Exp Brain Res (2016) 234:1325–1332
DOI 10.1007/s00221-016-4591-y

RESEARCH ARTICLE

When audiovisual correspondence disturbs visual processing

Sang Wook Hong[1,2] · **Won Mok Shim**[3,4,5]

Received: 6 May 2015 / Accepted: 30 January 2016 / Published online: 16 February 2016

Abstract Multisensory integration is known to create a more robust and reliable perceptual representation of one's environment. Specifically, a congruent auditory input can make a visual stimulus more salient, consequently enhancing the visibility and detection of the visual target. However, it remains largely unknown whether a congruent auditory input can also impair visual processing. In the current study, we demonstrate that temporally congruent auditory input disrupts visual processing, consequently slowing down visual target detection. More importantly, this cross-modal inhibition occurs only when the contrast of visual targets is high. When the contrast of visual targets is low, enhancement of visual target detection is observed, consistent with the prediction based on the principle of inverse effectiveness (PIE) in cross-modal integration. The switch of the behavioral effect of audiovisual interaction from benefit to cost further extends the PIE to encompass the suppressive cross-modal interaction.

✉ Sang Wook Hong
shong6@fau.edu

Won Mok Shim
Won.Mok.Shim@dartmouth.edu

1 Department of Psychology, Florida Atlantic University, 777 Glades Rd., BS 12, Room 209, Boca Raton, FL 33431, USA

2 Center for Complex Systems and Brain Sciences, Florida Atlantic University, Boca Raton, FL, USA

3 Department of Psychology and Brain Sciences, Dartmouth College, HB 6207 Moore Hall, Hanover, NH 03755, USA

4 Center for Neuroscience Imaging Research, Institute for Basic Science, Sungkyunkwan University, Suwon, Korea

5 Department of Biomedical Engineering, Sungkyunkwan University, Suwon, Korea

Keywords Multisensory · Audiovisual correspondence · Contrast · Cross-modal inhibition

Introduction

Multisensory integration is known to create a more robust and reliable perceptual representation of one's environment. For instance, temporally congruent auditory inputs can make a visual stimulus more salient irrespective of spatial congruency, thus enhancing the visibility and detection of a target (Fiebelkorn et al. 2011; Lippert et al. 2007; Noesselt et al. 2008; Stein et al. 1996). It was also shown that the response in early visual cortex increases when temporally congruent auditory signals are present (Convento et al. 2013; Romei et al. 2009, 2012). Such cross-modal modulation of neural responses in early visual cortex can be mediated by direct or indirect connections between primary sensory cortex and traditional multisensory areas, such as superior colliculus (SC) and posterior superior temporal sulcus (pSTS), as well as direct connections between primary sensory cortices (see review by Driver and Noesselt 2008). In addition, it was suggested that cortico-geniculate connections also play an important role in facilitatory interactions between congruent auditory and visual stimuli (Noesselt et al. 2010).

This multisensory enhancement has been shown to be most prominent when individual unimodal inputs are weak, which is described by the principle of inverse effectiveness (PIE: Meredith and Stein 1986; Stanford et al. 2005; Stein and Meredith 1993). While the PIE reflects the adaptive characteristic of multisensory interaction and the relationship between the stimulus strength and the effectiveness of multisensory enhancement, cross-modal interaction can also be suppressive. Recent human neuroimaging studies, as well as single-cell recordings in cats, found suppressive audiovisual interactions in multisensory areas, such as SC (Kadunce et al. 1997) and STS (Noesselt et al. 2007; Werner and Noppeney 2010). Further, it was reported that neural responses in primary visual cortex can be reduced by auditory inputs that are synchronously presented with visual inputs (Noesselt et al. 2007). However, the behavioral consequence of such suppressive interactions between auditory and visual stimuli and the effect of unimodal input strength remain unclear. As the strength of unisensory signal increases, the PIE predicts that the facilitatory effects of multisensory integration would decrease, but it does not provide a clear prediction regarding suppressive interactions between congruent inputs.

We hypothesized that suppressive multisensory interactions can occur when the intensity of unisensory inputs is high. However, the perceptual consequence of such suppressive interactions may not be apparent when a visual stimulus itself is salient, due to response saturation to the strong, suprathreshold visual stimulus. In continuous flash suppression (CFS), a target stimulus is presented to one eye, while dynamically changing, high-contrast patterns are presented to the other (Tsuchiya and Koch 2005). CFS is potent to suppress the target stimulus irrespective of its intensity and can thus provide an optimal paradigm to measure the effect of congruent sound on the processing of a high-intensity visual stimulus.

Here, we report a novel perceptual consequence of suppressive multisensory interaction through the use of CFS, in which detection of visual stimuli is impaired by a temporally congruent sound when the intensity of visual stimuli (i.e., luminance contrast) is high. As the PIE predicts, low-contrast visual stimuli detection is enhanced by synchronous sound presentation, compared with low-contrast visual stimuli detection without sound. Our finding suggests that multisensory interaction between temporally congruent unisensory inputs can switch from facilitation to inhibition as the intensity of unisensory signal increases.

General method

Observers

In exchange for course credit, 30 undergraduate students with normal or corrected-to-normal vision participated in both Experiments 1 and 2. All participants provided the informed consent approved by Florida Atlantic University Institutional Review Board.

Experimental rationale

We assessed the effect of temporally congruent auditory input on visual processing using continuous flash suppression (CFS). In CFS, one eye is presented with contour-rich, Mondrian-like patterns that continuously change, while the other eye is presented with a to-be-suppressed static image (Tsuchiya and Koch 2005; Tsuchiya et al. 2006). The continuously flashing patterns prevent the stimulus presented to the other eye (i.e., a grating in the current study) from being consciously perceived. CFS is particularly advantageous for the purpose of the current study, since it allows us to examine the effect of auditory input on the processing of a salient visual stimulus. In CFS, a visual stimulus is suppressed from visual awareness for a period of time from the stimulus onset, irrespective of its intensity. The time elapsed before breakup of interocular suppression has been used as an estimate of the sensitivity to the visual stimuli or strength of the neural representation of the suppressed visual stimuli. For example, high-contrast stimuli tend to overcome suppression faster than low-contrast stimuli (Tsuchiya and Koch 2005). We predict that a visual stimulus with temporally congruent sound will break through suppression faster than the same stimulus with no sound (or incongruent sound) if the temporally congruent sound has a facilitatory effect, but will break up suppression more slowly if the sound has an inhibitory effect. We tested two contrast levels of the visual stimulus in order to determine whether the auditory influence on visual processing can also be suppressive when the strength of the visual stimulus is sufficiently high.

Stimuli

Stimuli were presented on a Sony CPD-G520, 21″ CRT monitor (1024 × 768 pixels, 100 Hz frame rate), and the collection of behavioral responses was controlled by the Psychophysics Toolbox (Brainard 1997; Pelli 1997). Stimuli were presented in a dark room to observers positioned 90 cm from the CRT monitor whose R, G and B guns were calibrated using a light meter (IL-1700) and a luminance meter (Minolta LS100), creating a linearized lookup table (8 bits for each R, G and B guns). A four-mirror stereoscope was used to present stimuli dichoptically. On each trial, dynamically changing Mondrian-like patterns were presented within a 2° × 2° (visual angle) square aperture (with a 2.25° × 2.25° fusion contour) to one eye, and a target stimulus was presented to the other eye (Fig. 1a). Four nonius lines were presented to assure correct fusion of guidelines between the two eyes. The eye to which a target stimulus was presented was determined randomly. Auditory stimuli (pure tone, 3.5 kHz) were created in MATLAB and presented using acoustic noise-canceling headphones (Bose QuietComfort).

The target stimulus was a sinusoidal grating (three cycles/deg) with two different orientations, ±45° from vertical. The grating was masked by a Gaussian envelope

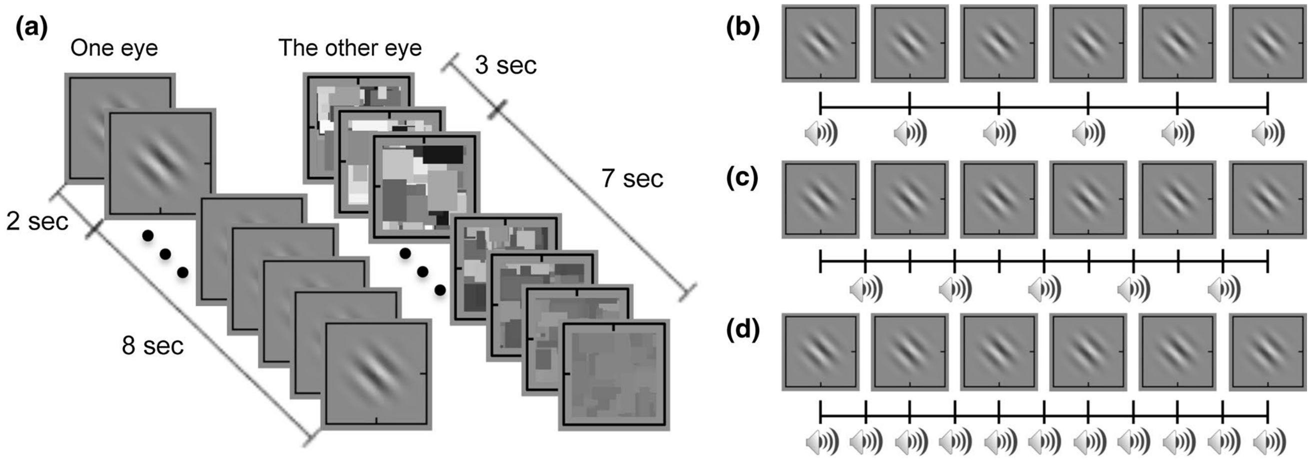

Fig. 1 Stimuli and procedure. **a** Schematic illustration of stimulus presentation in one trial. An oriented target grating was presented to one eye, and continuously changing Mondrian-like patterns (suppressor) were presented to the other. Contrast of the target was ramping up during the initial 2 s, and stayed at its maximum for the remaining 8 s. Overall contrast of the suppressor was at its maximum for the initial 3 s and was ramping down to 0 during the remaining 7 s. **b** Schematic illustration of the in-phase sound condition, in which auditory beeps were presented synchronously with contrast reversal of visual target. **c** Schematic illustration of the out-of-phase sound condition, in which auditory beeps were presented in-between contrast reversal of visual target. **d** Double-sound condition (Experiment 2), in which auditory beeps were presented both in-phase and out-of-phase with contrast reversal of visual target

to fit within a 2° square aperture with a smooth outline. The contrast of the grating was set to 60 % Michelson contrast with a fixed mean luminance (55 cd/m^2) for the high-contrast condition, and 10 % for the low-contrast condition. The target grating was presented at the center of the square apertures to one eye. The contrast of the target grating was reversed either every 1 s or 500 ms, which resulted in flickering with a temporal frequency of .5 or 1 Hz. The suppressors, dynamically changing Mondrian-like patterns, were presented to the other eye. Each suppressor was composed of 200 rectangular patches with random sizes. The luminance of each patch was randomly assigned, but within a predetermined range whose maximum and minimum values were used to compute the contrast of the suppressors. The mean luminance of the suppressors was fixed at 55 cd/m^2, which was identical to the luminance of the background. Sixty Mondrian-like patterns were created and presented in random order every 100 ms (10 Hz). The presentation of the suppressor was not synchronized with either contrast reversal of the visual target or the presentation of sound stimuli in order to avoid a possible effect of phase synchronization between the suppressor and sound.

Three audiovisual conditions for each contrast level were tested in Experiment 1. In the in-phase condition (Fig. 1b), a beep sound was accompanied by every contrast reversal of the target grating. In the out-of-phase condition (Fig. 1c), a beep sound was presented between every contrast reversal (500 ms before and after the contrast reversal in the .5 Hz condition and 250 ms before and after the contrast reversal in the 1 Hz condition). In the silent condition, the target grating was presented without any auditory input. In Experiment 2, only two conditions (double sound vs. silent) were tested. In the double-sound condition, the temporal frequency of sound was twice as the temporal frequency of contrast reversal in the suppressed grating. Thus, the auditory beeps were once presented in-phase with the contrast reversal of the target grating and next presented out-of-phase with the contrast reversal (double-sound condition, Fig. 1d). The silent condition was identical to Experiment 1.

Procedure

During the first two seconds of stimulus presentation, the overall contrast of the target stimulus was increased from 0 to its maximum contrast (either 10 or 60 %), while dark and bright parts of the grating were reversed in the temporal frequency of .5 or 1 Hz (Fig. 1a). This ramping technique was used to guarantee initial suppression of the target grating. The overall contrast then remained constant until the target was detected by breaking up suppression or until the end of the presentation time (total 10 s). The contrast of the Mondrian suppressor stimuli was 60 % for the first three seconds and then slowly decreased to 0 over the remaining 7 s. Reducing the contrast of the suppressor does not guarantee that suppression will be broken during the 10-s stimulus presentation. However, it was expected to decrease the number of trials in which the target stimulus did not break-through suppression, since the relative strength of the target stimulus increased through the reduction in the contrast of the suppressor.

Each trial began with pressing the space bar on the keyboard. Participants were instructed to report the orientation of the target grating as quickly and accurately as possible whenever they could identify the orientation. To report the orientation of the grating, participants pressed the left-arrow key on the keyboard for a left-tilted grating and the right-arrow key for a right-tilted grating. A trial in which an observer incorrectly reported the orientation of the target grating was counted as an error trial. Observers received feedback (a high-tone beep) in the occurrence of an error trial. A trial was terminated either by reporting target orientation (breakup of suppression) or after 10 s (no breakup of suppression). Two temporal frequencies of contrast reversal, two contrast levels and three sound conditions (two sound conditions for Experiment 2) were repeatedly tested 12 times in a completely random order. The response time (RT), defined as the time from stimulus onset to the breakup of suppression, was measured on each trial. The RTs were averaged for each condition and used for further analysis.

Results and discussion

Experiment 1 was conducted to examine whether the effect of temporally congruent sounds on visual processing varies depending on the intensity of visual stimuli. We were particularly interested in whether temporally congruent sound stimuli can inhibit the processing of concurrently presented visual stimuli when the intensity of visual stimuli is high.

A total of nine observers were excluded from the analysis due to the high error rate (over 20 % of trials, seven observers) or high no-breakup rate (over 20 % of trials, two observers). The average error rate of the remaining 21 observers was 9.5 %, and the average no-breakup rate was 4.8 %. The error trials and no-breakup trials were excluded from the analysis. The RTs were subjected to a three-way repeated measures analysis of variance (ANOVA). A three-way ANOVA (temporal frequency × sound congruency × contrast level) revealed that no temporal-frequency-related effect (neither the main effect nor interactions) was significant ($ps > .18$), indicating that the influence of sound on RTs was not affected by the temporal frequency of contrast reversal in visual stimuli. Thus, we aggregated the data from the two temporal frequency conditions and then conducted a two-way ANOVA (sound by contrast level). The RTs for target identification as a function of sound conditions and contrast levels are shown in Fig. 2a. We found a significant main effect of the contrast level of visual stimuli [$F(1,20) = 33.86$, $p < .001$]. The RTs were slower for a low-contrast target, which confirms that a weak visual stimulus breaks up suppression more slowly than a stronger stimulus.

It is notable that the orientation of the target could be identified even before it reached its maximum contrast level in the high-contrast condition (60 %). High-contrast targets broke through CFS after about 1.6 s on average, at which the contrast of the target was still ramping up. Since the contrast of the target stimuli in the high-contrast condition was ramped up from 0 to 60 % over 2 s, the contrast of the target at the mean RT would be higher than 45 %, which is already about the saturation level of contrast for the cortical contrast response function (Albrecht and Hamilton 1982). Furthermore, the transient neural signals caused by contrast-reversing target stimuli would be potent enough to induce early breakup of suppression.

Crucially, we found a significant interaction between the sound conditions and the contrast of the visual stimuli [$F(2,40) = 16.99$, $p < .001$], indicating that the effect of temporal congruency of auditory inputs varies depending on the intensity of visual stimuli. Planned contrast tests for the high-contrast condition revealed that the RTs were slower in the in-phase condition than in the both out-of-phase [$F(1,20) = 13.32$, $p = .002$] and silent conditions [$F(1,20) = 6.17$, $p = .022$] when the contrast of visual stimuli was high (60 %), indicating suppressive audiovisual

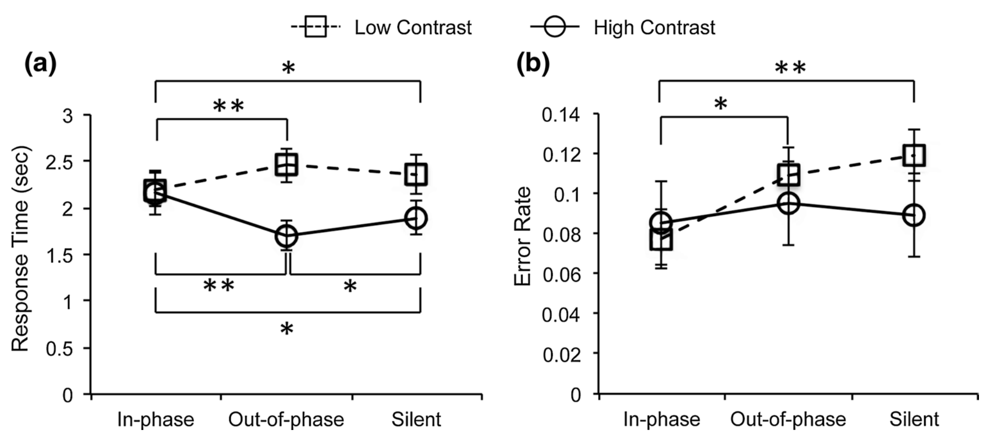

Fig. 2 **a** Results from Experiment 1. *Circles with solid line* represent the response times for the high-contrast condition, and *squares* with a *dashed line* represent the response times for the low-contrast condition. The *x*-axis represents three different sound conditions, and *y*-axis represents the response times (sec). The *error bars* are ±1 standard error (**$p < .01$, *$p < .05$). **b** *Error rates* in Experiment 1

interactions when the intensity of visual stimuli is high. On the contrary, facilitation of visual detection by congruent auditory stimuli was found when the contrast of the visual stimuli was low (10 %). In the low-contrast condition, the RTs in the congruent sound condition were significantly faster than in the both incongruent [$F(1,20) = 38.57$, $p < .001$] and silent conditions [$F(1,20) = 7.61$, $p = .012$]. Faster breakup of suppression with congruent auditory inputs with weak visual signals replicates the well-known inverse effectiveness in multisensory integration. Our results demonstrated that temporal congruency between multisensory inputs could interrupt unimodal sensory processing and that the effect of multisensory integration was manifested differently depending on the intensity of unimodal sensory input.

To test whether the facilitatory effect in the low-contrast condition and the inhibitory effect in the high-contrast condition were caused by the speed–accuracy trade-off, we analyzed the errors for each condition. A two-way ANOVA revealed that there was a significant main effect of the sound condition on the error rate [$F(2,40) = 3.687$, $p < .05$], but there was no significant interaction between the contrast and the sound conditions ($p = .214$). Planned contrast analyses revealed that the main effect of the sound condition was driven by the low-contrast condition. While there was no significant effect of the sound condition in the high-contrast condition ($p = .813$), the error rate for the in-phase condition was significantly lower than both the out-of-phase [$F(1,20) = 6.448$, $p < .05$] and the silent conditions [$F(1,20) = 10.00$, $p < .01$] in the low-contrast condition. This result indicates that the faster RTs in the low-contrast/in-phase condition may be due to the facilitatory effect of multisensory integration rather than the speed–accuracy trade-off.

In Experiment 2, we examined whether contrast-dependent cross-modal interaction persists even when temporally incongruent sound coexists with congruent sound. Previous studies showed that the facilitatory effect of audiovisual integration could be abolished when the congruent sound was perceptually grouped together with a sequence of identical sounds presented before and after the multisensory event (Keetels et al. 2007; Vroomen and de Gelder 2000). These results suggest that the presence of incongruent sound can interrupt pairing of temporally congruent audiovisual events. While keeping the temporal frequency of visual stimuli the same as Experiment 1, we doubled the temporal frequency of auditory signals. As depicted in Fig. 1d, half of auditory signals were presented synchronously with the contrast reversal of the visual stimuli and the remaining half were presented between the contrast reversals. We hypothesized that, if the effect of auditory stimuli is solely determined by the temporal coincidence between auditory and visual stimuli, irrespective of additional incongruent

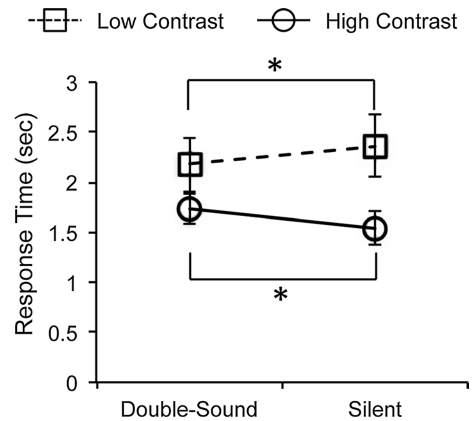

Fig. 3 Results from Experiment 2. *Circles with a solid line* represent the response times for the high-contrast condition, and *squares* with a *dashed line* represent the response times for the low-contrast condition. The *error bars* are ±1 standard error (*$p < .05$)

sounds, the facilitatory effect of sound on low-contrast visual stimuli and the inhibitory effect of sound on high-contrast visual stimuli would persist even when both congruent and incongruent auditory inputs coexist.

A total of nine observers were excluded from the analysis due to the high error rate (over 20 % of trials, eight observers) or high no-breakup rate (over 20 % of trials, one observer). The average error rate of the remaining 21 observers was 6 % and the average no-breakup rate was . 7 %. The error trials and no-breakup trials were excluded from the analysis. The averaged RTs for each condition are shown in Fig. 3. A three-way ANOVA again revealed that neither the main effect of temporal frequency of contrast reversal ($p > .135$) nor any interaction effects ($ps > .757$) were significant. Therefore, we aggregated the data between two temporal frequencies and performed a two-way ANOVA (sound congruency by contrast level). Consistent with Experiment 1, we found a significant main effect of contrast [$F(1,20) = 11.18$, $p = .003$] and interaction effect between the stimulus contrast and sound conditions [$F(1,20) = 16.83$, $p = .001$]. Planned contrast tests revealed that the RTs were significantly faster in the double-sound condition than in the silent condition when the contrast of the visual target was low [$F(1,20) = 7.18$, $p = .014$]. In contrast, the RTs were slower in the double-sound condition than in the silent condition when the contrast of visual stimuli was high [$F(1,20) = 6.80$, $p = .017$], replicating the inhibitory effect of sound on the processing of temporally congruent high-contrast visual stimuli, even when it is mixed with auditory stimuli that are incongruent with visual inputs. These results indicate that temporal synchronization between auditory and visual signals determines the cross-modal inhibitory interaction, even with the presence of a distracting sound. This finding seems to be inconsistent with the previous studies showing that auditory grouping precedes multisensory integration (Keetels et al. 2007; Vroomen and de Gelder 2000). The discrepancy between the current result and previous studies may be due to the difference in audiovisual stimuli. Both auditory and visual stimuli are periodic in the current study, but only auditory stimuli were periodic in the previous studies, which might promote the perceptual grouping of the auditory inputs. Repeated presentation of identical audiovisual stimuli for an extended duration might result in stronger audiovisual grouping rather than intramodal grouping.

General discussion

The facilitatory effect of multisensory integration is known to be strong when each unimodal signal is weak, and is known to decrease as the intensity of each unimodal signal increases, yielding little or no multisensory enhancement (Diederich and Colonius 2004; Stanford et al. 2005; Stein and Meredith 1993). The experiments reported here demonstrate that multisensory interaction could result in suppression of unimodal sensory processing, rather than a lack of cross-modal facilitation, which clarifies the reason for the limit in the PIE. The apparent strength of visual stimuli, inferred by the time to access to visual awareness by breaking up interocular suppression, was weakened by the presence of temporally congruent auditory stimuli when the physical strength (contrast) of the visual stimuli was high. We also found the cross-modal facilitation effect when the contrast of the gratings was low (10 %). This contrast-dependent multisensory interaction, either facilitatory or inhibitory, was robust even when both congruent and incongruent auditory inputs were present.

Previous studies using CFS have shown the facilitatory effect of temporally or semantically congruent auditory inputs during visual suppression. The lip movements of a talking face break CFS faster when presented with matched speech sound than when presented alone or presented with mismatched speech sound (Alsius and Munhall 2013; Plass et al. 2014). These studies suggest that audiovisual correspondences in speech perception can be processed even under visual suppression. Not only temporally congruent auditory stimuli but semantically congruent auditory inputs also speed up the breakup of suppression of dynamic visual stimuli during CFS (Cox and Hong 2015). However, the use of time to break interocular suppression to measure unconscious processing has recently been questioned because the response time in CFS can be affected by response biases or partial awareness (Gayet et al. 2014; Kang et al. 2011; Stein and Sterzer 2014). Indeed, a recent study measuring visual detection threshold during CFS, instead of time to break CFS, failed to replicate a well-known effect of congruent sound on detecting visual looming stimuli (Moors et al. 2015). Moreover, it is possible that the target in the low- and high-contrast condition becomes visible at the same time, but the orientation discrimination of a low-contrast stimulus may take more time than that of a high-contrast stimulus. This additional time to perform discrimination can explain slower RTs for the low-contrast condition than for the high-contrast condition. Thus, it may not be conclusive whether the audiovisual congruency effect observed in the current study indicates subconscious multisensory processing due to the limitation of the measurement. However, the current finding is not likely to be explained by response biases or differences in time to perform discrimination between the high- and low-contrast conditions because our study revealed both facilitatory and suppressive effects of congruent auditory stimuli depending on the intensity of visual stimuli. If the response bias or the time differences in discrimination determine response times, it should operate in the same direction irrespective of the intensity of the stimuli.

Prior work using binocular rivalry also consistently showed that either temporally congruent (Guzman-Martinez et al. 2012; Lunghi et al. 2014; van Ee et al. 2009) or semantically congruent auditory stimuli (Chen et al. 2011; Conrad et al. 2010, 2012) affect the rivalry dynamics of two competing visual stimuli. Although the influence of auditory input on binocular rivalry dynamics is known to result from audiovisual interaction during the dominance period rather than during the suppression period of the congruent visual stimuli, some evidence for multisensory integration outside of visual awareness can be found in recent studies on tactile–visual interaction during binocular rivalry. Lunghi et al. (2010) and Lunghi and Alais (2013) showed that congruent haptic stimulation affects binocular rivalry by shortening the duration of suppression as well as by prolonging the dominance duration. It was also shown that temporally (frequency) and spatially (orientation) congruent tactile stimulation decreases visual detection threshold, in comparison with the presentation of visual stimuli only, during rivalry suppression (Lunghi and Alais 2015), which more directly supports multisensory interaction outside of visual awareness.

Neural mechanisms mediating the impaired behavioral performance on congruent audiovisual stimuli are not yet clear. One possible mechanism is based on feedback from traditional multisensory areas to early sensory cortices. It has been shown that temporally congruent auditory inputs enhance neural responses in primary visual cortex via feedback from STS (Noesselt et al. 2007). Further, a recent fMRI study showed that neural responses in STS to audiovisual inputs are primarily suppressive for clear suprathreshold visual stimuli but are switched to be additive (being facilitatory) for noise-masked stimuli (Werner and

Noppeney 2010), which is consistent with the current finding. On the other hand, Iurilli et al. (2012) recently showed that neural responses of visual neurons in the mouse primary visual cortex are suppressed by sound (a noise burst) via GABAergic inhibition triggered by cortico-cortical connections from auditory cortex. This study suggests that a direct inhibitory connection between auditory and visual cortices is another possible mechanism for the suppressive audiovisual congruency effect observed in the high-contrast condition.

Attention can also be involved in the facilitatory effect of congruent auditory inputs on visual processing under CFS. Attention is known to affect perceptual selection during visual competition. Directing attention away from competing visual stimuli can slow perceptual switches in both binocular rivalry and bistable Necker cube (Alais et al. 2010). During binocular rivalry, paying attention to one of the monocular stimuli can increase the predominance duration of the attended stimulus (Chong and Blake 2006; Chong et al. 2005; Paffen et al. 2006; van Ee et al. 2009). This attentional modulation in visual awareness can be achieved by temporally congruent auditory input (van Ee et al. 2009), although such multisensory enhancement is observed only when the congruent visual stimulus is in dominant states. Indeed, it was suggested that attention can be more easily oriented toward multisensory stimuli than toward unisensory stimuli (Talsma et al. 2010; Van der Burg et al. 2008). Considering that audiovisual integration can occur even when visual stimuli are rendered invisible by CFS (Alsius and Munhall 2013; Plass et al. 2014), attention to congruent audiovisual stimuli may result in early breakup of suppression for the low-contrast condition in the current study. However, attentional modulation is not likely to account for the inhibitory effect on congruent audiovisual stimuli shown in the high-contrast condition. Although attention can accelerate neural adaptation by enhancing neural responses to a visual stimulus (Ling and Carrasco 2006), it is not apparent why attentional modulation in neural responses results in facilitation in one case (low-contrast condition) and suppression in the other (high-contrast condition).

Despite the cumulative evidence for the facilitatory effect of multisensory integration on our perceptual experience, it is still largely unknown how multisensory integration disrupts unisensory perceptual processing. The current study demonstrates the perceptual consequence of cross-modal inhibition between temporally congruent audiovisual stimuli when the unisensory signal intensity is high. The switch from facilitatory to suppressive multisensory interactions depending on the intensity of unisensory signals further extends the explanatory power of the PIE to include a full spectrum of multisensory interactions.

Compliance with ethical standards

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

References

Alais D, van Boxtel JJ, Parker A, van Ee R (2010) Attending to auditory signals slows visual alternations in binocular rivalry. Vis Res 50:929–935

Albrecht DG, Hamilton DB (1982) Striate cortex of monkey and cat: contrast response function. J Physiol 48:217–237

Alsius A, Munhall KG (2013) Detection of audiovisual speech correspondences without visual awareness. Psychol Sci 24(4):423–431

Brainard DH (1997) The psychophysics toolbox. Spatial Vis 10:433–436

Chen YC, Yeh SL, Spence C (2011) Crossmodal constraints on human perceptual awareness: auditory semantic modulation of binocular rivalry. Front Psychol 2:212. doi:10.3389/fpsyg.2011.00212

Chong SC, Blake R (2006) Exogenous attention and endogenous attention influence initial dominance in binocular rivalry. Vis Res 46:1794–1803

Chong SC, Tadin D, Blake R (2005) Endogenous attention prolongs dominance durations in binocular rivalry. J Vis 5:1004–1012

Conrad V, Bartels A, Kleiner M, Noppeney U (2010) Audiovisual interactions in binocular rivalry. J Vis 10:27

Conrad V, Vitello MP, Noppeney U (2012) Interactions between apparent motion rivalry in vision and touch. Psychol Sci 23:940–948

Convento S, Vallar G, Galantini C, Bolognini N (2013) Neuromodulation of early multisensory interactions in the visual cortex. J Cogn Neurosci 25(5):685–696

Cox D, Hong SW (2015) Semantic-based crossmodal processing during visual suppression. Front Psychol 6:722. doi:10.3389/fpsyg.2015.00722

Diederich A, Colonius H (2004) Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. Percept Psychophys 66:188–1404

Driver J, Noesselt T (2008) Multisensory interplay reveals crossmodal influences on "sensory specific" brain regions, neural responses, and judgments. Neuron 57:11–23

Fiebelkorn IC, Foxe JJ, Butler JS, Molholm S (2011) Auditory facilitation of visual-target detection persists regardless of retinal eccentricity and despite wide audiovisual misalignments. Exp Brain Res 213:167–174

Gayet S, Van der Stigchel S, Paffen CLE (2014) Breaking continuous flash suppression: competing for consciousness on the pre-semantic battlefield. Front Psychol 5:460. doi:10.3389/fpsyg.2014.00460

Guzman-Martinez E, Ortega L, Grabowecky M, Mossbridge J, Suzuki S (2012) Interactive coding of visual spatial frequency and auditory amplitude-modulation rate. Curr Biol 22:383–388

Iurilli G, Ghezzi D, Olcese U, Lassi G, Nazzaro C, Tonini R, Tucci V, Benfenati F, Medini P (2012) Sound-driven synaptic inhibition in primary visual cortex. Neuron 73:814–828

Kadunce DC, Vaughan JW, Wallace MT, Benedek G, Stein BE (1997) Mechanisms of within- and cross-modality suppression in the superior colliculus. J Neurophysiol 78(6):2834–2847

Kang M-S, Blake R, Woodman GF (2011) Semantic analysis does not occur in the absence of awareness induced by interocular suppression. J Neurosci 31:13535–13545

Keetels M, Stekelenburg J, Vroomen J (2007) Auditory grouping occurs prior to intersensory pairing: evidence from temporal ventriloquism. Exp Brain Res 180:449–456

Ling S, Carrasco M (2006) When sustained attention impairs perception. Nat Neurosci 9:1243–1245

Lippert M, Logothetis NK, Kayser C (2007) Improvement of visual contrast detection by a simultaneous sound. Brain Res 1173:102–109

Lunghi C, Alais D (2013) Touch interacts with vision during binocular rivalry with a tight orientation tuning. PLoS One 8:e58754

Lunghi C, Alais D (2015) Congruent tactile stimulation reduces the strength of visual suppression during binocular rivalry. Sci Rep 5:9413. doi:10.1038/srep09413

Lunghi C, Binda P, Morrone MC (2010) Touch disambiguates rivalrous perception at early stages of visual analysis. Curr Biol 20:R143–R144

Lunghi C, Morrone MC, Alais D (2014) Auditory and tactile signals combine to influence vision during binocular rivalry. J Neurosci 34(3):784–792

Meredith MA, Stein BE (1986) Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. J Neurophysiol 56:640–662

Moors P, Huygelier H, Wagemans J, de Wit L, van Ee R (2015) Suppressed visual looming stimuli are not integrated with auditory looming signals: evidence from continuous flash suppression. Perception 6:48–62

Noesselt T, Rieger JW, Schoenfeld MA, Kanowski M, Hinrichs H, Heinze HJ, Driver J (2007) Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. J Neurosci 27:11431–11441

Noesselt T, Bergmann D, Hake M, Heinze HJ, Fendrich R (2008) Sound increases the saliency of visual events. Brain Res 1220:157–163

Noesselt T, Tyll S, Boehler CN, Heinze HJ, Driver J (2010) Sound-induced enhancement of low-intensity vision: multisensory influences on human sensory-specific cortices and thalamic bodies relate to perceptual enhancement of visual detection sensitivity. J Neurosci 30:13609–13623

Paffen CL, Alais D, Verstraten FA (2006) Attention speeds binocular rivalry. Psychol Sci 17:752–756

Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spatial Vis 10:437–442

Plass J, Guzman-Martinez E, Ortega L, Grabowecky M, Suzuki S (2014) Lipreading without awareness. Psychol Sci 25(9):1835–1837

Romei V, Murray MM, Cappe C, Thut G (2009) Preperceptual and stimulus-selective enhancement of low-level human visual cortex excitability by sounds. Curr Biol 19:1799–1805

Romei V, Gross J, Thut G (2012) Sounds reset rhythms of visual cortex and corresponding human visual perception. Curr Biol 22:807–813

Stanford TR, Quessy S, Stein BE (2005) Evaluating the operations underlying multisensory integration in the cat superior colliculus. J Neurosci 25:6499–6508

Stein BE, Meredith MA (1993) The merging of the senses. MIT Press, Cambridge

Stein T, Sterzer P (2014) Unconscious processing under interocular suppression: getting the right measure. Front Psychol 5:387. doi:10.3389/fpsyg.2014.00387

Stein BE, London N, Wilkonson LK, Price DD (1996) Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis. J Cogn Neurosci 8(6):497–506

Talsma D, Senkowski D, Soto-Faraco S, Woldorff MG (2010) The multifaceted interplay between attention and multisensory integration. Trends Cogn Sci 14(9):400–410

Tsuchiya N, Koch C (2005) Continuous flash suppression reduces negative afterimages. Nat Neurosci 8(8):1096–1101

Tsuchiya N, Koch C, Gilroy L, Blake R (2006) Depth of interocular suppression associated with continuous flash suppression, flash suppression, and binocular rivalry. J Vis 6:1068–1078

Van der Burg E, Olivers CNL, Bronkhorst AW, Theeuwes J (2008) Pip and Pop: nonspatial Auditory Signals Improve Spatial Visual Search. J Exp Psychol Hum Percep Perform 34:1053–1065

van Ee R, van Boxtel JJ, Parker AL, Alais D (2009) Multisensory congruency as a mechanism for attentional control over perceptual selection. J Neurosci 29:11641–11649

Vroomen J, de Gelder B (2000) Sound enhances visual perception: cross-modal effects of auditory organization on vision. J Exp Psychol Hum Percep Perform 26(5):1583–1590

Werner S, Noppeney U (2010) Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization. Cereb Cortex 20:1829–1842