visual modality is available at an earlier point in time than information from the auditory modality. This visual information would enable subjects to start processing the place information at an earlier point in time, leading to faster overall reaction times. However, the fact that the variation along the Voicing dimension continues to interfere with the decision along the Place dimension indicates that the auditory voicing information becomes available before the analysis of the visual place information is completed.

These findings indicate that there is a mutual influence between the auditory voicing and the visual place information for classification along either dimension. Also of interest is the question of whether the influence is symmetric. That is, whether the auditory voicing has as much influence on the classification of visual place, as the visual place does on the auditory voicing. An examination of Table 2 shows that the increase in the mean RT for the Orthogonal over the Control conditions was very similar for the two dimensions (90 and 88 ms for the Voicing and Place dimensions, respectively). A paired t test on the mean difference scores between the Orthogonal and Control conditions for each subject revealed no significant difference between the two dimensions, $t(11) = .04$, $p > .5$. This result demonstrates that the influence between the auditory Voicing and the visual Place dimensions is both mutual and symmetric.

Even though the overall increase in RT from the Control to the Orthogonal conditions was symmetric for the two dimensions, it is possible that the proportion of increase was not, given the fact that the RTs in the Control condition for the Place dimension were significantly shorter than the RTs for the Voicing dimension. To examine that possibility, the proportion of each subject's Orthogonal to Control condition RT was calculated for the Place and Voicing dimensions separately. The mean proportion for these two dimensions was 1.16 and 1.13, respectively. A paired t test on these data indicated no significant difference, $t(11) = .65$, $p > .5$. Thus, analyzing the data to take into account the different baseline

RTs for the Place and Voicing dimensions still demonstrates symmetric interference between the two dimensions.

Next, consider the results for the Correlated conditions. Unlike the results from Experiment 1, there is a significant decrease in RT between the Correlated and the Control conditions for the Voicing dimension ($p < .01$), although not for the Place dimension ($p > .1$). Further analysis however, indicated that the correlated Voicing condition was no faster than either the Correlated or the Control conditions for the Place dimension ($p > .1$). Therefore, this finding probably does not represent a real redundancy gain. Rather, it is most likely attributable to selective serial processing in which the subjects take advantage of the correlation between the two dimensions to classify the stimuli along the faster (in this case, the Place) dimension (see Eimas et al., 1978, for other instances).

Finally, consider the question of whether those AV tokens in which auditory and visual information conflicted demonstrated a different pattern of responding from those tokens in which the auditory and visual information was congruent. The mean bilabial and alveolar responses for the Place dimensions are presented in Table 5. Recall that the bilabial responses are for stimulus pairs in which the auditory and visual information were congruent with respect to place of articulation such as auditory /b/ and visual /b/. The alveolar (i.e., illusory) responses are for stimulus pairs that were incongruent, such as auditory /b/ and visual /g/. Post-hoc analysis indicated that responses to both types of stimuli show a significant increase in mean RT between the Control and Orthogonal conditions ($p < .05$). Thus, even though the incongruent or illusory stimuli produced overall slower responding in the Control conditions than did the congruent or nonillusory stimuli, both types of stimuli produced an interference in the Orthogonal condition.

A similar pattern of responses occurred for the Voicing dimension. Both the voiced and the voiceless stimuli produced significant increases in RT between the Control and the Orthogonal conditions ($p < .05$). However, both types of stimuli included AV pairs that were both congruent and incongruent with respect to place of articulation. As indicated in Table 5, this resulted in little difference in the RTs between the voiced and voiceless stimuli in each of the experimental conditions. Thus, whether information from the two modalities was congruent or conflicted had no influence on the overall pattern of responding between the Control and Orthogonal conditions for either the Place or Voicing dimension.

In summary, the results in this second experiment were remarkably similar to the findings in the first experiment in that the variation along the voicing dimension interfered with the classification of the place information, and the variation along the place dimension interfered with the classification of the voicing information. In this experiment, however, the voicing information varied only in the auditory modality, whereas the place information varied only in the visual modality. Therefore, even when the information for these two dimensions is presented in separate modalities, subjects process the information in an integral manner.

Finally, the results from this study also demonstrated that even though subjects are unaware of any discrepancy between the auditory and visual information during the McGurk effect, they are still affected by the incongruency in the information between the two modalities. Specifically, our subjects' place and voicing decision times were longer when the auditory and visual information conflicted than when they were congruent. What is important is that the incongruent stimuli were still processed in the same integral manner as were the congruent stimuli. Thus, the incongruency between the auditory and visual modalities slowed down the processing of the information but did not alter it in any fundamental way.

Experiment 3: Auditory Nonspeech and Visual Speech Information

The results of the previous experiment indicate that place and voicing information are integrally processed, even when the variation along the two dimensions occurs in separate modalities. This finding, however, raises the following question: Are the visual place and auditory voicing information processed as an integral stimulus because they happen to be different dimensions of the same stimulus (albeit in different modalities) or will any variation in one modality interfere with the classification along a stimulus dimension in another modality?

Studies have shown that when two stimulus dimensions in different modalities (such as loudness and brightness) are combined in a redundant manner, large redundancy gains do occur in an absolute-judgment task even though the stimulus dimensions are not integral (Lockhead, 1970). Experiments by Marks (1987) indicate that such dimensions can produce crossmodal interference in a speeded discrimination task. Finally, research by Melara (1989) has shown that crossmodal dimensions of pitch and color can produce mutual and symmetric interference in a speeded classification task such as used in the current study. Therefore, the purpose of this last experiment was to pair the visual stimuli used in the previous experiment with nonspeech auditory stimuli that varied along a different dimension, one clearly not related to the visual Place dimension. The particular auditory dimension selected for this experiment was Pitch (high vs. low). The specific question addressed by this experiment was whether there would be any interference between the visual Place and the auditory Pitch dimensions in the Orthogonal conditions.

Table 5
Reaction Time (ms) for the Place and Voicing Dimensions as a Function of the Phonetic Decision (Experiment 2)

Phonetic dimensions	Phonetic decision	Experimental condition		
		Correlated	Control	Orthogonal
Place	bilabial (nonillusory)	617	670	992
	alveolar (illusory)	706	749	976
Voicing	voiced	756	893	991
	voiceless	689	898	975

Method

Subjects

Twelve new subjects participated in this experiment. They all received course credit for their participation, and no one reported any history of a speech or hearing disorder. All had normal or corrected vision.

Materials

Auditory stimuli. Two new auditory tokens were created. Each token consisted of two tones of different frequencies, separated by a brief silent interval. The first tone had a frequency of 1,000 Hz and was 164 ms in duration. The second tone had a frequency of either 500 (low tone) or 2,000 Hz (high tone), depending on the token, and was 525 ms in duration. The silent interval separating the two tones was 126 ms in duration. These durations were chosen to match the durations of the first syllable, silent interval, and second syllable of the /ibi/ and /ipi/ stimuli used in the previous experiments.

Each tone was created using a program that allowed the frequency and amplitude of the tone to be varied over time. These tones were created at a 10 KHz sampling rate and stored in separate files on a microcomputer (LSI 11/73). These files were then appended together with the appropriate silent interval to make the high- and low-tone tokens. Each tone was created with a 10 ms rise time at onset and a 10 ms fall time at offset. To equate the 2,000 and 500 Hz tones for loudness, the overall amplitude of the 2,000 Hz tone was 12 dB lower than the amplitude of the 500 Hz tone.

Visual stimuli. The visual stimuli consisted of the same videotape of /ibi/ and /igi/ that were used in Experiments 2 and 3.

Auditory-visual stimuli. Four auditory-visual stimuli were created by pairing each of the high- and low-tone auditory tokens with the visual /ibi/ and /igi/ stimuli. The auditory stimuli were output at a 10 KHz sampling rate, low-pass filtered at 4.9 KHz, and dubbed onto the videotape using the procedure described in Experiment 2. Because of their durational characteristics, the tonal stimuli were perfectly synchronized to the mouth movements on the videotape. However, phenomenally, they were not perceived as speech or even speech-like.

Procedure

The apparatus and procedure were similar to those of the preceding experiments. In this experiment, however, subjects were asked to classify the stimuli as either high or low, depending on whether the second tone was either higher or lower than the first tone, or as bilabial or velar.

Results

As in the previous experiments, subjects in this experiment had no trouble performing the task. The subjects averaged better than 96% correct in each of the experimental conditions, for each of the two stimulus dimensions. The pattern of RT in this experiment, however, was quite different from that of the previous experiments. An analysis of the geometric mean RTs indicated that the effect of Experimental Condition was not significant, $F(2,22) = .877$, $p > .4$, although there was a significant effect of Stimulus Dimension, $F(1,11) = 8.15$, $p < .0001$, and Condition × Dimension, $F(2,22) = 10.53$, $p < .001$. The group means for this experiment are presented in Table 2.

A post-hoc analysis of the mean RTs, using Duncan's Multiple Range test, indicated no significant difference between the Control and Orthogonal conditions for either the Place or the Pitch dimensions ($p > .1$). Therefore, unlike the previous experiments, subjects in this experiment were able to selectively process the information underlying the different stimulus dimensions. Apparently, the information from these particular stimulus dimensions is not processed in an integral fashion.

Analysis of the remaining experimental conditions revealed a pattern of responses that was similar to the previous experiment. There was a significant reduction in the mean RTs for the Pitch-Correlated condition over the Control condition ($p < .01$) but no significant difference between the Place-Correlated and Control conditions ($p > .1$). And, the mean RTs for the Place dimension were significantly faster than the mean RTs for the Pitch dimension in each of the three experimental conditions.

In summary, the results of this experiment provided no evidence of any influence between the visual speech stimuli that varied along the dimension of Place of articulation and the auditory stimuli that varied along the nonspeech dimension of Pitch. Although there is evidence that certain auditory and visual dimensions are integrated during perception (Melara, 1989; O'Leary & Rhodes, 1984), the results of the current experiment demonstrate some limitations on the auditory and visual dimensions that are processed in an integral fashion. This finding supports the notion that the auditory Voicing and visual Place dimensions influence each other because both are dimensions of the same underlying speech stimulus.

Discussion

The purpose of the current research was to examine whether phonetic information presented in the auditory modality is processed separately or integrally with phonetic information in the visual modality. To address that issue, auditory-visual speech tokens that simultaneously varied along the dimensions of Place and Voicing were presented to subjects for speeded classification along the two dimensions.

The results of these experiments, when combined with Green and Kuhl's (1989) previous finding, provide strong support for the idea that the interference between the auditory Voicing and visual Place dimensions is not unidirectional but instead is both mutual and symmetric. Classification along the Voicing dimension was influenced by variation along the Place dimension and classification along the Place dimension was influenced by variation along the Voicing dimension, even though the Voicing dimension was specified only in the auditory domain and the Place dimension was specified only in the visual domain. In other words, processing of the two dimensions was integral despite the fact that the information was derived from two separate modalities. Finally, the results indicated that the interference between the auditory and visual modalities does not occur between any arbitrary combination of dimensions; it is limited to those dimensions that are part of the same underlying speech stimulus.

What kind of model would account for the observed interferences that occur in the classification of these AV speech stimuli? Clearly, at the earliest stages of processing, the auditory and visual speech information must be processed separately simply because different sensory systems are involved (cf. Roberts & Summerfield, 1981). The results of the third experiment indicate that auditory nonspeech and visual speech are processed separately and therefore demonstrate that the integral processing of the speech tokens is not the result of some universal integration of complex auditory and visual signals occurring at some later stage of processing. And yet, despite the fact that the voicing and place information were varied along dimensions in different modalities, these two dimensions were still processed in an integral manner. Thus, even under optimal conditions, subjects are unable to separately access the underlying dimensions of Place and Voicing.

There are two types of models that have been proposed to account for the integral processing of underlying feature dimensions of objects. One class of models, proposed by Lockhead (1972), claims that objects are initially processed or perceived as holstic units or "blobs." After recognition has occurred, the underlying feature dimensions of the object become available for further analysis. An object may then be broken down along the lines of these feature dimensions, depending on the task requirements of the subject (see also Shepp, 1989). Dimensions are integral when the analysis along one dimension necessitates the analysis along another dimension. According to this model, subjects would first recognize a phoneme in some complex multidimensional space. After recognition, if the task were to require judgments along one of the feature dimensions, the segment would be analyzed along the relevant dimension. The interaction from the other dimension would be taken as evidence of integral processing of the two dimensions.

With regard to the results from the current study, the phoneme would first be recognized using whatever information was available, auditory only or auditory–visual. Thus, the auditory and visual information would be integrated for recognition purposes. Because the component dimensions of the phoneme are integral in nature (as demonstrated by the AO results), any breakdown of the phoneme after recognition will result in integral processing even when the information is presented in different modalities. Although the results from the current study are consistent with such a model, several important questions need to be addressed, including an explanation of the mechanism used to recognize the object or phoneme in the first place. The results from the current study suggest that the phoneme must be represented for recognition purposes along some amodal metric that allows the information from both modalities to be mapped on to it in a consistent fashion. This would imply that there are no modality-specific representations of phonemes or their underlying feature dimensions. The feature dimensions themselves must also be amodal in nature.

An alternative class of models claims that the underlying feature dimensions are processed in a parallel fashion before recognition of the phoneme. For example, in a model proposed by Massaro and his colleages (Massaro, 1987; Massaro & Cohen, 1983; Massaro & Oden, 1980; Oden & Massaro, 1978), perceptual recognition is the result of a mapping process between stimulus information and stored memory representations called prototypes. Specifically, different sources of information or "cues" are evaluated with respect to the various dimensions underlying the different speech prototypes stored in memory. This process occurs independently for each cue and results in an evaluation of the degree to which each cue matches the corresponding dimension. The evaluation is graded between zero and one and thus produces a continuous, as opposed to a categorical, classification of each cue. More important, the cues may be the result of information coming from a variety of sources (some top-down, others bottom-up), including information from other modalities, such as the visual modality.

After evaluation, the cues are integrated and compared to the dimensions of the stored prototypes. The integration function is complex in that it allows those cues that are least ambiguous (as determined by the evaluation process) to have a greater influence on the result of the comparison. The outcome of the integration stage is the degree to which each prototype matches the combined cues. That prototype with the best, relative match is then selected as the representation.

Massaro and Oden (1980) have shown that such a model is capable of accounting for the integration of auditory place and voicing information, whereas Massaro and Cohen (1983) have shown that this model can also account for the integration of place information when the information is derived through a combination of auditory and visual sources, as in the McGurk effect. The results of the current study are compatible with this model in that, like Green and Kuhl (1989), they demonstrate that visual place and auditory voicing information are integrated during perception. However, the results of this study extend this research by indicating that the integration function must produce a mutual, rather than an asymmetric, influence between the visual Place and auditory Voicing dimensions.

One important assumption of Massaro's model is that the auditory and visual cues are processed independently and any observed interactions are the result of the integration process, which maps the different cues onto the underlying dimensions of the prototypes. Although Massaro's own research and that of Roberts and Summerfield (1981) support the view that the auditory and visual information for the Place dimension is initially processed independently, other studies suggest that this may not be the case for other speech dimensions. For example, Breeuwer and Plomp (1986) presented subjects with an auditory signal consisting of just the fundamental frequency of various syllables, and visual information from a talker's face (see also Grant, Ardell, Kuhl, & Sparks, 1985). As indicated by the subjects' errors in Auditory-Only and Visual-Only conditions, neither the auditory or the visual signals alone carried information about manner and voicing of consonants. However, when the auditory and visual signals were presented together, identification of the consonants was accurate, especially for consonants like /b,p,m/, which differ only in voicing and nasality. Bernstein (1989) has pointed out that the evaluation of voicing in this situation depends on a relative comparison of the onset of the voicing (indicated by the presence of fundamental frequency information) and the time of lip opening. Such a comparison seems to run counter to Massaro's assumption of an independent evaluation of the auditory and visual sources with regard to voicing.

An alternative approach, which emphasizes a more interactive view of the processing of the information has been proposed by Eimas et al. (1978, 1981). Eimas et al. carried out an extensive examination of the interferences that occur among various speech dimensions using the speeded classification paradigm. On the basis of their findings, they proposed that (a) different channels of analysis process the speech signal along a limited set of dimensions; (b) the analysis mechanisms subserve both speech and nonspeech processing and are most likely part of a general auditory system; and (c) the mechanisms are arranged hierarchically, and operate in a parallel and interactive manner. Eimas et al. attributed their interference effects to the interactions among the different analysis mechanisms. This approach is similar to other interactive models of speech processing like that proposed by McClelland and Elman (1986).

The results of the current study indicate that the mutual interference that occurs during the classification of place and voicing information cannot be solely the result of spectral interactions in the auditory system. Either there are visual analysis mechanisms that operate in parallel and interact with the auditory mechanisms, or the auditory and visual information for a particular dimension are integrated before phonetic analysis, perhaps by mapping them onto underlying dimensions that are amodal in nature. Although both possibilities are compatible within the framework of the model proposed by Eimas et al., further research is needed to differentiate them. Regardless of the outcome of such research, the results of the present study indicate that place and voicing information are processed as an integral unit, even when the variation along the two dimensions occurs in separate modalities.

References

Bernstein, L. E. (1989). Independent or dependent feature evaluation: A question of stimulus characteristics. *Behavioral and Brain Sciences, 12,* 756–757.

Biederman, I., & Checkosky, S. F. (1970). Processing redundant information. *Journal of Experimental Psychology, 83,* 486–490.

Binnie, C. A., Montgomery, A. A., & Jackson, P. M. (1974). Auditory and visual contributions to the perception of consonants. *Journal of Speech and Hearing Research, 17,* 619–630.

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

Breeuwer, M., & Plomp, R. (1986). Speechreading supplemented with auditorily presented speech parameters. *Journal of the Acoustical Society of America, 79,* 481–499.

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,* 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.

Fitch, H. L., Halwes, T., Erickson, D. M., & Liberman, A. M. (1980). Perceptual equivalence of two acoustic cues for stop-consonant manner. *Perception & Psychophysics, 27,* 343–350.

Garner, W. R. (1979). *The processing of information and structure,* Hillsdale, NJ: Erlbaum.

Gielen, S., Schmidt, R. & Van Den Heuvel, P. (1983). On the nature of intersensory facilitation of reaction time. *Perception & Psychophysics, 34,* 161–168.

Grant, K. W., Ardell, L. H., Kuhl, P. K., & Sparks, D. W. (1985). The contribution of fundamental frequency, amplitude envelope, and voicing duration cues to speechreading in normal-hearing subjects. *Journal of the Acoustical Society of America, 77,* 671–677.

Green, K. P., & Kuhl, P. K. (1989). The role of visual information in the processing of place and manner features in speech perception. *Perception & Psychophysics, 45,* 34–42.

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

Kirk, R. E. (1968). *Experimental design procedures for the behavioral sciences,* Belmont, CA: Wadsworth.

Lockhead, G. R. (1970). Identification and the form of multidimensional discrimination space. *Journal of Experimental Psychology, 85,* 1–10.

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

MacDonald, J., & McGurk, H. (1978). Visual influences on speech perception processes. *Perception & Psychophysics, 24,* 253–257.

Marks, L. W. (1987). On cross-modal similarity: Auditory-visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 384–394.

Massaro, D. W. (1987). *Speech perception by ear and eye.* Hillsdale, NJ: Erlbaum.

Massaro, D. W., & Cohen, M. M. (1983). Evaluation and integration of visual and auditory information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 753–771.

Massaro, D. W., & Oden, G. C. (1980). Evaluation and integration of acoustic features in speech perception. *Journal of the Acoustical Society of America, 67,* 996–1013.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1–86.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264,* 746–748.

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

Miller, J. L. (1981). Phonetic perception: Evidence for context-dependent and context-independent processing. *Journal of the Acoustical Society of America, 69,* 822–831.

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological Review, 85,* 172–191.

O'Leary, A., & Rhodes, G. (1984). Cross-modal effects on visual and auditory object perception. *Perception & Psychophysics, 35,* 565–569.

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin, 92,* 81–110.

Roberts, M., & Summerfield, Q. (1981). Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory. *Perception & Psychophysics, 30,* 309–314.

Shepp, B. E. (1989). On perceiving objects: Holistic versus featural properties. In B. E. Shepp & S. Ballesteros (Eds.), *Object perception,* Hillsdale, NJ: Erlbaum.

Summerfield, Q. (1983). Audio-visual speech perception, lipreading, and artificial stimulation. In M. E. Lutman & M. P. Haggard (Eds.), *Hearing science and hearing disorders*, London, England: Academic Press.

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.

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

Wood, C. C., & Day, R. S. (1975). Failure of selective attention to phonetic segments in consonant–vowels syllables. *Perception & Psychophysics, 17*, 346–350.

Received April 24, 1989
Revision received March 21, 1990
Accepted March 21, 1990 ■