



Crossmodal constraints on human perceptual awareness: auditory semantic modulation of binocular rivalry

Yi-Chuan Chen^{1*}, Su-Ling Yeh² and Charles Spence¹

¹ Department of Experimental Psychology, University of Oxford, Oxford, UK

² Department of Psychology, National Taiwan University, Taipei, Taiwan

Edited by:

Nadia Bolognini, University
Milano-Bicocca, Italy

Reviewed by:

Toemme Noesselt, Otto von Guericke
University, Germany
Davide Bottari, University of
Hamburg, Germany

*Correspondence:

Yi-Chuan Chen, Department of
Experimental Psychology, University
of Oxford, South Parks Road, Oxford
OX1 3UD, UK.

e-mail: yi-chuan.chen@queens.
oxon.org

We report a series of experiments utilizing the binocular rivalry paradigm designed to investigate whether auditory semantic context modulates visual awareness. Binocular rivalry refers to the phenomenon whereby when two different figures are presented to each eye, observers perceive each figure as being dominant in alternation over time. The results demonstrate that participants report a particular percept as being dominant for *less* of the time when listening to an auditory soundtrack that happens to be semantically congruent with the other alternative (i.e., the *competing*) percept, as compared to when listening to an auditory soundtrack that was irrelevant to both visual figures (Experiment 1A). When a visually presented word was provided as a semantic cue, no such semantic modulatory effect was observed (Experiment 1B). We also demonstrate that the crossmodal semantic modulation of binocular rivalry was robustly observed irrespective of participants' attentional control over the dichoptic figures and the relative luminance contrast between the figures (Experiments 2A and 2B). The pattern of crossmodal semantic effects reported here cannot simply be attributed to the meaning of the soundtrack guiding participants' attention or biasing their behavioral responses. Hence, these results support the claim that crossmodal perceptual information can serve as a constraint on human visual awareness in terms of their semantic congruency.

Keywords: multisensory, audiovisual interaction, semantic congruency, consciousness, attention, stimulus contrast

INTRODUCTION

When viewing a scene, visual background context provides useful semantic information that can improve the identification of a visual object embedded within it, such as when the presentation of a kitchen scene facilitates a participant's ability to identify a loaf of bread, say (e.g., Biederman, 1972; Palmer, 1975; Davenport and Potter, 2004; though see Hollingworth and Henderson, 1998). Importantly, however, our environments typically convey contextual information via several different sensory modalities rather than just one. So, for example, when we are at the seaside, we perceive not only the blue sea and sky (hopefully), but also the sound of the waves crashing onto the beach, not to mention the smell of the salty sea air. Do such non-visual contextual cues also influence the visual perception of semantically related objects? In the present study, we investigated whether the semantic context provided by stimuli presented in another sensory modality (in this case, audition) modulate the perceptual outcome in vision; namely, visual awareness.

The phenomenon of binocular rivalry provides a fascinating window into human visual awareness (e.g., Crick, 1996). Binocular rivalry occurs when two dissimilar figures are presented to corresponding regions of the two eyes. Observers typically perceive one of the figures as dominant (while often being unaware of the presence of the other figure); after a while, the dominance of the figures may reverse and then keep alternating over time. This perceptual alternation has been attributed to the fact that the

visual system receives ambiguous information from the two eyes and tries to find a unique perceptual solution, and therefore the information presented to each eye competes for control of the current conscious percept (see Alais and Blake, 2005, for a review). The fact that a constantly presented dichoptic figure induces alternating perceptual experiences in the binocular rivalry situation demonstrates the dynamic way in which the brain computes sensory information, a process that gives rise to a specific percept (e.g., Leopold and Logothetis, 1996).

Several researchers have tried to understand how visual awareness emerges in the binocular rivalry situation. According to an early view put forward by Helmholtz (1962), the alternation of perceptual dominance is under voluntary attentional control. Subsequently, researchers suggested that the phenomenon occurs as a result of competition between either two monocular channels (Levelt, 1965; Tong and Engel, 2001) or else between two pattern representations, one presented to each eye (Leopold and Logothetis, 1996; Logothetis et al., 1996; Tong et al., 1998). More recent models (e.g., Tong et al., 2006) have suggested that the mechanisms underlying binocular rivalry include not only competition at multiple levels of information processing (for reviews, see Tong, 2001; Blake and Logothetis, 2002), but also some form of excitatory connections that facilitate the perceptual grouping of visual stimuli (Kovacs et al., 1996; Alais and Blake, 1999), as well as top-down feedback, including attentional control and mental imagery (Meng and Tong, 2004; Mitchell et al., 2004; Chong et al., 2005;

van Ee et al., 2005; Pearson et al., 2008). That said, the underlying mechanisms giving rise to conscious perception in the binocular rivalry situation, while starting from interocular suppression, extend to a variety of different neural structures throughout the visual processing hierarchy.

Given that the phenomenon of binocular rivalry is, by definition, visual, one might have expected that the perceptual outcome for ambiguous visual inputs should thus be generated entirely within the visual system (cf. Hupé et al., 2008). On the other hand, however, some researchers have started to investigate whether visual awareness can be modulated by the information presented in another sensory modality. So, for example, it has recently been demonstrated that concurrently presented auditory cues can help to maintain the awareness of visual stimuli (Sheth and Shimojo, 2004; Chen and Yeh, 2008). Similar evidence has emerged from a binocular rivalry study demonstrating that the dominance duration of a looming (or rotating) visual pattern can be extended temporally when the rate of change of the visual stimulus happens to be synchronous with a series of pure tones or vibrotactile stimuli (or their combination, see van Ee et al., 2009). In addition, the directional information provided by the auditory modality can enhance the dominance duration of the moving random-dot kinematogram which happens to be moving in the same direction (Conrad et al., 2010).

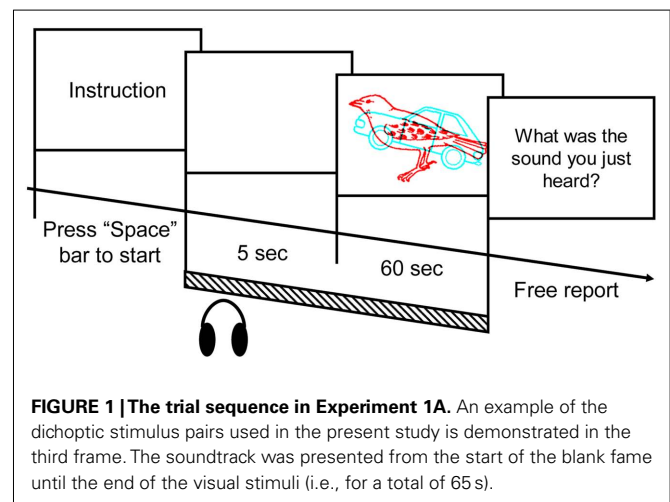
Considering the seaside example outlined earlier, the meaning of a background sound (or soundtrack) plausibly provides a contextual effect on human information processing, which may, as a result, modulate the perceptual outcome that a person is aware of visually. Semantic congruency, which relies on the associations picked-up in daily life, provides an abstract constraint other than physical consistency between visual and auditory stimuli (such as coincidence in time or direction of motion mentioned earlier). This high-level factor has started to capture the attention of researchers interested in multisensory information processing (e.g., Greene et al., 2001; Molholm et al., 2004; van Atteveldt et al., 2004; Taylor et al., 2006; Iordanescu et al., 2008; Noppeney et al., 2008; Schneider et al., 2008; Chen and Spence, 2010; for a recent review, see Spence, 2011). On the other hand, modulations resulting from the presentation of semantically meaningful information have recently been documented by researchers studying unimodal binocular vision (Jiang et al., 2007; Costello et al., 2009; Ozkan and Braunstein, 2009). In the present study, we therefore investigated whether the semantic context provided by an auditory soundtrack would modulate human visual perception in the binocular rivalry situation.

Our first experiment was designed to test the crossmodal semantic modulatory effect on the dominant percept under conditions of binocular rivalry, while attempting to minimize or control any possible response biases elicited by the meaning of the sound. After first establishing this crossmodal effect, we then go on to explore the ways in which auditory semantic context modulates visual awareness in the binocular rivalry situation. Two visual factors, one high-level (selective attention) and one low-level (stimulus contrast) which have been shown to modulate visual perception in the binocular rivalry situation (Meng and Tong, 2004), are used to probe behaviorally the underlying mechanisms by which the auditory semantic context modulating visual

awareness occurred in terms of current models of binocular rivalry (Tong et al., 2006).

EXPERIMENT 1

In our first experiment, we investigated whether the semantic context of a background soundtrack would modulate the dominance of two competing percepts under the condition of binocular rivalry. The participants viewed a dichoptic figure consisting of a bird and a car (see **Figure 1**) while listening to a soundtrack. When studying audiovisual semantic congruency effects, the possibility that participants' responses are based on their utilizing a strategy designed to satisfy a particular laboratory task has to be avoided (see de Gelder and Bertelson, 2003). That is, there is a danger that the participants might merely report the stimulus that happened to be semantically congruent with the soundtrack rather than the percept that happened to be more salient (or dominant). In order to reduce the likelihood that the above-mentioned response bias would affect participants' performance, a novel experimental design was used in Experiment 1A: the participants only had to press keys to indicate the start and the end time of the perceptual dominance of the pre-designated figure (e.g., "bird") during the test period, while they listened to either the soundtrack that was *incongruent* with the visual target (i.e., a car soundtrack, in this case) or the sound that was *irrelevant* to both figures (i.e., a soundtrack recorded in a restaurant). A parallel task in which the pre-designated target figure was the car was also conducted. The participants listened to either the bird soundtrack in the incongruent condition or to the restaurant soundtrack in the irrelevant condition. Thus, the soundtrack was never congruent with the visual target that participants had to report. This aspect of the experimental design was introduced in order to reduce the likelihood that participants would simply report their perceptual dominance in accordance with whichever soundtrack they happened to hear. On the other hand, if the auditory semantic context can either prolong the dominance of the visual percept that happens to be semantically congruent with the soundtrack, or else shorten the dominance duration of the percept that happens to be semantically incongruent with the soundtrack, in the binocular rivalry situation, the dominance duration of the visual



target should be shorter in the *incongruent* than in the *irrelevant* condition.

EXPERIMENT 1A

Participants

Twelve volunteers (including the first author, three males, with a mean age of 26 years old) took part in this experiment in exchange for a £10 (UK Sterling) gift voucher or course credit. The other 11 participants were naïve as to the specific purpose of the study. They all had normal or corrected-to-normal vision and normal hearing by self report. The participants were tested using depth-defined figures embedded in red–green random-dot stereograms to ensure that they had normal binocular vision. The study has been approved by the ethic committee and human participant recruit system in Department of Experimental Psychology, University of Oxford. All of the participants were informed of their rights in accordance with the ethical standards laid down in the 1990 Declaration of Helsinki and signed a consent form.

Apparatus and stimuli

The visual stimuli were presented on a 15" color CRT monitor (75 Hz refresh rate). The participants sat at a viewing distance of 58 cm from the monitor in a dimly lit experimental chamber. The visual test stimuli consisted of the outline-drawings of a bird ($4.44^\circ \times 2.76^\circ$) and car ($4.41^\circ \times 2.27^\circ$) taken from Bates et al. (2003). The two figures were spatially superimposed, with the bird presented in red [CIE (0.621, 0.341)] and the car in cyan [CIE (0.220, 0.347)], or vice versa, against a white background [CIE (0.293, 0.332)]. These two color versions of the visual pictures (bird in red and car in cyan, or bird in cyan and car in red) were used to balance the influence of participants' dominant eye when viewing dichoptic figures. The participants wore glasses with a red filter on the left eye and a cyan filter on the right eye during the course of the experiment.

Three sound files, *bird* (consisting of birds singing in a forest), *car* (consisting of car horn and engine-revving sounds in a busy street), and *restaurant* (consisting of the sound of tableware clattering together in a restaurant), which had been recorded in realistic environments (downloaded from www.soundsnap.com on 06/11/2008) were used as the auditory soundtracks. The sound files were edited so that the auditory stimulus started from the beginning of the sound file and lasted for 65 s. The sounds were presented over closed-ear headphones and ranged in loudness from 55 to 68 dB SPL.

Design and procedure

Two factors, semantic congruency (incongruent or irrelevant) and visual target (bird or car), were manipulated. Each participant reported the dominance of either the bird or the car percept in separate sessions in a counterbalanced order. Under those conditions in which the visual target was the bird, the participants were instructed to press the "1" key as soon as the image of the bird became dominant. The participants were informed that the criterion for responding that the bird was dominant was that they were able to see every detail, such as the texture of the wings, of the figure of the bird. As soon as any part of the bird figure became vague or else started to be occupied by the features of the car

figure, they had to press the "0" key as soon as possible, to indicate that the image of the bird was no longer completely dominant. This criterion enabled us to estimate the dominance duration of the bird percept more conservatively, since it excluded those periods of time when the car percept being dominant as well as when participants experienced a mixed percept. Similarly, under those conditions in which the visual target was the car, the participants had to press "1" and "0" to indicate when they started and stopped perceiving the car percept as being dominant.

The participants initiated each trial by pressing the "SPACE" bar. A blank screen was presented for 5 s, followed by the presentation of the dichoptic figures for a further 60 s. The participants were instructed to fixate the area of the bird's wing and car door and to start reporting the dominance of the target figure as soon as the dichoptic figures were presented. They had to monitor the dominance of the target picture continuously during the test period. The participants were also instructed to pay attention to the context of the sound as well (in order to ensure that the soundtracks were processed; see van Ee et al., 2009, Experiment 4). At the end of the trial, the question "What sound did you just hear?" was presented on the monitor, and the participants had to enter their answer (free report) using the keyboard. The sound was presented from the onset of the blank frame until the offset of the visual stimuli, in order to allow participants sufficient time to realize what the semantic context conveyed by the soundtrack was.

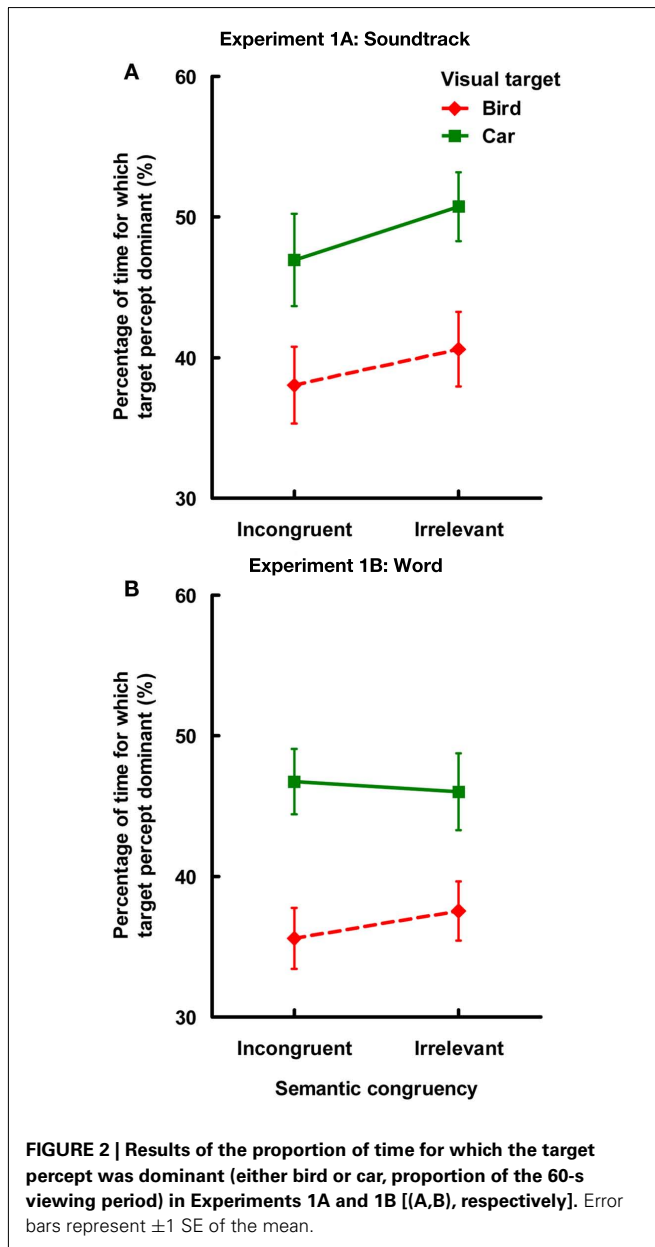
In both visual tasks (i.e., when the visual target was a bird and when it was a car), a block of 12 trials was presented (consisting of two sound conditions \times two color versions of visual pictures, each conditions were repeatedly tested three times). The order of presentation of these 12 trials was randomized. Prior to the completion of the experimental block of trials, a practice block containing six no-sound trials was presented in order to familiarize the participants with the task. The participants were instructed to establish their criterion for reporting the exclusive dominance of the target picture, and to try and hold this criterion constant throughout the experiment. The experiment lasted for approximately 1 h.

Results

The *proportion of time for which the target percept was dominant* was calculated by dividing the sum of each dominance duration of the target percept by 60 s. Note that the participants may have occasionally pressed the "1" or "0" key twice. In such cases, the shorter duration (i.e., the duration from the second "1" keypress to the first "0" keypress) was used.

A two-way analysis of variance (ANOVA) was conducted with the factors of semantic congruency (incongruent or irrelevant) and visual target (bird or car; see **Figure 2A**)¹. The results revealed significant main effects of both semantic congruency

¹ For reasons that are unknown, one participant stopped reporting the dominance of the target figure during the first 25 s in a trial in the incongruent condition, while he/she kept reporting the dominance percept until the end in all of the other trials. This trial, as well as a matched color version trial in the irrelevant condition, was excluded from the data analysis. By doing this, we were able to ensure that the data in the incongruent and irrelevant conditions came from equal number of trials from the two color versions, so that the factor of participants' eye dominance could be matched.



[$F(1,11) = 25.68$, $MSE = 0.0005$, $p < 0.0005$, $\eta_p^2 = 0.71$] and visual target [$F(1,11) = 11.50$, $MSE = 0.01$, $p < 0.01$, $\eta_p^2 = 0.51$]. There was, however, no interaction between these two factors [$F(1,11) = 0.60$, $MSE = 0.001$, $p = 0.46$, $\eta_p^2 = 0.07$]. The planned simple main effect of the semantic congruency factor revealed that the proportion of dominance of the target picture was lower when listening to the *incongruent* soundtrack than when listening to the *irrelevant* soundtrack both when the visual target was the bird [$F(1,22) = 6.34$, $MSE = 0.001$, $p < 0.05$, $\eta_p^2 = 0.15$], as well as when it was the car [$F(1,22) = 13.96$, $p < 0.005$, $\eta_p^2 = 0.33$]. In addition, the magnitude of the auditory modulatory effect (incongruent vs. irrelevant) was not significantly different in the bird and car target conditions [$F(1,11) = 0.60$, $MSE = 0.002$, $p = 0.45$, $\eta_p^2 = 0.06$].

EXPERIMENT 1B

Two further possibilities regarding the crossmodal semantic modulation reported in Experiment 1A need to be considered. First, the presented soundtrack may have accessed its associated abstract semantic representation and then modulated the dominant percept in the binocular rivalry situation. In this case, the semantic modulation constitutes a form of top-down semantic modulation rather than a form of audiovisual interaction. Second, even though the design of Experiment 1A effectively avoids the bias that the participants strategically reported the percept that is congruent with the meaning of the soundtrack as being dominant, it is important to note that a second type of bias should also be considered. That is, it could be argued that the presentation of the incongruent soundtrack may have provided a cue that discouraged the participants from reporting the target percept as being dominant, as compared to the presentation of the irrelevant soundtrack.

Experiment 1B was designed to control for the possibility that the crossmodal semantic modulation effects observed thus far might simply have resulted from the participants holding an abstract concept in mind, as well as the response bias elicited by the presentation of a cue that was incongruent with the identity of the visual target. Rather than presenting a soundtrack, the name of one of the soundtracks was presented on the monitor for 5 s prior to the presentation of the dichoptic figures (during this period, a blank frame had been presented in Experiment 1A). That is, the participants were provided with a word (the associated name of the soundtracks used in Experiment 1A) that was either incongruent with or irrelevant to the visual target, while they were tested in silence during the subsequent test period. The participants were instructed to retain the word in memory and to report it at the end of each trial, in order to ensure that they had maintained this semantic cue during the course of the test period. The word therefore provided an abstract semantic cue to the participants. In addition, the presentation and retention of this semantic cue in memory by participants would be expected to elicit a similar response bias in the incongruent (as compared to the irrelevant) condition. Our prediction was that if an abstract semantic cue or the response bias elicited by the incongruent cue (rather than the audiovisual semantic interaction) was sufficient to induce the semantic effect in the binocular rivalry situation, the significant difference between incongruent and irrelevant conditions should still be observed.

Two factors, semantic congruency (incongruent or irrelevant) and visual target (bird or car), were manipulated in this experiment. When the visual target was the bird, the words “car” and “restaurant” were presented in the incongruent and irrelevant conditions, respectively. Similarly, when the visual target was the car, the words “bird” and “restaurant” were presented in the incongruent and irrelevant conditions, respectively. The other experimental details were exactly the same as in Experiment 1A.

Participants

Twelve volunteers (including the first author, six males, with a mean age of 24 years old) took part in this experiment. The other 11 participants did not attend Experiment 1A and they were naive as to the goal of the study. The other details are the same as in Experiment 1A.

Results

The data were analyzed in the same manner as in Experiment 1A. The participants misreported the word in six trials (out of total 288 trials). These trials, as well as matched color version trials in the other word condition, were excluded from the analysis (4.2% of total trials). A two-way ANOVA was conducted with the factors of semantic congruency (incongruent or irrelevant) and visual target (bird or car; see **Figure 2B**). Once again, the results revealed a significant main effect of visual target [$F(1,11) = 7.99$, $MSE = 0.01$, $p < 0.05$, $\eta_p^2 = 0.42$]. Critically, however, neither the main effect of semantic congruency [$F(1,11) = 0.96$, $MSE = 0.0005$, $p = 0.35$, $\eta_p^2 = 0.08$], nor the interaction term [$F(1,11) = 1.49$, $MSE = 0.001$, $p = 0.25$, $\eta_p^2 = 0.11$] was significant. The planned simple main effect of the semantic congruency factor revealed that the proportion of dominance of the target picture was *not* significant as a function of whether the condition was incongruent or irrelevant when the visual target was either the bird [$F(1,22) = 2.39$, $MSE = 0.001$, $p = 0.15$, $\eta_p^2 = 0.09$], or the car [$F(1,22) = 0.33$, $p = 0.57$, $\eta_p^2 = 0.01$].

Comparison of Experiments 1A and 1B

In order to verify that the semantic modulation on the dominant percept in the binocular rivalry situation was significant in Experiment 1A but not in Experiment 1B, a three-way ANOVA on the factor of cue type (soundtrack or word), semantic congruency (incongruent or irrelevant), and visual target (bird or car) was conducted. The between-participants factor was cue type while the latter two factors were varied on a within-participants basis. The results revealed significant main effects of semantic congruency [$F(1,22) = 18.47$, $MSE = 0.0005$, $p < 0.0005$, $\eta_p^2 = 0.47$] and visual target [$F(1,22) = 18.75$, $MSE = 0.01$, $p < 0.0005$, $\eta_p^2 = 0.46$]. Critically, the interaction between cue type and semantic congruency was significant [$F(1,22) = 8.43$, $MSE = 0.0005$, $p < 0.01$, $\eta_p^2 = 0.29$]. The simple main effect of the semantic congruency factor was significant when the cue was a soundtrack [$F(1,22) = 25.93$, $MSE = 0.0005$, $p < 0.0001$, $\eta_p^2 = 0.55$], but not when the cue was a word [$F(1,22) = 0.97$, $p = 0.34$, $\eta_p^2 = 0.04$]. The magnitude of the auditory modulatory effect (incongruent vs. irrelevant) was submitted to a two-way ANOVA on the factor of cue type and visual target. Only the main effect of cue type reached significance [3.2 vs. 0.6% for soundtrack and word condition, respectively, $F(1,22) = 8.39$, $MSE = 0.001$, $p < 0.01$, $\eta_p^2 = 0.28$]. The other main effect of visual target [$F(1,22) = 0.26$, $MSE = 0.002$, $p = 0.61$, $\eta_p^2 = 0.02$] and the interaction term [$F(1,22) = 2.06$, $MSE = 0.002$, $p = 0.17$, $\eta_p^2 = 0.09$] were not significant.

Discussion

The results of Experiment 1A therefore demonstrate a crossmodal modulation on the proportion of dominance measure resulting from the auditory semantic context that was present in the binocular rivalry situation. Note that the soundtrack to which the participants listened during the test period was *never* congruent with the visual target. That is, the crossmodal semantic modulation of binocular rivalry was *indirect* in terms of the meaning of the sound either increasing the amount of time for which the

participant perceived the non-target visual picture and/or decreasing the amount of time for which they perceived the target picture. Both possibilities would have led to a reduction in the proportion of dominance of the target picture. This feature of the design means that it was not the case that the meaning of the sound directly biased the participants' response to report the semantically congruent visual stimulus as being dominant.

On the other hand, the results in Experiment 1B, demonstrated that simply maintaining a word in memory during the test period did not bias the participants' visual perception or responses. Note that the comparison of the results of Experiments 1A and 1B is meaningful based on the *a priori* assumption that the presentation of a semantically congruent (though task-irrelevant) word can prime the participants' performance regarding the picture (e.g., Glaser and Glaser, 1989). Hence, the modulatory effect of auditory semantic content reported in Experiment 1A cannot simply be attributed to a semantic priming effect elicited by activating an abstract concept regarding one of the pictures (cf. Balceitis and Dale, 2007), nor to any response bias that was potentially elicited by the presentation of a soundtrack that was incongruent with the visual target (i.e., congruent with the competing percept). In addition, due to the fact that the participants were continuously receiving the auditory information during the test period in Experiment 1A while simply provided a semantic cue before the test period in Experiment 1B, we suggest that the crossmodal semantic congruency effect should be perceptual in nature (i.e., depending on the input of sensory information) rather than simply a conceptual effect (depending on the prior acquired knowledge). These results therefore highlight a significant crossmodal modulation of perceptual dominance in the binocular rivalry situation. This result can be attributed to the semantic context embedded in the auditory soundtrack that the observers were listening to.

In both experiments, the results revealed that the proportion of dominance was larger when the car was the target than when the bird was the target. Note that the wing of the bird contained small individual elements constituting the texture of a feather, so the individual elements may have disappeared occasionally (Kovacs et al., 1996). On the other hand, most of the lines making up the figure of the car were connected, and so they should group into a unitary element (such as a car door). Since the participants were instructed to report the target figure as being dominant *only* when they could see all of its features, the well-grouped figure (i.e., the car) should have reached this criterion more easily than the less-well-grouped figure (i.e., the bird).

In van Ee et al.'s (2009) study, it was reported that an auditory stimulus enhanced the dominance duration of a synchronous visual pattern only when participants happened to attend to that visual pattern in the binocular rivalry situation (see their Experiment 1). In Experiment 1A of the present study, given that the participants had to simply monitor the dominance of one of the two figures, their goal-directed attention should presumably have been focused voluntarily (i.e., endogenously) on the target stimulus. Nevertheless, in order to further investigate the interplay between crossmodal semantic congruency and the participants' selective attention on the perception of binocular rivalry, these two factors were manipulated independently in Experiment 2A. In addition, given that the visual competition mechanism in binocular rivalry

started from a low-level interocular suppression, binocular rivalry is presumably susceptible to stimulus saliency (Mueller and Blake, 1989). We therefore decided to test the interplay between cross-modal semantic congruency and the manipulation of the stimulus contrast in Experiment 2B.

EXPERIMENT 2

We designed two further experiments in order to measure whether the crossmodal semantic congruency effect is robust when simultaneously manipulating visual factors that have previously been shown to modulate participants' perception in the binocular rivalry situation, such as participants directing their selective attention to a specific percept or increasing the stimulus contrast of one of the dichoptic images (Meng and Tong, 2004; van Ee et al., 2005). Besides, knowing whether the modulations of crossmodal semantic congruency and either visual factor (i.e., selective attention or stimulus contrast) work additively or interactively to influence human visual perception would help us understand the possible mechanism underlying the crossmodal effect of auditory semantic context. We therefore manipulated auditory semantic congruency and visual selective attention in Experiment 2A, and auditory semantic congruency and visual stimulus contrast in Experiment 2B, respectively.

In Experiments 2A and 2B, the participants performed a typical binocular rivalry experiment reporting the percept that was subjectively dominant. That is, the participants had to press the "1" key whenever the image of the bird was dominant, and the "0" key whenever the image of the car was dominant. Three performance indices were used: the *proportion of time for which the bird percept was dominant* was calculated by dividing the sum duration of each bird percept by the sum duration of both the bird and car percept within the test period (thus, the proportion of time for the dominance of the bird and car views were reciprocally related). Accordingly, this measure would be expected to increase following any experimental manipulation that favored the bird percept (i.e., maintaining the bird percept or the bird figure by means of its higher contrast), whereas it should be decreased by the manipulation favoring the car percept (i.e., maintaining the car percept or the car figure by means of its higher contrast). The second index consisted of the average *number of switches* between the bird and car percept that took place during the test period in each sound condition. This index links closely to the idea of voluntary control in the binocular rivalry situation: in particular, when participants try to maintain a particular percept, they are able to delay the switch to the other percept (van Ee et al., 2005). It has been suggested that the combination of an increase in the proportion of dominance duration as well as a reduction in the number of switches can be considered as the signature of selective attention in the binocular rivalry situation (see van Ee et al., 2005). The third index was the number of times that the *first percept* was the bird out of six trials in each sound condition. This index can be considered as the result of initial competition between the images presented to each eye.

EXPERIMENT 2A

In Experiment 2A, the target of participants' selective attention over the dichoptic figure was manipulated independently of the

meaning of the sound. That is, the participants were instructed to maintain the bird percept, to maintain the car percept, or to view the figures passively in the control condition (see Meng and Tong, 2004; van Ee et al., 2005). Meanwhile, the participants either heard the birds singing or else the revving car engine soundtracks.

Participants

Seven participants (including the first author, three males, with a mean age of 24 years old) took part in Experiment 2A. All of them had prior experience of binocular rivalry experiments, and three of them had taken part in Experiment 1A. However, the other six participants (except the author) were naïve regarding the goal of the present experiment.

Design and procedure

Two factors, sound (bird and car) and selective attention (passive, maintain bird, or maintain car) were manipulated. The 5-s blank frame presented prior to the visual stimulus display now contained the instruction to "just look at the figures PASSIVELY," "try to maintain the percept of the BIRD as long as possible" or "try to maintain the percept of the CAR as long as possible." Note that in all three conditions, the participants had to report their current dominant percept (either bird or car). The bird or car soundtrack started at the onset of the attention instruction frame. The figures were larger than those used in Experiment 1 (bird: $7.85^\circ \times 6.39^\circ$; car: $8.34^\circ \times 3.95^\circ$).

Three blocks of experimental trials were presented. There were 12 trials (two sound conditions \times three selective attention conditions \times two color versions of the pictures) presented in a randomized order in each block. A practice block containing six no-sound trials, two for each selective attention condition, was conducted prior to the main experiment.

Results

A two-way ANOVA was conducted on the factors of sound and selective attention for each index separately. In the analysis of the data concerning the proportion of dominance data (see **Figure 3A**), there were significant main effects of sound [$F(1,6) = 17.25$, $MSE = 0.004$, $p < 0.01$, $\eta_p^2 = 0.75$], and selective attention [$F(2,12) = 12.63$, $MSE = 0.04$, $p < 0.005$, $\eta_p^2 = 0.68$]. A *post hoc* Tukey's test revealed that there was a significant difference between the *maintain bird* and *maintain car* conditions ($p < 0.01$). The interaction between sound and selective attention was, however, not significant [$F(2,12) = 1.39$, $MSE = 0.002$, $p = 0.29$, $\eta_p^2 = 0.18$]. In order to examine whether the cross-modal semantic modulation of binocular rivalry was significant in all three selective attention conditions, planned simple main effects on the sound factor were conducted. The results revealed that the bird percept was dominant for a larger proportion of the time when presenting the bird soundtrack than when presenting the car soundtrack in the maintain bird condition [$F(1,18) = 5.57$, $MSE = 0.002$, $p < 0.05$, $\eta_p^2 = 0.13$], in the passive condition [$F(1,18) = 16.65$, $p < 0.001$, $\eta_p^2 = 0.37$], and in the maintain car condition [$F(1,18) = 4.91$, $p < 0.05$, $\eta_p^2 = 0.11$]. The magnitude of the auditory semantic modulation effect (bird- vs. car-sound) was not significantly different in the three selective attention conditions [$F(2,12) = 1.40$, $MSE = 0.004$, $p = 0.29$, $\eta_p^2 = 0.19$].

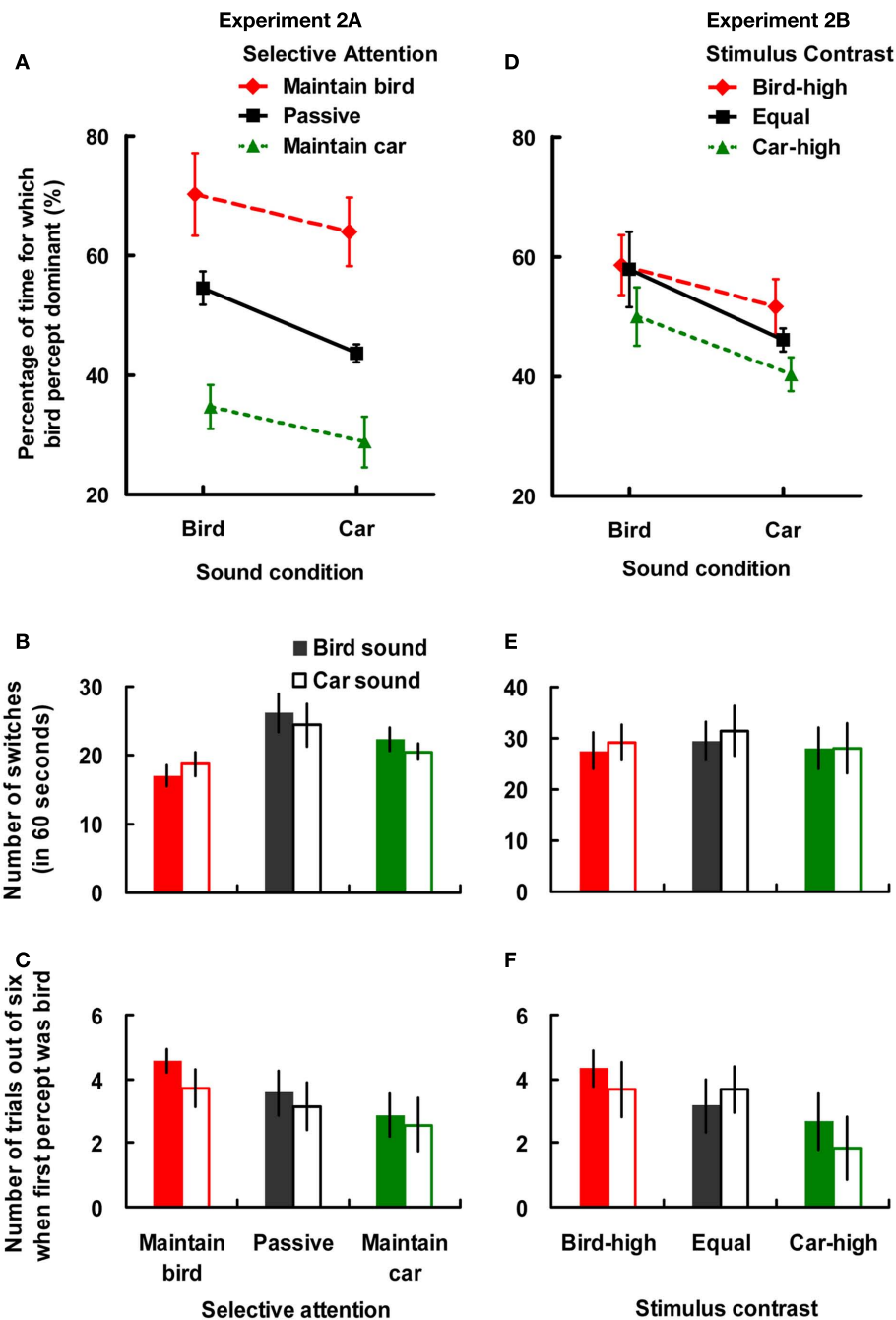


FIGURE 3 | Results of the bird- and car-sound conditions under the manipulation of either selective attention (A–C) in Experiment 2A or stimulus contrast (D–F) in Experiment 2B in the binocular rivalry situation. The upper panels (A,D) depict the proportion of time that the bird percept was dominant (proportion to total dominance duration of both bird and car percepts); the middle panels (B,E) show the average number of switches that took place in each trial (during the course of 60 s); the lower

panels (C,F) show the number of times that participants reported the bird as the first percept (out of six trials). (A) The proportion of time that the bird percept was dominant in the maintain bird (dashed line), passive viewing (solid line), and maintain car (dotted line) conditions of Experiment 2A. (D) The proportion of time that the bird percept was dominant in the bird-high (dashed line), equal (solid line), and car-high (dotted line) conditions in Experiment 2B. Error bars represent ± 1 SE of the mean.

Analysis of the number of switches in each condition (see Figure 3B) revealed significant main effects of both sound [$F(1,6) = 8.19$, $MSE = 0.43$, $p < 0.05$, $\eta_p^2 = 0.58$] and selective attention [$F(2,12) = 3.92$, $MSE = 48.05$, $p < 0.05$, $\eta_p^2 = 0.40$].

A *post hoc* Tukey's test revealed that the number of switches was higher in the passive condition than in the maintain bird condition ($p < 0.05$). The interaction between sound condition and selective attention was also significant [$F(2,12) = 5.27$, $MSE = 2.57$,

$p < 0.05$, $\eta_p^2 = 0.47$]. The simple main effect revealed that the number of switches was *smaller* in the *bird-sound* condition than in the *car-sound* condition when the participants had to try and maintain their view of the bird [$F(1,18) = 5.39$, $MSE = 1.85$, $p < 0.05$, $\eta_p^2 = 0.16$]. By contrast, the number of switches was *smaller* in the *car-sound* condition than in the *bird-sound* condition when the participants passively viewed the figures [$F(1,18) = 5.09$, $p < 0.05$, $\eta_p^2 = 0.15$], or when they had to try and maintain the view of the car [$F(1,18) = 6.02$, $p < 0.05$, $\eta_p^2 = 0.17$].

The results of the first percept in each condition (see **Figure 3C**) revealed that the participants reported the bird as the first percept somewhat more frequently when they were either listening to the bird sound or when trying to maintain the bird view, as compared to when they were either listening to the car sound or else trying to maintain their view of the car. However, neither of the main effects, nor their interaction, reached statistical significance (all $F_s < 3.27$, $p_s > 0.12$, $\eta_p^2 = 0.35$).

Discussion

These results demonstrate that the participants reported the bird percept as being dominant for more of the time (i.e., the proportion of dominance was larger) when they heard the sound of birds singing than when they heard the sound of cars revving their engines. In addition, selective attention also modulated the proportion of dominance of the bird percept when the participants tried to maintain their view of the bird or car, consistent with the results reported by both Meng and Tong (2004) and van Ee et al. (2005). Nevertheless, the interaction between these two factors was not significant. Critically, the modulation of auditory semantic context was robustly observed in the maintain bird, passive, and maintain car conditions, and what is more, the magnitude of the crossmodal modulatory effect was similar in these three conditions. In other words, the crossmodal semantic modulation observed in Experiment 2A cannot be attributed solely to the meaning of the sound guiding participants' attentional selection, either voluntarily or involuntarily. If it had been the case, similar results in terms of the proportion of dominance measure should have been observed in the bird-sound/maintain bird and bird-sound/passive conditions because they both depend on the participants' devoting attention to the bird percept. In the same vein, similar results should have been observed in the car-sound/maintain car and car-sound/passive conditions because they both depend on the participants' devoting attention to the car percept. As a result, we should have observed that the effect of auditory semantic context was reduced or eliminated when the participants had to attend to a specific percept during the test period (see Hsiao et al., 2010).

On the other hand, the crossmodal semantic modulation of visual awareness in the binocular rivalry situation may be mediated (or enhanced) by selective attention as demonstrated by van Ee et al. (2009). These researchers reported a crossmodal modulation of binocular rivalry perception by following the presentation of series of beeps. However, this effect was only observed when the participants happened to attend to the temporally-coincident percept rather than when they passively viewed the dichoptic figure (Experiment 1 in their study). By contrast, we observed the crossmodal semantic congruency effect on the participants'

proportion of dominance measure in the passive condition, while the magnitude was *no* larger in the maintain bird and maintain car conditions. Nevertheless, a possible explanation for this result is that this measure had almost reached ceiling (or floor) in the bird-sound/maintain bird and the car-sound/maintain car conditions.

In terms of the number of switches, the results reveal that selective attention effectively reduced the frequency of perceptual switches during the test period (see also van Ee et al., 2005). This result is in line with the fact that the occurrence of switches from one percept to the other under conditions of binocular rivalry can be modulated by attention (see Lumer et al., 1998; Leopold and Logothetis, 1999). We therefore observed that selective attention modulated both the proportion of dominance and number of switches measures (van Ee et al., 2005). Note that the attentional effect was more obvious in the maintain bird condition than in the maintain car condition. This result may have been due to the fact that it is harder for the bird percept to dominate (see Experiment 1). Consequently, more attentional effort should be devoted in the maintain bird condition to maintain it. It should, however, also be noted that the auditory semantic context reported in Experiment 2A somehow assisted visual attentional control over perceptual switching. This is evidenced by the fact that when the participants were listening to the bird soundtrack and were instructed to try and maintain the bird percept in awareness, the number of switches was smaller than when they heard the car soundtrack (see the opposite patterns modulated by the sound in the maintain bird and the other two conditions). It is possible that since attentional control over a given object representation relies on holding that target in working memory (see Desimone and Duncan, 1995), the presentation of a semantically congruent auditory soundtrack may have helped the participants to hold the target in mind during the test period.

In summary, the crossmodal semantic modulation was robust in the binocular rivalry situation irrespective of the participants' state of selective attention in terms of determining the proportion of dominance measure. We therefore suggest that attentional control over a specific percept is not a necessary condition for the crossmodal modulation by auditory semantic context in the binocular rivalry situation (cf. van Ee et al., 2009). In addition, we also observed that crossmodal semantic modulation and visual attentional control interacted in terms of the switch times measure.

EXPERIMENT 2B

The final experiment in the present study addressed the question of whether the modulatory effect of auditory semantic context would interact with low-level visual factors in determining the consequences of binocular rivalry. The luminance contrast of a figure provides a bottom-up (i.e., stimulus-driven) factor. That is, a higher luminance contrast figure will likely win the initial competition and be perceived for more of the time than a figure with a lower luminance contrast (Mueller and Blake, 1989; Meng and Tong, 2004). The participants in this experiment heard either the bird or car soundtrack while presented with one of three levels of luminance contrast (see below).

Participants

Six of the participants who took part in Experiment 2A (one dropped out) were tested.

Design and procedure

Two factors, sound (bird and car) and stimulus contrast (equal, bird-high, car-high) were manipulated. In the equal condition, the dichoptic figures used in the previous experiments were presented. In the bird-high condition, the luminance contrast of the bird figure was constant (the Michelson Contrast value measured through the color filter was 85.4% for red and 70.7% for cyan), whereas the luminance contrast of the car figure was reduced (the Michelson Contrast value was 81.0% for red and 65.9% for cyan), and vice versa in the car-high condition. In each trial, the frame that normally provides the attentional instruction was now left blank (just as in Experiment 1A). The participants were instructed to view the figures passively.

Three blocks of experimental trials were presented. There were 12 trials (two sound conditions \times three stimulus contrast conditions \times two color versions of visual pictures) presented in a randomized order in each block. A practice block containing six no-sound trials, two for each stimulus contrast condition, was conducted before the main experiment. The other details were the same as Experiment 2A.

Results

A two-way ANOVA was conducted with the factors of sound and stimulus contrast for each index separately. Analysis of the proportion of dominance data (see **Figure 3D**) revealed significant main effects of sound [$F(1,5) = 7.31$, $MSE = 0.01$, $p < 0.05$, $\eta_p^2 = 0.59$] and stimulus contrast [$F(2,10) = 11.76$, $MSE = 0.003$, $p < 0.005$, $\eta_p^2 = 0.71$]. A *post hoc* Tukey's test revealed that there was a significant difference between the bird-high and car-high conditions ($p < 0.01$), and between the equal and car-high conditions ($p < 0.05$). The interaction between sound and stimulus contrast was, however, not significant [$F(2,10) = 1.14$, $MSE = 0.002$, $p = 0.36$, $\eta_p^2 = 0.20$]. The planned simple main effect of the sound factor revealed that the proportion of dominance of the bird percept was higher when participants heard the bird soundtrack than when they heard the car soundtrack in the equal luminance condition [$F(1,15) = 8.82$, $MSE = 0.005$, $p < 0.01$, $\eta_p^2 = 0.27$], in the car-high condition [$F(1,15) = 5.97$, $p < 0.05$, $\eta_p^2 = 0.18$], while failing to reach significance in the bird-high condition [$F(1,15) = 3.03$, $p = 0.10$, $\eta_p^2 = 0.09$]². Note that the magnitude of the auditory modulation effect (bird- vs. car-sound) was not significantly different across the three levels of stimulus contrast [$F(2,10) = 1.12$, $MSE = 0.003$, $p = 0.37$, $\eta_p^2 = 0.18$]. No significant differences were observed in the analysis of the number of switches (see **Figure 3E**); all $F_s < 4.01$, $p_s > 0.05$, $\eta_p^2 < 0.44$.

Analysis of the first percept data (see **Figure 3F**) revealed a significant main effect of stimulus contrast [$F(2,10) = 7.18$, $MSE = 1.33$, $p < 0.05$, $\eta_p^2 = 0.59$]. A *post hoc* Tukey's test revealed a significant difference between the bird-high and car-high conditions ($p < 0.05$). The other main effect, that of sound [$F(1,5) = 0.83$, $MSE = 1.20$, $p = 0.40$, $\eta_p^2 = 0.14$],

and the interaction between these two factors [$F(2,10) = 2.02$, $MSE = 0.78$, $p = 0.18$, $\eta_p^2 = 0.29$], were not significant.

Discussion

The results of Experiment 2B once again demonstrate that auditory semantic context can increase the proportion of dominance of a semantically congruent percept under conditions of binocular rivalry. Besides, we also replicated the finding that the proportion of dominance is influenced by the relative luminance contrast of the two visual figures (Meng and Tong, 2004). Critically, there was no interaction between the modulation by sound and stimulus contrast on the proportion of dominance measure. This result also indicates that even when the participants were listening to the bird (or car) soundtrack, they were still sensitive to the low-level visual properties (i.e., stimulus contrast in this experiment) of the dichoptic figures during the test period.

On the other hand, the results of Experiment 2B reveal that only the stimulus contrast determined the first percept whereas the auditory semantic context did not (see also Experiment 2A; though see Rommetveit et al., 1968; Costello et al., 2009). That is, the figure that had the higher contrast was perceived first. It should be noted that the measurement of which picture reached awareness first (i.e., the first percept) may merely reflect the result of dichoptic masking rather than genuine binocular rivalry (as indexed by the proportion of dominance duration of a given percept during the test period, see Blake, 1988, p. 140; Noest et al., 2007). That said, the results of Experiment 2B revealed that both visual stimulus contrast and auditory semantic context can modulate the perceptual outcome of binocular rivalry (in an additive fashion), while the former was more dominant than the latter in terms of determining the perceptual outcome of dichoptic masking. We therefore suggest that these two factors, visual stimulus contrast and auditory semantic context, can be dissociated in terms of both the proportion of dominance and first percept measures. On the other hand, the results reported here also suggest that, even though dichoptic masking and binocular rivalry may involve a similar mechanism of interocular suppression (since they were both sensitive to visual stimulus contrast), binocular rivalry seems to involve the later stages of visual processing (perhaps including the semantic level) as well (see Noest et al., 2007; van Boxtel et al., 2007; Baker and Graf, 2009).

GENERAL DISCUSSION

The results of the two experiments reported in the present study demonstrate that a participant's visual awareness in the binocular rivalry situation can be modulated by the semantic context provided by a concurrently presented auditory soundtrack. In Experiment 1A, the proportion of dominance measure of the target percept was smaller when the participant listened to a soundtrack that was incongruent (i.e., that was congruent with the competing percept) than to a soundtrack that was irrelevant to both percepts (i.e., the restaurant soundtrack in the present study). Besides, our results also highlighted the fact that the proportion of dominance measure was unaffected by the instruction to maintain a word in working memory (rather than continuously hearing a soundtrack) during the test period (Experiment 1B). In terms of the proportion of dominance results, we further demonstrated

²It should be noted that in the bird-high condition, all six participants consistently reported the bird percept for more of the time when listening to the bird sound than when listening to the car sound. A repeated measure *t*-test revealed this difference to be significant [$t(5) = 2.22$, $p < 0.05$, one-tailed].

that the modulation by auditory semantic context was additive with that resulting from visual selective attention (Experiment 2A) and additive with that resulting from visual luminance contrast as well (Experiment 2B). Each of the three factors, however, may influence other aspects of participants' performance in the binocular rivalry situation. So, for example, visual selective attention effectively reduced the switch times during the test period (see Experiment 2A; see also van Ee et al., 2005). On the other hand, visual stimulus contrast effectively modulated the participants' first percept (Experiment 2B). Note that the auditory semantic context only modulated switch times when it was simultaneously manipulated with visual selective attention (in Experiment 2A, but not in Experiment 2B), and never modulated the first percept. Considering all these three indices allows us to suggest that the crossmodal modulation of which by auditory semantic context can be dissociated from that by visual selective attention and visual stimulus contrast to a certain extent.

One of the more novel observations to emerge from the results reported here is that the semantic context of an auditory soundtrack can effectively modulate participants' visual awareness. As an extension to the previous studies reported by Sheth and Shimojo (2004), van Ee et al. (2009), and Conrad et al. (2010) in which the visibility of stimuli undergoing visual competition was maintained by the presentation of an auditory stimulus in terms of its physical properties (such as temporal synchrony or direction of motion), the results reported here demonstrate that the crossmodal modulation of binocular rivalry by sound can also extend to its semantic context. The results of the present study therefore provide important evidence that the factors modulating binocular rivalry can reach the semantic level (Engel, 1956; Yang and Yeh, 2011; though see Zimba and Blake, 1983; Blake, 1988), and critically, can occur crossmodally.

It should be noted that since our participants were instructed to attend to the soundtrack during the test period, the soundtrack may be not automatically processed (cf. Schneider et al., 1984; see also van Ee et al., 2009). Here we would rather suggest that the soundtrack should be "selected in," rather than simply "filtered out," in the early stages of auditory information processing (e.g., Treisman and Riley, 1969). Nevertheless, it is possible that once the sound had been processed, the auditory semantic context then unavoidably interacts with any relevant visual information (see Treisman and Davies, 1973; Brand-D'Abrescia and Lavie, 2008; Yuval-Greenberg and Deouell, 2009).

Recently, Pearson et al. (2008) demonstrated that generating the visual mental image of the percepts in a binocular rivalry situation can increase the possibility of that percept winning the competition to reach awareness. Can the crossmodal semantic modulation reported here have been the result of participants generating a visual mental image corresponding to the soundtrack that they happened to be listening to? Mental imagery can be considered as providing a top-down means of modulating a particular object representation, though the time required to generate a mental image is much longer than that required to execute a shift of selective attention (see Pearson et al., 2008, Experiment 3). Note, however, that Pearson et al. (2008) also reported that when the background was 100% illuminated (i.e., a white background, as in the present study), their participants performed similarly under

the conditions of viewing passively and generating a mental image. Besides, Segal and Fusella's (1969, 1970) early studies demonstrate that a person's sensitivity to detect a visual (or auditory) target was lowered when he/she imagined that stimulus in the same sensory modality. Such modality-specific suppression during mental image generation has recently been observed in primary sensory areas in humans (i.e., visual and auditory cortices, see Daselaar et al., 2010). It is true that in most of the experiments reported by Pearson et al. (2008) there was no visual stimulus presented during the imagery period. Therefore, it seems unlikely that a person's ability to invoke mental imagery can be used to enhance a particular percept in the binocular rivalry situation where the visual background was white and visual and auditory stimuli were continuously presented, as in the present study.

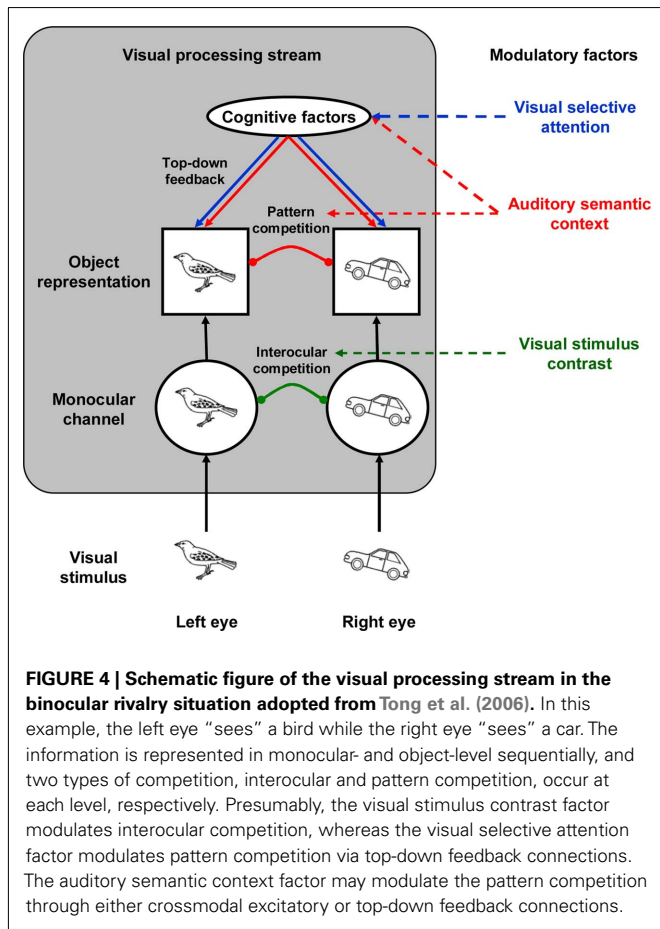
POSSIBLE MECHANISMS OF THE MODULATION BY AUDITORY SEMANTIC CONTEXT ON VISUAL AWARENESS

The results of Experiments 2A and 2B demonstrated that auditory semantic context, visual selective attention, and visual stimulus contrast, all modulated participants' visual perception under conditions of binocular rivalry. All three factors effectively modulated the typical index of proportion of dominance duration of a given percept. Considering the fact that three indices we used, the factors of auditory semantic context and visual stimulus contrast can be dissociated; however, while the modulation of auditory semantic context was not necessarily mediated by visual selective attention, these two factors may interact to some degree.

Let us then consider how the audiovisual semantic congruency effect reported here could be implemented in the model of binocular rivalry based on the three mechanisms (inhibitory, lateral excitatory, and feedback connections) proposed by Tong et al. (2006; see Figure 4). The modulation by visual stimulus contrast can be accounted for by interocular inhibition (Tong, 2001), whereas the modulation elicited by visual selective attention can be accounted for by feedback connections (Tong et al., 2006). Auditory semantic context likely enhanced the representation of semantically congruent visual object representation (Iordanescu et al., 2008; Chen and Spence, 2010) which, as a result, was more likely to win the visual competition. This crossmodal facilitation may be mediated by mid-level lateral excitation, or by top-down feedback connections (Tong et al., 2006). The mid-level lateral excitatory effect can be compared to perceptual grouping (Kovacs et al., 1996; Alais and Blake, 1999; Alais et al., 2006) or the contextual constraints (Treisman, 1962; Shimojo and Nakayama, 1990; Watson et al., 2004) on the perception in the binocular rivalry situation but, in this case, occurring crossmodally. The top-down feedback connection, though, is perhaps the mechanism that auditory semantic context and visual selective attention interactively modulate the visual perception in binocular rivalry.

HUMAN PERCEPTUAL AWARENESS: UNISENSORY OR MULTISENSORY?

Hupé et al. (2008) recently demonstrated that the perceptual outcomes of simultaneously presented visual and auditory bistable stimuli were generated separately. This result may imply that the sites where conscious perception emerges may be separate for different sensory modalities in terms of the traditional view that each sensory modality has its own processing module (e.g.,



Pylyshyn, 1999; Zeki, 2003). Here, on the contrary, we observed that auditory semantic context modulated visual awareness under conditions of binocular rivalry, which is in line with a view of the brain as a closely connected multisensory network: in terms of the neurophysiology, we now know that massive amounts of information is continually being communicated between those brain areas that used to be considered as being sensory-specific. Consequently, many researchers now no longer consider brain regions as being structured as discrete unimodal modules (see Ghazanfar and Schroeder, 2006; Driver and Noesselt, 2008). In terms of psychological functioning, more generally, it is worth considering the powerful constraints that semantics places on the perceptual system as it tries to infer the nature of the environmental stimulation (see Hohwy et al., 2008). That is, audiovisual semantic congruency

REFERENCES

- Alais, D., and Blake, R. (1999). Grouping visual features during binocular rivalry. *Vision Res.* 39, 4341–4353.
- Alais, D., and Blake, R. (eds). (2005). *Binocular Rivalry*. Cambridge, MA: MIT Press.
- Alais, D., Lorenceau, J., Arrighi, R., and Cass, J. (2006). Contour interactions between pairs of Gabors engaged in binocular rivalry reveal a map of the association field. *Vision Res.* 46, 1473–1487.
- Baker, D. H., and Graf, E. W. (2009). On the relation between dichoptic masking and binocular rivalry. *Vision Res.* 49, 451–459.
- Balçetis, E., and Dale, R. (2007). Conceptual set as a top-down constraint on visual object identification. *Perception* 36, 581–595.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–660.
- Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645.
- Bates, E., D’Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., Herron, D., Lu, C. C., Pechmann, T., Pleh, C., Wicha, N., Federmeier, K., Gerdjikova, I., Gutierrez, G., Hung, D., Hsu, J., Iyer, G., Kohnert, K., Mehotcheva, T., Orozco-Figueroa, A., Tzeng, A., and Tzeng, O. (2003). Timed picture naming in seven languages. *Psychon. Bull. Rev.* 10, 344–380.
- Biederman, I. (1972). Perceiving real-world scenes. *Science* 177, 77–80.
- Blake, R. (1988). Dichoptic reading: the role of meaning in binocular rivalry. *Percept. Psychophys.* 44, 133–141.

can provide heuristics, or prior knowledge, on multisensory integration that modulate what we experience on an everyday basis in the real world (i.e., see the literature on the unity assumption, Welch and Warren, 1980; Spence, 2007). The accumulating evidence demonstrating crossmodal semantic interactions in human perception implies that the semantic representations for different sensory modalities are not independent (see McCarthy and Warrington, 1988). However, it is still unclear whether the semantic systems are either completely amodal (e.g., Pylyshyn, 1984), or else the semantic systems for each sensory modality may be highly connected while still retaining some modality-specific information (e.g., Shallice, 1988; Barsalou, 1999; Plaut, 2002). Furthermore, the interplay between perceptual systems and higher-level semantic systems may also imply that perception and cognition share common representation systems, as proposed by the view of grounded cognition (Barsalou, 1999, 2008).

CONCLUSION

The experiments reported here provide empirical support for the claim that auditory semantic context modulates visual perception in the binocular rivalry situation. The results demonstrate that the effect of auditory semantic context is dissociable from the previously reported effects of visual selective attention and visual stimulus contrast (Meng and Tong, 2004). Recently, the crossmodal modulation of visual perception in the binocular rivalry situation has been demonstrated by the concurrent presentation of both tactile (Lunghi et al., 2010) and olfactory stimuli (Zhou et al., 2010). However, the modulation reported in the former case was based on congruency defined in terms of the direction of motion (see Conrad et al., 2010, for the audiovisual case), while in the latter case it was based on odorant congruency that is comparable to the semantic factor investigated in the present study. We therefore suggest that when considering how the dominant percept in binocular rivalry (and so, human visual awareness) emerges, information from other sensory modalities also needs to be considered; and, in turn, that multisensory stimulation provides a novel means other than unimodal stimulation to probe the contextual constraints on human visual awareness.

ACKNOWLEDGMENTS

This research was supported by a joint project funded by the British Academy (CQ ROKK0) and the National Science Council in Taiwan (NSC 97-2911-I-002-038). Yi-Chuan Chen was supported by the Ministry of Education in Taiwan (SAS-96109-1-US-37). Su-Ling Yeh receives supports from NSC 96-2413-H-002-009-MY3 and NSC 98-2410-H-002-023-MY3.

- Blake, R., and Logothetis, N. K. (2002). Visual competition. *Nat. Rev. Neurosci.* 3, 1–11.
- Brand-D'Abrescia, M., and Lavie, N. (2008). Task coordination between and within sensory modalities: effects on distraction. *Percept. Psychophys.* 70, 508–515.
- Chen, Y.-C., and Spence, C. (2010). When hearing the bark helps to identify the dog: semantically-congruent sounds modulate the identification of masked pictures. *Cognition* 114, 389–404.
- Chen, Y.-C., and Yeh, S.-L. (2008). Visual events modulated by sound in repetition blindness. *Psychon. Bull. Rev.* 15, 404–408.
- Chong, S. C., Tadin, D., and Blake, R. (2005). Endogenous attention prolongs dominance durations in binocular rivalry. *J. Vis.* 5, 1004–1012.
- Conrad, V., Bartels, A., Kleiner, M., and Noppeney, U. (2010). Audiovisual interactions in binocular rivalry. *J. Vis.* 10, 27.
- Costello, P., Jiang, Y., Baartman, B., McGlennen, K., and He, S. (2009). Semantic and subword priming during binocular suppression. *Conscious. Cogn.* 18, 375–382.
- Crick, F. (1996). Visual perception: rivalry and consciousness. *Nature* 379, 485–486.
- Daselaar, S. M., Porat, Y., Huijbers, W., and Pennartz, C. M. A. (2010). Modality-specific and modality-independent components of the human imagery system. *Neuroimage* 52, 677–685.
- Davenport, J. L., and Potter, M. C. (2004). Scene consistency in object and background perception. *Psychol. Sci.* 15, 559–564.
- de Gelder, B., and Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *Trends Cogn. Sci. (Regul. Ed.)* 7, 460–467.
- Desimone, R., and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18, 193–222.
- Driver, J., and Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on “sensory specific” brain regions, neural responses, and judgments. *Neuron* 57, 11–23.
- Engel, E. (1956). The role of context in binocular resolution. *Am. J. Psychol.* 69, 87–91.
- Ghazanfar, A. A., and Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends Cogn. Sci. (Regul. Ed.)* 10, 278–285.
- Glaser, W. R., and Glaser, M. O. (1989). Context effects in Stroop-like word and picture processing. *J. Exp. Psychol. Gen.* 118, 13–42.
- Greene, A. J., Easton, R. D., and LaShell, L. S. R. (2001). Visual-auditory events: cross-modal perceptual priming and recognition memory. *Conscious. Cogn.* 10, 425–435.
- Helmholtz, H. V. (1962). *Helmholtz's Treatise on Physiological Optics*, New York, NY: Dover Publications.
- Hohwy, J., Roepstorff, A., and Friston, K. (2008). Predictive coding explains binocular rivalry: an epistemological review. *Cognition* 108, 687–701.
- Hollingworth, A., and Henderson, J. M. (1998). Does consistent scene context facilitate object perception? *J. Exp. Psychol. Gen.* 127, 398–415.
- Hsiao, J. Y. Chen, Y.-C., Spence, C., and Yeh, S.-L. (2010). Semantic congruency, attention, and fixation position modulate conscious perception when viewing a bistable figure. *J. Vis.* 10, 867.
- Hupé, J. M., Joffo, L. M., and Pressnitzer, D. (2008). Bistability for audiovisual stimuli: perceptual decision is modality specific. *J. Vis.* 8, 1.1–15.
- Iordanescu, L., Guzman-Martinez, E., Graboweky, M., and Suzuki, S. (2008). Characteristic sound facilitates visual search. *Psychon. Bull. Rev.* 15, 548–554.
- Jiang, Y., Costello, P., and He, S. (2007). Processing of invisible stimuli: advantage of upright faces and recognizable words in overcoming interocular suppression. *Psychol. Sci.* 18, 349–355.
- Kovacs, I., Paphothomas, T. V., Yang, M., and Feher, A. (1996). When the brain changes its mind: interocular grouping during binocular rivalry. *Proc. Natl. Acad. Sci. U.S.A.* 93, 15508–15511.
- Leopold, D. A., and Logothetis, N. K. (1996). Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature* 379, 549–553.
- Leopold, D. A., and Logothetis, N. K. (1999). Multistable phenomena: changing views in perception. *Trends Cogn. Sci. (Regul. Ed.)* 3, 254–264.
- Levelt, W. J. M. (1965). *On Binocular Rivalry*. The Netherlands: Royal VanGorcum.
- Logothetis, N. K., Leopold, D. A., and Sheinberg, D. L. (1996). What is rivalling during binocular rivalry? *Nature* 380, 621–624.
- Lumer, E. D., Friston, K. J., and Rees, G. (1998). Neural correlates of perceptual rivalry in the human brain. *Science* 280, 1930–1934.
- Lunghi, C., Binda, P., and Morrone, M. C. (2010). Touch disambiguates rivalrous perception at early stages of visual analysis. *Curr. Biol.* 20, R143–R144.
- McCarthy, R. A., and Warrington, E. K. (1988). Evidence for modality-specific meaning systems in the brain. *Nature* 334, 428–430.
- Meng, M., and Tong, F. (2004). Can attention selectively bias bistable perception? Differences between binocular rivalry and ambiguous figures. *J. Vis.* 4, 539–551.
- Mitchell, J. F., Stoner, G. R., and Reynolds, J. H. (2004). Object-based attention determines dominance in binocular rivalry. *Nature* 429, 410–413.
- Molholm, S., Ritter, W., Javitt, D. C., and Foxe, J. J. (2004). Multisensory visual-auditory object recognition in humans: a high-density electrical mapping study. *Cereb. Cortex* 14, 452–465.
- Mueller, T. J., and Blake, R. (1989). A fresh look at the temporal dynamics of binocular rivalry. *Biol. Cybern.* 61, 223–232.
- Noest, A. J., van Ee, R., Nijs, M. M., and van Wezel, R. J. A. (2007). Percept-choice sequences driven by interrupted ambiguous stimuli: a low-level neural model. *J. Vis.* 7, 10.
- Noppeney, U., Josephs, O., Hockings, J., Price, C. J., and Friston, K. J. (2008). The effect of prior visual information on recognition of speech and sounds. *Cereb. Cortex* 18, 598–609.
- Ozkan, K., and Braunstein, M. L. (2009). Predominance of ground over ceiling surfaces in binocular rivalry. *Atten. Percept. Psychophys.* 71, 1305–1312.
- Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Mem. Cognit.* 3, 519–526.
- Pearson, J., Clifford, C. W. G., and Tong, F. (2008). The functional impact of mental imagery on conscious perception. *Curr. Biol.* 18, 982–986.
- Plaut, D. C. (2002). Graded modality-specific specialization in semantics: a computational account of optic aphasia. *Cogn. Neuropsychol.* 19, 603–639.
- Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behav. Brain Sci.* 22, 341–423.
- Pylyshyn, Z. W. (1984). *Computation and Cognition*. Cambridge, MA: MIT Press.
- Rommertveit, R., Toch, H., and Svendsen, D. (1968). Effects of contingency and contrast contexts on the cognition of words: a study of stereoscopic rivalry. *Scand. J. Psychol.* 9, 138–144.
- Schneider, T. R., Engel, A. K., and Debener, S. (2008). Multisensory identification of natural objects in a two-way crossmodal priming paradigm. *Exp. Psychol.* 55, 121–132.
- Schneider, W., Dumais, S. T., and Shiffrin, R. M. (1984). “Automatic and control processing and attention,” in *Varieties of Attention*, eds. R. Parasuraman and R. Davies (New York: Academic Press), 1–27.
- Segal, S. J., and Fusella, V. (1969). Effects of imaging on signal-to-noise ratio, with varying signal conditions. *Br. J. Psychol.* 60, 459–464.
- Segal, S. J., and Fusella, V. (1970). Influence of imagined pictures and sounds on detection of visual and auditory signals. *J. Exp. Psychol.* 83, 458–464.
- Shallice, T. (1988). Specialization within the semantic system. *Cogn. Neuropsychol.* 5, 133–142.
- Sheth, B. R., and Shimojo, S. (2004). Sound-aided recovery from and persistence against visual filling-in. *Vision Res.* 44, 1907–1917.
- Shimojo, S., and Nakayama, K. (1990). Real world occlusion constraints and binocular rivalry. *Vision Res.* 30, 69–80.
- Spence, C. (2007). Audiovisual multisensory integration. *Acoust. Sci. Technol.* 28, 61–70.
- Spence, C. (2011). Crossmodal correspondences: a tutorial review. *Atten. Percept. Psychophys.* 73, 971–995.
- Taylor, K. I., Moss, H. E., Stamatakis, E. A., and Tyler, L. K. (2006). Binding crossmodal object features in perirhinal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 103, 8239–8244.
- Tong, F. (2001). Competing theories of binocular rivalry: a possible resolution. *Brain Mind* 2, 55–83.
- Tong, F., and Engel, S. A. (2001). Interocular rivalry revealed in the human cortical blind-spot representation. *Nature* 411, 195–199.
- Tong, F., Meng, M., and Blake, R. (2006). Neural bases of binocular rivalry. *Trends Cogn. Sci. (Regul. Ed.)* 10, 502–511.
- Tong, F., Nakayama, K., Vaughan, J. T., and Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron* 21, 753–759.
- Treisman, A. (1962). Binocular rivalry and stereoscopic depth perception. *Q. J. Exp. Psychol.* 14, 23–37.
- Treisman, A., and Davies, A. (1973). “Dividing attention to ear and eye,” in *Attention and Performance IV*, ed. S. Kornblum (New York: Academic Press), 101–117.
- Treisman, A., and Riley, J. G. A. (1969). Is selective attention selective

- perception or selective response? A further test. *J. Exp. Psychol.* 79, 27–34.
- van Atteveldt, N., Formisano, E., Goebel, R., and Blomer, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron* 43, 271–282.
- van Boxtel, J. J. A., van Ee, R., and Erkelens, C. J. (2007). Dichoptic masking and binocular rivalry share common perceptual dynamics. *J. Vis.* 7, 3.1–11.
- van Ee, R., van Boxtel, J. J. A., Parker, A. L., and Alais, D. (2009). Multimodal congruency as a mechanism for willful control over perceptual awareness. *J. Neurosci.* 29, 11641–11649.
- van Ee, R., van Dam, L. C. J., and Brouwer, G. J. (2005). Voluntary control and the dynamics of perceptual bi-stability. *Vision Res.* 45, 41–55.
- Watson, T. L., Pearson, J., and Clifford, C. W. G. (2004). Perceptual grouping of biological motion promotes binocular rivalry. *Curr. Biol.* 14, 1670–1674.
- Welch, R. B., and Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychol. Bull.* 3, 638–667.
- Yang, Y. H., and Yeh, S.-L. (2011). Accessing the meaning of invisible words. *Conscious. Cogn.* 20, 223–233.
- Yuval-Greenberg, S., and Deouell, L. Y. (2009). The dog's meow: asymmetrical interaction in cross-modal object recognition. *Exp. Brain Res.* 193, 603–614.
- Zeki, S. (2003). The disunity of consciousness. *Trends Cogn. Sci. (Regul. Ed.)* 7, 214–218.
- Zhou, W., Jiang, Y., He, S., and Chen, D. (2010). Olfaction modulates visual perception in binocular rivalry. *Curr. Biol.* 20, 1356–1358.
- Zimba, L. D., and Blake, R. (1983). Binocular rivalry and semantic processing: out of sight, out of mind. *J. Exp. Psychol. Hum. Percept. Perform.* 9, 807–815.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 17 June 2011; accepted: 15 August 2011; published online: 12 September 2011.

Citation: Chen Y-C, Yeh S-L and Spence C (2011) Crossmodal constraints on human perceptual awareness: auditory semantic modulation of binocular rivalry. *Front. Psychology* 2:212. doi: 10.3389/fpsyg.2011.00212

This article was submitted to *Frontiers in Perception Science*, a specialty of *Frontiers in Psychology*.

Copyright © 2011 Chen, Yeh and Spence. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.