Perception & Psychophysics
1987, 42 (6), 587-593

The perception of speaking rate using visual information from a talker's face

KERRY P. GREEN
Northeastern University, Boston, Massachusetts

It has been well documented that listeners are able to estimate speaking rate when listening to a talker, but almost no work has been done on perception of rate information provided by looking at a talker's face. In the present study, the method of magnitude estimation was used to collect estimates of the rate at which a talker was speaking. The estimates were collected under four experimental conditions: auditory only, visual only, combined auditory-visual, and inverted visual only. The results showed no difference in the slope of the functions relating perceived rate to physical rate for the auditory only, visual only, and combined auditory-visual presentations. There was, however, a significant difference between the normal visual-only and the inverted-visual presentations. These results indicate that there is visual rate information available on a talker's face and, more importantly, suggest that there is a correspondence between the auditory and visual modalities for the perception of speaking rate, but only when the visual information is presented in its normal orientation.

Numerous studies have demonstrated that auditory prosodic information plays an important role in spoken language processing. One prosodic variable that has been shown to be important in phonetic perception is speaking rate. Listeners' perception of speaking rate has been examined in several studies (see Miller, 1981, and Grosjean & Lane, 1981, for reviews), which have shown that listeners can make systematic and orderly judgments of speaking rate on the basis of the auditory signal produced by a talker (Grosjean, 1977; Grosjean & Lane, 1974, 1976; Lane & Grosjean, 1973). More importantly, studies have also shown that listeners take into account a talker's rate of speech during speech perception. For example, a change in the physical rate at which a syllable is articulated can influence a listener's identification of its initial consonant (Miller, Aibel, & Green, 1984; Miller & Liberman, 1979; Summerfield, 1981). Moreover, the rate of a precursor phrase can influence the identification of a test syllable whose own rate is held constant (Miller, Green, & Schermer, 1984; Miller & Grosjean, 1981; Port, 1976; Summerfield, 1981). Thus, these studies clearly demonstrate the importance of auditory rate information, extracted from individual segments or from connected discourse, during language processing.

In a normal face-to-face conversation, a listener has access not only to auditory information, but also to visual information provided by the talker's mouth and face. Several studies have shown that this visual information can play a role in spoken language processing, particularly in aiding perception of speech presented in noise or against background conversation (Dodd, 1977; Erber, 1969, 1975; Sumby & Pollack, 1954; Summerfield, 1979). Studies have also shown that the perception of even well-specified auditory information can be influenced by conflicting visual information (Green & Miller, 1985; MacDonald & McGurk, 1978; Massaro & Cohen, 1983; McGurk & MacDonald, 1976; Summerfield, 1979; Summerfield & McGrath, 1984). However, very little is known about the nature of the visual information and about how it is integrated with the auditory information during spoken language processing. Most studies in this area have focused on visual information that is available for the perception of consonants or vowels (e.g., Berger, 1970; Binnie, Montgomery, & Jackson, 1974; Lowell, 1974; see Summerfield, 1983, for a review). Very few studies have addressed the issue of whether visual prosodic information is also available on a talker's face. Fisher (1969) demonstrated that visual information was available for discriminating sentences that ended with a falling intonation contour from sentences that ended with a rising contour. More recently, Bernstein, Eberhardt, and Demorest (1986) have shown that information about stress, as well as intonation, is available from a talker's face.

There has been only one study in which the perception of visual rate information was examined. Green and Miller (1985) showed that visual information about speaking rate is available at the segmental level. More importantly, their

This research was conducted as part of a doctoral dissertation at Northeastern University by the author and was supported in part by NIH Grant NS 14394 to Joanne L. Miller and by NIH BRSG RR 07143 to Northeastern University. The write-up of the research was done while the author was supported by NIH Grant HD 07239 to the University of Washington. I would like to thank the members of my dissertation committee—Joanne Miller, Harlan Lane, Adam Reeves, and William Warren—for their guidance and encouragement. I would also like to thank François Grosjean, Pat Kuhl, Joanne Miller, and William Warren for their valuable comments on earlier versions of the manuscript, and Ted Kozlowski for his assistance in data collection. Correspondence should be sent to Kerry P. Green at the Child Development and Mental Retardation Center, WJ-10, University of Washington, Seattle, WA 98195.

results also demonstrated that visual rate information is utilized during the perception of phonetic segments. The question of the existence of visual information that specifies global speaking rate across an entire phrase or sentence has remained unanswered. And the answer to this question is important to our understanding of the kind of visual information that may be used during language processing and for determining whether the influence of visual rate information on phonetic perception is similar to that of auditory rate information. The experiments described below addressed this question by assessing whether or not people were sensitive to visual rate information available during connected discourse.

EXPERIMENT 1

The purpose of Experiment 1 was to examine perception of visual speaking rate by extending the work of Grosjean and his colleagues on the perception of auditory speaking rate. Using a magnitude estimation task, Grosjean (1977; Grosjean & Lane, 1974, 1976; Lane & Grosjean, 1973) has shown that listeners' perception of speaking rate grows as a power function of the physical rate with an exponent of approximately 1.88. Furthermore, listeners' judgments of speaking rate are influenced more by articulation rate of the utterance than by pause rate (Grosjean & Lane, 1974, 1976; Grosjean & Lass, 1977; Lane & Grosjean, 1973). Finally, the slope of the function relating perceived rate to physical rate is not affected by the listener's linguistic knowledge of the language (Grosjean, 1977). That is, naive speakers will produce functions relating perceived rate to physical rate with slopes that are similar to those produced by native speakers of a language. This is not to say that the judgments made by the naive listeners are identical to those of the native speakers of the language. As Grosjean (1977) has shown, the actual rate judgments of the naive listeners are significantly higher than the rate judgments of the native speakers, indicating that for a particular speaking rate, a naive listener will judge the talker to be speaking at a faster rate than would a native speaker.

In the current experiment, a magnitude estimation task was used to collect subjects' estimates of perceived speaking rate under three experimental conditions: (1) auditory information only, in which the subjects listened to the speaker talking at various rates of speech; (2) visual information only, in which the subjects watched a videotape (without sound) of the same person talking at various rates; and (3) combined visual and auditory presentation, in which the subjects watched and listened to the videotape of the person talking at various rates. The first question addressed was whether the subjects' estimates would vary with changes in speaking rate, indicating that people were indeed sensitive to the visual rate information. If there is visual information available on the face that can specify speaking rate, then the function relating perceived rate to physical rate will have a slope greater than zero.

A second, and related, question addressed was how subjects' judgments of speaking rate would compare when based solely on visual information and when based solely on auditory information. This question was of interest because several previous studies had suggested that perception of temporal rate was different for the auditory and visual modalities (Gebhard & Mowbray, 1959; Myers, Cotton, & Hilp, 1981; Welch, DuttonHurt, & Warren, 1986). For example, Welch et al. (1986) have shown that subjects' magnitude estimates of temporal rate for a light flashing at a specified frequency are consistently greater than estimates of a tone pulsing at the same frequency.

Method

Subjects. The subjects were 30 students from an introductory psychology course who were given course credit as an incentive to participate in the experiment. All subjects reported normal hearing and normal or corrected-to-normal vision. The subjects were assigned randomly to one of the three experimental conditions: (1) auditory information only (AO); (2) visual information only (VO); and (3) combined auditory and visual information (AV).[1]

Stimuli. The stimuli consisted of five different productions of a portion of the Rainbow Passage (see Grosjean & Lass, 1977), each spoken at a different speaking rate.[2] The stimuli were generated using a magnitude production procedure (Lane, Catania, & Stevens, 1961). A male speaker was asked to read the passage at a normal rate of speech. To this rate, the experimenter assigned a numerical value of 10. The experimenter then presented the speaker with a series of values (2.5, 5, 10, 20, 30). Each value was presented a total of four times in random order, and the speaker was asked to read the passage at a rate that stood in the same proportion to the normal rate as the value stood to 10. The passage was written on a blackboard mounted directly behind a videocamera. The productions were recorded onto videotape using a color camera (Sony, DXC-1640), a microphone (AKG-D200E), and a ¾-in. video cassette recorder (Sony, VO-2611). Each production was examined and five samples that contained no hesitations or mispronunciations were chosen, each representative of one of the five different rates. There were no such distinguishing visual characteristics as sudden movements of the head, uncharacteristic eye blinks, or facial expressions that could distinguish a particular sample. The rates for these five samples, from slowest to fastest, were 172.2, 228.6, 304.2, 355.2, and 412.8 syllables per minute (spm), respectively. The ratio of the slowest rate to the fastest rate was 2.4:1, which was identical to that used by Grosjean and Lass (1977) for the same passage. Thus, the AO condition replicated the Grosjean and Lass study and allowed us to make a direct comparison of our results with those of previous studies.

Procedure. Each of the five different samples was presented four times, for a total of 20 trials. These 20 trials were randomly intermixed and copied onto a test videotape. The stimuli were presented in a quiet, dimly illuminated room using a 12-in. color monitor (Sony, CVM-1250) and a ¾-in. video cassette deck (Sony, VO-2611). During the AO condition, the audio signal was presented over the monitor's built-in loudspeaker with the video signal disconnected. During the VO condition, the video was turned on and the sound on the monitor was turned off. During the AV condition, both audio and video were presented over the video monitor. In addition, for the VO and AV conditions, a piece of black tape was placed on the monitor so that it covered the talker's eyes. This was done to prevent the subjects from picking up any extraneous cues from the talker's eye movements that may have occurred during the reading of the passage.

Each subject was seated at a table about 4 ft in front of the monitor, and instructed to watch, listen, or watch and listen, depending

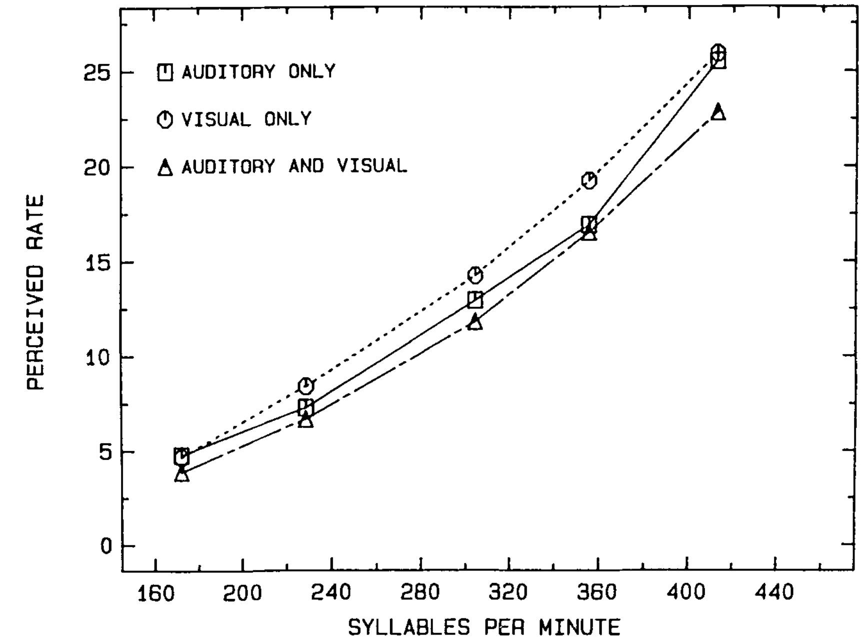

Figure 1. Mean rate estimates as a function of the physical rate of the utterance for the auditory-only (AO), visual-only (VO), and auditory-visual (AV) conditions (Experiment 1).

on the experimental condition. Each subject was allowed to read a written transcript of the test passage before the start of the experiment.

The method of magnitude estimation was used to collect the subjects' judgments of speaking rate. The experimenter played two trials of the talker speaking at his normal rate of speech, in the format appropriate for the experimental condition, and assigned them a numerical value of 10. This rate (304.2 spm), called the standard, was also one of the five rates used in the experiment. Next, the test videotape, consisting of the 20 trials, was presented and the subjects were asked to assign, on each trial, a number that corresponded to the speaker's apparent rate as a proportion of the standard rate of 10. For example, if the subject thought the speaker was talking twice as fast as the standard, he or she was instructed to assign the number 20, half as fast, 5, and so on. The subjects responded by writing their estimates on a response sheet.

Results

Figure 1 displays perceived rate as a function of physical rate for each condition. Each data point represents the mean of 40 responses (4 trials × 10 subjects). The means and standard deviations for the three experimental conditions are presented in Table 1. Linear regression was used to fit a straight line to the log of each subject's individual data as a function of the log of the physical rate. The correlation between the *X* and *Y* values of each function provides a measure of how well each subject's regression line fits the data. The average correlations across the 10 subjects were .99 for each of the three conditions (with ranges of .98 to .99, .97 to .99, and .98 to .99 for the AO, VO, and AV conditions, respectively). Thus, in each condition, the subjects gave responses that were well fit by straight lines on a log-log scale.[3]

The slope or exponent of each regression line provides a measure of how each subject's perceived rate changed as a function of a change in the physical rate. The mean exponents for the AO, VO, and AV conditions were 1.89, 1.94, and 2.02, respectively. The first thing to note is that

Table 1
Mean Estimates and Standard Deviations for Five Speaking Rates in Four Experimental Conditions Across Experiments 1 and 2

	Speaking Rate (Syllables per Minute)											
	177.2		228.6		304.2		355.2		412.8		Exponent	
Condition	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
					Experiment 1							
AO	4.8	2.1	7.3	2.3	12.9	2.5	16.9	3.5	25.5	6.2	1.94	.6
VO	4.7	1.7	8.4	1.6	14.2	2.6	19.2	3.2	25.9	5.5	1.97	.4
AV	3.9	1.0	6.7	1.8	11.8	1.3	16.5	3.2	22.8	4.7	2.02	.4
					Experiment 2							
IV	5.2	1.0	10.0	2.4	13.7	2.9	17.0	2.1	22.2	2.8	1.61	.2

the exponent of 1.89 for the AO condition is practically identical to the exponent of 1.88 found by Grosjean and Lass (1977) for the same passage. Thus, the AO condition successfully replicated the Grosjean and Lass study, and we can be confident that our subjects in the AO condition were judging the speaking rate of the talker much as the subjects in previous studies had. The second thing to note is that the exponent of 1.94 for the VO condition is substantially greater than zero. This result indicates that there is visual information for speaking rate available on the face. Finally, the exponents for the three conditions are all very similar. These exponents were analyzed using a one-way analysis of variance. No significant differences among the mean exponents of the three conditions were obtained [$F(2,27) = .12$, $p = .89$].

The mean estimates were examined to determine if there were any consistent differences among the three experimental conditions. Recall, for example, that Welch et al. (1986) found that the estimates of subjects judging temporal rate were significantly higher under a VO condition than under an AO condition. The mean estimates for the three experimental conditions were analyzed using a two-way analysis of variance with rate and experimental condition as the main effects. As expected, there was a significant rate effect [$F(4,108) = 259.3$, $p < .0001$], indicating that as the physical rate increased, so did the subjects' perceived rate estimates. Interestingly, there was no significant effect for experimental condition [$F(2,27) = 2.04$, $p > .10$], indicating that there was no difference in the mean estimates among the three experimental conditions. Finally, there was no significant condition × rate interaction [$F(8,108) = .58$, $p > .10$], indicating that the pattern of increased perceived rate with increased physical rate was consistent across all three conditions. Apparently, then, unlike the perception of temporal rate, the perception of speaking rate does not produce significantly higher estimates in the VO condition than in the AO (or AV) condition.

Although there were no differences in the actual mean estimates, it remained a possibility that there might be consistent differences in the variability of the subjects' responses across the three experimental conditions. For example, although the estimates given by subjects in the VO and AO conditions were similar, it seemed possible that the subjects in the VO condition were more variable in their responding. We examined this possibility in two ways. First, we analyzed the standard deviations of each subject's responses to a particular rate, using an analysis of variance with rate and experimental condition again as the main effects. As expected, there was a significant rate effect [$F(4,108) = 8.63$, $p < .001$], because the subjects were using larger numbers for the faster rates, thus producing greater amounts of variability. There was no significant effect for experimental condition [$F(2,27) = 2.37$, $p > .10$], indicating that there was no difference in the variability of the mean estimates among the three experimental conditions, and there was no significant condition × rate interaction [$F(8,108) = 1.21$, $p > .1$].

Second, an F test for homogeneity of variance was performed on the 15 possible pair combinations for the 5 speaking rates × 3 experimental conditions (see Table 1). Of the 15 combinations, only 3 showed significant differences, and the pattern was not consistent. (For the rate of 177.2 spm, the AV condition was significantly lower than the AO condition [$p < .05$], and for the rate of 304.2 spm, the AV condition was significantly lower than either the AO condition or the VO condition [$p < .05$].) Thus, it appears that there were no consistent differences in the variability of the subjects' responses.

It was tempting to conclude from these results that there was no difference in the perception of speaking rate for the visual and auditory modalities. However, before coming to such a conclusion, an alternative interpretation had to be ruled out. It seemed possible that the subjects were not actually judging speaking rate in the VO condition. Rather, they may have been estimating some such other quality as overall duration, which was correlated with speaking rate and could therefore have resulted in estimates that were similar to those in the AO condition.[4]

One way to address this issue would be to present the visual stimuli in a way that would maintain such global physical properties as overall duration but disrupt the subjects' processing of the speaking rate information. The approach taken in the next experiment was to present the visual stimuli in an inverted orientation, that is, upside down.

EXPERIMENT 2

There is considerable evidence that inverted objects, in particular inverted faces, are more difficult to recognize and remember than their upright counterparts (Carey & Diamond, 1977; Rock, 1974; Yin, 1969). Furthermore, young infants (4 to 7 months of age) have a difficult time in discriminating expressions on inverted faces, although they show no difficulty with normally oriented faces (Oster, 1980; Walker, 1982). These findings have led some researchers (e.g., Hochberg, 1968; Walker, 1982) to consider the perception of facial expression as using dynamic configurational information that is available only in a normal orientation.

Many of the features used in the perception of facial expressions (e.g., size and direction of mouth opening, orientation of the lips) are also involved in lipreading and, presumably, the visual perception of speaking rate. Therefore, if, when estimating speaking rate, the subjects were perceiving the visual information as a speech event, their estimates would be affected by an inversion of the visual information. Thus, in a fourth experimental condition, we replicated our VO condition with one important difference: the video monitor, turned upside down, displayed the talker's face in an inverted orientation.

Method

Subjects. The subjects were 10 new students from an introductory psychology course who were given course credit as an incen-

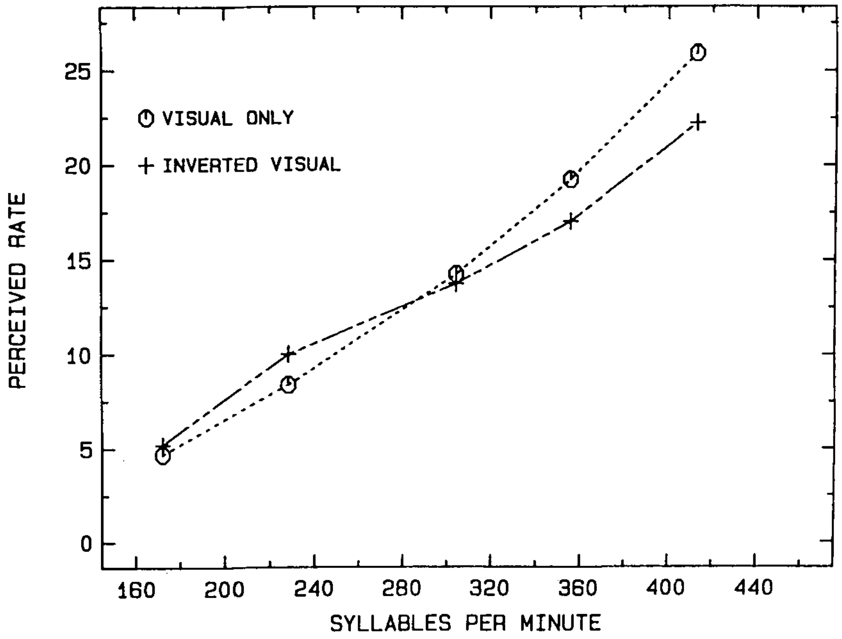

Figure 2. Mean rate estimates as a function of the physical rate of the utterance for the inverted-visual (IV) condition (Experiment 2) and visual-only (VO) condition (Experiment 1).

tive to participate in the experiment. All subjects reported normal hearing and normal or corrected-to-normal vision.

Stimuli and Procedure. The stimuli consisted of the same videotape used in the first experiment. The procedure for the inverted visual (IV) condition was the same as that for the VO condition in the first experiment, except that the monitor was turned upside down.

Results

Figure 2 displays the results of the IV condition, along with the results from the original VO condition. The means and standard deviations for the IV condition are given in Table 1. As in Experiment 1, linear regression was used to fit straight lines to each subject's individual data. The average correlation for the IV condition was .96 (with a range of .90 to .99), again indicating that the subjects gave responses that were well fit by straight lines on a log-log scale.[5]

The exponents for each subject's individual data were calculated and compared with the exponents obtained from the VO condition in Experiment 1. A t test indicated that the mean exponent for the inverted visual condition (1.58) was significantly less than the mean exponent for the VO condition (1.94) [$t(18) = 2.68, p < .02$]. The mean estimates for the IV condition were compared with the mean estimates from the VO condition in Experiment 1, again using a two-way analysis of variance. As expected, there was a significant rate effect [$F(4,72) = 170.0$, $p < .0001$]. Although there was no significant condition effect [$F(1,18) = 1.01, p > .1$], there was a significant condition × rate interaction [$F(4,72) = 3.32, p < .02$]. Further post hoc analyses indicated that the mean estimates for the two fastest rates were significantly lower in the IV condition ($p < .05$). Finally, the variability of the subjects' responses in the IV and VO conditions was compared, using the same procedure as in the first experiment. An analysis of variance on the subjects' standard deviations revealed a significant rate effect [$F(4,72) = 16.2, p < .001$] but again no significant experimental condition effect [$F(1,18) = .93, p > .1$] or condition × rate interaction [$F(4,72) = .41, p > .1$]. Furthermore, an F test for homogeneity of variance indicated that the variances were reliably different for only one rate condition (412.8 spm). Thus, these results indicate that there are reliable differences between the estimates of speaking rate that subjects give in the IV condition and those they give in the VO condition, although there are no consistent differences in the variability of the subjects' responses.

These results demonstrate that the subjects in the normal VO condition were not basing their estimates on a simple physical metric such as the overall duration of the passages. If they had, then there should have been no difference in either the exponents or the mean estimates of the IV and VO conditions, since, except for orientation, the physical characteristics of the stimuli were identical; the only difference between the two was the orientation of the visual image.

GENERAL DISCUSSION

The present experiments were intended to determine whether information on speaking rate could be obtained from the speaker's face—that is, whether a speaker's face could impart such information—and, if so, whether there was a difference in subjects' judgments of speaking rate when the information was presented in different modalities. With regard to the first issue, the results from Experiment 1 demonstrate that subjects can make systematic judgments of speaking rate from information presented in the visual modality. The exponents relating perceived rate to physical rate for the VO condition were substantially greater than zero. Thus, it appears that there is information available from the face that can indicate the talker's rate of speech.

As for the second issue, the results from Experiments 1 and 2, taken together, suggest that there are no inherent differences in the perception of speaking rate for the auditory and visual modalities. These results are of interest in light of the recent work by Welch et al. (1986) on the perception of temporal rate. As mentioned earlier, Welch et al. found that subjects gave reliably higher estimates to a repeating visual stimulus than they did to an auditory stimulus that repeated at the same rate. Furthermore, when the auditory and visual stimuli were presented simultaneously, the auditory stimulus biased the subjects' estimates more than did the visual stimulus. Welch et al. concluded that audition dominates peoples' perception of temporal rate. The fact that no reliable differences were obtained among the mean estimates of the three normal experimental conditions in the present study suggests that such a conclusion may be premature. One possible explanation for the difference between the results reported here and Welch et al.'s is that audition dominates the perception of repetition rate but does not dominate the perception of speaking rate.

An alternative explanation lies in the nature of the stimuli used in this experiment. Here, auditory and visual stimuli were the product of the same event, a talker reading aloud a passage at different rates of speech, whereas in Welch et al.'s study, the stimuli consisted of an artificial pairing of a flashing light and a repeating tone. It remains a possibility that the discrepancy they found in the rate estimates obtained in the auditory and visual modalities was due to the fact that the auditory and visual information actually specified two very different events to the subjects. The reason for the emphasis on events is the assumption that organisms evolved to perceive environmental events, and furthermore, that events are perceived amodally without reference to the sensory modality in which the information is available (E. J. Gibson & Spelke, 1983; J. J. Gibson, 1966). Following this line of reasoning, if auditory and visual stimuli do specify different events to the observer, it is possible that the different events are perceived as having different rates. Further studies will be necessary to help differentiate between these two explanations.

REFERENCES

Berger, K. W. (1970). Vowel confusions in speech reading. *Ohio Journal of Speech & Hearing*, **5**, 123-128.

Bernstein, L. E., Eberhardt, S. P., & Demorest, M. E. (1986). Judgments of intonation and contrastive stress during lipreading. *Journal of the Acoustical Society of America*, Suppl. 1, **80**, S78.

Binnie, C. A., Montgomery, A. A., & Jackson, P. L. (1974). Auditory and visual contributions to the perception of selected English consonants for normally hearing and hearing-impaired listeners. In H. Birk Nielsen and E. Kampp (Eds.), *Visual and audio-visual perception of speech* (Scandinavian Audiology Suppl. 4, pp. 182-209). Stockholm: Almquist & Wiksell Periodical Co.

Carey, S., & Diamond, R. (1977). From piecemeal to configurational representation of faces. *Science*, **195**, 312-314.

Dodd, B. (1977). The role of vision in the perception of speech. *Perception*, **6**, 31-40.

Erber, N. P. (1969). Interaction of audition and vision in the recognition of oral speech stimuli. *Journal of Speech & Hearing Research*, **12**, 423-425.

Erber, N. P. (1975). Auditory-visual perception of speech. *Journal of Speech & Hearing Disorders*, **40**, 481-492.

Fisher, C. G. (1969). The visibility of terminal pitch contour. *Journal of Speech & Hearing Research*, **12**, 379-382.

Gebhard, J., & Mowbray, G. (1959). On discriminating the rate of visual flicker and auditory flutter. *American Journal of Psychology*, **72**, 521-529.

Gibson, E. J., & Spelke, E. (1983). The development of perception. In P. Mussen (Ed.), *Carmichael's manual of child psychology* (pp. 1-76). New York: Wiley.

Gibson, J. J. (1966). *The senses considered as perceptual systems.* Boston: Houghton Mifflin.

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

Grosjean, F. (1977). The perception of rate in spoken and signed languages. *Perception & Psychophysics*, **22**, 408-413.

Grosjean, F., & Lane, H. (1974). Effects of two temporal variables on the listener's perception of reading rate. *Journal of Experimental Psychology*, **102**, 893-896.

Grosjean, F., & Lane, H. (1976). How the listener integrates the components of speaking rate. *Journal of Experimental Psychology: Human Perception & Performance*, **2**, 538-543.

Grosjean, F., & Lane, H. (1981). Temporal variables in the perception and production of spoken and sign languages. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech* (pp. 207-238). Hillsdale, NJ: Erlbaum.

Grosjean, F., & Lass, N. (1977). Some factors affecting the listener's perception of reading rate in English and French. *Language & Speech*, **20**, 198-208.

Hochberg, J. (1968). The mind's eye. In R. N. Haber (Ed.), *Contemporary theory and research in visual perception*. New York: Holt, Rinehart, & Winston.

Lane, H., Catania, A., & Stevens, S. S. (1961). Voice level: Autophonic scale, perceived loudness and effects of sidetone. *Journal of Acoustical Society of America*, **33**, 160-167.

Lane, H., & Grosjean, F. (1973). Perception of reading rate by speakers and listeners. *Journal of Experimental Psychology*, **97**, 141-147.

Lowell, E. L. (1974). Auditory and visual perception of different units of speech. In H. Birk Nielsen & E. Kampp (Eds.), *Visual and audio-visual perception of speech* (Scandinavian Audiology Suppl. 4, pp. 31-37). Stockholm: Almquist & Wiksell Periodical Co.

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

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

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

Miller, J. L. (1981). 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., Aibel, I., & Green, K. (1984). On the nature of rate-dependent processing during phonetic perception. *Perception & Psychophysics*, **35**, 5-15.

Miller, J. L., Green, K., & Schermer, T. M. (1984). A distinction between the effects of sentential speaking rate and semantic congruity on word identification. *Perception & Psychophysics*, **36**, 329-337.

Miller, J. L., & Grosjean, F. (1981). How the components of speaking rate influence the perception of phonetic segments. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 208-215.

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

Myers, A. K., Cotton, B., & Hilp, H. A. (1981). Matching the rate of concurrent tone bursts and light flashes as a function of flash surround luminance. *Perception & Psychophysics*, **30**, 33-38.

Oster, H. (1980). Infants' responses to emotional expressions. In M. E. Lamb & L. R. Sherrod (Eds.), *Infant social cognition*. Hillsdale, NJ: Erlbaum.

Port, R. F. (1976). *The influence of speaking tempo on the duration of stressed vowel and medial stop in English trochee words*. Unpublished doctoral dissertation, University of Connecticut.

Rock, I. (1974). The perception of disoriented figures. *Scientific American*, **230**, 78-85.

Sumby, W. G., & Pollack, I. (1954). Visual contributions to speech intelligibility in noise. *Journal of the Acoustical Society of America*, **26**, 212-215.

Summerfield, Q. (1979). Use of visual information for phonetic perception. *Phonetica*, **36**, 314-331.

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

Summerfield, Q. (1983). Audio-visual speech perception, lipreading and artificial stimulation. In M. E. Lutman & M. P. Haggard (Eds.), *Hearing science and hearing disorders* (pp. 132-182). London: Academic Press.

Summerfield, Q., & McGrath, M. (1984). Detection and resolution of audio-visual incompatibility in the perception of vowels. *Quarterly Journal of Experimental Psychology*, **36A**, 51-74.

Walker, A. S. (1982). Intermodal perception of expressive behaviors by human infants. *Journal of Experimental Child Psychology*, **33**, 514-535.

Welch, R., DuttonHurt, L., & Warren, D. (1986). Contributions of vision and audition to temporal rate perception. *Perception & Psychophysics*, **39**, 294-300.

Yin, R. (1969). Looking at upside down faces. *Journal of Experimental Psychology*, **81**, 141-145.

NOTES

[1] A between-subjects design was used in order to obviate the possibility that any knowledge from previous experimental conditions would influence the subjects' responses in a particular condition. For example, if subjects were given the AO (or AV) condition before the VO condition, they would have had enough exposure to the test passage to enable them to memorize it. Thus, during the VO condition, they might have tried to lip-read the passage from memory, thus influencing their estimates. Using a between-subjects design ensured that the judgments made in the VO or AV condition reflected the subjects' actual estimates of speaking rate.

[2] The text of the passage used in this experiment is as follows:

> The rainbow is a division of white light into many beautiful colors. These take the shape of a long round arch, with its path high above, and its two ends apparently beyond the horizon. There is, according to legend, a boiling pot of gold at one end. People look, but no one ever finds it.

[3] Correlations on the raw judgments were also calculated. The average correlations were again high in all three conditions (means of .98, .98, and .97 for the AO, VO, and AV conditions, respectively).

[4] It is unlikely that the subjects in the AO and AV conditions were also basing their estimates of speaking rate on overall duration. The results of Grosjean and Lane (1976) are relevant to this issue. Grosjean and Lane collected listeners' estimates of speaking rate for a passage in which the articulation time, number of pauses, and duration of pauses were all independently manipulated. Varying the number and duration of pauses had a much greater effect on the average speaking rate (and therefore the passage duration) than did articulation time. If one assumes that listeners were simply using the duration of the passage to make their estimates, then one would expect the number and duration of pauses to have the greatest influence on listeners' estimates of speaking rate. Grosjean and Lane found, however, that articulation time had the greatest influence. Their results demonstrated that listeners use a complex combination of articulation rate and pause rate information when estimating overall speaking rate.

[5] Correlations on the raw judgments were also calculated for the IV condition and again a similar result (mean of .95) was obtained.

(Manuscript received February 17, 1987;
revision accepted for publication May 19, 1987.)