# Face and mouth inversion effects on visual and audiovisual speech perception

Lawrence D. Rosenblum, Deborah A. Yakel
University of California, Riverside
Kerry P. Green
University of Arizona

Published as: Rosenblum, L.D., Yakel, D.A., & Greene, K.G. (2000). Face and mouth inversion affects on visual and audiovisual speech perception. Journal of Experimental Psychology: Human Perception and Performance. 26(3), 806-819.

Abstract

Three experiments examined whether image manipulations known to disrupt face perception also disrupt visual speech perception. Research has shown that an upright face with an inverted mouth looks strikingly grotesque while an inverted face and an inverted face containing an upright mouth are perceived as looking relatively normal. The current study examined whether a similar sensitivity to an upright facial context plays a role in visual speech perception. Visual and audiovisual syllable identification tasks were tested under four presentation conditions: upright face-upright mouth; inverted face-inverted mouth; inverted face-upright mouth; upright face - inverted mouth. Results revealed that for some visual syllables only the upright face-inverted mouth image disrupted identification. These results suggest that an upright facial context can play a role in visual speech perception. A follow-up experiment testing upright and inverted isolated mouths supported this conclusion.

There is mounting evidence that visual speech perception is an important component of the general speech perception process. While it is clear that speechreading (lipreading) can be useful for the hearing impaired, visual speech is also used by individuals with good hearing when faced with a noisy environment (e.g., MacLeod & Summerfield, 1987). Visual speech information is also useful to listeners attempting to understand a speaker with a heavy foreign accent, or when speech conveys complicated subject matter (Reisberg, McLean, & Goldfield, 1987). Access to visual speech is also necessary for normal speech development (Mills, 1987). Finally, compelling evidence for the importance of visual speech is evidenced in the McGurk effect (McGurk and MacDonald, 1976). In this effect, normal listeners reporting hearing audiovisually discrepant syllables as some combination of the auditory and visual syllables (e.g., auditory /ba/ + visual /ga/ are perceived as /da/) or as a syllable dominated by the visual segments (e.g., auditory /ba/ + visual /va/ are perceived as /va/). The McGurk effect demonstrates the importance of visual speech information by showing that its integration is automatic and mandatory for most observers.

The phenomenon of visual speech perception also poses some interesting theoretical questions. Although a good deal is known about auditory speech perception and the visual recognition of faces, relatively little is known about how perceivers extract speech information from seeing a face. One question concerns the degree to which information and/or operations for face recognition might be similar to those involved in visual speech. While both functions clearly make use of facial information, there is neuropsychological evidence for a separation of the functions (e.g., Bruce and Young, 1986; Campbell, Landis, and Regard, 1986; Ellis, 1989; but see Baynes, Funnell, & Fowler, 1994; Campbell, 1986; Damasio, 1989; Diesch, 1995; Hillger & Koenig, 1991; Sergent, 1982; and Rosenblum and Saldaña, 1998). In addition, the speech and face perception functions have been discussed as paradigmatic examples of domain-specific, informationally-encapsulated cognitive modules (e.g., Fodor, 1985; for speech Liberman and Mattingly, 1985) suggesting a clear division between the functions.

However, a series of recent behavioral studies has revealed that a specialized type of facial information might be used for both functions (e.g., Bertelson, Vroomen, Wiegeraad, & de Gelder, 1994; Green, 1994; Jordan & Bevan, 1997; Massaro & Cohen, 1996). It has long been known that image inversion is much more detrimental to face recognition than it is to recognition of non-face images (e.g., houses) (e.g., Carey and Diamond, 1977; Scapinello and Yarmey, 1970; Yin, 1969; and see Valentine, 1988 for a review). This finding has been interpreted as support that an upright facial context facilitates recognition in a manner specialized for faces (e.g., Yin, 1969; however, see Valentine, 1988, and Diamond and Carey, 1986). Related findings reveal that an upright face context can facilitate discrimination of facial features (Rhodes, Brake, and Atkinson, 1993; Tanaka and Farah,1993).

Does an upright facial context also facilitate visual speech perception? Recently, a number of researchers have addressed this question by testing performance with inverted face images in both speechreading and audiovisual speech perception (McGurk effects) (e.g., Bertelson, Vroomen, Wiegeraad, & de Gelder, 1994; Green, 1994; Jordan & Bevan, 1997; Massaro & Cohen, 1996). Each of these papers report some inversion disruptions of speechreading accuracy and/or decreases in the McGurk effect. These results have been considered evidence that common specialized information is used for face and visual speech perception (e.g., Green, 1994; Massaro & Cohen, 1996). However, there are many reasons why inverting faces disrupts visual speech perception. For example, inverting a speaking face also violates the natural (gravitational) dynamical influences on the face which are known to be important for accurate event identification (cf. Green, 1994; Bingham, Rosenblum, and Schmidt, 1995).

Given this uncertainty regarding the source of inversion disruptions, it may pay to turn to other face perception effects for exploring whether upright facial context information (beyond the mouth) might be used for visual speech perception. In the 'Margaret Thatcher effect', an inverted face and an inverted face containing upright lips and eyes are both perceived as looking normal, but an upright face with inverted lips and eyes looks strikingly grotesque (Bartlett and Searcy, 1993; Parks, 1983; Parks, Coss, and Coss, 1985; Sjoberg and Windes, 1992; Thompson, 1980; Valentine and Bruce, 1985). Explanations of why inverted features look gruesome have been offered (e.g., Valentine, 1988) and will be discussed below. However, the more salient aspect of the effect is that mis-oriented features are disruptive only when presented in the context of an upright face. Similar results have been obtained with facial distortions involving vertical elongation between upright features (Bartlett & Searcy, 1993) and with face matching tasks. In addition, recent evidence shows that the Thatcher manipulation can inhibit *recognition* of upright famous faces but has no comparable effect when it is applied to inverted faces (Lander, Rosenblum, & Bruce, in preparation). These effects demonstrate that there is some special characteristic of the upright face context, not present for inverted faces, that heightens sensitivity to facial distortions. In this sense, the Thatcher-type effects extend the facial inversion effects in supporting a facilitatory property of upright face perception.

Thatcher-type effects provide a novel way to examine visual speech inversion effects. If visual speech and face perception share upright facial-context facilitation, then Thatcher-type effects should occur for visual speech perception. Thus, for a Thatcher-type stimulus (upright face with inverted features), the upright facial context could heighten sensitivity to feature inversion, which could in turn, disrupt visual speech perception. Although it would not be surprising to find that inverting facial features—particularly the mouth—disrupts visual speech perception, the true test is whether featural inversion is more disruptive in the context of an upright face. It is not clear how other interpretations, for example, one based on dynamical/gravitational disruptions, could account for such results.

The question of whether visual speech and face recognition are facilitated when the face is upright was investigated by examining the influence of image and feature inversion on visual (and audiovisual) speech perception. The experiments differ from previous research by utilizing the Thatcher effect to test visual speech perception with stimuli comprised of upright and inverted mouths and faces[1]. Experiments 1 and 2 test four image conditions: an upright face with upright mouth; an inverted face with an inverted mouth; an inverted face with an upright mouth; and an

upright face with inverted mouth (Thatcher stimulus) (See Figure 1). Experiment 3 further explores this question by testing the effects of presenting an isolated mouth in upright and inverted orientations.

Throughout, performance on both speechreading and on audiovisual speech perception will be evaluated with the different image conditions. Evaluating performance on both tasks will permit for comparisons with studies that have tested facial inversion effects (Bertelson, et al., 1994; Green, 1994; Jordan & Bevan, 1997; Massaro & Cohen, 1996). More importantly, incongruent audiovisual (McGurk effect) stimuli provide a more sensitive test of how visual speech information is used. Past research has shown that McGurk-type integration tasks are more easily disrupted by image manipulations than are straight speechreading tasks or tasks with audiovisually congruent stimuli (e.g., Jordan and Bevan, 1997; McGrath, 1985; Rosenblum and Saldaña, 1996; Summerfield, et al., 1989). Relatedly, other researchers have considered the audiovisual task a more robust test of the use of visual speech information in that its influence is observed without asking for explicit judgments about the visual information itself (subjects are asked what they *hear*). Whereas audiovisually congruent speech can provide similar task constraints, incongruent McGurk-type stimuli provide an easy way to observe whether visual speech influences what listeners report hearing. In this sense, a McGurk-type task may ensure that perceptual functions are being examined rather than some type of post-perceptual, decision/problem solving processes which may underlie general speechreading (e.g., Liberman and Mattingly, 1985; McGrath, 1985; Summerfield, MacLeod, McGrath, and Brooke, 1989; Rosenblum and Saldaña, 1992).

# EXPERIMENT 1

Experiment 1 tests speechreading and audiovisual integration using the four mouth-face inversion conditions outlined above. If the sensitivity to features afforded by the upright facial context affects visual speech perception, then the distortions of the Thatcher stimulus (upright face with inverted mouth) should be more disruptive than the other image conditions. While the motivating question concerns the Thatcher stimulus manipulation, testing these image conditions also allows us to address the facial inversion issue. By comparing performance with the upright face-upright mouth image to performance with the inverted face-inverted mouth image, we can examine whether face inversion affects visual speech perception with our stimuli (cf. Jordan and Bevan, 1997).

For Experiment 1, the syllables tested included audio and visual /ba/ and /va/. These syllables were chosen because previous research in our laboratory has shown that in a McGurk stimulus context, an audio /ba/ - visual /va/ is heard as /va/ up to 98% of the time (Rosenblum and Saldaña, 1992; 1996; Saldaña and Rosenblum, 1993; and 1994).

## Method

### Participants

Fourteen undergraduates at the University of California, Riverside, were compensated $5 for their participation in the experiment. All reported normal or corrected vision, good hearing, and were native speakers of English.

### Stimuli

The stimuli were prepared by recording a speaker in a fully lit room with no alteration to the speaker's face. A Panasonic PVS350 camcorder and SM57 microphone were used to record the initial audiovisual tape. The speaker was seated five feet in front of the camera. His head was placed in a wire head brace to inhibit movement. The camera was centered so that the recorded image consisted of the middle of the speaker's forehead to about one inch below his chin. The speaker was recorded articulating the syllables /ba/, and /va/, eight times each.

Using a MacIIci computer and a Panasonic 7500A video recorder, one visual /ba/ and one visual /va/ were transferred to the computer. These tokens were edited so that each was three seconds (90 frames) long. Four tokens with different orientations were created for each of the /ba/ and /va/ visual tokens. These consisted of an upright face with upright mouth, an inverted face

with an inverted mouth, an inverted face with an upright mouth and an upright face with inverted mouth (Thatcher stimulus) (see Figure 1).

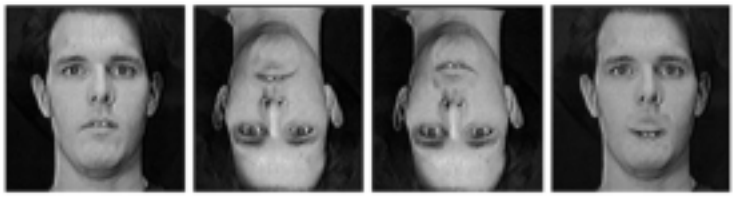

Figure 1. Photographs of the four face-mouth orientation conditions used as stimuli for all experiments. From left to right: Upright Face – Upright Mouth; Inverted Face – Inverted Mouth; Inverted Face - Inverted Mouth; Upright Face – Inverted Mouth (Thatcher condition). The speaker is shown initiating a /va/ syllable.

An NIH Image software program was used to modify each frame to create the inverted face and inverted mouth tokens. Using this program, a box-shaped region was isolated around the mouth approximately 5 mm from the corners of the upper and lower lips. The size of the box for each frame varied depending on the movement of the mouth but the parameters mentioned above were generally maintained for all frames of each token. The box was then inverted vertically to create an upright face-inverted mouth for all 90 frames of both the /ba/ and /va/ tokens. The full frames of the upright face-inverted mouth tokens were then inverted vertically to create the inverted face-upright mouth tokens. Full frame inversion was also performed with the frames of the upright face-upright mouth tokens to create the inverted face-inverted mouth tokens. The eight visual tokens were then transferred back onto a videotape using the MacIIci computer and the Panasonic 7500A video recorder.

An AMC 486/33 computer and two Panasonic 7500A video recorders were used for dubbing the audiovisual tokens. A good exemplar of an audio /ba/ and an audio /va/ which had been sampled for a previous experiment (Rosenblum & Saldaña, 1992), were used to dub the visual tokens. These audio tokens matched the videotaped speaker's visible articulation and were of good quality. These /ba/ and /va/ tokens were dubbed synchronously onto the eight visual /ba/ and /va/ tokens so that acoustic onset matched the visible release. In order to ensure that the audiovisual synchrony was subjectively equivalent for all eight tokens, two procedures were implemented. First, during the dubbing procedure itself, one of the two video monitors was inverted so that all of the tokens could be viewed with an upright and inverted mouth. In addition, after dubbing, five pilot subjects judged the synchrony of all the stimuli as equally good.

The audiovisual combination conditions for all four image orientation conditions consisted of a congruent audio /ba/ - visual /ba/, a congruent audio /va/ - visual /va/, an incongruent audio /va/ - visual /ba/, and an incongruent audio /ba/ - visual /va/. This produced a total of 16 audiovisual stimuli.

The presentation tape consisted of 260 tokens. The 16 audiovisual stimuli, eight visual-alone stimuli (/ba/ and /va/ for the four image orientation conditions), and two audio-alone stimuli (/ba/ and /va/) were all presented 10 times each. All 260 tokens were randomized and then separated into six blocks of 40 trials and one block of 20 trials. All tokens were separated by a 3 second ISI and the blocks were separated by 12 seconds.

Procedure

Subjects were run in groups of one to three, and were seated at a table five feet in front of a Panasonic 21 inch video monitor. The audio stimuli were presented through two Infinity SL speakers positioned directly beneath the monitor. The only source of illumination was the television monitor and one small light which was focused away from the monitor and illuminated subjects' response sheets.

Subjects were told that syllable stimuli would be presented auditorily, visually, and audiovisually in random order. They were also informed that the face and/or mouth of the speaker would be inverted on some of the trials. For the audio and audiovisual trials, subjects were asked to write down the first letter of whatever syllable they *heard* after each token was presented and then look back up to the monitor for the next presentation (a free response task). For the visual trials they were instructed to write down the first letter of the syllable they saw the speaker produce. An experimenter was present to make certain that all subjects were watching the presentation of each token and that they were not looking at other subjects' response sheets.

Subjects were given a five minute break after the first four blocks were presented. The experiment lasted approximately 40 minutes for each subject.

## Results and Discussion

The mean percentages of incorrect responses for each token and image type are shown in Table 1. For the visual alone stimuli, an incorrect response was recorded whenever a subject responded with an initial consonant other than the one presented visually. For both the auditory alone and audiovisual trials, an incorrect response was recorded whenever a subject responded with an initial consonant other than the consonant presented on the audio portion of the tape. In the case of incongruent stimuli, a visual influence is demonstrated if the percentage of incorrect responses is higher than in the audio alone condition and the responses are based on the visual information presented. In the present experiment, all incorrect responses to the audiovisual stimuli consisted of the visual token being reported rather than the correct audio token. In order to evaluate the general influence of the visual syllables in the audiovisually incongruent condition, paired comparisons were performed testing incorrect responses to the audiovisually incongruent stimuli against the corresponding alone conditions. (For all analyses in this and remaining experiments, the Greenhouse-Geisser epsilon correction was used to adjust for heterogeneity of variance.) These tests revealed that performance was significantly poorer for the audiovisually incongruent stimuli ($p < .001$ for each) indicating that the visual stimuli did influence perception of the incongruent auditory speech, even when participants were asked to base their judgments on what they heard.

Table 1

Mean Percentage Incorrect responses to Visual, Audiovisual Stimuli in Experiment 1

| | Orientation | | | |
|---|---|---|---|---|
| Presentation | UF-UM | IF-IM | IF-UM | UF-IM |
| Visual Only | | | | |
| ba | 1 | 1 | 10 | 3 |
| va | 0 | 4 | 1 | 39 |
| Audio/Visual (incongruent) | | | | |
| ba/va | 95 | 84 | 89 | 45 |
| va/ba | 73 | 67 | 50 | 61 |
| Audio/Visual (congruent) | | | | |
| ba/ba | 1 | 1 | 2 | 1 |
| va/va | 1 | 3 | 3 | 1 |

Note. U = upright, I = inverted, F = face, M = mouth,
Audiovisual stimuli are listed audio first/visual second.

Comparisons were made specifically to address the two questions outlined above. The first question examined whether a Thatcher-type effect was evident for these stimuli. The Thatcher effect was defined as occurring only when performance with the Thatcher stimulus (upright face - inverted mouth) was significantly worse than for each of the other three image orientation condition stimuli. Initial omnibus ANOVAs revealed significant relevant main and interaction effects to justify the critical contrast tests (see Table 3). Three means comparison contrasts were conducted for each visual and audiovisual syllable comparing performance on the Thatcher stimulus against performance on each of the other three face-mouth orientation conditions. For the visual-alone /va/, there were significantly fewer correct responses to the Thatcher stimulus than to each of the other three orientation stimuli (Thatcher stimulus vs.: upright face - upright mouth, $F(1,13) = 36.87$, $p < .0001$; inverted face - upright mouth, $F(1,13) = 30.36$, $p < .0001$; inverted face - inverted mouth, $F(1,13) = 35.52$, $p < .0001$). For the visual influence of /va/ in the incongruent audio /ba/-visual /va/ stimulus, there were significantly fewer incorrect responses on the Thatcher stimulus than the other three (Thatcher stimulus vs.: upright face - upright mouth, $F(1,13) = 52.53$, $p < .0001$; inverted face - upright mouth, $F(1,13) = 41.37$, $p < .0001$; inverted face - inverted mouth, $F(1,13) = 31.38$, $p < .0001$). For the congruent audio /va/-visual /va/ token, performance with the Thatcher stimulus was not significantly different from any of the three of the other stimulus conditions. For the visual /ba/, performance with the Thatcher image was not significantly different from all three of the other image conditions. This fact also held for the audio /va/-visual /ba/ stimulus, and the congruent audio/ba/-visual /ba/.

Thus, for the visual /va/ syllable, the Thatcher stimulus significantly reduced speechreading performance and the visual influence in the McGurk effect (but not perception of audiovisually congruent speech). The Thatcher stimulus did not however, disproportionately disrupt the speechreading and visual influence of the /ba/ syllable. Still, the fact that the Thatcher stimulus did disproportionately disrupt visual and incongruent audiovisual speech perception is important with regard to our general question about the influence of an upright facial context on visual speech. In that disruption occurred for the inverted mouth only in the context of an upright face (and not for the inverted mouth - inverted face stimulus), it is not clear how these results could be explained by other interpretations, for example, a violation of apparent gravitational influences. This question will be examined more closely in Experiments 2 and 3.

The second question was whether a general facial inversion effect occurred with these stimuli in visual and audiovisual speech perception contexts. For this contrast, performance with the upright face-upright mouth image was compared to performance with the inverted face-inverted mouth image. This comparison was not significant for speechreading performance with visual /va/, $F < 1$, for the visual influence of the audio /ba/-visual /va/ stimulus, $F(1,13) = 1.881$, $p = .1781$, or for the congruent audio /va/-visual /va/, $F < 1$. Comparisons of performance with upright and inverted face images were also not significant for speechreading /ba/, $F(1,13)=2.96$, $p = .11$), or perception of the congruent audio /ba/-video /ba/, $F < 1$, but was significant with regard to the visual influence of the audio /va/-visual /ba/ stimulus, $F(1,13) = 30.095$, $p < .0001$, with the upright face showing greater visual influence (see Table 1). Thus, whereas perception of visual /va/ was not disrupted by facial inversion, perception of visual /ba/ was disrupted when in the incongruent audiovisual test. It should be noted that the finding of an inversion effect on the visual influence of /b/ but not on speechreading of /b/ itself, replicates the results of Jordan and Bevan (1997).

Thus, the Thatcher and Inversion manipulations had differential effects on the two visual syllables. Whereas the Thatcher manipulation disrupted perception of visual /va/, it had no effect on visual /ba/. In contrast, facial inversion only disrupted perception of visual /ba/, and only in the context of the McGurk effect. It is unclear why the syllables were differentially influenced. However, it is the case that in each of the studies that have examined inversion effects on speech, effects have been dependent on visual-segment (see below). To explain this fact, Jordan and Bevan (1997) suggest that when the syllable's production looks similar in upright and inverted orientations—it possesses *vertical symmetry*—inversion should be less likely to disrupt visual and audiovisual speech perception. For example, whereas /ba/ involves the two lips meeting and separating, production of /va/ involves the upper teeth visibly over the lower lip and would

therefore be considered less vertically symmetric than /b/. In fact, screen measurements of our particular tokens revealed an asymmetry in the lip movements of /ba/ and /va/. During the production of the /ba/ token, the upper lip moved up about .5 in while the lower lip moved down about 1.25 in. For the /va/ token, the upper lip moved up about .25 in while the lower lip moved downward 1.5 in. This greater relative asymmetry in /va/ lip movements may account for why the inverted mouth /va/ in the Thatcher context was disruptive. At the same time however, this explanation would seem at odds with the result that inverted face visual /ba/ was more disruptive than inverted face visual /va/.

Given the differential effects observed for /ba/ and /va/, Experiment 2 was designed to further explore what varied effects the image manipulations might have on other visual syllables and incongruent audiovisual combinations. For Experiment 2, the auditory and visual syllables were /ba/ and /ga/. Beyond providing an additional test of visual segments with different vertical symmetry, the incongruent audiovisual stimuli of audio /ba/-visual /ga/ and audio /ga/-visual /ba/ are known to induce percepts that are not visually dominated in contrast to /ba/-/va/ pairs. Where audio /ba/-visual /ga/ often induces a 'fusion' percept of 'da', audio /ga/-visual /ba/ often induces a 'combination' percept of 'bga' (Green, Kuhl, Meltzoff, and Stevens, 1991; Massaro, 1987; McGurk and MacDonald, 1976). Thus, using /ba/-/ga/ pairs will provide tests of the image manipulations with other types of audiovisual integration effects.

Although it is unclear exactly how the Thatcher and Inversion manipulations will affect these visual syllables, intuitive predictions can be derived from the findings of Jordan and Bevan (1997) and ourselves. Jordan and Bevan (1997) observed a significant effect of inversion on visual /ba/, but not /ga/, in audiovisually incongruent contexts. No inversion effects were observed in visual-alone and audiovisually congruent contexts for these syllables. The inversion findings of Experiment 2 should turn out similarly.

Turning to the Thatcher effects, Experiment 1 revealed no Thatcher influence on visual /ba/. This finding might replicate in Experiment 2, even though the visual /ba/ will be paired with an audio /ga/, rather than audio /va/, in the audiovisually incongruent condition. It is less clear what will happen with the Thatcher influence on the visual /ga/. Relative vertical symmetry may explain the syllable-specificity of the Thatcher effects in Experiment 1. This factor might also bear on the Thatcher stimulus influence in Experiment 2. As described by Jordan and Bevan (1997), a visual /ga/ has even more vertical symmetry than visual /ba/. In fact, the visual /ga/ recorded for Experiment 2 did show relatively more vertical symmetry in lip movements (upper lip moved up about .25 in; while the lower lip moved down about 1 in) than our /ba/ and /va/ tokens. Potentially then, the visual /ga/ of Experiment 2 will show less susceptibility to the Thatcher manipulation than /ba/. Based on these intuitions, neither of the visual syllables tested in Experiment 2 should show Thatcher stimulus effects.

# EXPERIMENT 2

## Method

### Participants

Fourteen undergraduates at the University of California, Riverside participated for partial fulfillment of a class requirement. All reported normal or corrected vision, good hearing, and were native speakers of English. None of the participants of Experiment 1 participated in Experiment 2.

### Stimuli

The stimuli were produced with the same speaker from Experiment 1 and were produced in the same manner. The (new) syllables /ba/ and /ga/ were recorded.

The four different mouth-face orientations for each visual /ba/ and visual /ga/ were created using the same methods and equipment from Experiment 1. The audio stimuli consisted of the audio /ba/ and audio /ga/ sampled from the videotape used to derive the visual tokens.

### Procedure

In general, the same methods were used as those in Experiment 1. However for this experiment, the participants were given a five-alternative forced choice for their responses. The

five choices were: b, d, th, bg, and g. A forced choice procedure was used because pilot participants reported difficulty identifying 'bg' as an initial 'consonant' (see also, Green, 1994; Green, Kuhl, and Meltzoff, 1991). When presented with 'bg' as a possible choice, participants were able to overcome this orthographic bias. The rest of the procedure for Experiment 2 was the same as in Experiment 1.

Results and Discussion

The mean percentage of incorrect responses was calculated for each token type (see Table 2). Paired comparisons revealed that the percentage of incorrect responses to all audiovisually incongruent stimuli was significantly higher than the percentage of incorrect responses in the audio alone conditions, $p < .001$. These results show that the visual stimuli influenced auditory speech judgments.

Table 2 Mean Percentage Incorrect Responses to Visual, Audiovisual Stimuli in Experiment 2

| | Orientation | | | |
|---|---|---|---|---|
| Presentation | UF-UM | IF-IM | IF-UM | UF-IM |
| Visual Only | | | | |
| ba | 2 | 1 | 16 | 20 |
| ga | 28 | 38 | 33 | 46 |
| Audio/Visual (incongruent) | | | | |
| ba/ga | 70 | 63 | 53 | 54 |
| ga/ba | 44 | 41 | 44 | 45 |
| Audio/Visual (congruent) | | | | |
| ba/ba | 9 | 6 | 9 | 4 |
| ga/ga | 1 | 1 | 1 | 1 |

Note. U = upright, I = inverted, F = face, M = mouth, Audiovisual stimuli are listed audio first/visual second.

The two comparisons examined in Experiment 1 were also examined for this experiment. Again, initial omnibus ANOVAs revealed significant relevant main and interaction effects to justify the critical contrast tests (see Table 3). To test the Thatcher effect, performance with the Thatcher stimulus was compared to that on each of the other three image conditions. The Thatcher stimulus significantly reduced speechreading accuracy with the visual /ga/ syllable: Thatcher stimulus vs.: upright face - upright mouth, $F(1,13)=23.09$, $p < .0001$; inverted face - upright mouth, $F(1,13)=12.42$, $p < .002$; inverted face - inverted mouth, $F(1,13)=5.03$, $p = .037$. However, the Thatcher stimulus did not reduce the visual influence of visual /ga/ on audio /ba/ relative to all three other mouth-face orientation conditions. Similarly, the Thatcher stimulus did not significantly disrupt performance with the congruent audio /ga/-visual /ga/ relative to all three other face-mouth condition stimuli. With regard to the visual /ba/, the Thatcher image condition did not significantly reduce performance relative to all three of the other mouth-face orientation images in the speechreading, audiovisually incongruent, or audiovisually congruent contexts.

Table 3 Results of Significant ANOVA Tests for Experiments 1 and 2

| | Experiment | |
|---|---|---|
| Effect | 1 | 2 |
| Visual Alone Stimuli | | |
| Syllable | $F(1,13)=7.186$, $p=.0189$ | $F(1,13)=6.940$, $p=.0188$ |
| Orientation | $F(3,39)=18.617$, $p=.0006$ | $F(3,39)=5.365$, $p=.007$ |

| | | |
|---|---|---|
| Syllable x orientation | F(3,39)=15.84, p=.0004 | F(3,39)=2.701, p=.0568 |
| Audiovisual Stimuli | | |
| Congruency | F(1,13)=223.034, p<.0001 | |
| Orientation | F(3,39)=22.844, p<.0001 | F(3,39)=3.129, p=.0431 |
| Congruency x orientation | F(3,39)=15.761, p<.0001 | |
| Syllable x orientation | F(3,39)=13.247, p<.0001 | F(3,39)=3.253, p=.0318 |
| Syllable x orientation x congruency | F(3,39)= 11.962, p<.0001 | |

Thus, while the visual syllable /ga/ is somewhat susceptible to disruptions from the Thatcher manipulation, the visual /ba/ is not. This latter finding is consistent with the results from Experiment 1. The fact that /ga/, but not /ba/, was affected by the Thatcher manipulation is not compatible with predictions based on the vertical symmetry hypothesis. This hypothesis predicted that the more vertically symmetric /ga/ should be less influenced by the Thatcher stimulus than the less symmetric /ba/ (as defined by Jordan and Bevan, 1997). However, it could be that more than lip movements play a role in the apparent vertical symmetry of these segments. In fact, /ga/ did involve a visible and vertically asymmetric movement of the tongue not visible for the /ba/ production. The relative salience of tongue and lip movements for perceived vertical symmetry is a topic for future research.

Analyses were also conducted to examine the influence of facial inversion. These tests compared performance on the upright face - upright mouth condition to that of the inverted face - inverted mouth condition. Inversion disrupted the visual /ga/ syllable's influence on audio /ba/, $F(1,13)=4.3$, $p < .05$, but not the identifiability of visual-alone /ga/, or congruent audio /ga/-visual /ga/. In contrast, facial inversion did not disrupt performance with the visual /ba/ in speechreading, audiovisually incongruent, or audiovisually congruent contexts.

Thus, perception of the syllable /ga/ was disrupted by inversion in the discrepant audiovisual context, while perception of /ba/ was not. It should be noted that these results are opposite to those of Jordan and Bevan (1997) who found inversion effects on visual /ba/, but not /ga/. These results are not compatible with a vertical symmetry hypothesis based on lip movements. However, it is likely that our visual syllables were different from those of Jordan and Bevan, possibly in tongue movement asymmetry (see above) which might account for the different results. This point will be elaborated in the General Discussion.

The results of Experiment 2 also contrast somewhat with those of Experiment 1, in which inversion did decrease the influence of visual /ba/ on audio /va/. This difference might be related to the fact that a different visual /ba/ was used in the two experiments or to the relative ease with which the audio /va/ and /ga/ syllables can be influenced. Alternatively, it might be that inversion influences audiovisual integrations in which the visible articulation dominates more so than integrations underlying combination responses. Future research can be designed to examine these alternatives.

To summarize the Thatcher stimulus effects thus far, for at least the visual /va/ (and to some degree, /ga/) stimuli, the Thatcher image disrupted visual speech perception. For these syllables then, an upright facial context may heighten sensitivity to facial distortions, which in turn, disrupt visual speech perception. If this is true, then a related prediction can be made. If Thatcher effects in visual speech perception are a result of upright facial context inducing a distorted interpretation of the mouth, then mouths presented in isolation (without the face context) should not produce the same effects. This prediction is examined in Experiment 3, which tests visual speech perception from isolated mouths.

Experiment 3 will also test an alternative explanation for the observed Thatcher effects. Hypothetically, it might be that the effects are based on the Thatcher stimulus being comprised of an inverted mouth located in the lower portion of the screen. While it is unclear how the screen position of the mouth could bear on visual speech perception, this possibility can be easily examined with isolated mouth stimuli. Accordingly, upright and inverted isolated mouth images

will be presented in both upper and lower screen positions. If the previously observed Thatcher stimulus effects are based on upright facial context information, then no Thatcher effect would be expected from the isolated mouth stimuli. If, on the other hand, the observed effects are based on some interaction of mouth inversion and (lower) screen position, then mouth-alone stimuli should show a similar inhibition of performance.

Using mouth-alone stimuli in Experiment 3 will permit an additional test of upright facial context information: the influence of image inversion using mouth-alone stimuli. Recall that the previous visual speech inversion effects have been attributed to disruption of upright facial context information (e.g., Jordan and Bevan, 1997). If the basis for the inversion effects is upright face information (rather than, for example, disruption of perceived gravitational influences on articulation) then perception from an inverted isolated mouth should not be disrupted.

Experiment 3 will examine these questions with /ba/-/va/ and /ba/-/ga/ visual syllables and audiovisual combinations and will include tests of the same full-face conditions tested in Experiments 1 and 2. Retesting these conditions will allow for more direct comparisons between full-face and mouth-alone stimulus performance as well as serve as replication tests of the previous experiments.

# EXPERIMENT 3

## Method

### Participants

Twenty-eight undergraduates at the University of California, Riverside participated for partial fulfillment of a class requirement. Thirteen subjects responded to the /ba/-/va/ stimuli and fifteen subjects responded to the /ba/-/ga/ stimuli. All reported normal or corrected vision, good hearing, and were native speakers of English. None of the subjects in Experiments 1 or 2 participated in Experiment 3.

### Stimuli

The sixteen full-face visual, audiovisual, and audio stimuli that were used in Experiments 1 and 2 were used in this experiment. To create mouth-alone tokens, the mouth was isolated for each of the four different orientations that were created for /ba/ and /va/ in Experiment 1 and for /ba/ and /ga/ in Experiment 2. This was performed using the MacIIci computer and NIH Image software program. To isolate the mouth, each frame of each of the four images (per syllable) was darkened except where the box had been outlined around the mouth as had been established for Experiments 1 and 2. This created four new stimuli for each of the four visual syllables: an upright mouth in the lower portion of the screen; an inverted mouth in the upper portion of the screen; an upright mouth in the upper portion of the screen; an inverted mouth in the lower portion of the screen (see Figure 2).

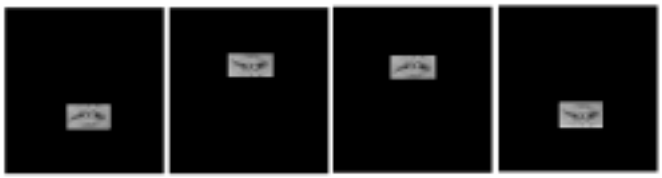

Figure 2. Photographs of the four mouth-alone conditions used in Experiment 3. From left to right: Upright Mouth – Lower Screen Position; Inverted Mouth – Upper Screen Position; Upright Mouth – Upper Screen Position; Inverted Mouth – Lower Screen Position. The mouth is shown initiating a /va/ syllable.

The 16 mouth-alone visual stimuli were then transferred back onto videotape using the MacIIci computer and a Panasonic 7500A video recorder. The auditory tokens used in Experiments 1 and 2 were then dubbed synchronously onto the twelve mouth-alone visual tokens. Both congruent and incongruent tokens were dubbed using the same procedure used in Experiments 1 and 2. After dubbing, two presentation videotapes were created. The first tape used the syllables /ba/ and /va/ from Experiment 1 together with the new /ba/ and /va/ mouth-alone stimuli. The second tape used the /ba/ and /ga/ syllables from Experiment 2 along with the new mouth-alone /ba/ and /ga/ stimuli.

Each presentation tape consisted of 356 tokens. For each tape, the 16 full-face audiovisual tokens and the 16 mouth-alone audiovisual tokens were presented eight times each. The eight full-face visual tokens, the eight mouth-alone visual tokens and the two audio tokens were presented 5 times each. For each tape, the 356 tokens were completely randomized and then separated into eight blocks of 40 trials and one block of 36 trials. All tokens were separated by a 3 s ISI and the blocks were separated by 12 s.

Procedure

The procedures for the group that was tested using the /ba/-/va/ presentation tape were the same as for Experiment 1(free response task). The procedures for the group that was tested using the /ba/-/ga/ presentation tape were the same as for Experiment 2 (five-alternative forced choice task). The experiment lasted approximately 50 minutes for each subject.

## Results and Discussion

Table 4 lists the mean percentage of incorrect responses for each token type. Incorrect responses are reported in the same way as for Experiments 1 and 2. For both tapes, the percentage of incorrect responses in the audiovisual conditions for both full-face and mouth-alone conditions was significantly higher than the percentage of incorrect responses in the audio alone conditions, $p < .001$, indicating that all of the visual stimuli influenced auditory speech judgments.

Table 4 Mean Percentage Incorrect Responses to Visual, Audiovisual Stimuli in Experiment 3

| | Full Face | | | | Mouth Only | | | |
|---|---|---|---|---|---|---|---|---|
| Presentation | UF-UM | IF-IM | IF-UM | UF-IM | UF-UM | IF-IM | IF-UM | UF-IM |
| Visual Only | | | | | | | | |
| ba | 2 | 6 | 1 | 5 | 0 | 5 | 0 | 6 |
| va | 0 | 0 | 0 | 45 | 0 | 3 | 0 | 1 |
| Audio/Visual (incongruent) | | | | | | | | |
| ba/va | 99 | 98 | 94 | 62 | 97 | 98 | 100 | 98 |
| va/ba | 90 | 55 | 77 | 85 | 92 | 57 | 88 | 58 |
| Audio/Visual (congruent) | | | | | | | | |
| ba/ba | 1 | 6 | 0 | 5 | 0 | 7 | 0 | 7 |
| va/va | 0 | 1 | 0 | 13 | 0 | 1 | 0 | 0 |
| Presentation | UF-UM | IF-IM | IF-UM | UF-IM | UF-UM | IF-IM | IF-UM | UF-IM |
| Visual Only | | | | | | | | |
| ba | 4 | 8 | 9 | 7 | 7 | 11 | 1 | 8 |
| ga | 31 | 15 | 32 | 25 | 39 | 33 | 22 | 23 |
| Audio/Visual (incongruent) | | | | | | | | |
| ba/ga | 73 | 64 | 74 | 66 | 72 | 67 | 61 | 63 |
| ga/ba | 65 | 70 | 64 | 64 | 63 | 60 | 54 | 69 |
| Audio/Visual (congruent) | | | | | | | | |
| ba/ba | 2 | 0 | 1 | 0 | 3 | 1 | 3 | 2 |
| ga/ga | 1 | 4 | 1 | 4 | 3 | 0 | 3 | 3 |

Note. U = upright, I = inverted, F = face, M = mouth,
Audiovisual stimuli are listed audio first/visual second.

The remaining data analyses will be presented in two sections. First, data from the full-face stimuli responses, based on the Thatcher and image inversion comparisons, will be presented for replication purposes. The second section will present analogous analyses for the mouth-alone stimuli. Again, results of omnibus ANOVAs for both /ba/-/va/ and /ba/-/ga/ stimulus tapes justified the critical contrast tests (see Table 5).

Table 5 Results of Significant ANOVA Tests for Experiment 3

| | Stimulus Tape | |
|---|---|---|
| Effect | /ba/-/va/ | /ba/-/ga/ |
| Visual Alone Stimuli | | |
| Syllable | | F(1,14)=14.38, p=.0020 |
| Facial Completeness | F(1,12)=17.400, p=.0013 | |
| Orientation | F(3,36)=29.451, p<.0001 | |
| Syllable x Completeness | F(1,12)=11.219, p=.0058 | |
| Completeness x Orientation | F(3,36)=20.183, p=.0002 | F(3,42)=3.445, p= .0250 |
| Syllable x Orientation | F(3,36)=9.100, p=.0042 | |
| Completeness x syllable x orientation | F(3,36)=14.348, p=.0016 | |
| Audiovisual Stimuli | | |
| Congruency | F(1,12)=506.93, p<.0001 | F(1,14)=80.922, p<.0001 |
| Syllable | F(1,12) = 5.606, p=.0355 | |
| Orientation | F(3,36) = 9.083, p<.0001 | |
| Completeness x congruency | F(1,12) = 6.110, p=.0294 | F(1,14)=5.321, p<.0369 |
| Completeness x syllable | F(1,12) = 7.734, p=.0166 | |
| Congruency x syllable | F(1,12)=10.202, p=.0077 | |
| Congruency x orientation | F(3,36)=18.355, p<.0001 | |
| Syllable x orientation | F(3,36)=9.944, p< .0001 | F(3,42)=3.947, p=.0144 |
| Syllable x orientation x congruency | F(3,36)=9.347, p<.0001 | |
| Syllable x orientation x completeness x congruency | F(3,36)=17.404, p<.0001 | |

Replications of full-face image manipulations

To test the Margaret Thatcher effect, performance in the Thatcher condition (upright face - inverted mouth) was compared with that in the other three orientations for the visual and audiovisual conditions. First, for the /ba/-/va/ stimulus tape, the Thatcher comparison was significant for the visual /va/: Thatcher stimulus (upright face - inverted mouth) vs.: upright face - upright mouth, $F(1,12)=45.26$, $p < .0005$; inverted face - upright mouth, $F(1,12)=45.26$, $p < .0005$; inverted face - inverted mouth, $F(1,12)=45.26$, $p < .0005$. This comparison also held for the audio /ba/-visual /va/ (Thatcher stimulus vs.: upright face - upright mouth, $F(1,12)=29.01$, $p < .0009$; inverted face - upright mouth, $F(1,12)=22.08$, $p = .0023$; inverted face - inverted mouth, $F(1,12)=27.48$, $p = .0011$) and for the congruent audio /va/-visual /va/ (Thatcher stimulus vs.: upright face - upright mouth, $F(1,12)=15.71$, $p < .009$; inverted face - upright mouth, $F(1,12)=15.71$, $p < .009$; inverted face - inverted mouth, $F(1,12)=13.33$, $p < .01$).

The Thatcher comparison was not significant for visual /ba/, nor for the congruent audio /ba/-visual /ba/, nor for the incongruent audio /va/-visual/ba/. These results are generally similar to the results obtained in Experiment 1 with the exception of the significant Thatcher manipulation effect on the congruent /va/-/va/ syllable. Considered with results of Experiment 1, our results indicate that perception of visual /va/ is disrupted in the Margaret Thatcher condition, whereas perception of /ba/ is not.

The Thatcher comparison was also examined for the /ba/-/ga/ presentation tape. This comparison was not significant for any of the six comparisons conducted (visual /ga/; audio /ga/-visual /ga/; audio /ba/-visual /ga/; visual /ba/; audio /ba/-visual /ba/; audio /ga/-visual /ba/). These results were generally similar to those of Experiment 2 differing only in that Experiment 2 revealed a significant Thatcher influence on the visual /ga/ condition. It should be noted that all three experiments indicate that the Thatcher image manipulation does not disrupt perception of a visual /ba/ in speechreading or audiovisual contexts.

The influence of general facial inversion was also examined with the full-face stimuli of Experiment 3. As before, this question was tested by comparing performance with the upright face-upright mouth condition to performance with the inverted face-inverted mouth condition. For this comparison, neither the visual /va/, the audio /ba/-visual /va/ or congruent audio /va/-visual /va/ were significant. Likewise, inversion effects were not observed for the visual /ba/ and congruent audio /ba/-visual /ba/ stimuli. However, an inversion effect was observed for the visual influence with the audio /va/-visual /ba/ stimulus, $F(1,12)=120.36$, $p < .0001$. These results are completely consistent with the results of Experiment 1 and further suggest that the syllable /ba/ can be susceptible to facial inversion effects in an incongruent audiovisual context.

Facial inversion effects were also investigated for the /ba/-/ga/ presentation tape. Results revealed that the inversion effect was significant for the visual /ga/ stimulus, $F(1,14)=29.01$, $p = .0053$, as well as the influence of that stimulus on audio /ba/, $F(1,14)=6.88$, $p = .024$, but not for the congruent audio /ga/-visual /ga/. In contrast, the inversion effect did not occur for the visual /ba/, nor did it affect the visual influence for the audio/ga/-visual/ba/ stimulus, or congruent audio /ba/-visual /ba/. These results generally replicate the effects found in Experiment 2 showing some inversion effect for the syllable /ga/ but not for the syllable /ba/ (see Table 2).

In general, these results replicate the findings from Experiments 1 and 2 (21 of 24 comparisons were replicated). The few disparities might be related to differences in procedures between the first two, and this third experiment.

Mouth-alone image tests

Tests were conducted to determine whether the observed Thatcher image effects were based on the upright facial context rather than some interaction of mouth inversion and screen position. For these purposes, a Thatcher comparison was conducted on the mouth-alone stimuli. Again for these tests, a Thatcher effect was defined as occurring only when performance with the Thatcher stimulus (for the mouth-alone images: inverted mouth in lower screen position) was significantly worse than for each of the other three image condition stimuli. For all 12 of the syllable and syllable combination stimuli tested (visual /va/; audio /va/-visual /va/; audio /ba/-visual /va/; visual /ba/; audio /ba/-visual /ba/; audio /va/-visual /ba/; visual /ga/; audio /ga/-visual /ga/; audio /ba/-visual /ga/; visual /ba/; audio /ba/-visual /ba/; audio /ga/-visual /ba/), no significant Thatcher effects were observed with the mouth-alone stimuli. These results contrast with the full-face Thatcher results for particularly, the visual /va/ stimulus. Thus, although the performance of subjects in Experiment 3 was disrupted with the full-face Thatcher /va/ image, no disruption occurred for these same subjects when the mouth was presented in isolation. These contrasting results suggest that the observed full-face Thatcher effects *were* based on the upright facial context intrinsic to the Thatcher stimulus, and not on extraneous factors such as an interaction of screen position with mouth inversion.

Beyond providing tests of the Thatcher effect for full-face and mouth-alone stimuli, Experiment 3 also allowed for straight inversion comparisons for the two stimulus types. As outlined above, the full-face images induced significant inversion effects for the audio /va/-visual /ba/, visual /ga/, and audio /ba/-visual /ga/ stimuli in Experiment 3 (which replicated the results of Experiments 1 and 2). Inversion comparisons were also conducted with the mouth-alone stimuli. These tests involved comparing performance between the images analogous to those used for the full-face tests. Accordingly, performance with the upright mouth-lower screen position image was compared to performance with the inverted mouth-upper screen position image. Of the 12 syllable and syllable combination stimuli tested, a significant mouth-alone inversion effect was observed

only for the visual /ba/, $F(1,14)=9.0$, $p < .01$, audio /ba/-visual /ba/, $F(1,14)=6.245$, $p = .028$, and audio /va/-visual /ba/, $F(1,14)=172.88$, $p < .0001$, all from the /ba-va/ stimulus tape.

Recall the hypothesis that if visual speech inversion effects are based on upright facial context information, then perception should not be disrupted when the mouth is isolated from the face. This hypothesis was supported by the findings that only in the full-face context did the visual /ga/ and audio /ba/-visual /ga/ stimuli display inversion effects. Possibly for these tokens, the information available in an upright face facilitates speechreading so that when it is inverted, the visual influence is disrupted. Further implications of these results will be discussed in the General Discussion.

On the other hand, a significant inversion effect *was* observed for both the full-face and mouth alone audio /va/-visual /ba/. According to the hypothesis, this might suggest that the inversion-based disruption was not based on a disruption of upright facial context information. Perhaps for this syllable combination, inversion might have distorted other informational dimensions such as the natural gravitational dynamics acting on the articulations as specified in both full-face and mouth-alone displays.

According to the above ANOVAs, the mouth-alone visual /ba/ and audio /ba/-visual /ba/ stimuli displayed a significant inversion effect, even though analogous effects did not occur for the full-face stimuli. However, these statistical outcomes are likely based on the low variance in these scores. For example, for the audio /ba/-visual /ba/, the difference in means (100% correct for the upright mouth; 93% correct for the inverted mouth) was attributable to the inverted mouth performance of just three of thirteen subjects (i.e., ten of thirteen subjects scored perfectly on this task). In fact, when evaluated as trials, subjects mis-identified the inverted audio /ba/-visual /ba/ on a total of only 7 of 104 trials. With regard to the visual /ba/ effect (100% correct for the upright mouth; 95% correct for the inverted mouth), it was also attributable to the inverted mouth performance of three of thirteen subjects. Subjects mis-identified the inverted visual /ba/ on a total of only 3 of 65 trials. It would seem then, that the statistical significance observed for inversion of the audio /ba/-visual /ba/, and visual /ba/ is based on the low variance of these data and does not indicate any substantial condition effects.

## General Discussion

The results of these experiments show that visual speech perception can be disrupted by a Thatcher-type manipulation. These results go beyond the inversion effects in showing that an upright facial context facilitates visual speech perception in a way similar to facilitation for face perception and recognition (Lander, et al., in preparation).

The extant inversion findings cannot determine whether disruption is a consequence of defeating the facilitatory effect of upright faces, or the influence of apparent gravity on articulatory dynamics. Two of our results bear on this question. First, Experiments 1 and 3 reveal that for at least the visual /va/, inverting the mouth is most disruptive to visual (and audiovisual) speech perception when it is seen in the context of an upright face. The same inverted mouth is not disruptive when seen in the context of an inverted face. Additionally, Experiment 3 revealed that this disruptive influence did not occur when the (inverted) mouth was presented in isolation. Thus for some visual speech, the disruptive aspects of apparent gravity on the mouth's articulatory dynamics is not, in and of itself, the cause of inversion effects for visual speech perception. Instead, there does seem to be some facilitatory effect of an upright facial context.

The evidence for upright face facilitation in speechreading bears on the relation between visual speech and face perception. Before discussing these theoretical implications, the nature of the effects' segment-specificity will be addressed along with considerations of the previous findings on perceiving speech from inverted faces.

### Segment and task-specific effects

The Thatcher and inversion image manipulations differentially influenced perception of the three visual syllables tested. A sense of these differences for the full-face stimuli can be gained by surveying the percentage of image manipulation effects for tokens containing a visual /va/, /ga/, or

/ba/ calculated across all experiments. (It should be noted that the visual /ba/ was tested twice as often as the other two syllables.) The Thatcher manipulation disrupted the perception of tokens containing a visual /va/ 83% of the time; a visual /ga/ 17% of the time; and a visual /ba/ 0% of the time. Inversion affected the perception of tokens containing a visual /va/ 0%, a visual /ga/ 50%, and a visual /ba/ 17% of the time. The situation is further complicated by the mouth-alone stimuli: image inversion manipulations only influenced the tokens containing a visual /ba/ 17% of the time, while the (analog) 'Thatcher' manipulation did not influence perception of any mouth-alone stimuli.

It is unclear why the image manipulations would differentially affect the visual segments so dramatically. As stated earlier, the differences may relate to the vertical symmetry of the consonantal productions. Alternatively, the segments might be differentially influenced by the disjointed components of mouth and face when in the Thatcher stimulus context.

It should be noted that the extant visual speech inversion research reveals a great deal of segment variability across studies. Inversion effects have been observed for the visual segments /n/ (Bertelson, et al., 1994), /b/ (Green, 1994; Jordan and Bevan, 1997; the current paper), /g/ (Green, 1994; the current paper), /v/ (Massaro and Cohen, 1996), / / (Massaro and Cohen, 1996), and /m/ (Jordan and Bevan, 1997). At the same time however, research has shown instances for which inversion effects *do not* occur for the segments /m/ (Bertelson, et al., 1994), /b/ (Massaro and Cohen, 1996; the current paper), /d/ (Massaro and Cohen, 1996), /g/ (Green, 1994; Jordan and Bevan, 1997), and /t/ (Jordan and Bevan, 1997). Thus, there is substantial discrepancy across experiments, which is likely related to differences in speaker characteristics (cf. Green, 1994; Jordan and Bevan, 1997). Future research, possibly using synthetic faces (cf. Massaro and Cohen, 1996), can be designed to further explore the segment-specificity of the Thatcher and inversion manipulations.

Turning to task effects, the image manipulations had a more consistent influence on the speechreading and McGurk tasks. Pooling across the three experiments, of the 14 significant image manipulation effects, seven occurred in the McGurk visual influence task, five in the visual-alone speechreading task, and two in the audiovisually congruent task. Also, if an image manipulation did disrupt speechreading of a visual syllable, it also disrupted that syllable's influence on a discrepant auditory syllable. Only in one case did this fact not occur (Thatcher image manipulation effect on visual /ga/), but this finding did not replicate in Experiment 3. Thus, while many past studies have shown that McGurk visual influence tests are substantially more sensitive to image manipulations than are speechreading tasks (e.g., Green, 1994; Jordan and Bevan, 1997; McGrath, 1985; Rosenblum and Saldaña, 1996; Summerfield, et al., 1989), our results revealed only a small difference between speechreading and visual influence tasks (see also Massaro and Cohen, 1996). This may mean that the image distortion effects that did occur in our experiments were relatively strong.

Disruptions of the audiovisually congruent stimuli occurred only twice: for the audiovisual /va/-/va/ in the full-face condition of Experiment 3 (disrupted by the Thatcher manipulation); and for the audiovisual /ba/-/ba/ in the mouth-alone condition of Experiment 3 (disrupted by the inversion manipulation). Interestingly, both of these pairs involved visual segments whose perception was also disrupted (by the same image manipulation) in the visual-alone and visual influence contexts of Experiment 3. This suggests that, at least for the Experiment 3 subjects, the respective image manipulations had a large disruptive influence on perception of these two visual segments. The fact that the audiovisually congruent stimuli showed the least disruptions by the image manipulations is similar to findings of most other visual speech image manipulation studies (e.g., Jordan and Bevan, 1997; Green, 1994; Rosenblum and Saldaña, 1996; but see Massaro and Cohen, 1996). The robustness of audiovisually congruent stimuli may be related to the facts that: a) they provide redundant segment information; and b) they provide the most natural and common conditions of any of those involving visual speech.

## Bases of Thatcher Effects

Our findings provide a new demonstration of a Thatcher-type effect. This effect is novel in that it occurred with a dynamic stimulus in the context of a visual speech task. Thus, our findings have implications for current explanations of Thatcher effects.

The Thatcher effect shows that sensitivity to facial distortions is much greater when a face is seen upright. Bartlett and Searcy (1993) have discussed three explanations for the effect. First, in that it is the Thatcher image's 'gruesome expression' that is apparent in an upright context, inversion could specifically inhibit expression recognition (Valentine, 1988). Secondly the effect could be based on the relationship between *object* (e.g., face) and *non-object* (e.g., external; gravitational) frames of reference (Parks, 1983; Parks, et al., 1985; Rock, 1974). These two frames might interact to determine how orientation is assigned to image features. When the Thatcher stimulus is upright, the external and object (facial) frames force an interpretation of the mouth and eyes as upright causing their gruesome appearance. When the Thatcher stimulus is inverted, the two frames compete, thereby making interpretation of the features less stable and hence, less strikingly grotesque. It follows from this that an image that is missing the object frame will rely solely on the external frame for assigning featural orientation (Parks, et al., 1985).

A third explanation of the Thatcher effect is based on the relationship between *component* and *holistic* (or configural) facial information (e.g., Bartlett and Searcy, 1993; Sergent, 1984; Valentine and Bruce, 1988; Yin, 1969). A face can provide information from individual component features (e.g., eyes, nose, mouth) as well as how those features are configured to comprise the whole image. When a face is inverted, perception of holistic (and not component) information is disrupted. Regarding the Thatcher image, its inversion disrupts the viewer's ability to interpret the configuration of the inverted mouth and eyes relative to the facial frame (e.g., Rock, 1988). The grotesqueness of the features are most noticeable when the image is upright and holistic/configural information is available.

Our speech findings bear on these explanations. First, we found evidence for a Thatcher-type effect using a task not involving judgments of facial expression. Although it is likely that some of our stimuli looked grotesque, judgments of visual speech were influenced without explicit reference to expression. This provides evidence against the expression-based explanation. Our results would also seem incompatible with the frame of reference account. This account maintains that isolated features should be assigned an orientation based on the non-object, external frame. Thus, isolated inverted features should be interpreted in the same way as inverted features in the context of an upright face. However, our Experiment 3 revealed that for an inverted mouth /va/, an upright facial context influenced visual speech in a way different from when the mouth was in isolation. Thus, our results provide evidence against the frames of reference and expression explanations, leaving support for the holistic information account.

Other recent evidence has revealed the importance of holistic facial information (e.g., Barlett and Searcy, 1993). Tanaka and Farah (1993) as well as Rhodes, Brake, and Atkinson (1993), have found differential influences of image inversion depending on whether facial features were seen in the context of a facial frame or in isolation. These findings can be seen as analogous to those observed in our Experiment 3 in supporting the involvement of holistic information.

It must be acknowledged that if holistic information is used for visual speech, its use is dependent on visual segment. The idea that holistic information is usable—but not mandatory—for visual speech perception is congruent with the speculations of Jordan and Bevan (1997). They suggest that holistic information might play some role in visual speech perception, but not as great a role as for face perception. Based on their observed segment-specific effects, they speculate that holistic information might be more important for extraction of more extreme visual speech movements (e.g., for Jordan and Bevan: visual /b/ and /m/). Perception of these segments might make use of the configural information which specifies mouth movements relative to the more stable nose and cheeks. Jordan and Bevan's speculation is supported by our Thatcher results: the extreme (and vertically asymmetric) /v/ articulation was the segment most influenced by the holistically-disrupted Thatcher condition. (However, our *inversion* results are not supportive of this speculation in that this type of holistically-disruptive manipulation was most influential on the relatively subtle articulations of /ga/.) An analogous explanation has been offered

by Bartlett and Searcy (1993) in discussing how Thatcher-effects interact with different types (and extremes) of expression.

Implications for the relation between face and visual speech perception

In revealing a Thatcher-type effect with visual speech, our results buttress the inversion effects in demonstrating the specific influence of upright facial information on visual speech perception. These findings add to the evidence that similar information may be used in and/or operations may be applied to both visual speech and face perception. Clearly, the influence of common factors on both functions need not imply that the functions are related. However, the specific common factor we have shown to be influential—upright facial context information—has been considered idiosyncratic to face recognition and perception. In this sense, the visual speech Thatcher results bear on questions of the independence and information encapsulation of the speech and face functions. A number of other recent considerations further motivate a re-examination of this issue.

Historically, most conceptions of both face and visual speech perception, as well as of modularity, have considered the functions separate (e.g., Bruce and Valentine, 1986; Fodor, 1983). Early support for the separation was provided by neuropsychological data (e.g., Bruce and Young, 1986; Campbell, Landis, and Regard, 1986; Ellis, 1989). There is some lesion evidence for a double-dissociation of the functions (Campbell, Landis, & Regard, 1986), as well as normal subject data suggesting a left visual field/right-hemisphere (LVF/RH) advantage for face recognition, and a right visual field/left-hemisphere (RVF/LH) advantage for speechreading (e.g., Campbell, de Gelder, and de Haan 1996; Smeele, in press; Young, 1984; Young, Hay, McWeeney, Ellis, & Barry, 1985; and see Ellis, 1989, for a review). More recently however, both sets of neuropsychological data have been challenged by contradictory results (e.g., Baynes, Funnell, & Fowler, 1994; Campbell, 1986; Diesch, 1995; Damasio, 1989; Hillger & Koenig, 1991; Sergent, 1982) as well as through critiques of methodology (e.g., Rosenblum and Saldaña, 1998). Thus, the neuropsychological evidence is ambiguous with regard to the dissociation of speechreading and face recognition functions.

As for the behavioral literature, some recent reports hint at a contingency between the functions. De Gelder, Vroomen, and van der Heide (1991) have found a correlation between speechreading and face identification performance in individual subjects suggesting some relationship between functions. Also, Yakel, Rosenblum, and Fourtier (under review) have found evidence that observers perform better when speechreading sentences from a single-speaker vs. multiple-speaker presentation tape. Relatedly, Schweinberger and Soukup (1998) have found that RTs for identifying vowels portrayed in facial photographs are faster when observers are familiar with the faces. Finally, Sheffert and Fowler (1995) have found that visual speaker information can be retained along with audiovisually presented words. These findings are consistent with the notion that familiarity with the face of the speaker facilitates speechreading performance.

Next, Walker, Bruce, and O'Malley (1995) found that subjects who were familiar with the faces of the speakers portrayed in the stimuli were less susceptible to the McGurk effect when the faces and voices were incongruent in speaker identity. Walker, et al. (1995) speculate that the audiovisually discrepant stimuli violate perceiver expectations more obviously with familiar speakers. They conclude that facial identity and facial speech perception are not independent functions.

What might be the basis of the contingencies between visual speech and face perception? One possibility suggested by the current findings, is that the contingencies are based on both functions' use of common (e.g., upright facial) information. Interestingly, Remez and his colleagues (Remez, Fellowes, & Rubin, 1995; 1998) suggest an analogous explanation for contingencies observed between *auditory* speech and speaker recognition (e.g., Church & Schacter, 1994; Mullennix, Pisoni, & Martin, 1989; Nygaard, Sommers, & Pisoni, 1994; Palmeri, Goldinger, & Pisoni, 1993; Remez, Fellowes, & Rubin, 1998; Saldaña and Rosenblum, 1994). Remez, et al. have found evidence that isolated phonetic attributes of a speech signal can be used for both linguistic and talker recognition. They argue that the observed contingencies between auditory speech and speaker recognition might be based, in part, on the use of this common information.

With regard to the visual modality, the observed contingencies between face and visual speech perception might also be due, in part, to both functions' ability to use common visual information. Historically, the visual attributes for face recognition and visual speech recovery have been considered to be very different. Information for face perception is thought to be comprised of the shape and configuration of the facial features as well as skin tone, hair, and the overall shape of the face (see Bruce, 1988, for a review). In contrast, the features for speechreading are construed as articulatory dimensions such as place of constriction; open, closed, or rounded lips; and visible teeth (e.g., Montgomery and Jackson, 1983; Summerfield and McGrath 1984; McGrath, 1985) or some time-varying form of these dimensions (see Rosenblum and Saldaña, 1998, for a review). However, the current results, along with the inversion effect findings (e.g., Bertelson, et al, 1994; Green, 1994; Jordan & Bevan, 1997; Massaro & Cohen, 1996), suggest that there are circumstances for which both functions can make use of upright facial context information. Future research can examine whether other informational dimensions might be useful for both functions and whether this commonality underlies the contingencies between visual speech and face perception.

## References


Bartlett, J.C., & Searcy, J. (1993). Inversion and configuration of faces. Cognitive Psychology, 25(3), 281-316.

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(6), 633-641.

Bertelson, P., Vroomen, J., Wiegeraad, G., & de Gelder, B. (1994). Exploring the relation between McGurk interference and ventriloquism. Proceedings of the 1994 International Conference on Spoken Language Processing (ICLP94), 2, 559-562.

Bingham, G. P., Schmidt, R. C., Rosenblum, L. D. (1995). Dynamics and the orientation of kinematic forms in visual event recognition. Journal of Experimental Psychology: Human Perception & Performance, 21(n6), 1473-1493.

Bruce, V. (1988). Recognising faces. Erlbaum; Hove, England.

Bruce, V., & Young, A. (1986). Understanding face recognition. British Journal of Psychology, 77, 305-327.

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

Campbell, R. T., Landis, T., & Regard, M. (1986). Face recognition and lip-reading: A neurological dissociation. Brain, 109, 509-521.

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

Church, B.A., & Schacter, D.L. (1994). Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20(3), 521-533.

Damasio, A.R. (1989). Neural mechanisms. In A. Young, A. Ellis (Eds.) Handbook of research on face processing. North Holland, Elsevier, p. 405-425.

Diamond, R., & Carey, S. (1986). Why faces are not special: An effect of expertise. Journal of Experimental Psychology: General, 115, 107-117.

Diesch, E. (1995). Left and right hemifield advantages of fusions and combinations in audiovisual speech perception. Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 48(2), 320-333.

Ellis, H.D. (1989). Processes underlying face recognition. In R. Bruyer (Ed.), Neuropsychology of face perception and facial expression. New Jersey: Erlbaum.

Fodor, J. A. (1983). Modularity of mind. Cambridge, MA: Bradford Books.

Green, K.P. (1994). The influence of an inverted face on the McGurk effect. Poster presented at the Spring, 1994, meeting of the Acoustical Society of America, Cambridge, Massachusetts. Journal of the Acoustical Society of America, 95, 3014 (Abstract).

Green, K.P., Kuhl, P.K., Meltzoff, A.M., & Stevens, E.B. (1991). Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk effect. Perception & Psychophysics, 50, 524-536.

Jordan, T.R., & Bevan, K. (1997). Seeing and hearing rotated faces: Influences of facial orientation on visual and audio-visual speech recognition. Journal of Experimental Psychology: Human Perception and Performance, 23(2). 388-403.

Lander, K. Rosenblum, L.D., & Bruce, V. (In preparation.) Recognizing 'Thatcherized' faces: Influences of inverted and dynamic presentations. To be submitted to Journal of Experimental Psychology: Human Perception and Performance.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. Cognition, 21(1), 1-36.

MacLeod, A. and Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. British Journal of Audiology, 21, 131-141.

Massaro, D. W. (1987). Speech perception by ear and eye. In B. Dodd & R. Campbell (Eds.), Hearing by eye: The psychology of lip-reading (pp. 53-83). London: Lawrence Erlbaum Associates, Inc.

Massaro, D.W. & Cohen, M.M. (1996). Perceiving speech from inverted faces. Perception & Psychophysics, 58, 1047-1065.

McGrath, M. (1985). An examination of cues for visual and audio-visual speech perception using natural and computer-generated faces. Ph.D. thesis, University of Nottingham, England.

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

Mills, A.E. (1987). The development of phonology in the blind child (pp. 145-162). In B. Dodd and R. Campbell Eds, Hearing by eye: The psychology of lip reading. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Montgomery, A.A. & Jackson, P.L. (1983). Physical characteristics of the lips underlying vowel lipreading performance. Journal of the Acoustical Society of America, 73, 2134-2144.

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. Journal of the Acoustical Society of America, 85(1), 365-378.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. Psychological Science, 5(1), 42-46.

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. Journal of Experimental Psychology: Learning, Memory, and Cognition, 19(2), 309-328.

Parks, T. E. (1983). Letter to the editor. Perception, 12, 88.

Parks, T.E., Coss, R.G., & Coss, C.S. (1985). Thatcher and the Cheshire cat: context and the processing of facial features. Perception, 14, 747-754.

Reisberg, D., Mclean, J., & Goldfield, A. (1987). Easy to hear but hard to understand: A lipreading advantage with intact auditory stimuli. In B. Dodd & R. Campbell (Eds.), Hearing by ear and eye: The psychology of lipreading. Hillsdale, New Jersey: Lawrence Earlbaum Associates, Inc.

Remez, R. E., Fellowes, J. M. & Rubin, P. E. (1995). Perceiving the sex of a talker without natural voice timbre. Barnard College Technical Report: Speech Perception Laboratory, November issue.

Remez, R. E., Fellowes, J. M. & Rubin, P. E. (1997). Talker identification based on phonetic information. Journal of Experimental Psychology: Human Perception and Performance, 23 (n3). 651-666

Rhodes, G., Brake, S., & Atkinson, A.P. (1993). What's lost in inverted faces? Cognition, 47. 25-57.

Rock, I. (1974). The perception of disoriented figures. Scientific American, 230(1), 78-85.

Rosenblum, L.D., and Saldaña, H.M. (1992). Discrimination tests of visually-influenced syllables. Perception and Psychophysics. 52, 461-473.

Rosenblum, L.D., & Saldaña, H.M. (1996). An audiovisual test of kinematic primitives for visual speech perception. Journal of Experimental Psychology: Human Perception and Performance, 22, 318-331.

Rosenblum, L.D., & Saldaña, H.M. (1998). Time-varying information for visual speech perception. To appear in R. Campbell, B. Dodd, D. Burnham (Eds), Hearing by Eye: Part 2, The Psychology of Speechreading and Audiovisual Speech. Earlbaum: Hillsdale, NJ.

Rosenblum, L.D., Saldaña, H. M. (1992). Discrimination tests of visually influenced syllables. Perception & Psychophysics, 52(4), 461-473.

Saldaña, H.M. and Rosenblum, L.D. (1993). Visual influences on auditory pluck and bow judgments. Perception and Psychophysics. 54 (3), 406-416.

Saldaña, H.M. and Rosenblum, L.D. (1994). Selective adaptation in speech perception using a compelling audiovisual adaptor. Journal of the Acoustical Society of America, 95 (6) 3658-3661.

Scapinello, K. F.; Yarmey, A. D. (1970). The role of familiarity and orientation in immediate and delayed recognition of pictorial stimuli. Psychonomic Science, 21(6), 329-331.

Sergent, J. (1982). About face: Left-hemisphere involvement in processing physiognomies. Journal of Experimental Psychology: Human Perception and Performance, 8, 1-14.

Sergent, J. (1984). An investigation into component and configural processes underlying face perception. British Journal of Psychology, 75, 221-242.

Sheffert, S.M. & Fowler, C.A. (1995). The effects of voice and visible speaker change on memory for spoken words. Journal of Memory and Language, 34, 665-685.

Smeele, P.M.T., Massaro, D.W., Cohen, M.M., and Sittig, A.C., (1998). Laterality in visual speech perception. Journal of Experimental Psychology: Human Perception and Performance, 24 (4), 1232-1242.

Summerfield, Q., MacLeod, P., McGrath, M., & Brooke, N.M. (1989). Lips, teeth, and the benefits of lipreading. In Handbook of Research on Face Processing, Young and Ellis (eds.) Elsevier, 1989, pp. 223-233.

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

Tanaka, J.W. & Farah, M.J. (1993). Parts and wholes in face recognition. Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 46A, 225-245.

Thompson, P. (1980). Margaret Thatcher: A new illusion. Perception, 9, 483-484.

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. British Journal of Psychology, 79(4), 471-491.

Valentine, T., & Bruce, V. (1988). Mental rotation of faces. Memory & Cognition, 16(6), 556-566.

Walker, S., Bruce, V., & O'Malley, C. (1995). Facial identity and facial speech processing: Familiar faces and voices in the McGurk effect. Perception & Psychophysics, 57, 1124-1133.

Yakel, D.A., Rosenblum, L.D., & Fourtier, M.A. (under revision). Effects of talker variability on speechreading. Submitted to Perception & Psychophysics.

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

Young, A.W. (1984). Right hemisphere superiority for recognizing the internal and external features of famous faces. British Journal of Psychology, 75, 161-169.

Young, A. W., Hay, D. C., McWeeny, K. H., Ellis, A. W., & Barry, C. (1985). Familiarity decisions for faces presented to the left and right cerebral hemispheres. Brain and Cognition, 4, 439-450.

## Author Notes

We gratefully acknowledge the assistance of Chad Audet, Chantel Bosely, and Rebecca Vasquez as well as the helpful comments of Vicki Bruce, Tim Jordan, Dominic Massaro, Dan Ozer, and the UCR Cognitive Science group. During revision of this manuscript, the third author,

Kerry Green, passed away. The first two authors would like to dedicate this paper to Kerry, valued colleague and friend.

This research was supported by NSF Grants DBS-9212225 and SBR-9617047 awarded to the first author and a Research and Training grant P60 DC-01409 from the National Institute on Deafness and Other Communication Disorders to the National Center for Neurogenic Communication Disorders awarded to the third author.

Requests for reprints should be sent to Lawrence D. Rosenblum, Department of Psychology, University of California, Riverside, Riverside, California, 92521, rosenblu@citrus.ucr.edu.

## Footnotes

1 Whereas the original Thatcher illusion inverted the mouth *and* eyes relative to the facial frame, Parks, et al. (1985) found that the illusion worked well when the mouth alone was inverted. We chose to invert only the mouth in our stimuli because it was important that our subjects concentrate on the mouth area and not be distracted by distortions of extraneous features.