Journal of Experimental Psychology:
Human Perception and Performance
1998, Vol. 24, No. 4, 1232–1242

0096-1523/98/$3.00

Laterality in Visual Speech Perception

Paula M. T. Smeele
Delft University of Technology

Dominic W. Massaro and Michael M. Cohen
University of California, Santa Cruz

Anne C. Sittig
Delft University of Technology

The lateralization of visual speech perception was examined in 3 experiments. Participants were presented with a realistic computer-animated face articulating 1 of 4 consonant–vowel syllables without sound. The face appeared at 1 of 5 locations in the visual field. The participants' task was to identify each test syllable. To prevent eye movement during the presentation of the face, participants had to carry out a fixation task simultaneously with the speechreading task. In one study, an eccentricity effect was found along with a small but significant difference in favor of the right visual field (left hemisphere). The same results were found with the face articulating nonlinguistic mouth movements (e.g., kiss). These results suggest that the left-hemisphere advantage is based on the processing of dynamic visual information rather than on the extraction of linguistic significance from facial movements.

Laterality studies attempt to localize brain activity of certain psychological functions and to examine possible hemispheric differences in these functions. The present study focuses on laterality effects for the perception of visible speech. A general finding is that the left hemisphere is superior to the right hemisphere in the processing of auditorily presented speech (Kimura, 1961, 1967; Studdert-Kennedy & Shankweiler, 1970). Listeners identify speech sounds presented to their right ear (left hemisphere) more accurately than those presented to their left ear (right hemisphere). The left hemisphere has also been found to be dominant for the recognition of written letters and words (see Bryden, 1965, for a review of lateralization of spoken and written language; see also Bryden, 1982; Mishkin & Forgays, 1952). This result is particularly interesting because written language involves detection and interpretation of visual shape configurations (i.e., letters), which are thought to be processed primarily in the right hemisphere.

Paula M. T. Smeele and Anne C. Sittig, Department of Industrial Design Engineering, Delft University of Technology, Delft, The Netherlands; Dominic W. Massaro and Michael M. Cohen, Department of Psychology, University of California, Santa Cruz. Paula M. T. Smeele is now at the Human Factors Research Institute of the Netherlands Organization for Applied Scientific Research (TNO), Soesterberg, The Netherlands.

This research was supported by grants from the Public Health Service (PHS R01 NS 20314) and the National Science Foundation (BNS 8812728), by the graduate division of the University of California, Santa Cruz, and by grants from the Netherlands Organization for Scientific Research (NWO) and the Delft University Fund. The experiments were carried out at the University of California, Santa Cruz.

We thank Tim Jordan and two anonymous reviewers for their comments.

Correspondence concerning this article should be addressed to Dominic W. Massaro, Department of Psychology, Social Sciences II, 1156 High Street, University of California, Santa Cruz, California 95064. Electronic mail may be sent to massaro@fuzzy.ucsc.edu.

Lateralization of visual speech perception is a rather new area of investigation. Facial movements naturally accompany the production of speech sounds, and infants use these two sources of speech information early in life (Dodd, 1979; Kuhl & Meltzoff, 1982). Visual information from the face, tongue, and lip movements of a talker provides information about the spoken message and enhances the intelligibility of auditory speech (Binnie, Montgomery, & Jackson, 1974; Summerfield, 1979). Visual speech predominantly provides information about the place of articulation (Binnie et al., 1974; Grant & Walden, 1995; Smeele, 1994; Smeele & Sittig, 1991a; van Son, 1993). The influence of visual speech occurs in a wide variety of conditions: in noisy environments (Binnie et al., 1974; Breeuwer & Plomp, 1984; Sumby & Pollack, 1954), with highly complex sentences (Reisberg, McLean, & Goldfield, 1987), with conflicting auditory and visual speech (Green & Kuhl, 1989; Massaro, 1987; McGurk & MacDonald, 1976), and with asynchronous auditory and visual speech information (Campbell & Dodd, 1980; Smeele & Sittig, 1991b). Similar recency and suffix effects are found with heard and speechread lists, indicating that there is a close connection between the processing of auditory and visual speech (Campbell, 1987b; Campbell & Dodd, 1982; Gathercole, 1987; Vroomen, 1992). If speechreading is predominantly a linguistic task, a straightforward prediction is that the left hemisphere would be the dominant hemisphere. On the other hand, a right-hemisphere advantage is usually found in the perceptual processing of faces (usually the delayed matching of one face with another; Sergent, 1995). If speechreading requires facial processing, then one would expect to find right-hemisphere dominance (see also Moscovitch, Scullion, & Christie, 1976).

The spatial aspect of visual speech is also present in sign language, and the recognition of signs has shown left-hemisphere dominance (Bellugi & Klima, 1993). Although sign language is dependent on spatial information, several studies have shown that processing sign language is different from other visuospatial functions in the brain (Poizner, Bellugi, & Klima, 1990; Söderfeldt, Rönnberg, & Risberg, 1994). For sign language, it seems that the linguistic character of the stimuli (signs) overrides the visuospatial one. The question is whether this is also true for visual speech information.

To date, hemispheric dominance in visible speech perception has not been conclusively demonstrated. Campbell (1986) examined lateralization of static lip shapes. Normal right-handed participants were presented with an auditory syllable preceding a photograph of a person's face pronouncing the same or a different syllable and indicated whether the spoken syllable matched the syllable seen on the photograph. A right-hemisphere advantage was found in that participants responded faster in the matching task when the visual stimulus was presented to the participants' left visual field (LVF) than to their right visual field (RVF). Baynes, Funnell, and Fowler (1994) studied the influence of a mouthed word on the identification of a dubbed acoustic word (i.e., the McGurk effect). They found that participants reported more McGurk illusions, thus showing more visual influence, when a talker's face was presented to their LVF than to their RVF.

However, left-hemisphere dominance for visuospatial speech information is suggested by other studies. Campbell and colleagues (Campbell, 1987a; Campbell et al., 1990; Campbell, Landis, & Regard, 1986) tested patients with either right-hemisphere or left-hemisphere lesions. In their first study (Campbell et al., 1986), 2 patients were contrasted: one with a right-hemisphere and one with a left-hemisphere lesion. The patient with right-hemisphere damage showed severely impaired recognition of familiar faces but had normal speechreading performance. However, the patient with left-hemisphere lesions showed poor speechreading skills. This pattern of results indicated a dissociation between the processing of a face and of speech movements and supported the view that the left hemisphere is dominant for speechreading. Unfortunately, the Campbell et al. (1990) study uncovered a person with right-hemisphere damage who could not speechread. More recently, Campbell, De Gelder, and De Haan (1996) found a left-hemisphere advantage when photographs of speaking faces were matched on the basis of perceived mouth shape. No such advantage was found when the same images were matched for identity. Finally, using videotaped speech, Diesch (1995) observed different results for different types of auditory–visual syllables. The visible component of spoken McGurk syllables was presented 3.5° to the left or right of fixation. A face was presented on each side of the fixation, but only one of them was articulating a syllable. The responses revealed a small right-hemisphere advantage for fusion stimuli (e.g., auditory /ba/ paired with visual /ga/) and a small left-hemisphere advantage for combination stimuli (e.g., visual /ba/ paired with auditory /ga/).

The different outcomes of these studies might be partially due to the fact that different groups of participants (normal vs. neurological) were tested or that different experimental designs were used. In studies involving normal participants, a right-hemisphere dominance was found (Baynes et al., 1994; Campbell, 1986), whereas studies involving neurological patients indicated a dominant role of the left hemisphere (Campbell, 1987a; Campbell et al., 1990, 1986). To illustrate differences in designs, participants in Campbell's (1986) study had to match auditory speech with static pictures of faces, participants in the Campbell et al. (1990) study had to identify conflicting auditory and visual speech information, and photographs were matched to each other in the Campbell et al. (1996) study.

We conducted three experiments to further investigate laterality effects for visual speech perception in the normally functioning brain. A computer-animated talker was used rather than still photographs. The animation enabled flexible and exact control over the location of the face. In the study by Baynes et al. (1994), the face was always presented in the same location, and the places on which participants had to fixate varied. With this design, participants could predict exactly where the face would occur on each trial, making the possibility of eye movements problematic. Given the fact that the fixation mark appeared for more than 1,800 ms, there was enough time for making eye movements in the direction of the face presentation. To reduce the likelihood of occurrence of eye movements, we had participants maintain a central fixation and varied the location of the face. To discourage participants from making eye movements away from the fixation to look toward the face, we asked them to perform a visual counting task. Successful performance on the task required them to maintain eye position at the point of fixation during the critical part of the visible syllable presentation.

As noted earlier, we used computer-animated speech rather than still pictures or a natural talking face (Cohen & Massaro, 1990, 1993). The animation provided the dynamic aspect of visual speech that is missing in still pictures. The use of an animated face also allowed us to achieve exact control over the facial movements. There is evidence that talkers show an asymmetry in their articulations (Graves, Goodglass, & Landis, 1982; Wolf & Goodale, 1987). They tend to open the right side of their mouths faster and wider than the left side. The explanation is that the left hemisphere exercises more control over speech production than does the right hemisphere, and therefore more of the right side of the face is moved. Thus, a real face might show a perceptual advantage presented to the RVF (left hemisphere) because there is more information on the talker's right side. If the face is presented to the right of fixation, its right side would be closer to foveal or central vision. If the face is presented to the left of fixation, its right side would be farther from central vision. Thus, the more informative part of the face would be closer to central vision when presented in the RVF than in the LVF. The animated face was made to be symmetrical, which precluded this potential confounding.

Thus, in our experiments, there were several valuable improvements in methodology. In Experiments 1 and 2, right-handed participants speechread the synthetic face articulating consonant–vowel (CV) syllables without sound. The face appeared at the fixation point or at one of four peripheral locations. In Experiment 3, the speech aspect of the visual identification task was eliminated. Right-handed participants were presented with nonlinguistic facial movements to investigate whether hemispheric contributions would be similar to those found for linguistically meaningful movements.

Experiment 1: Visible Speech

Method

Participants. Twenty native speakers of American English participated in this experiment. They were all students at the University of California, Santa Cruz. Their ages ranged from 18 to 21 years. They all reported being right-handed and having normal hearing and normal or corrected-to-normal vision. They received course credit or were paid for their participation.

Stimuli. Synthetic visible speech stimuli were used as test items. Using computer animation, we programmed a synthetic face (Cohen & Massaro, 1990, 1993) to articulate each of the four syllables /ba/, /va/, /ða/, and /da/ without sound. The face appeared at one of five locations on a video display monitor: one location in the center, two locations in the left field of the screen (at a visual angle of 3.5° and 7° measured from the center to the foveal edge of the face), and two locations in the right field (at a visual angle of 3.5° and 7° from the center). The face was approximately 8.1 cm wide (9.3° of visual angle) and the mouth 2.5 cm wide on the screen (2.9° of visual angle). The syllables were articulated in citation form, with a duration of about 1,100 ms. A 100-ms, 1000-Hz warning tone preceded each stimulus by a random interval between 600 and 1,000 ms.

Simultaneously with the animated face, a central fixation dot was presented on the screen. The dot either grew slightly larger once or twice during the presentation of the visual speech stimulus or remained constant in size. These size variations were visible as flickers. The size change occurred during the first 333 ms, the last 333 ms, or during both intervals of the speech stimulus. The normal size of the dot was 2.0 mm, and the increased size was 2.5 mm (corresponding to visual angles of 0.23° and 0.29°, respectively). Pilot work demonstrated that the dot task could not be performed accurately unless the dot was in foveal vision. Figure 1 shows a trial with the synthetic face appearing in the RVF.

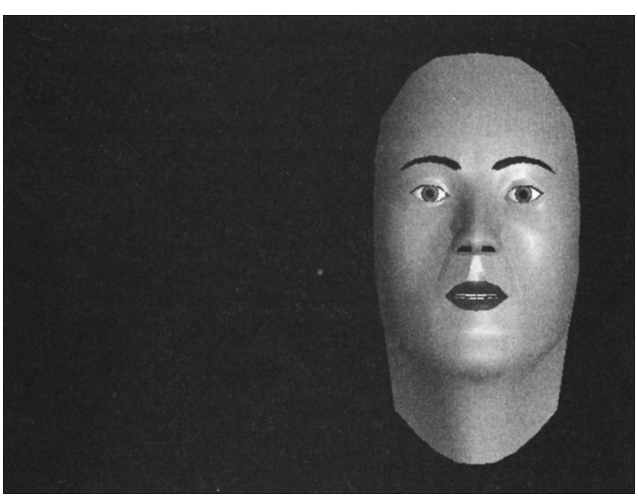

Figure 1. Illustration of a trial. The synthetic face appears 7° to the right of the central fixation dot.

Design and procedure. Four different syllables, five locations of the face, and three dot conditions were used. To measure how well participants could detect the size variations of the dot, we added a control condition to the design in which only the central dot (with no synthetic face) was presented. Thus, there were 4 (syllables) × 6 (five face locations and no face) × 3 (dot-size variations) = 72 independent conditions. Eight random blocks were generated, giving 8 × 72 = 576 test trials for each participant.

The participants were instructed to identify the synthetic face as /ba/, /va/, /ða/, or /da/. To prevent the participants from making eye movements toward the face, they had to perform the dot-counting task simultaneously with the speechreading task. While fixating the central dot, they had to detect and count its size variations (0, 1, or 2). They first made their syllable responses by pressing a key labeled *b, v, th,* or *d* on a terminal keyboard. They then pressed a key labeled *0, 1,* or *2* for the dot-counting task. In the control condition with no face, only the dot-counting task was required. To counterbalance response hand, 10 participants responded with their right hand and 10 with their left.

Up to 4 participants could be tested simultaneously in individual sound-attenuated rooms. The stimuli were presented on a 12-in (30.48 cm) color monitor (NEC Model C12-202A). Each participant received 30 practice trials, and eight blocks of 72 test trials, with a short break after the first four blocks. The pace of the experiment was participant driven in that the next trial did not occur until all participants had responded to the previous trial. The interstimulus interval was approximately 5 s. The experiment took about 90 min.

Results

We first analyzed the participants' performance on the central dot-counting task (fixation task). When no visual speech stimulus was presented, the participants correctly detected the central dot-size changes on 85% of the trials. This result indicated that the fixation task was not too easy; even when the participants could attend to the dot, performance was not perfect. The percentages of correct counting were 99 with no, 71 with one, and 85 with two dot-size changes.

The accuracy of counting dot-size changes dropped to an average of 74% when a visual speech stimulus was presented. The decrease from 85% to 74% was partially due to the face overlapping the dot when it was presented at the central location and partially due to dual-task interference. Given that pilot results indicated that accuracy on the dot-counting task required central fixation, this level of performance indicated that the participants maintained their fixation as instructed. Analyses of the errors, shown in Table 1 (Experiment 1), indicated that the participants reported predominantly fewer flickers than had actually appeared. In other words, the participants did not appear to guess randomly but reported flickers only if they saw them. Correct performance on the fixation task with simultaneous visible speech amounted to 95% with no, 62% with one, and 64% with two dot-size changes. An analysis of variance (ANOVA) was carried out on the percentage of correct counts as a function of the six presentation conditions: the fixation task alone and the five locations of the face presentation. The effect of presentation condition was significant, $F(5, 90) = 19.16$, $p < .001$. The accuracy was 85% with no visible speech, 73% and 72% (face at 7° and 3.5° in

Table 1
Counting Responses to the Dot-Size Changes (Fixation) Task When No Face Was Presented in Experiments 1–3

No. of changes	Counting response	% response in Experiment 1	% response in Experiment 2	% response in Experiment 3
0	0	99.7	92	95
	1	0.3	7	5
	2	0	1	0
1	0	26	21	16
	1	71	70	77
	2	3	9	6
2	0	4	3	1
	1	11	13	9
	2	85	83	90

the LVF), 67% (face in center), and 69% and 76% (face at 3.5° and 7° in the RVF).

In the analysis of the speechreading task, we maximized the proportion of trials on which participants maintained fixation by including only those trials in which participants correctly counted the dot-size changes. Such a selection did not eliminate those trials on which the participants did not fixate on the dot but nevertheless performed the fixation task correctly. However, the evidence just presented showed this was necessarily a small proportion of trials (see Table 1). Excluding the trials on which participants did fixate on the dot but did not perform the fixation task correctly only reduced the number of trials taken into account. Note that, although this generally led to an uneven number of trials in each condition, performance was averaged over trials in each condition, so any difference in numbers of trials would not have mattered in the analyses. There were 24 replications for each participant at each of the 20 (4 syllables × 5 locations of the face) conditions. Excluding trials in which the participant was incorrect on the fixation task never reduced the number of observations for a participant in a given condition below 8. The average minimum number was about 15. Thus, the estimated proportions for each participant at each condition should be fairly reliable.

Figure 2 shows the average proportion of correctly recognized visual speech syllables for each location of the synthetic face (speechreading task) only for those trials on which participants correctly counted the dot-size changes (i.e., an average of 74% of the trials). Note that in the figure, the presentation of the face in the RVF corresponds with direct presentation of information to a participant's left hemisphere, and presentation of the face in the LVF with direct presentation to a participant's right hemisphere. As can be seen, performance was best at fixation and decreased as the face was moved into the periphery.

For each participant, accuracy of identification was computed for each of the four syllables at each of the five locations. To address the hemispheric specialization issue, only the four eccentric locations were analyzed in the ANOVA. In addition, these four locations were partitioned into two variables: LVF or RVF and distance (3.5° or 7° in Experiment 1 and 4.6° and 9.1° in Experiment 2). Thus, the ANOVA had three factors: syllable, left or right location, and distance. This ANOVA was carried out on all of the results as well as the results of just those trials on which the participant was accurate on the dot-counting task. The two analyses gave equivalent results, so we report only the analysis based solely on trials on which the participant was accurate in the dot-counting task.

As can be seen in Figure 2, there was a large difference in accuracy across the four syllables, $F(3, 57) = 17.77$, $p < .001$. Best performance was on the syllable /va/ and poorest on /da/, with the other two syllables falling in between. Performance averaged 81%, 90%, 84%, and 65% for the syllables /ba/, /va/ /ða/, and /da/, respectively. Performance also decreased with increases in distance into the periphery, $F(1, 19) = 14.43, p = .002$. The effect of visual field was not significant, $F(1, 19) = 0.77$, nor were any of its interactions. Identification scores were 76% (7° LVF) and 82% (3.5° LVF) versus 77% (7° RVF) and 80% (3.5° RVF). Thus, the results did not indicate a laterality effect. There was no significant difference in speechreading performance between the 10 participants who responded with their right hand and the 10 who responded with their left hand, $F(1, 18) = 0.05, p = .82$.

Discussion

Although the participants' performance on the speechreading task decreased significantly with peripheral location of the face, no significant laterality effect was found. Thus, in contrast to hemispheric asymmetries reported previously (Baynes et al., 1994; Campbell, 1986; Campbell et al., 1986), Experiment 1 did not reveal hemispheric dominance.

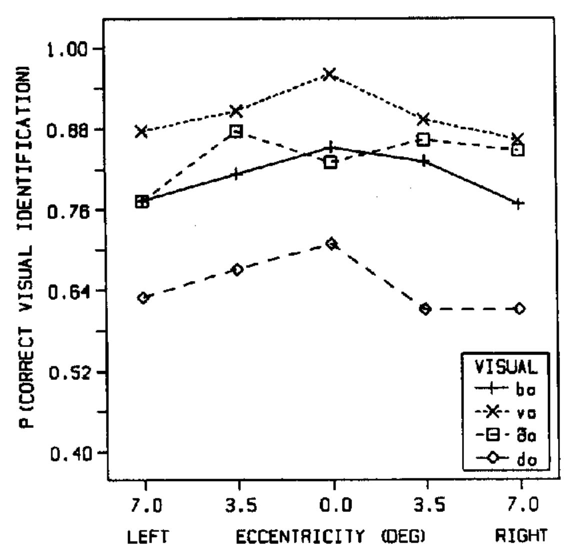

Figure 2. Proportion of correctly identified visual syllables as a function of eccentricity of the synthetic face. Averaged results are from 20 participants. Only those trials on which the participants had correctly performed the fixation task are included. LEFT = positions of the face in the left visual field (right hemisphere); RIGHT = positions in the right visual field (left hemisphere).

It is possible that visual field effects were obscured because the analyses included only the trials on which (dot) fixation was correct. Participants might have tended to fixate faces in one visual field or another when they did not fixate the dot correctly. To investigate this, we performed the same ANOVA on data from the trials on which the participant was incorrect on the fixation task. Speechreading performance averaged 76% (7° RVF), 81% (3.5° RVF), 83% (face in center), 80% (3.5° LVF), and 82% (7° LVF). The analysis revealed no significant face eccentricity effect, $F(1, 19) = 0.34$, $p = .57$, and no effect of visual field, $F(1, 19) = 1.51$, $p = .23$. These results suggest that the participants did not tend to fixate a specific visual field. Therefore, our conclusion about the outcome of Experiment 1 remained unchanged.

One possible explanation for finding no hemispheric differences is that the experimental task was too easy because of the relatively long duration of the syllables. Syllables spoken slowly do not require fast information extraction or processing, and therefore the potential dominance of one hemisphere cannot be observed. Some evidence for this possibility is that there was only a fairly small decrement of 4% in overall accuracy between the central and peripheral presentations of the test syllables. In Experiment 2, the task was made more difficult by using a faster speaking rate as well as face positions farther into the periphery.

Experiment 2: Visible Speech

Method

Participants. Twenty native speakers of American English participated in this experiment. They were students at the University of California, Santa Cruz. Their ages ranged from 18 to 22 years. They all reported being right-handed and having normal hearing and normal or corrected-to-normal vision. They received course credit or were paid for their participation.

Stimuli, design, and procedure. As in Experiment 1, the syllables /ba/, /va/, /ða/, and /da/ were used. Again, a synthetic face spoke these syllables without sound. However, a faster speaking rate was used. The rate was a factor of 1.50 faster than in Experiment 1. Stimulus duration now was about 700 ms. Pilot testing revealed that this rate decreased accurate identification but that speechreading was still possible. The lateral positions of the face were at visual angles of 9.1° (extreme position) and 4.6° (position halfway), as compared with 7° and 3.5° in Experiment 1. The size variations of the central dot appeared during the first 267 ms or in the last 267 ms of the speech stimulus. Otherwise, the experimental design and procedure were identical to those of Experiment 1.

Results

Analysis of the participants' performance on the dot-counting task (fixation task) revealed that they correctly detected the size changes of the central dot in 82% of the trials when no face was presented. Accuracy was 92% with no, 70% with one, and 83% with two dot-size changes. Table 1 shows the responses to the fixation task for the various conditions in Experiment 2.

When the participants also had to identify the visual speech stimulus simultaneously, they correctly detected the dot-size changes on 69% of the trials. Correct performance was 80% with no, 69% with one, and 58% with two dot-size changes. The effect of presentation condition was significant, $F(5, 90) = 21.84$, $p < .001$. The accuracy was 82% with no face, 70% and 67% (face at 9.1° and 4.6° LVF), 65% (face in center), and 62% and 68% (face at 4.6° and 9.1° RVF).

The data were analyzed in the same manner as in Experiment 1. Figure 3 shows speechreading performance for each location of the synthetic face only for those trials on which the participants correctly detected the dot-size changes (i.e., 69% of the trials). Excluding trials in which the participant was incorrect on the fixation task never reduced the number of observations for a participant in a given condition below 5. The average minimum number was about 13.

As can be seen in Figure 3, there was a large difference in accuracy across the four syllables, $F(3, 57) = 14.23$, $p < .001$. Performance averaged 67%, 81%, 77%, and 53% for the syllables /ba/, /va/, /ða/, and /da/, respectively, giving roughly the same order of difficulty as was found in Experiment 1. As in Experiment 1, speechreading performance significantly decreased when the face was presented farther into the periphery, $F(1, 19) = 22.74$, $p < .001$. In contrast to Experiment 1, the effect of visual field was significant, $F(1, 19) = 4.72$, $p = .041$. The interaction between syllable and visual field was also significant, $F(3, 57) = 2.81$, $p = .047$. As can be seen in Figure 3, the RVF advantage occurred only for the syllables /va/ and /ða/. There was no interaction between visual field and face eccentricity, $F(1, 19) = 1.32$, $p = .26$. Thus, the results of Experiment 2 showed a laterality effect in favor of the left hemisphere.

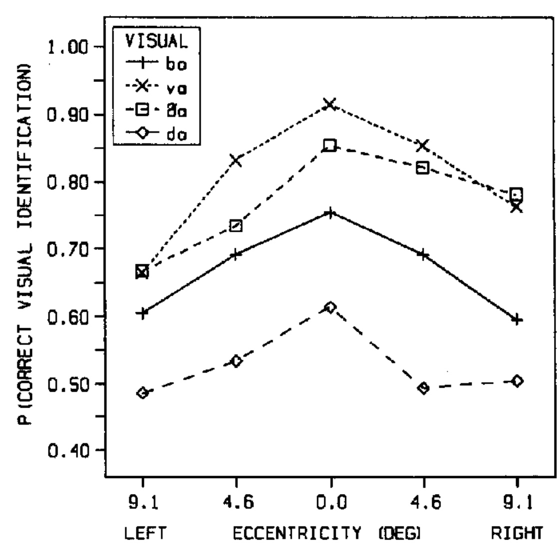

Figure 3. Proportion of correctly identified visual syllables as a function of eccentricity of the synthetic face. Averaged results are from 20 participants given correct performance on the fixation task. LEFT = left visual field (right hemisphere); RIGHT = right visual field (left hemisphere).

Having observed a hemispheric difference, it was important for us to know whether correct performance on the fixation task was critical for this outcome. To investigate this, we performed the same ANOVA as in Experiment 1 on the results from those trials on which the participant was incorrect on the fixation task. Given that errors on the fixation task were less frequent than correct judgments, there were some conditions with no observations for 3 of the 20 participants. The data from these participants were eliminated from the analysis. In contrast to the analysis when fixation performance was correct, there was no significant effect of visual field. Speechreading performance averaged about 67% in both visual fields. The analysis did reveal a significant face eccentricity effect, $F(1, 16) = 13.56$, $p = .002$. Thus, accuracy on the fixation task seemed essential to observing a visual field difference. Consistent with the results of Experiment 1, speechreading performance did not differ significantly between the 10 participants who responded with their left hand and the 10 who responded with their right hand ($F < 1$).

Discussion

The results of Experiment 2 show that a faster speaking rate and larger eccentricities produced a laterality effect in favor of the left hemisphere. One might question the effectiveness of the dot task in keeping the participants fixated on the fixation point. It is possible that participants were able to make eye movements to the face and identified it after they fixated on it. The laterality effect could have resulted from a larger number of movements when the face was presented to the right of fixation than to the left. If this were the case, then there should have been a larger laterality effect when participants were inaccurate on the fixation task. However, the results show no effect of visual field when participants were inaccurate on the fixation task. Thus, it seems unlikely that an eye-movement bias could account for the laterality effect.

Another possible explanation for the laterality effect is that participants might have moved their eyes after the first phase of the fixation task, that is, after they had accurately fixated on the fixation point for roughly 0.33 s. Participants would appear to require about 200 ms after the onset of presentation of the face to move their eyes from the dot to the face (Groner, 1988). Thus, it was of interest to determine what part of the test syllable was informative to assess how quickly the face had to be fixated for an accurate identification. If the critical information is presented early in the syllable, then it is less likely that an eye movement would be effective because the critical information in the syllable would no longer be present by the time it is fixated.

To locate the informative part of the syllable, we conducted a control experiment in which the initial 133, 200, 267, 333, 400, or 467 ms of the syllables were eliminated. Performance on these truncated syllables would indicate how much of the initial part of the syllable could be eliminated without preventing correct identification of the syllable. Six new participants were presented with the truncated versions of the four speech syllables. Each of the six truncated versions occurred 12 times, for a total of 288 trials. The syllables were presented at the center of the screen. The participants' task was to identify the speech syllables. There was no fixation task. Figure 4 shows speechreading performance as a function of the amount of visual information presented.

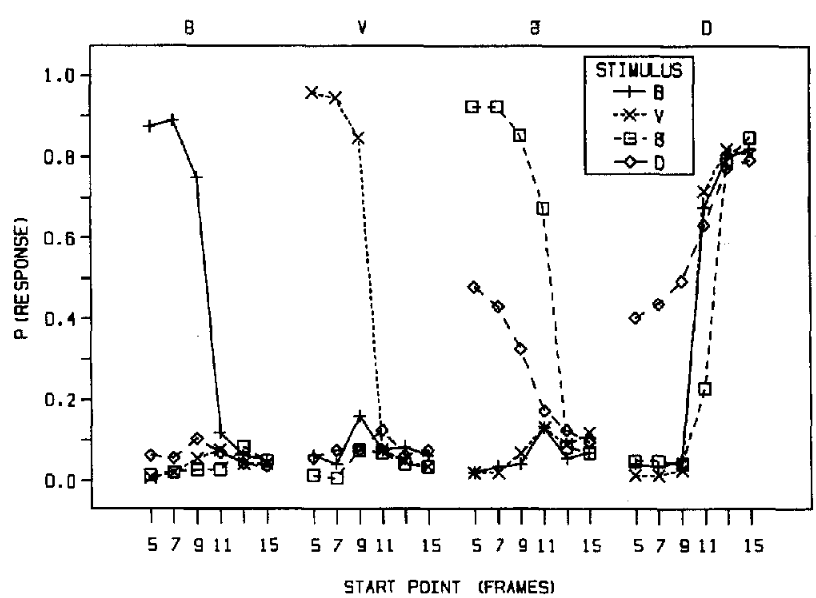

Figure 4. Proportion of responses to the visual syllables as a function of the amount of information presented. The abscissa indicates the onset of the stimuli in frames (1 frame = 33.3 ms) relative to the original ones in Experiment 2. The stimuli are indicated by various lines, and responses are placed over the figure. Averaged results are from 6 participants.

For /ba/, /va/, and /ða/, correct identification performance was high, roughly between 75% and 96%, when the part removed from the original syllable did not exceed the initial 267 ms. With longer parts removed, performance decreased dramatically to about 12% or less, except for /ða/, which remained high (67%) until more than 333 ms was removed. With these shorter syllables (i.e., more than 267 ms removed from onset), there was a striking increase in the proportion of /da/ responses (from 67% up to 85%). The fact that /d/ was the dominant response here is likely to be related to the resemblance between the visual information left in the truncated syllables and the mouth movements for a normal /da/. Correct identification for /da/ was relatively poor (about 40%) when less than the initial 267 ms were removed from the syllable. In these cases, /ða/ and /da/ responses were about equally likely and together made up more than 85% of the total responses. With longer parts removed from the stimulus onset, /da/ responses prevailed. These results indicate that the crucial information needed to distinguish the four consonants is in the initial 267 ms of the syllables. If we assume that participants had to maintain fixation for at least 267 ms after trial onset to perform the fixation task correctly and that 200 ms would be needed to move their eyes, they could not arrive at the location of the face before the informative part of the syllable had already occurred. This control experiment demonstrated that it is unlikely that the participants could have moved their eyes and recognized the syllables correctly on the basis of information obtained after the syllable was fixated. It can thus be concluded that performance in Experiment 2 was determined by peripherally presented information and therefore addressed the issue of hemispheric specialization.

A possible conclusion from Experiment 2 is that speechreading is a linguistic task that requires primarily left-hemisphere processing. The finding that speechreading skills are lateralized to the left hemisphere is consistent with the results reported by Campbell et al. (1986) involving patients. However, our results are not consistent with those of Campbell (1986) and Baynes et al. (1994). They found right-hemisphere dominance when normal participants were presented with faces producing speech sounds.

Does this mean that in certain cases the left hemisphere and in others the right hemisphere is specialized for extracting linguistic information from faces? One important difference between our Experiment 2 and that of Campbell's (1986) is that we used moving faces, whereas Campbell used still pictures. We assume that this changed how the visual speech stimuli were processed. Mouth movements occur during speech. If the left hemisphere is more involved in the processing of temporal sequences and the right hemisphere in the processing of static images, then the two sets of results are not necessarily contradictory (Bradshaw & Nettleton, 1981). However, the results reported by Baynes et al. (1994) are not consistent with this static–moving hypothesis. They found a right-hemisphere advantage for moving faces.

The finding of a right-hemisphere dominance in the studies of Campbell (1986) and Baynes et al. (1994) might be related to the fact that the face was presented together with speech sounds. Speechreading (without sound) might give rise to a different pattern of hemispheric dominance than the integration of visual and auditory speech. It remains unclear, however, why right-hemisphere presentation of the face led to better performance when participants identified audiovisual speech information, particularly given the general finding of a left-hemisphere advantage for auditory speech. Furthermore, the study by Campbell et al. (1990) provides evidence against a hypothesis that the identification of audiovisually presented speech information engages primarily right-hemisphere processes. Their study indicated that the left hemisphere played the dominant role in the integration of seen and heard speech. More recently, Johnson and Rosenblum (1996) found similar left-hemisphere advantages for both speechreading without sound and the integration of audible and visible speech.

A third hypothesis is that hemispheric asymmetries do not depend so much on the linguistic content of the visual stimulus but mainly on its visuospatial properties (Hellige, 1993). It has been shown that certain changes in characteristics of a visual stimulus can reduce or even reverse laterality effects (Sergent, 1984; Sergent & Bindra, 1981). For example, shortening the stimulus exposure duration can change the finding of no hemispheric difference (Moscovitch et al., 1976) into a right-hemisphere advantage (see Sergent & Bindra, 1981). If the left hemisphere is more involved in motion perception, then the left-hemisphere advantage could have been attributable to the dynamic character of the test syllables. Thus, we do not know if left-hemisphere dominance is due to the dynamic or linguistic character of the test syllables. To test between these hypotheses, we asked whether different results would be obtained when visual nonlinguistic facial movements were presented (i.e., mouth movements that do not carry phonetic information). If these nonlinguistic mouth movements provide a hemispheric advantage similar to the linguistic gestures, this would suggest that it is dynamic, not linguistic, information that is critical for a left-hemisphere dominance.

Experiment 3: Visible Nonspeech

Method

Participants. Twenty native speakers of American English participated in this experiment. They were students at the University of California, Santa Cruz. Their ages ranged from 18 to 23 years. They all reported being right-handed and having normal hearing and normal or corrected-to-normal vision. They received course credit or were paid for their participation.

Stimuli, design, and procedure. The synthetic face produced mouth movements that resembled a kiss, a snarl, a tongue protrusion, or a cough. These four productions were chosen to be roughly the nonspeech counterparts for /ba/, /va/, /ða/, and /da/, respectively, with regard to the position and movements of the articulators. The duration of the four stimuli, face positions, and all other experimental details were similar to those in Experiment 2. The kiss involved a closing and protrusion of the lips, some jaw rotation, and the corners of the mouth coming together. The snarl involved a raising of the upper lip and a lowering of the lower lip, consequent bearing of the teeth, and some jaw rotation. There was extreme tongue movement between the teeth for the tongue protrusion. Finally, the mouth corners moved centrally, and there was jaw rotation for the cough. Although these nonspeech movements were made to be analogous to the speech items, they were seen as nonspeech by the participants. The only nonspeech gesture that came close to speech was the cough because it was relatively nondistinctive and could resemble a nondistinctive syllable such as /da/.

Results

When no face was presented, participants correctly detected the size variations of the central dot on 87% of the trials. The percentage of correct responses reached 95% with no, 77% with one, and 90% with two dot-size changes. Table 1 shows the fixation responses for the various conditions in Experiment 3.

When the participants had to identify the visual stimulus simultaneously with performance of the fixation task, they correctly detected the dot-size changes in 74% of the trials. Correct fixation was 87% with no, 71% with one, and 63% with two dot-size changes. The effect of presentation condition on performance of the fixation task was significant, $F(5, 90) = 25.65, p < .001$. The percentages were 74% with no face, 71% and 71% (face at 9.1° and 4.6° LVF), 68% (face in center), and 69% and 73% (face at 4.6° and 9.1° RVF).

As in Experiments 1 and 2, for calculating participants' identification of the nonspeech mouth movements, we excluded those trials on which they did not correctly perform the fixation task. Figure 5 shows identification of each of the nonspeech mouth movements for each location of the synthetic face. Excluding trials in which the participant was incorrect on the fixation task never reduced the number of observations for a participant in a given condition below 6. The average minimum number was about 16.

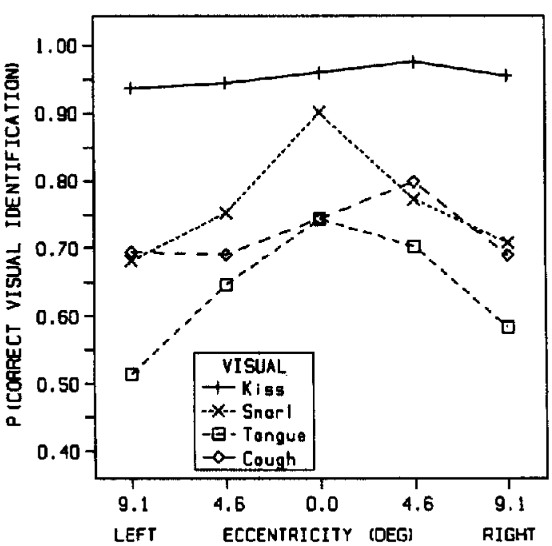

Figure 5. Proportion of correctly identified visual nonspeech stimuli as a function of eccentricity of the synthetic face. Averaged results are from 20 participants given correct performance on the fixation task. LEFT = left visual field (right hemisphere); RIGHT = right visual field (left hemisphere).

To investigate whether the processing of nonspeech mouth movements was lateralized, we evaluated participants' identification of each of the test items as a function of VF and eccentricity. There was a significant difference in identification performance of the four nonspeech stimuli, $F(3, 57) = 16.22, p < .001$. Performance averaged 95%, 76%, 64%, and 72% for the kiss, snarl, tongue protrusion, and cough, respectively. Identification of the kiss was high. Contrary to our expectations, tongue protrusion (highly visible tongue) was identified less accurately than the cough.

Performance decreased with increases in distance into the periphery, $F(1, 19) = 39.64, p = .001$. As in Experiment 2 with speech stimuli, the effect of visual field was significant, $F(1, 19) = 16.85, p < .001$, with performance higher with RVF presentation. The interaction between nonspeech stimulus and distance was significant, $F(3, 57) = 4.01, p = .012$, reflecting the finding that the decrease in performance with increasing eccentricity was largest for the tongue and smallest for the kiss. The interaction between nonspeech stimulus and visual field was not significant ($F < 1$).

As in the speech case, it was important to know whether correct performance on the fixation task was critical for this outcome. The ANOVA on the results of 20 participants from those trials in which they were incorrect on the fixation task showed no difference between the LVF and RVF, $F(1, 19) = 0.21$. Speechreading performance averaged about 74% in both visual fields. Although performance was somewhat better on the less peripheral displays, this result failed to reach statistical significance, $F(1, 19) = 3.09, p = .09$. As in the previous comparable speech task (Experiment 2), accuracy on the fixation task seemed to be essential to observing a visual field difference for nonspeech. Performance on the nonspeech stimuli did not differ significantly between the participants who responded with their left hand and those who responded with their right hand ($F < 1$).

Discussion

A left-hemisphere advantage was found with both speech (Experiment 2) and nonspeech (Experiment 3) facial movements. Figure 6 plots the average performance for speech and nonspeech as a function of spatial location. An ANOVA was carried out as previously but now with the addition of speech versus nonspeech as a factor. The effects of visual field and eccentricity remained significant and did not interact with stimulus type. The similarity in the results suggest that hemispheric asymmetries depend on visual (physical) properties rather than on the linguistic content of the stimulus. It seems that the dynamic aspect of the articulatory movements caused an increase in the LH participation.

Jordan, Patching, and Milner (in press) measured fixation behavior using an eye-tracking device. When their participants were instructed to fixate on a stimulus, the fixations were somewhat inaccurate with somewhat more fixations to the right than to the left of the fixation stimulus. However, the fixation itself was fairly accurate (within 1° of visual angle). In fact, about 90% of the fixations were within 0.5° of the fixation point. Although this small difference might be

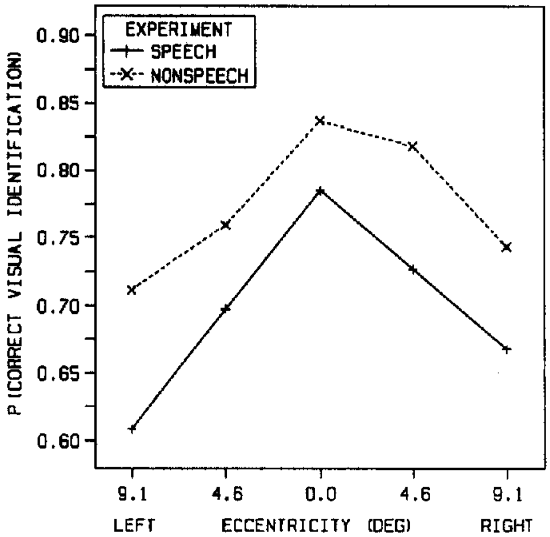

Figure 6. Proportion of correctly identified visual speech and nonspeech stimuli as a function of eccentricity of the synthetic face. Averaged results are from 20 participants in each group given correct performance on the fixation task. LEFT = left visual field (right hemisphere); RIGHT = right visual field (left hemisphere).

important for small displays, it would not have been in our task, in which some displays were presented more than 9° away from the fixation. Jordan et al. also found that participants showed a bias to fixate to the right of the fixation point. They found an overall bias of about 0.5° to the right. The average results in Figure 6 show that a slight 0.5° bias on some of the trials could not account for the asymmetry that was observed. Performance on the display presented 9.1° to the right of the fixation point was about equal to performance on the display presented 4.6° to the left. Thus, the advantage for the RVF (left hemisphere) does not seem to have been caused by a slight fixation bias.

General Discussion

Laterality effects for mouth movements were investigated in three experiments. For identification of speech mouth movements, no hemispheric differences were found in Experiment 1, but changes in speaking rate and face eccentricities led to a small left-hemisphere advantage in Experiment 2. Experiment 3, in which nonspeech mouth movements were studied, also showed a left-hemisphere superiority. Our conclusion is that this superiority is related to the better performance of the left hemisphere in processing the dynamic properties (the temporal frequencies) of the facial movements rather than on their linguistic content.

Concerns might be raised about the nature of the fixation task. It might be interpreted as involving the detection of movement (or better, change) in a dot, just as the main task involved interpretation of (mouth) movements. One might argue that any apparent hemispheric specialization may reflect attentional priming of movement detection via the fixation task. We do not think that priming occurred. Changes in size are different from actual movements in that the dot stayed in one position. This was supported by the participants' reports of perceiving the size changes as flickers.

Although we eliminated this potential confounding with our symmetrical animated face, the asymmetry in speech production raises an intriguing explanation for the small left-hemisphere advantage. Viewers might have learned to attend more to the right side of a talker's face (i.e., the side seen on the perceiver's left). With peripheral viewing, then, participants would be attending farther in the periphery when the face was presented in the LVF than in the RVF. Given that the attended part of the face would be closer to foveal viewing when it was presented in the RVF, this would necessarily produce an RVF (left-hemisphere) advantage.

More knowledge is required to understand how hemispheric contributions vary with changes in stimulus characteristics. As we have seen, the visual stimuli of the studies discussed earlier differed considerably from each other. Therefore, uncontrolled aspects of the visual displays might have contributed to the hemispheric asymmetries that were observed. For example, the fact that the visual stimuli were static pictures of faces might account for the failure of finding a left-hemisphere dominance in the study by Campbell (1986). These static images might have caused increased participation of the right hemisphere.

Further research on hemispheric asymmetries might include studies using a variant of a dichotic listening method (Broadbent, 1954; Kimura, 1961). Dichotic listening is the simultaneous presentation of different stimuli to the left and right ear. Applying this technique to the visual case, a face would appear in both visual fields producing different mouth movements. Participants could be instructed to identify the stimuli in both visual fields or the stimulus in one specific visual field. Laterality effects might show up stronger with the simultaneous presentation of two stimuli.

One promising approach is to measure brain activity while participants are performing both linguistic and nonlinguistic tasks using positron emission tomography, magnetic resonance imaging, or magnetoencephalography. These techniques represent a more direct way to localize cortical activity than the methods used in behavioral studies. Results should show whether different cortical areas are involved in processing linguistic and nonlinguistic facial movements. In addition, the magnetoencephalography technique might be able to provide information about the time course of processing (Sams et al., 1991; Sams & Levänen, 1996).

To summarize, the present study showed a left-hemisphere advantage for processing visual information from both linguistically meaningful and nonlinguistic mouth movements. Speechreading involves various skills ranging from visual shape recognition to linguistic interpretation. One might argue that researchers have merely emphasized linguistic aspects of speechreading, aspects that most likely tap into left-hemisphere resources. However, our results with nonspeech mouth movements indicate that linguistic aspects alone cannot be responsible for a left-hemisphere advantage. A strong conclusion would be that the dynamic aspect of the mouth movements is sufficient to produce increased participation of the left hemisphere.

References

Baynes, K., Funnell, M. G., & Fowler, C. A. (1994). Hemispheric contributions to the integration of visual and auditory information in speech perception. *Perception & Psychophysics, 55,* 633–641.

Bellugi, U., & Klima, E. S. (1993). Language, spatial cognition and neuronal plasticity [Abstract]. *Proceedings of the 16th Annual Meeting of the European Neuroscience Association* (Madrid, Spain), p. 106.

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 & E. Kampp (Eds.), *Visual and audio-visual perception of speech* (Scandinavian Audiology Supplement 4, pp. 181–209). Stockholm, Sweden: Almquist & Wiksell.

Bradshaw, J. L., & Nettleton, N. C. (1981). The nature of hemispheric specialization in man. *Behavioral and Brain Sciences, 4,* 51–91.

Breeuwer, M., & Plomp, R. (1984). Speechreading supplemented with frequency-selective sound-pressure information. *Journal of the Acoustical Society of America, 76,* 686–691.

Broadbent, D. E. (1954). The role of auditory localization in attention and memory span. *Journal of Experimental Psychology, 47,* 191–196.

Bryden, M. P. (1965). Tachistoscopic recognition, handedness, and cerebral dominance. *Neuropsychologia, 3,* 1–8.

Bryden, M. P. (1982). *Laterality: Functional asymmetry in the intact brain.* New York: Academic Press.

Campbell, R. (1986). The lateralization of lip-read sounds: A first look. *Brain and Cognition, 5,* 1–21.

Campbell, R. (1987a). The cerebral lateralization of lip-reading. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 215–226). London: Erlbaum.

Campbell, R. (1987b). Lip-reading and immediate memory processes or on thinking impure thoughts. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 243–255). London: Erlbaum.

Campbell, R., De Gelder, B., & De Haan, E. (1996). The lateralization of lip-reading: A second look. *Neuropsychologia, 34,* 1235–1240.

Campbell, R., & Dodd, B. (1980). Hearing by eye. *Quarterly Journal of Experimental Psychology, 32,* 85–99.

Campbell, R., & Dodd, B. (1982). Some suffix effects on lipread lists. *Canadian Journal of Psychology, 36,* 509–515.

Campbell, R., Garwood, J., Franklin, S., Howard, D., Landis, T., & Regard, M. (1990). Neuropsychological studies of auditory-visual fusion illusions: Four case studies and their implications. *Neuropsychologia, 28,* 787–802.

Campbell, R., Landis, T., & Regard, M. (1986). Face recognition and lipreading: A neurological dissociation. *Brain, 109,* 509–521.

Cohen, M. M., & Massaro, D. W. (1990). Synthesis of visible speech. *Behavioral Research Methods and Instrumentation, 22,* 260–263.

Cohen, M. M., & Massaro, D. W. (1993). Modeling coarticulation in synthetic visual speech. In N. M. Thalmann & D. Thalmann (Eds.), *Models and techniques in computer animation* (pp. 139–156). Tokyo: Springer-Verlag.

Diesch, E. (1995). Left and right hemifield advantages of fusions and combinations in audiovisual speech perception. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 48A,* 320–333.

Dodd, B. (1979). Lipreading in infants: Attention to speech presented in and out of synchrony. *Cognitive Psychology, 11,* 478–484.

Gathercole, S. E. (1987). Lip-reading: Implications for theories of short-term memory. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 227–241). London: Erlbaum.

Grant, K. W., & Walden, B. E. (1995). Predicting auditory-visual speech recognition in hearing-impaired listeners. *Proceedings of the XIIIth International Congress of Phonetic Sciences, 3,* 122–129.

Graves, R., Goodglass, H., & Landis, T. (1982). Mouth asymmetry during spontaneous speech. *Neuropsychologia, 20,* 371–381.

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.

Groner, R. (1988). In G. Luer, U. Lass, & J. Shallo-Hoffmann (Eds.), *Eye movement research: Physiological and psychological aspects* (pp. 295–319). Toronto: Hogrefe.

Hellige, J. B. (1993). *Hemispheric asymmetry: What's right and what's left.* Cambridge, MA: Harvard University Press.

Johnson, J. A., & Rosenblum, L. D. (1996, December). *Hemispheric differences in perceiving and integrating dynamic visual speech information.* Poster presented at the 132nd Meeting of the Acoustical Society of America, Honolulu, HI.

Jordan, T. R., Patching, G. R., & Milner, A. D. (in press). Ensuring central fixations in divided visual field studies of cerebral asymmetry: Instructions are not enough. *Quarterly Journal of Experimental Psychology.*

Kimura, D. (1961). Cerebral dominance and the perception of verbal stimuli. *Canadian Journal of Psychology, 15,* 166–171.

Kimura, D. (1967). Functional asymmetry of the brain in dichotic listening. *Cortex, 3,* 163–178.

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science, 218,* 1138–1141.

Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry.* Hillsdale, NJ: Erlbaum.

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

Mishkin, M., & Forgays, J. D. (1952). Word recognition as a function of retinal locus. *Journal of Experimental Psychology, 43,* 43–48.

Moscovitch, M., Scullion, D., & Christie, D. (1976). Early vs. late stages of processing and their relation to functional hemispheric differences in face recognition. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 401–416.

Poizner, H., Bellugi, U., & Klima, E. S. (1990). Biological foundations of language: Clues from sign language. *Annual Review of Neuroscience, 13,* 283–307.

Reisberg, D., McLean, J., & Goldfield, A. (1987). Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 97–113). London: Erlbaum.

Sams, M., Aulanko, R., Hämäläinen, M., Hari, R., Lounasmaa, O. V., Lu, S., & Simola, J. (1991). Seeing speech: Visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters, 127,* 141–145.

Sams, M., & Levänen, S. (1996). Where and when are the heard and seen speech integrated: Magnetoencephalographical (MEG) studies. In D. G. Stork & M. E. Hennecke (Eds.), *Speechreading by humans and machines* (pp. 233–238). Berlin: Springer-Verlag.

Sergent, J. (1984). Configural processing of faces in the left and the right cerebral hemispheres. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 554–572.

Sergent, J. (1995). Visualizing the working cerebral hemispheres. In F. L. Kitterle (Ed.), *Hemispheric communication: Mechanisms and models* (pp. 189–211). Hillsdale, NJ: Erlbaum.

Sergent, J., & Bindra, D. (1981). Differential hemispheric processing of faces: Methodological considerations and reinterpretation. *Psychological Bulletin, 89,* 541–544.

Smeele, P. M. T. (1994). *Perceiving speech: Integrating auditory and visual speech.* Unpublished doctoral dissertation, Delft University of Technology, Delft, The Netherlands.

Smeele, P. M. T., & Sittig, A. C. (1991a). The contribution of vision to speech perception. *Proceedings of the Second European Conference on Speech Communication and Technology* (Eurospeech 91, Genova, Italy), pp. 1495–1497.

Smeele, P. M. T., & Sittig, A. C. (1991b, August). *Effects of desynchronization of vision and speech on the perception of speech: Preliminary results.* Paper presented at the CCITT (Communitee Consulative Internationale de Telecommunication et Telegraphic) Brazil Conference 91 (Contribution D. 81).

Söderfeldt, B., Rönnberg, J., & Risberg, J. (1994). Regional cerebral blood flow in sign language users. *Brain and Language, 46,* 59–68.

Studdert-Kennedy, M., & Shankweiler, D. (1970). Hemispheric specialization for speech perception. *Journal of the Acoustical Society of America, 48,* 579–594.

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

Summerfield, A. Q. (1979). Use of visual information in phonetic perception. *Phonetica, 36,* 314–331.

van Son, N. (1993). *Speech processing strategies for the profoundly hearing impaired.* Unpublished doctoral dissertation, Utrecht University, Utrecht, The Netherlands.

Vroomen, J. H. M. (1992). *Hearing voices and seeing lips: Investigations in the psychology of lipreading.* Unpublished doctoral dissertation, Katholieke Universiteit Brabant, Brabant, The Netherlands.

Wolf, M. E., & Goodale, M. A. (1987). Oral asymmetries during verbal and nonverbal movements of the mouth. *Neuropsychologia, 25,* 375–396.

Received November 7, 1994
Revision received June 13, 1997
Accepted July 10, 1997 ■