Perceptual dominance during lipreading

RANDOLPH D. EASTON and MARYLU BASALA
Boston College, Chestnut Hill, Massachusetts

Two experiments were performed under visual-only and visual-auditory discrepancy conditions (dubs) to assess observers' abilities to read speech information on a face. In the first experiment, identification and multiple choice testing were used. In addition, the relation between visual and auditory phonetic information was manipulated and related to perceptual bias. In the second experiment, the "compellingness" of the visual-auditory discrepancy as a single speech event was manipulated. Subjects also rated the confidence they had that their perception of the lipped word was accurate. Results indicated that competing visual information exerted little effect on auditory speech recognition, but visual speech recognition was substantially interfered with when discrepant auditory information was present. The extent of auditory bias was found to be related to the abilities of observers to read speech under nondiscrepancy conditions, the magnitude of the visual-auditory discrepancy, and the compellingness of the visual-auditory discrepancy as a single event. Auditory bias during speech was found to be a moderately compelling conscious experience, and not simply a case of confused responding or guessing. Results were discussed in terms of current models of perceptual dominance and related to results from modality discordance during space perception.

When considering the perceptual accomplishments of a person moving and orienting in space, it seems clear that vision is the dominant perceptual system, both in terms of the pickup of visual information for its own sake and in terms of its apparent "tuning" of the other perceptual systems (Lee, 1978; Turvey, 1977). Laboratory research over the past 2 decades has demonstrated that when visual information and nonvisual information regarding the layout of space are artificially made to conflict, vision dominates the perceptual experience and behavior (e.g., Lee & Lishman, 1975; Pick, Warren, & Hay, 1969). More recent investigations have also revealed, however, that vision does not completely or inevitably dominate the processing of nonvisual information. If one instructs or permits subjects to attend to nonvisual information, visual dominance can be reduced, or even eliminated (Easton, in press; Warren & Schmitt, 1978). Also, if the precision of perceptual judgments of nonvisual information is sufficiently enhanced, visual dominance is found to decrease (Easton, in press; Welch, Widawski, Harrington, & Warren, 1979). Finally, if nonvisual information is appropriate or ecologically valid for the task at hand, visual bias can be lessened (Lederman, 1979). The approaches and hypotheses of these studies are not mutually exclusive: the consensus that has emerged based on the empirical findings

is that the relations among the perceptual systems are not fixed, but can shift, depending upon stimulus and receptor variables or the cognitive control at the disposal of an observer. It is also important to note that modality bias appears to be a genuine perceptual effect rather than being attributable to response effects or postperceptual decision processes (Bertelson & Radeau, 1976, 1981; Welch & Warren, 1980).

Most of the research on perceptual dominance has dealt with spatial localization tasks. All three hypotheses regarding modality dominance outlined above would predict that if a task that did not favor vision for its successful completion were chosen or designed, nonvisual dominance effects should occur. Bimodal speech perception in the presence of discrepant auditory dubs suggests itself as an appropriate task in this regard. It has been argued by some investigators that auditory-verbal processing of information is of an entirely different order from that of visual processing, since it is extended in time, whereas vision is primarily a spatial system (O'Connor & Hermeline, 1978). While this distinction may be questionable due to the critical role played by change or movement in visual perception (Gibson, 1979), it is true that the auditory system *only* processes information specifying change (Neisser, 1976). Since speech information consists of complex visual and auditory changes of stimulation, it is possible that auditory dominance effects could be demonstrated if a perceptual discrepancy that eliminated the ordinary redundancy between visual and auditory information during speech were created.

0031-5117/83/120562-09$01.15/0

It is also possible, of course, that auditory dominance would occur during perception of dubbed speech for reasons other than temporal change. There appears to be far less information created visually during speech than is created auditorily. It has been argued that there are only four sets of consonants that are visually contrastable (Woodward & Barber, 1960). Erber (1974) estimates, further, that the forty English consonant and vowel phonemes produce sixteen visually contrastable articulations on the face of a speaker. Put another way, 60% of speech sounds are estimated to be obscure or invisible on the face. Workers who teach lipreading also recognize what are referred to as homophenous sounds, sounds which do not sound alike but do look alike: it is argued that there is not a single consonant that has a characteristic lip or jaw movement (Jeffers & Barley, 1974). The simple point is that auditory information could dominate visual information during speech due to the fact that more auditory information is present.

Finally, another reason to expect an advantage of auditory over visual speech information is that speech may be more naturally auditory than visual. Speech may have developed without dependence on visual input, and today people apparently listen to speech much more than they watch it. The upshot is that auditory information logically could dominate during visual-auditory speech discrepancy for any or all of the reasons outlined above.

In one of the very few studies reported using discrepant visual-auditory information during speech perception, the argument has been advanced that the primary usefulness of visual speech information is that it is redundant with auditory speech information and thus useful in "noisy" environments (Dodd, 1977). In fact, Dodd has demonstrated visual bias effects when observers are required to perceive CVC morphemes spoken by the lips in the presence of a discrepant auditory dub (i.e., the visual information dominates the observer's auditory experience). But this perceptual task was performed in the presence of substantial white noise (sufficient to reduce auditory recognition to 50%-60%), and thus may have shifted the precision of judgment, ecological validity, or directed attention factors in the direction of vision. Under more natural signal-to-noise ratio conditions, audition may well be found to dominate vision during speech perception.

In the only other reported studies of visual-auditory speech discrepancy, the influence of vision during speech has also been emphasized (MacDonald & McGurk, 1978; McGurk & MacDonald, 1976). McGurk and MacDonald have demonstrated that visual information for speakers' lip movement profoundly modifies auditory perception of speech, an apparent visual biasing effect. As in Dodd's study, several conditions of these experiments require that the interpretation of the effects be considered carefully within the context of perceptual dominance. MacDonald and McGurk's (1978) subjects were required to perceive CV morphemes rather than complete words. Even more importantly, their subjects were required to report what they *heard* during bimodal speech perception. Logically, no effects other than *visual* bias could be demonstrated under these circumstances. In the studies to be reported in the present paper, the subjects were presented monosyllabic and compound words under visual-auditory discrepancy conditions and, depending upon their group, were asked to report what they *heard* or *saw*. Only the latter condition allows for a demonstration of auditory bias during bimodal speech perception.

In summary, while it has been demonstrated that the visual location of speaking lips will dominate the heard location of the spoken words (e.g., Aronson, & Rosenbloom, 1971; Thurlow & Jack, 1973), perceptual dominance during speech recognition in the presence of discrepant auditory-visual information is less well investigated and understood. In particular, the biasing effects of discrepant auditory dubs on visual speech perception, under conditions in which complete words are spoken without substantial auditory background noise, have not been studied. All three hypotheses currently discussed to account for dominance relations among the modalities (attention, precision, appropriateness) would lead logically to predictions of auditory bias.

In order to assess speech reading under modality discrepancy conditions, we devised a standard lipreading (SLR) and a dubbed lipreading (DLR) test. In Experiment 1, monosyllabic and spondaic (compound) words were used, and the relationship between visual and auditory information was systematically manipulated in order to vary the degree of discordance at the initial and final phonetic positions of words. Both identification and multiple choice responding were assessed. In Experiment 2, monosyllabic words were used and the compellingness of the visual-auditory discordance as a single speech event was manipulated in an attempt to assess the effect of observers' assumptions and beliefs on perceptual dominance during speech.

EXPERIMENT 1

Method

Procedure

Four different tests were devised: (1) standard lipreading (SLR) identification test, (2) standard lipreading (SLR) multiple choice test, (3) dubbed lipreading (DLR) identification test, and (4) dubbed lipreading (DLR) multiple choice test. In the identification tests, the subjects were instructed to identify the lipped

word and to write it down in the space provided on an answer sheet. In the multiple choice tests, the subjects identified the lipped word from a list of alternatives. The SLR consisted of only visual lip information, whereas the DLR included both visual lip information and an auditory dub.

The subjects were divided into two groups. Group 1 received the SLR identification and multiple choice tests as repeated measures, and Group 2 received the DLR identification and multiple choice tests. The identification and multiple choice tests were constructed with different lists of words, and the order of testing within each group was counterbalanced. This design allowed us to compare, across groups, the subjects' ability to read lips under standard versus dubbed conditions, and also to determine, within groups, the sensitivity of the identification and multiple choice methods of evaluating lipreading.

Stimulus items. The SLR and DLR tests consisted of 30 words each, 15 monosyllabic and 15 spondaic. Each subject under SLR or DLR conditions received two different tests (identification and multiple choice) and was thus tested on 60 different words. The same words were used for the SLR and the DLR tests, thus allowing us to compare the results from the two groups of subjects.

The dubs for the DLR varied systematically with the visual information on the speaker's lips and were chosen in accordance with the following five criteria: (1) SAME—dubbed word same as lipped word, (2) INITIAL—dub having same initial visual information as that of lipped word, (3) FINAL—dub having same final visual information as that of lipped word, (4) BOTH—dub having same initial and final visual information as lipped word, and (5) NEITHER—dub in which all visual information differed from that of the lipped word. Examples of these dubbing categories are presented in the upper portion of Table 1. It should be noted that when it is claimed that visual and auditory information correspond visually, it is not meant in the strictest sense. Rather, correspondence was based on visual speech confusability data collected by Jeffers and Barley (1974). Thus, for example, the word pair lord(visual)-swirl(auditory) possesses a final phoneme correspondence in terms of the visual confusability between /d/ and /l/. Dubbing in this manner allowed us to determine at which sequential position in a word a visual-auditory discrepancy would disrupt most strongly observers' abilities to attend to visual speech information. The SAME category was obviously not a discrepancy condition but was included to facilitate the possibility that observers would experience discrepant information on other trials as specifying a single speech event.

Within the categories of INITIAL and FINAL, a further distinction was made in order to manipulate the degree of perceptual discrepancy. By definition, when the INITIAL or FINAL positions of the lipped and dubbed words involved the same visual information, the final or initial phonemes, respectively, were discrepant. The degree of this discrepancy was manipulated by using lipped and dubbed words that possessed *visible* versus *obscure* consonant phonemes (Jeffers & Barley, 1974) at the initial or final positions of words when a discrepancy was present at that position. Visible consonant phonemes would represent a relatively large speech discrepancy, whereas obscure consonant phonemes would represent a relatively small discrepancy. Examples of these dubbing categories are presented in the lower portion of Table 1.

The auditory dubs were synchronized with the visual information by dubbing auditory information onto a prerecorded videotape of a speaking face. No specific dubbing apparatus (e.g., MacDonald, Dwyer, Ferris, & McGurk, 1978) other than the "dub-over" control of the video deck (Sony AV-3650) was used. Our goal was to create synchrony between the lips and dubs in the perceptual experience of observers. A group of judges screened the tapes, and dubs that were deemed too asynchronous were redone until satisfactorily "in synch."

Response alternatives. The response alternatives for the multiple choice tests were also related systematically to the visual lip information. In the SLR, the following five categories were used: (1) CORRECT—the word that was lipped, (2) INITIAL, (3) FINAL, (4) BOTH, and (5) NEITHER. (Categories 2-5 follow the same patterns as outlined above for auditory dubs.) The response list for DLR was extended to include the auditory dub as a sixth alternative and was referred to as the AUDITORY category.

Presentation of stimulus items. The lipping of the word lists was performed by a female college student with no previous lipreading or articulatory experience. Her voice was also used for dubbing in the DLR.

The subjects were administered two tests in a single session. In all tests, the words were presented with a 10-sec interval between trials. Another female speaker was used to count off the number of a given trial 1 sec prior to the initiation of word presentation. Thus, the subjects received warning signals so that their full attention would be directed toward the videotape screen at the time of word presentation. The words were articulated clearly at a normal pace.

Control and experimental groups. Six subjects were selected to form a control group. These subjects received only the DLR. They were instructed to watch the videotape screen while paying attention to what they were hearing (the experimenter insured that the subjects kept their eyes open and on the TV screen during each trial). Their task was to identify the auditory dub in both recall and recognition contexts. The control group in this experiment served to evaluate whether conflicting visual information would influence the subjects' auditory identifications of the presented words.

Forty subjects formed the experimental group, 20 taking the DLR and 20 the SLR. They were told that the experimenters were interested in lipreading skill: their task was to watch the video screen and identify what the lips were saying. DLR subjects were instructed further that they were to pay attention to the visual lips and to attempt to ignore simultaneously present auditory information, which might not always be the same as the visual information.

Apparatus

Testing was administered through the use of Sony video equipment. The subjects watched the articulation of the word lists on a 21-in. black-and-white TV screen, which was filled by the speaker's entire head and face. The subjects were seated 8 ft from the screen and were tested under normal lighting and sound conditions.

Table 1

	Visual (Lip)	Auditory (Dub)
A		
INITIAL position correspondence	Face	Fame
FINAL position correspondence	Teeth	Mouth
BOTH position correspondence	Word	Whirl
NEITHER position correspondence	Feel	Roam
SAME	Wild	Wild
B		
INITIAL visual discrepancy	Rough	Rub
INITIAL obscure discrepancy	Buzz	Bunch
FINAL visible discrepancy	Buff	Rough
FINAL obscure discrepancy	Chime	Time

Note—Top panel: Relation in terms of visual correspondence of dubbed (auditory) to lipped (visual) information at different word positions. Bottom panel: Magnitude of auditory-visual discrepancy was created in terms of visible vs. obscure consonants when the visual and auditory information did not correspond.

Subjects
The subjects consisted of a group of 46 randomly selected students who were fulfilling course requirements at Boston College. All were reportedly free of any visual or hearing disorders.

Results

Both the identification and multiple choice data were analyzed separately. Generally, the multiple choice data yielded patterns very similar to those obtained with the identification data, although, as expected, multiple choice responding resulted in much higher overall lipreading accuracy. For present purposes, we will present only the general finding for multiple choice testing, and present the identification data through more detailed analysis.

Control Group

The subjects who had been asked to pay attention to auditory information while keeping their eyes on the speakers lips identified the correct auditory word in the DLR with .99 accuracy. Observation by the experimenter verified that the subjects kept their eyes on the TV monitor and the speaking lips on each trial. Thus, discrepant visual speech information specifying complete words appears to exert little or no effect on auditory speech identification.

Percent correct. Table 2 presents means and standard deviations for accuracy across the major experimental conditions. Analysis of variance yielded significant effects for all three variables: SLR was superior to DLR [F(1,38) = 122.4, p < .001]; multiple choice performance was superior to identification [F(1,38) = 601.7, p < .001]; spondaic words resulted in greater accuracy of lipreading than monosyllabic words [F(1,38) = 7.7, p < .01]. There was also a significant interaction between type of lipreading test (SLR vs. DLR) and whether identification or multiple choice was required [F(1,38) = 58.1, p < .001]. The interaction was attributable to a smaller multiple choice/identification difference for the DLR, and a smaller decremental effect of dubbing under identification conditions. Nevertheless, simple-effect tests revealed significant performance decrements to be associated with dubbing and identification conditions in all cases (ps < .001).

Simply put, these data indicate that observers can read single words on the lips with about 20% accuracy, but if they are allowed to choose from among a set of alternatives, accuracy increases to about 90%. When discrepant auditory information is simultaneously present, however, lipreading accuracy drops appreciably, to about 5%. If a multiple choice procedure is used under visual-auditory discrepancy conditions, performance is improved to about 42%. In all cases, spondaic words are lipread more accurately than monosyllabic words.

Error analysis. As was seen in Table 2, total error for DLR was about 94% for identification testing. An initial analysis of these errors indicated that they were distributed equally across the four different dubbing categories. This pattern proved true for both monosyllabic and spondaic words. We then proceeded to assess auditory errors, or those occasions when the auditory dub was given as the response. Auditory error comprised 32% of the total error. The distribution of these errors across major experimental conditions is provided in Table 3. Analysis of variance indicated that monosyllabic words resulted in more auditory error than spondaics [F(1,19) = 16.82, p < .001]. Furthermore, auditory error was smallest when the initial and final visual and auditory phonemes did not correspond (NEITHER category) [F(3,57) = 4.09, p < .025].

Inspection of responses also revealed another type of error that occurred on an appreciable number of trials (12% of total error for monosyllabic words). We have chosen to call these errors combination errors, since they contained phonemes from *both* the visual and auditory words at corresponding positions of the word (initial, middle, final). As an example of a combination error, if the visual word was "next" and the auditory word was "chime," the perceived visual word was reported to be "chest," a combination of visual and auditory phonemes (see Table 6 for other examples from Experiment 2). The errors are of particular interest because they suggest a form of partial auditory bias, rather than complete auditory bias as apparently is the case for auditory error. Combination errors also appear to be related to what have been referred to by other investigators as fusion experiences during visual-

Table 2
Accuracy of Lipreading for Visual-Only (SLR) and Visual-Auditory Discrepancy (DLR) Conditions.

	Multiple Choice				Identification			
	Monosyllabic		Spondaic		Monosyllabic		Spondaic	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
SLR	.86	.09	.90	.06	.17	.12	.21	.11
DLR	.39	.16	.44	.23	.04	.05	.07	.07

Note—Both identification and multiple choice testing were assessed.

Table 3
Distribution of DLR Auditory Error (Dub Given as a Response) Across Word Type and
Relation Between Dubbed (Auditory) and Lipped (Visual) Information

	Initial		Final		Both		Neither	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Monosyllabic	.15	.12	.17	.12	.20	.12	.13	.11
Spondaic	.12	.09	.09	.08	.11	.12	.02	.05

Note—Auditory error was 32% of total error.

auditory speech discrepancy (Dodd, 1977; MacDonald & McGurk, 1978; McGurk & MacDonald, 1976). As indicated earlier, however, there are many differences between the experimental procedures employed in the present investigation and previous studies of visual-auditory speech discrepancy. Furthermore, it is not at all clear that the subjective experience is the same for combination and fusion errors; hence, we have chosen a different terminology. For now, we merely note the existence of combination errors in Experiment 1; as part of Experiment 2, the issue is taken up systematically by exploring in more quantitative terms the perceptual experience of observers who make combination errors during lipreading.

A final quantitative analysis of error patterns assessed total identification error as a function of the type of phoneme discrepancy at the initial and final positions of words (visible- vs. obscure-consonant discrepancy). Percent error as a function of word type, phoneme position, and discrepancy type is presented in Table 4. The only significant effect to emerge was associated with the discrepancy-type factor: obscure-consonant discrepancies resulted in more lipreading errors [$F(1,19) = 9.7$, $p < .01$].

Summary

General consideration of findings from Experiment 1 will be deferred until the General Discussion section. To summarize, however, it was found that the accuracy of perceived visual speech (lips) was substantially reduced by the presence of discrepant auditory information. The effect occurred for both identification and multiple choice responding. In addition, an analysis of identification errors, especially auditory (Dub) errors, revealed systematic audi-

tory biasing in the perceptual experience of observers. Monosyllabic words were more susceptible to the auditory biasing than were compound words, and the greater the visual-auditory discrepancy (visible-consonant), the smaller the auditory bias effect. In sharp contrast, discrepant visual information exerted virtually no effect on the perception of auditory speech.

EXPERIMENT 2

One important finding of Experiment 1 was that the larger the visual-auditory speech discordance (i.e., NEITHER dubbing category or visible-consonant discordance), the less the biasing effects of auditory information. What appear to be complementary effects have been reported in the case of spatial localization under visual-proprioceptive and visual-auditory discrepancy: with small amounts of discordance, visual bias is substantial, but with larger amounts of discordance, visual bias is found to be lessened (Warren & Cleaves, 1971; Warren, Welch, & McCarthy, 1981). These results suggest that the perceptual bias mechanisms that operate during the perception of space and speech share similar properties.

In the domain of space perception, Warren and his colleagues have also demonstrated recently that an important determinant of the amount of perceptual biasing by one perceptual system over another is the "compellingness" of the bimodal discordance as an illusory single event (Warren et al., 1981). In an experiment that made use of a small TV monitor and a displaced audio speaker, greater perceptual bias of spatial position occurred if a face of a speaking person was used as the visual stimulus than if auditorily modulated light was the only visual stimulus. Presumably, sub-

Table 4
DLR Percent Error as a Function of Word Type, Phoneme Position, and Discrepancy Type

	Initial Phoneme				Final Phoneme			
	Monosyllabic		Spondaic		Monosyllabic		Spondaic	
Discrepancy	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Visible	.75	.44	.93	.24	.95	.22	.93	.18
Obscure	1.00	.00	1.00	.00	1.00	.00	1.00	.00

jects' beliefs that discordant spatial information emanates from a single location play a critical role in determining perceptual bias.

We designed a second experiment to assess further the relation between perceptual bias during speech and space perception. In an attempt to strengthen the single-event assumption in our subjects, we created a test in which visual-auditory speech discrepancy was present on half the trials, but visual-auditory speech information corresponded on the remaining trials. Furthermore, in a high compelling condition, a speaking woman was used as the visual stimulus and her voice was dubbed over as the auditory stimulus. In a low compelling condition, a man's voice was dubbed over the videotape of the speaking woman. If perceptual bias during speech perception is similar to that studied during space perception, we would expect to find more auditory bias or lipreading disruption for the high compelling condition.

One further issue was addressed in Experiment 2. It could be argued that what we have been referring to as auditory bias during lipreading is not really a subjectively compelling experience (e.g., auditory errors and combination errors), as is reportedly the case during ventriloquism (Bertelson & Radeau, 1981). Since lipreading is such a difficult skill to begin with, subjects under dubbing conditions may simply be confused and guess at a correct response. In order to explore this issue further, we used only monosyllabic words in Experiment 2, since we found in Experiment 1 that both auditory and combination errors were more frequent under this condition. Furthermore, the subjects were required to rate on a five-point scale their degree of confidence that their answer was correct—that is, was the lipped word.

Method

Procedure

Three separate tests were devised for Experiment 2. The first was a standard lipreading (SLR) test in which monosyllabic words were presented to observers, who then were required to identify the lipped word. The second two tests involved dubbed lipreading (DLR). In a high compelling test, a woman speaker was videotaped as the visual stimulus and her voice was dubbed over as the auditory stimulus. In a low compelling test, a man's voice was used to dub the auditory information over the videotape of the speaking woman.

In each of the three tests, the same 32 monosyllabic words were used as the visual stimuli. For half these words, the dubbed auditory information was identical to the visual information; for the other half, discrepant auditory information was dubbed over the videotape. For the 16 visual-auditory discrepancy trials, half involved the BOTH dubbing category and half involved the NEITHER dubbing category (see Table 1).

In addition to being required to identify the lipped word, the subjects in Experiment 2 were also asked to rate how confident they were that their response was correct on a five-point Lickert-type scale (1 = complete guess, 3 = moderately confident, 5 = absolutely confident). The subjects were required to provide a response on each trial and to rate their confidence in it.

The three tests were administered to separate groups of 20 subjects each. All procedures were the same as those used in Experiment 1, except that the intertrial interval was increased to 15 sec to accommodate the added rating response.

Apparatus and Materials

The same video equipment used in Experiment 1 was used again. The subjects' response sheets contained 32 blanks for identification and a five-point rating scale adjacent to each identification blank.

Subjects

Sixty undergraduate students fulfilling a course requirement served as subjects. They were assigned randomly to the three groups.

Results

Mean accuracy for the SLR group was .13, which was determined to be significantly greater than zero [$t(19) = 8.5$, $p < .001$]. In addition, the mean confidence associated with correct responses was 3.0, while that associated with incorrect responses was 2.3. The difference between these average confidence values was significant [$F(1,19) = 23.7$, $p < .001$].[1]

For the DLR groups, the first analysis determined that lipreading accuracy averaged 81% on "same" trials and 7% on "discrepancy" trials [$F(1,38) = 512.5$, $p < .001$]. On the "same" trials, the low- and high-compellingness subjects did not differ in lipreading accuracy. Rated confidence in their correct responses was 3.7, and that associated with incorrect responses was 2.3 [$F(1,38) = 30.2$, $p < .001$].

The distribution of the DLR discrepancy data across compellingness and dub categories is presented in Table 5. ANOVA revealed a significant interaction between the compellingness and dub category variables [$F(1,38) = 3.5$, $p < .05$]. As can be seen, the interaction is attributable to the fact that accuracy for the low-compellingness group under the NEITHER dub category resulted in significantly greater accuracy (.14) than did the other three conditions (ps < .05). It was also determined that the accuracy of .14 was significantly greater than zero [$t(1,19) = 3.01$, $p < .01$]. This result indicates that low compellingness combined with a relatively large visual-auditory discordance results in the least disruption of lipreading performance.

An analysis of errors was conducted next. As was

Table 5
Accuracy of Lipreading Under High and Low Compelling
Conditions of Experiment 2

| | Dubbing Categories | | | | | |
| | Both | | | Neither | | |
	Mean	SD	C	Mean	SD	C
High Compelling	.04	.07	2.0	.04	.06	2.6
Low Compelling	.06	.08	3.5	.14	.18	3.7

Note—C = subjects' confidence in these correct responses.

Table 6
Combination Errors (E), Their Frequency of Occurrence (F),
and the Average Confidence (C) Associated With Each

Visual	Auditory	E	F	C
mail	but	bell	10	3.9
word	fair	wear	9	2.8
lamp	ring	lip	7	3.4
birth	rough	breath	6	3.5
light	keg	leg	6	3.4
whole	wet	old	6	3.5
word	fair	were	5	3.2

Note–Only combinations which were reported by at least 5 sub-jects are shown (n = 40). Combination error was 18% of total error.

the case in Experiment 1, auditory error (dub) occurred as a response on an appreciable number of trials (28% of total error). In addition, combination errors also occurred on an appreciable number of trials (18% of total error). As defined in Experiment 1, a combination error is a response that contains phonemes from both the visual and auditory words at the corresponding phonetic positions. Table 6 presents a list of these errors and their frequencies of occurrence (only combination errors that were reported by at least five subjects are shown). Table 7 presents a distribution of total error for the BOTH and NEITHER dubbing categories across error type—auditory, combination, other. The subjects' confidence that these erroneous responses were, in fact, correct is also presented. (These data are collapsed across the high- and low-compellingness distinction, since these groups did not differ in terms of the distribution of their total error or confidence ratings across error type.) ANOVA on the percent error data revealed a significant interaction between dubbing category and error type [$F(2,76)=9.5$, $p < .001$]. The interaction is attributable to the fact that (1) auditory errors occurred more frequently under the BOTH dubbing category, (2) combination errors occurred equally often across dubbing categories, and (3) the remaining errors occurred more frequently for the NEITHER dubbing category (simple effects, $ps < .01$).

ANOVA on the confidence data[1] revealed a significant effect for error type [$F(2,76)=14.5$, $p < .001$]. Greater confidence values were given for auditory errors than for combination errors, which in turn were

given greater confidence values than the other errors (Newman-Keuls, $ps < .01$). There also proved to be a significant interaction between error type and dubbing category. As can be seen, the interaction is attributable to the relatively high confidence associated with auditory error for the BOTH dubbing category.

The major finding to emerge from the error analyses is that greatest auditory error occurs for the BOTH category *and* subjects are relatively confident that their responses are correct. In fact, they are as confident of these incorrect responses as subjects are of correct SLR responses (3.4 vs. 3.0). Combination errors were also found to occur on an appreciable number of trials, and more confidence was associated with them than with the remaining errors (2.7 vs. 2.2).

GENERAL DISCUSSION

Based on the results from the control group of Experiment 1, it can be argued that discrepant visual speech information exerts little effect on auditory speech recognition (this finding appears to contradict findings reported by McGurk and MacDonald, but we will postpone consideration of this issue until later in the discussion). In contrast, observers' abilities to read lips visually were interfered with substantially when discrepant auditory information was simultaneously present. This effect occurred for identification and multiple choice testing. Furthermore, the auditory bias observed was not a general disruption of visual processing, but was related systematically to several factors, including people's abilities to read lips ordinarily (SLR), the phonetic relation between lipped and dubbed words, and the compellingness of the speech discordance as a single event.

In Experiment 1, auditory bias was evidenced by the fact that lipreading ability under dubbing conditions dropped from 88% to 42% for multiple choice and 19% to 5% for identification. Furthermore, during identification testing, the actual dub (auditory error) was given as a response about 30% of the time. The data on accuracy of lipreading under nondiscrepancy conditions are consistent with previous assessments, and also replicate the finding that spondaic words are lipread more accurately than monosyllabic words (Erber, 1974). A novel finding associated with the bimodal discordance or dubbing procedure was that word

Table 7
Distribution of Total Error Across Different Categories of Error for Experiment 2

	Auditory			Combination			Other		
	Mean	SD	C	Mean	SD	C	Mean	SD	C
Both	.39	.22	3.4	.17	.16	2.7	.44	.22	2.2
Neither	.16	.17	2.5	.19	.13	2.7	.65	.19	2.2

Note–C = subjects' confidence that these responses were, in fact, correct.

type did not interact with SLR or DLR tests. The extent to which spondaic words are lipread more accurately than monosyllabic words was the same whether discrepant dubs were present or not. Thus, the interfering or biasing effect of auditory information is closely related to the ability to read lips ordinarily. A reasonable explanation of this finding is that spondaic words as speech events possess a longer and more structured period of articulation and therefore provide the lipreader more information.

Another factor that proved to be related to auditory bias was the magnitude of the visual-auditory discrepancy. Several findings emerged from the experiments that demonstrate that as the discrepancy between visual and auditory phonetic information becomes larger, the biasing effect of audition increases: (1) spondaic words resulted in less total error and auditory error than monosyllabic words; (2) for both monosyllabic and spondaic words, auditory error was smallest when *neither* the final nor the initial phonemes of the lipped and dubbed words corresponded; and (3) visible-consonant discrepancies at the beginning or ending of words resulted in less total error than obscure consonant discrepancies. As noted earlier, complementary effects have been reported for perceptual bias during spatial discordance: increased discrepancy reduces visual bias (Warren & Cleaves, 1971; Warren et al., 1981).

A final factor that proved to be related to perceptual bias during speech perception was the compellingness of the speech discrepancy as a single event. In Experiment 2, lipreading accuracy was highest when the visual-auditory discrepancy was the least compelling. If a man's voice was dubbed over the videotaped face of a speaking woman, and both the initial and final visual and auditory phonemes of monosyllabic words did not correspond, auditory bias was decreased significantly, as indicated by increased lipreading accuracy. Again, similar effects have been reported for modality discordance during space perception: the higher the compellingness of the visual-auditory discordance as a single event, the greater the visual bias (Warren et al., 1981).

The results from space perception and speech perception both are consistent with a general model of (1980). According to their model, a perceiver's natural inclination is to process bimodal information as a single event. Bias is an attempt by the perceptual system to use its built-in flexibility to continue to perceive a single event even when discrepant bimodal information is present. However, if the discrepancy is too great or the compellingness of the illusory event is too low, the single-event assumption can no longer be maintained and observers may attempt to attend differentially to the two sources of information.

Despite the similarity among results from space and speech perception investigations, one might still reasonably question whether the bias effects from the two realms of study are really analogous. For instance, one important question that has not been addressed is whether the conscious experience during dubbed speech reading is as subjectively compelling as the experience during ventriloquism reportedly is (Bertelson & Radeau, 1976, 1981). The present data provide some suggestive evidence that, on certain occasions, observers remain unaware of a visual-auditory speech discordance. Rated confidence that auditory errors were, in fact, correct was equal to, if not greater than, the confidence associated with correct lipreading responses (SLR). It is difficult to explain why a subject would give the auditory information as a response during DLR, and do so with relative confidence, if a modality discordance had been experienced. Indeed, it may be that auditory error in these experiments represents a case of relatively complete auditory bias, under conditions in which discordance is not experienced.[2]

Auditory bias was not always complete, however, as evidenced by the occurrence of what we have chosen to call combination errors. Apparently, in the case of these errors, visual information is incorporated into an observer's experience, and auditory bias is only partial. The evidence also suggests that subjects may have experienced single events on some of these trials, since rated confidence that their combination errors were, in fact, correct was significantly higher than when they made other kinds of errors.

As indicated in Experiment 1, effects seemingly analogous to our combination errors have been reported by others (Dodd, 1977; MacDonald & McGurk, 1978; McGurk & MacDonald, 1976), and have been referred to as "fusions." The term fusion implies a blending or melting of two separate sources of information in consciousness. Although we have some indirect evidence that our subjects may have experienced combination errors as fusions, we are not certain that this was the case (rated confidence in these responses was, after all, a bit below "moderately confident"). There are also several other reasons that we have chosen a different terminology to describe this type of response. McGurk and MacDonald used morphemes that were not complete words, and they required subjects to report what they *heard* rather than what they *saw*. In fact, fusions like those reported by McGurk and MacDonald would have been expected to occur for the subjects in our control group of Experiment 1, who were required to attend to auditory speech information. However, there proved to be virtually no visual influence on auditory speech perception, and no combination errors were reported. A plausible explanation of these discrepant findings is that complete words were used in the present investigation, a condition obviously more typical of normal speech perception during connected discourse. The presence of additional speech information in

words, compared with phonemes—especially transitions to and from phonemes—evidently eliminated the occurrence of fusion experiences like those reported by McGurk and MacDonald among our control subjects.

To summarize, it seems intuitively true that people usually rely on auditory information during speech more than they do visual information. But we cannot properly assess the role of auditory information during normal speech conditions because visual information and auditory information are redundant. The use of auditory dubs to create a visual-auditory speech discrepancy demonstrated that auditory information does indeed dominate during speech recognition. The findings seem quite analogous to findings from investigations of spatial discordance and are consistent with all three models of perceptual dominance that are currently discussed—attentional allocation, modality precision, and ecological validity. This is not particularly surprising, since all three models have in common the underlying notion that if information processed in a particular modality is more informative, it will dominate intermodal processing.

REFERENCES

ARONSON, E., & ROSENBLOOM, S. Space perception in early infancy: Perception within a common auditory-visual space. *Science*, 1971, 172, 1161-1163.

BERTELSON, P., & RADEAU, M. Ventriloquism, sensory interaction, and response bias: Remarks on the paper by Choe, Welch, Gilford, and Joula. *Perception & Psychophysics*, 1976, 19, 531-535.

BERTELSON, P., RADEAU, M. Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception & Psychophysics*, 1981, 29, 578-584.

DODD, B. The role of vision in speech perception. *Perception*, 1977, 6, 31-40.

EASTON, R. D. The effect of head movements on visual and auditory dominance. *Perception*, 1983, in press.

ERBER, N. D. Visual perception of speech by deaf children: Recent developments and continuing needs. *Journal of Speech and Hearing Research*, 1974, 39, 178-185.

GIBSON, J. J. *The ecological approach to visual perception*. Boston: Houghton-Mifflin, 1979.

JEFFERS, J., & BARLEY, M. *Speechreading (lipreading)*. Springfield, Ill: Charles C Thomas, 1974.

LEDERMAN, S. J. Auditory texture perception. *Perception*, 1979, 8, 93-103.

LEE, D. N. The functions of vision. In H. L. Pick & E. Saltzman (Eds.), *Modes of perceiving and processing information*. New York: Erlbaum, 1978.

LEE, D. N., & LISHMAN, J. R. Visual proprioceptive control of stance. *Journal of Human Movement Studies*, 1975, 1, 87-95.

MACDONALD, J., DWYER, D., FERRIS, J., & McGURK, H. A simple procedure for accurately manipulating face-voice synchrony when dubbing speech onto videotape. *Behavior Research & Instrumentation*, 1978, 10, 845-847.

MACDONALD, J., & McGURK, H. Visual influence on speech perception. *Perception & Psychophysics*, 1978, 24, 253-257.

McGURK, H., & MACDONALD, J. Hearing lips and seeing voices: A new illusion. *Nature*, 1976, 264, 746-748.

NEISSER, U. *Cognition and reality*. San Francisco: Freeman, 1979.

O'CONNOR, N., & HERMELINE, B. *Seeing and hearing and space and time*. New York: Academic Press, 1978.

PICK, H. O., WARREN, D. H., & HAY, J. C. Sensory conflict in judgments of spatial direction. *Perception & Psychophysics*, 1969, 6, 203-205.

THURLOW, W. R., & JACK, C. E. Certain determinants of the "ventriloquism effect." *Perceptual and Motor Skills*, 1973, 36, 1171-1184.

TURVEY, M. T. Preliminaries to a theory of action with reference to vision. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting and knowing*. Hillsdale, N.J: Erlbaum, 1977.

WARREN, D. H., & CLEAVES, W. T. Visual-proprioceptive interaction under large amounts of conflict. *Journal of Experimental Psychology*, 1971, 90, 206-214.

WARREN, D. H., & SCHMITT, T. L. On the plasticity of visual-proprioceptive bias effects. *Journal of Experimental Psychology: Human Perception and Performance*, 1978, 4, 302-310.

WARREN, D. H., WELCH, R. B., & McCARTHY, T. J. The role of visual-auditory "compellingness" in the ventriloquism effect: Implications for transitivity among the spatial senses. *Perception & Psychophysics*, 1981, 30, 557-564.

WELCH, R. B., & WARREN, D. H. Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, 1980, 88, 638-667.

WELCH, R. B., WIDAWSKI, M. H., HARRINGTON, J., & WARREN, D. H. An examination of the relationship between visual capture and prism adaptation. *Perception & Psychophysics*, 1979, 25, 126-132.

WOODWARD, M. R., & BARBER, C. G. Phoneme perception in lipreading. *Journal of Speech and Hearing Research*, 1960, 3, 212-222.

NOTES

1. Statistical analysis consisted of averaging confidence values associated with correct versus incorrect responses (or other category distinctions) to arrive at a single score in each category for each subject. If a subject did not have a value in a given category (e.g., one subject did not lipread any word correctly), a confidence value that equaled the average of the other confidence values in the category was assigned.

2. We deliberately chose to have subjects rate the confidence they had in a response rather than to have them detect whether a discordance was present or not. We did so because we were attempting, in general, to strengthen the single event assumption. Thus, we do not have a direct measure of whether discordance was experienced or not.