Detection of Audiovisual Speech Correspondences Without Visual Awareness
Agnès Alsius and Kevin G. Munhall
Psychological Science 2013 24: 423 originally published online 5 March 2013
DOI: 10.1177/0956797612457378

The online version of this article can be found at:
http://pss.sagepub.com/content/24/4/423

Published by:
SAGE
http://www.sagepublications.com

On behalf of:
Association for Psychological Science

Additional services and information for Psychological Science can be found at:
Email Alerts: http://pss.sagepub.com/cgi/alerts
Subscriptions: http://pss.sagepub.com/subscriptions
Reprints: http://www.sagepub.com/journalsReprints.nav
Permissions: http://www.sagepub.com/journalsPermissions.nav

>> Version of Record - Apr 15, 2013

OnlineFirst Version of Record - Mar 5, 2013

What is This?
The ability to combine information presented in different sensory modalities into unified percepts is an important aspect of an organism’s successful interaction with the external world. Psychophysical studies have demonstrated that multisensory integration enhances perceptual sensitivity, reduces ambiguity, and shortens reaction time (RT; Bolognini, Frassinetti, Serino, & Ladavas, 2005; Frassinetti, Bolognini, & Ladavas, 2002). For instance, a synchronous sound can improve detection of a flash (McDonald & Ward, 2000; Noesselt, Bergmann, Hake, Heinze, & Fendrich, 2008), increase the perceived luminance of a foveal light (Stein, London, Wilkinson, & Price, 1996), and influence the perceptual organization of simple naturalistic visual scenes (Sekuler, Sekuler, & Lau, 1997; Shams, Ivaki, Chawala, & Bhattacharya, 2005; Vroomen & de Gelder, 2000) and complex naturalistic visual scenes (Arrighi, Marini, & Burr, 2009; Kim, Kroos, & Davis, 2010). Despite the many studies illustrating the consequences of multisensory integration at the behavioral and neural levels, the actual mechanisms of such cross-modal enhancement and the stages of processing at which it occurs remain largely unknown. In our experiments reported here, we explored whether conscious awareness of visual sensory signals is required for audiovisual integration to occur, or whether the visual and auditory signals interact to some extent at a preconscious level (i.e., prior to conscious visual sensory processing).

Some recent studies have indicated that when an audiovisual speech stimulus is presented, the observer’s awareness of the visual input (Munhall, ten Hove, Brammer, & Paré, 2009) or acoustic input (Tuomainen, Andersen, Tiippana, & Sams, 2005) is a key factor influencing audiovisual integration. These studies have involved the McGurk illusion (McGurk & MacDonald, 1976), in which the perception of an otherwise intelligible speech sound is dramatically altered if the observer is watching a phonologically conflicting visual (lip movement) utterance. To study whether audiovisual integration arises only when the visual stimulus is consciously processed as a face, Munhall et al. (2009) used a bistable percept, a dynamic version of Rubin’s (1915) classic face-vase illusion in which the face moved and seemingly articulated a syllable. The results showed that a simultaneous sound was more likely to be integrated with the visual speech stimulus when the face, rather than the vase, was consciously perceived as the dominant figure. Tuomainen et al. (2005) combined pseudowords synthesized as sine-wave speech with visual information from speaking lips. In sine-wave speech, the formants of the original speech are replaced with time-varying sinusoids; only...
observers informed of its speechlike nature perceive a sine-wave stimulus as speech. Tuomainen et al. found that incongruence between the visual and auditory information influenced categorization of the sine-wave speech (the McGurk effect) only when the sounds were heard as actual speech, whereas the visual influence was negligible when participants were not aware of the speechlike nature of the stimuli.

In contrast to the results of these studies, mounting neural evidence indicates that audiovisual interactions can take place at very early stages of processing, in primary sensory cortices (Shams et al., 2005; Watkins, Shams, Tanaka, Haynes, & Rees, 2006) or even subcortically (see Driver & Noesselt, 2008, for a review). Given that such cross-modal effects arise early (Klucharev, Möttönen, & Sams, 2003), it seems possible that at least some intersensory interactions occur at a pre-conscious level of processing. The detectability of simple visual signals can be enhanced by the presence of simultaneous task-irrelevant sounds (e.g., Bolognini et al., 2005). Auditory information also heightens perceptual awareness of visual stimuli (Frassinetti, Bolognini, Bottari, Bonora, & Ladavas, 2005) by, for instance, preventing adaptation to the visual information (Sheth & Shimojo, 2004) or extending the duration of a congruent visual scene’s dominance in binocular rivalry (Alais, van Boxtel, Parker, & van Ee, 2010; Chen, Yeh, & Spence, 2011; Conrad, Bartels, Kleiner, & Noppeney, 2010; van Ee, van Boxtel, Parker, & Alais, 2009; see Schwartz, Grimault, Hupé, Moore, & Pressnitzer, 2012, for a recent review on multistable perception and cross-modal binding). Finally, information perceived without awareness in one modality can sometimes bias the perceptual experience of sensory information presented in another modality, leading to some classic multisensory illusions (Bertelson, Francesco, Ladavas, Vroomen, & de Gelder, 2000; Dufour, Touzalin, Moessinger, Brochard, & Desprès, 2008; Leo, Bolognini, Passamonti, Stein, & Ladavas, 2008). Together, these findings suggest that intersensory interactions, at least for simple audiovisual pairings, can occur at early levels of processing, even when the visual or auditory modality is blocked from awareness.

Simple multisensory stimuli commonly used in multisensory experiments, such as beeps and light flashes, differ from naturally occurring events in a number of ways; for instance, they are brief and have minimal internal structure. The observer combines such stimuli on the basis of their similar onset timing, close proximity, or both, creating an artificial event. More-complex natural stimuli have inherent dynamics in each modality that reflect the source of the stimuli, and these dynamics must be matched for multisensory integration to occur. It is notable that for natural stimuli, such as spoken language, variables such as onset timing and location are of limited importance for integration (Munhall & Vatikiotis-Bateson, 2004).

The question of whether conscious awareness of sensory signals is necessary for the integration of complex natural signals remains unresolved. Within the speech domain, a few studies have shown that audiovisual conditions lower the detection threshold for masked auditory stimuli. For instance, Grant and Seitz (2000) showed that the ability to detect the presence of auditory speech in a background of noise was improved when observers also saw a face articulating the same utterance. Kim and Davis (2004) partially replicated this finding with a larger stimulus sample. However, both of these studies had limitations. Grant and Seitz used only three sentences repeatedly in their entire experiment, and Kim and Davis’s study contained no audiovisual control condition in which the auditory and visual information were incongruent.

The aim of the present study was to investigate whether the detectability of a complex visual signal, preconsciously processed, can be improved by a concurrent acoustic stream. The use of a visual detection task rather than an auditory detection task is advantageous because the visual signal in an audiovisual event characteristically precedes the auditory signal in time and is thought to have a priming effect on auditory processing (Sánchez-García, Alsius, Enns, & Soto-Faraco, 2011). Thus, local expectancy effects were minimized in our design. An interocular suppression technique called continuous flash suppression (CFS; Tsuchiya & Koch, 2005) was used to keep the visual stimulus (a talking face) outside of participants’ conscious awareness. In CFS, arrays of random overlapping rectangles of varying sizes and colors are presented successively to one eye at a fast pace, reliably suppressing from consciousness an image presented to the corresponding location of the other eye.

Several studies using the CFS paradigm have demonstrated that both low-level properties of visual stimuli, such as coherence in motion (Conrad et al., 2010; Yamada & Kawabe, 2011), and high-level aspects of visual stimuli can be processed during interocular suppression (Jiang, Costello, & He, 2007). For instance, the right fusiform face area is activated even when faces are rendered consciously invisible through CFS (Jiang & He, 2006).

In the current experiments, the articulatory movements of the suppressed talking face were either matched or mismatched with an auditory sentence. In Experiment 1, we compared how long it took for the talking face to overcome interocular suppression and become visible to participants in the matching and mismatching conditions (see Mudrik, Breska, Lamy, & Deouell, 2011, and Yang, Zald, & Blake, 2007, for similar paradigms). Differences between RTs in the two conditions would suggest that audiovisual correspondences can be detected before visual information is consciously perceived. In Experiment 2, to avoid anticipatory responses and ensure that participants were following the instructions, we followed the same procedure but included some catch trials in which the face was not presented. Finally, in a control experiment (Experiment 3), we excluded the possibility that any difference in performance between the matching and mismatching conditions was explained by representational dissociations occurring at the first stages of partial awareness of the stimuli (i.e., when the talking face started emerging into dominance,
rather than during effective suppression). If this were the case, faster RTs in one condition, compared with the other, could be attributed to differences in recognition speeds, motor responses, or response criteria (Kouider & Dehaene, 2007). Thus, in Experiment 3, we did not use interocular suppression; instead, the visual stimuli were digitally edited, and a gradual transition between the visual noise and the talking face (mimicking perceptions in Experiment 2) was presented.

**Method**

**Participants**

Experiment 1 had 16 participants (13 females, 3 males; mean age = 19.8 years), Experiment 2 had 37 participants (29 females, 8 males; mean age = 19.6 years), and Experiment 3 had 37 participants (28 females, 9 males; mean age = 20 years). All subjects were native speakers of English and reported having normal vision and no speech or hearing difficulties. The project was approved by the local ethics committee. Participants received monetary compensation.

**Apparatus and data acquisition**

The study was conducted in a dark sound booth. Stimuli were presented in stereo using a head-mounted display (HMD) with integrated headphones (nVisor SX60, NVIS, Reston, VA). The HMD contained two independent LCD screens, one for each eye, with a simultaneous refresh rate of 60 Hz. To calibrate the two displays, we measured the voltage drop across a photo resistor taped alternately onto each eyepiece; 16 levels of luminance equally spaced across the luminance range were individually adjusted until the voltage drops matched. The same photo resistor was used for all measurements, and the results were confirmed through several readings, with eyepieces alternated between measurements. The HMD was attached to a PC (Intel Core, Santa Clara, CA) running custom software that presented the stimuli synchronously to the two eyes. Video was controlled by a GeForce 8400 GS graphic card (NVIDIA, Santa Clara, CA). Screen tearing was avoided by rendering frames for both eyes to a back buffer and then presenting them upon screen refresh (i.e., page flipping). Responses were obtained with a response box connected to the PC. RTs (time from stimulus onset to when participants indicated they saw the face) were independently computed for each trial. Auditory speech stimuli were delivered through the headphones at an intensity of 73 dB(A) SPL.

**Stimuli**

The suppressors used in Experiments 1 and 2 were colorful, dynamic noise masks (i.e., overlapping rectangles, each measuring 0.83° × 0.83°) presented at a frequency of approximately 30 Hz. For all three experiments, audiovisual stimuli were prepared from digital video recordings of a female speaker (showing the lower part of the face) articulating a set of 84 English sentences. The recorded utterances were edited to last 6 s using Adobe Premiere 6.0 software. The videos (720 × 576 pixels at 29.97 frames/s; played using the Cinepack codec, Microsoft, Redmond, WA) were then converted to gray scale.

To ensure effective suppression of the face at the beginning of each trial, we reduced the contrast of the face by 84%, centered on the middle value of the luminance space. The contrast ramping during the trials was accomplished by computing the luminance of each pixel in each frame and scaling it by the contrast adjustment percentage (0.13%) away from the center of the luminance space. Therefore, in each frame, the contrast increased by 0.13%, so that at the end of the trial, the contrast of the face was 60% relative to the contrast of the original image.

The suppressor was first presented at full contrast; after the initial 1,000 ms, the contrast was decreased linearly by −0.6% per frame, so that by the end of the trial, the contrast had been reduced by 90%. The eye viewing the face and the eye viewing the suppressor were counterbalanced across trials and randomly intermixed throughout the experiment. Both the image of the talking face and the suppressor were circumscribed by a black rectangular border (19° × 16°).

In the trials that presented matching audiovisual information (50%), the visual sentence was presented together with its original auditory recording. To equate the amount of visual information in the matching and mismatching trials, and to ensure that any difference between conditions was not explained by low-level stimulus differences (such as motion rate or local contrast differences), we used the same visual sentences in both conditions. The mismatching stimuli (50% of trials) were created by taking an original audiovisual stimulus and replacing the auditory sentence with another sentence from the pool of 84 sentences. Each auditory sentence was presented only once to each participant.

For the catch trials in Experiments 2 and 3, we created a solid-gray video frame by computing the average luminance of the individual pixels that made up the face stimulus; the frame was presented instead of the face. Catch trials (22% of the trials in Experiments 2 and 3, 16 sentences) were randomly intermixed with genuine trials.

Experiment 3 was designed to simulate the visual stimuli in Experiment 2 but did not involve interocular suppression. Instead, the video recordings of talking faces used in Experiment 2 (including the contrast ramping) were digitally blended into the visual noise mask using the cross-dissolve transition function of Adobe Premiere 6.0, and the output displays were presented to both eyes (thus, on each trial participants saw a smooth transition between the visual noise and the face). This condition has been used in previous binocular-rivalry studies to control for partial awareness (Jiang et al., 2007; Mudrik et al., 2011), but it does not completely match the perceptual experience of rivalry in every aspect. Thus, the absolute magnitude of the RT is not expected to be the same in this condition as in the binocular-rivalry condition. Jiang et al. adjusted...
the contrast’s rate of change for their control condition to bring the RT in that condition artificially in line with the RT for their binocular-rivalry condition. Because absolute RT was not our concern, we did not make such adjustments. Partial awareness was possible, in principle, with the superimposed stimuli in this control condition. Thus, Experiment 3 allowed us to rule out partial awareness as an explanation for our results.5

If congruency effects were the result of processing occurring after the visual stimuli had already begun to overcome suppression, similar differences in RTs between the matching and mismatching trials would be expected in the CFS experiments (Experiments 1 and 2) and the control experiment (Experiment 3). Any effects occurring at the first stages of partial awareness of the face were expected to affect RTs equally in all experiments. Finding differences between the matching and mismatching trials only in the CFS experiments would allow us to reject the partial-awareness interpretation.

Procedure

Before we started the experimental sessions, we adjusted the HMD individually for each participant to account for interocular differences and to ensure correct visibility of both screens. Participants were instructed to close one eye at a time while reading a short sentence and to verbally report any difference in visual clarity between their two eyes. After this calibration phase, participants first completed eight practice trials (four matching, four mismatching) to familiarize themselves with the experimental paradigm. None of the auditory or visual sentences presented in the practice trials were used in the experimental trials. Effective suppression was assessed after the practice trials by participants’ subjective reports and was confirmed later with RT data. Participants who consistently saw the face at the beginning of the trials were excluded from analyses.

Figure 1 shows a schematic representation of the audiovisual stimuli used in the three experiments. Participants were instructed to maintain stable fixation at the center of the screen throughout the trials (no fixation point was presented) and to press a button on the response box as soon as they saw any part of the face. They were also instructed to ignore the auditory information. The recorded RT served as an index of the time required to bring the face into conscious perception. Participants were instructed to not press the button if they did not see a face (catch trials). After each trial, the experimenter triggered the onset of the following trial by pressing a key on the computer keyboard.

![Fig. 1. Illustration of the stimulus sequence in the three experiments. Experiments 1 and 2 used continuous flash suppression (CFS): Participants were presented with a dynamic, colorful noise mask to one eye and a video of a talking face to the other eye; the contrast of the mask decreased across frames, while the contrast of the face increased. In Experiment 3, a control experiment, the visual stimuli from Experiment 2 were digitally blended into the visual noise mask, and the output display was presented to both eyes, so participants saw a smooth transition between the visual noise and the face. Participants were instructed to press a button as soon as any part of the face became visible. In matching trials, the visual sentence stimulus was accompanied by the auditory recording of the same sentence. In mismatching trials, the visual stimulus was accompanied by the auditory recording of another sentence, selected randomly.](image-url)
Experiment 1 included 56 genuine trials (no catch trials), divided into two blocks of 28 trials; Experiments 2 and 3 each included 56 genuine trials plus 16 catch trials, divided into two blocks of 36 trials.

Results

Figure 2 summarizes overall RTs as a function of stimulus type (matching, mismatching) in the three experiments. RTs longer than 6 s (0.02%, 0.008%, and 0.002% of trials in Experiments 1, 2, and 3, respectively) were excluded from further analyses. In Experiment 1, a two-tailed paired t test revealed a significant effect of stimulus type on RT, \( t(15) = -2.83, p = .01, d = -0.708 \); average RT was shorter in the matching condition \((M = 2,629 \text{ ms})\) compared with the mismatching condition \((M = 2,764 \text{ ms})\).\(^5\) Experiment 2 replicated this effect: Even though RTs were longer overall than in Experiment 1 (possibly because participants performed an extra perceptual visual verification process to avoid false alarms in catch trials), there was still a reliable difference between the matching \((M = 2,898 \text{ ms})\) and mismatching \((M = 2,988 \text{ ms})\) conditions, \( t(36) = -2.10, p < .05, d = 2.12 \). The fact that congruency between auditory and visual information modulated RTs suggests that audiovisual correspondence is detected even when the visual information is suppressed and therefore preconsciously processed.

The analysis of participants’ RTs revealed substantial individual differences across participants, with some participants detecting the face early in the sequence and others much later. Note, however, that 12 (75%) of the 16 subjects in Experiment 1 and 24 (65%) of the 37 subjects in Experiment 2 had a shorter RT for the matching stimuli, compared with the mismatching stimuli, and there was no systematic pattern relating this difference in RT to how early the recognition of the face occurred (i.e., how early or late participants detected the face did not predict the difference between their RTs on matching trials and their RTs on mismatching trials).

In contrast to the results for the CFS experiments, the results for Experiment 3, the control experiment, did not reveal differences in RTs between the matching \((M = 4,550 \text{ ms})\) and mismatching \((M = 4,547 \text{ ms})\) conditions, \( t(36) = 0.30, p = .766, d = 0.313 \). Furthermore, an analysis of variance with experiment (Experiment 2, Experiment 3) as a between-participants factor and stimulus type (matching, mismatching) as a within-participants factor revealed a significant Experiment \( \times \) Stimulus Type interaction, \( F(1, 72) = 4.49, p < .05 \), thus reinforcing the finding that between-condition differences in performance in the CFS experiments are not likely to have been caused by partial awareness. In Experiments 2 and 3, subjects were very accurate in withholding responses in the catch trials \((M = 97\% \text{ and } 95\%, \text{ respectively})\), \( F(1, 73) = 2.76, p = .10, d = 0.386 \).

Discussion

The main finding of these three experiments is that a complex visual stimulus rendered invisible by interocular suppression can emerge into conscious perception more quickly when it is combined with a correlated acoustic signal. That is, congruency between auditory and visual information affected the time the visual stimulus remained suppressed, with matching visual speech stimuli consistently breaking through interocular suppression more quickly than mismatching stimuli did. The effect cannot be attributed to differences between the matching and mismatching visual stimuli (e.g., differences in contrast or motion speed) because the same stimuli were used in the two conditions. Furthermore, the effect is unlikely to be accounted for by partial awareness because no differences in RTs between the matching and mismatching trials were found in the control experiment (Experiment 3).

Together, our results extend to the cross-modal binding domain previous findings showing that visual grouping can occur preconsciously (e.g., Mitroff & Scholl, 2005). Our results also suggest that audiovisual interactions transpire at relatively early levels of processing, where neural information associated with the suppressed visual stimuli is still registered. These findings add to growing evidence that the detectability of a visual stimulus undergoing conscious suppression can be enhanced by a concurrent sound (Alais et al., 2010; Bolognini et al., 2005; Chen et al., 2011; Conrad et al., 2010; Frassinetti et al., 2002; Frassinetti et al., 2005; Sheth & Shimojo, 2004; van Ee et al., 2009) and show that such cross-modal gains can generalize to real-world, naturalistic stimuli that change dynamically over time.

Previous studies examining whether awareness is a critical factor for the audiovisual integration of speech stimuli have yielded mixed results (Grant & Seitz, 2000; Kim & Davis, 2004; Munhall et al., 2009; Tuomainen et al., 2005), possibly because of differences in tasks (e.g., identification of phonetic
content vs. detection of the presence of speech). Whereas studies that required phonetic categorization of the speech stimuli (Munhall et al., 2009; Tuomainen et al., 2005) have pointed to a crucial role of consciousness in audiovisual integration, studies that involved detection tasks (including the present study) have provided evidence for preconscious binding of correlated cross-modal information.

These differing results are not necessarily incompatible, however. The auditory and visual sensory systems interact at multiple levels of processing during speech perception (see Eskelund, Tuomainen, & Andersen, 2011; Hertrich, Mathiak, Lutzenberger, Menning, & Ackermann, 2007; Klucharev et al., 2003). It is possible, therefore, that the encoding of phonetic information requires integration over a longer time scale than the initial interaction of vision and audition requires (Munhall et al., 2009). Indeed, electroencephalography studies have shown that the earliest interactions between the auditory and visual properties of an audiovisual stimulus occur approximately 85 ms after stimulus onset, whereas the first phonetic audiovisual interactions are observed 155 ms after onset of an auditory stimulus (Klucharev et al., 2003). Further evidence for the multistage account of audiovisual speech integration has come from studies showing that brain areas responsive to the processing of cross-modal coincidence of physical events do not overlap with those involved in cross-modality fusion (Miller & D’Esposito, 2005). Still more evidence has come from studies showing that the stage underlying the audiovisual detection advantage is not specific to speech, whereas the stage in which the phonetic information of an audiovisual stimulus is encoded (e.g., the McGurk illusion) is specific to speech (Eskelund et al., 2011). Therefore, we believe it is possible that early audiovisual interactions (e.g., early detection of audiovisual spatiotemporal correspondences) occur prior to conscious visual awareness, but that audiovisual interactions affect phonetic categorization only after the conscious processing of sensory signals (see Munhall et al., 2009, and Tuomainen et al., 2005).

The observed detection benefits for audiovisual stimuli could be derived from structural coupling between auditory and visual events resulting from the inherent relationship between acoustic speech sounds and their visible generators (Fowler, 2004). That is, in addition to matching in phonological content, the concurrent streams of auditory and visual speech information are usually invariant in their onsets and offsets, rhythmic patterning, duration, intensity, and overall amplitude contour (Chandrasekaran, Trubanova, Stillittano, Caplier, & Ghazanfar, 2009). Mouth movements (e.g., peak opening and closing) correlate with peaks in the acoustic waveform, and the visual kinematics correspond with the timing of changes in the auditory envelope in great detail (see Yehia, Rubin, & Vatikiotis-Bateson, 1998). This detailed correspondence appears to be important for audiovisual speech integration.

Although in our study, the visual stimuli were identical in the matching and mismatching audiovisual conditions, variations in the timing of changes in the mouth aperture and in the onsets and offsets of movement were temporally related to the patterning of the acoustical energy in the matching condition but not in the mismatching condition. It is plausible, therefore, that such comodulation of the auditory and visual signals in the matching condition was registered by systems responsible for establishing cross-modal correspondence (Bernstein, Auer, & Takayanagi, 2004; Eskelund et al., 2011; Grant & Seitz, 2000; Kim & Davis, 2004; Kim et al., 2010). Increased brain activations in response to temporally corresponding audiovisual streams have been found in a broadly distributed neural system encompassing the superior colliculus, fusiform gyrus, lateral occipital cortex, extrastriate visual cortex, and superior temporal regions (Stevenson, Altieri, Kim, Pisoni, & James, 2010). The results of the present study suggest that at least some of these areas may be activated even when visual information is preconsciously processed. The activation in these areas would subsequently amplify the neural response in visual regions responsible for the perceptual representation of the face, enhancing the signal-to-noise ratio derived from visual rivalry. Such amplification of the neuronal response in visual cortices is possibly mediated by bidirectional cortical interactions between higher-order association regions and early sensory areas (Schroeder & Foxe, 2002) or by lateral circuitry between the auditory and visual regions (Falchier, Clavagnier, Barone, & Kennedy, 2002).

Because the auditory stimuli were comprehensible to the participants in our study, perhaps they used high-level linguistic information (semantic, syntactic, lexical, or phonological) to generate probabilistic hypotheses and thus anticipate representations of the visual words in advance of their appearance (Sánchez-García et al., 2011). Participants may also have used visual imagery to predict the upcoming visual signal (see Pearson, Clifford, & Tong, 2008). Chen et al. (2011), for example, showed that when subjects were instructed to attend to an audio soundtrack, the soundtrack’s semantic content influenced their visual awareness of static pictures in a binocular-rivalry paradigm.

However, we do not believe that the effects found in our studies can be explained solely by such high-level expectancies. First, participants were explicitly instructed to ignore the auditory information, and the auditory sentence in a given trial was not predictive of the type of trial being presented (22% of the trials were catch trials, and the remaining trials were equally split between the matching and mismatching conditions). However, it is possible that participants in fact did not ignore the irrelevant auditory information and used it to anticipate the upcoming mouth movements. Note, however, that if this were the case, the same expectancy effects likely would have resulted in faster responses in the matching condition than in the mismatching condition of the control experiment, but they did not.

Second, it has been shown that even in the case of a language that participants do not know, visually guided enhancements of auditory speech perception can occur when the masked auditory
signals are accompanied by visual sentences (Kim & Davis, 2003); presumably, such visual enhancements of auditory stimuli are due to local correspondences between the two signals. Further research examining whether an effect similar to the one reported here arises when the suppressed speech is in an unknown language will allow researchers to determine whether the enhancement of the visual signal occurs at a prephonological level or at higher levels of linguistic processing.

In conclusion, our findings constitute novel evidence that a temporally dynamic, complex visual speech stimulus that is only preconsciously processed can be integrated with suprathreshold auditory information, and that such integration heightens the robustness of residual signals associated with the suppressed visual speech stimulus. Although the neural mechanism that mediates the enhancement of the visual signal remains open for investigation, this behavioral finding suggests that acoustic signals modulate brain activity at an early stage of visual processing, possibly by amplifying the visual signal and thus facilitating its conscious processing (Shams et al., 2005).

Declaration of Conflicting Interests
The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding
This research was supported by the Natural Sciences and Engineering Research Council of Canada. A. A. was supported by a postdoctoral fellowship from Spain’s Ministry of Education.

Notes
1. All visual-angle measurements in this article are based on the specifications of the product and varied slightly among participants because of individual adjustments.
2. Image manipulations were based on the luminance of the original images, and therefore only relative values (i.e., percentages) can be provided here. The specific ramping percentages we used were established on the basis of previous pilot data regarding the threshold at which the flash stimulus could reliably suppress the face at the beginning of the trial and the contrast at which the face would become visible before the end of the trial.
3. For the visual noise ramping, the same algorithm was applied to reduce the contrast over the course of the trial; the contrast-adjustment percentage (0.6%) reduced the contrast range toward the center of the luminance range. The contrast adjustment for any given frame of a visual sequence can be computed by multiplying the time (in seconds) by the frame duration (29.97 ms) by the step size (0.13 for the face, 0.6 for the noise) and adding or subtracting (as appropriate) the obtained value from the initial contrast adjustment of that stimulus (84% for the face, 0% for the visual noise).
4. The weight of each pixel was gradually changed from 100% of its weight in the visual noise stream (initial frame) to 100% of its weight in the talking-face stream (final frame).
5. This interpretation, however, rests on the assumption that the time window over which the partial-awareness effects would be observed (the near-threshold stimulation) was similar in the control and the CFS conditions. Note, however, that this may not necessarily have been the case. The rate of transition between dominance of the flash stimulus and the face could have been shorter in the control condition because of the very straightforward changes in stimulus contrast that occurred; in comparison, the transitions in the CFS condition were much more variable.
6. All d values were corrected for dependence between means, using Morris and DeShon’s (2002) Equation 8.

References
Fowler, C. A. (2004). Speech as a supramodal or amodal phenomenon. In G. Calvert, C. Spence, & B. E. Stein (Eds.), The


