Behavioral/Cognitive

Vision Dominates at the Preresponse Level and Audition Dominates at the Response Level in Cross-modal Interaction: Behavioral and Neural Evidence

Qi Chen[1] **and Xiaolin Zhou**[2,3,4]

[1]Center for Studies of Psychological Application and Department of Psychology, South China Normal University, Guangzhou 510631, China, and [2]Center for Brain and Cognitive Sciences and Department of Psychology, [3]Key Laboratory of Machine Perception (Ministry of Education), and [4]PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing 100871, China

There are ongoing debates on the direction of sensory dominance in cross-modal interaction. In the present study, we demonstrate that the specific direction of sensory dominance depends on the level of processing: vision dominates at earlier stages, whereas audition dominates at later stages of cognitive processing. Moreover, these dominances are subserved by different neural networks. In three experiments, human participants were asked to attend to either visual or auditory modality while ignoring simultaneous stimulus inputs from the other modality. By manipulating three levels of congruency between the simultaneous visual and auditory inputs, congruent (C), incongruent at preresponse level (PRIC), and incongruent at response level (RIC), we differentiated the cross-modal conflict explicitly into preresponse (PRIC > C) and response (RIC > PRIC) levels. Behavioral data in the three experiments consistently suggested that visual distractors caused more interference to auditory processing than vice versa (i.e., the typical visual dominance) at the preresponse level, but auditory distractors caused more interference to visual processing than vice versa (i.e., the typical auditory dominance) at the response level regardless of experimental tasks, types of stimuli, or differential processing speeds in different modalities. Dissociable neural networks were revealed, with the default mode network being involved in the visual dominance at the preresponse level and the prefrontal executive areas being involved in the auditory dominance at the response level. The default mode network may be attracted selectively by irrelevant visual, rather than auditory, information via enhanced neural coupling with the ventral visual stream, resulting in visual dominance at the preresponse level.

Introduction

Although constantly bombarded with streams of information from multiple sensory modalities, organisms are able to "look without seeing" while attending to auditory information and to "listen without hearing" while attending to visual information. The prefrontal areas are known to be involved in resolving cross-modal conflicts (Weissman et al., 2004; Macaluso and Driver, 2005; Mayer et al., 2009; Orr and Weissman, 2009), but there are ongoing debates regarding the extent to which information from one particular modality dominates or interferes with the processing of information from another modality and whether there are asymmetries in such sensory dominance (Yuval-Greenberg and Deouell, 2007; Yuval-Greenberg and Deouell, 2009).

Received April 24, 2012; revised Jan. 3, 2013; accepted Feb. 11, 2013.

Author contributions: Q.C. and X.Z. designed research, Q.C. performed research, Q.C. analyzed data, and Q.C. and X.Z. wrote the paper.

This work was supported by the Natural Science Foundation of China (Grants 31070994, 30070260, 30470569, 60435010, 30970895, 30970889, and 30110972) and the Ministry of Science and Technology of China (Grant 2010CB833904). Q.C. was also supported by the Foundation for the Author of National Excellent Doctoral Dissertation of China (Grant 200907) and by the Program for New Century Excellent Talents in the University of China.

The authors declare no competing financial interests.

Correspondence should be addressed to Dr. Qi Chen, Department of Psychology, South China Normal University, 510631 Guangzhou, P. R. China. E-mail: qi.chen27@gmail.com.

DOI:10.1523/JNEUROSCI.1985-12.2013

 0270-6474/13/337109-13$15.00/0

Cross-modal conflict may occur at different levels of cognitive processing, including early perceptual processing, postperceptual (e.g., semantic) representation, and response selection (Marks, 2004). The asymmetry of functional dominance between modalities during cross-modal interaction may vary according to the level of processing. Previous evidence suggested that vision might dominate at the preresponse level, whereas audition might dominate at the response level. For example, participants fail more frequently in responding to the auditory component of bimodal audiovisual stimuli than to the visual component, indicating visual dominance in cross-modal interaction (i.e., the Colavita effect; Colavita, 1974). A signal detection theory study on the Colavita effect showed that perceptual sensitivity to auditory stimuli is reduced significantly when the auditory stimuli are paired with visual stimuli, but the response criterion is unchanged, implying that visual dominance may occur only at the preresponse level (Koppen et al., 2009). In contrast, in motor-related tasks (e.g., rhythmic tapping), auditory distractors cause more interference with tapping in synchrony to visual signals than vice versa (Repp and Penel, 2004; Kato and Konishi, 2006; Mayer et al., 2009), implying auditory dominance at the response level.

To test the above hypothesis, we created three levels of cross-modal congruency between simultaneously presented visual and auditory stimuli according to whether the bimodal inputs referred to

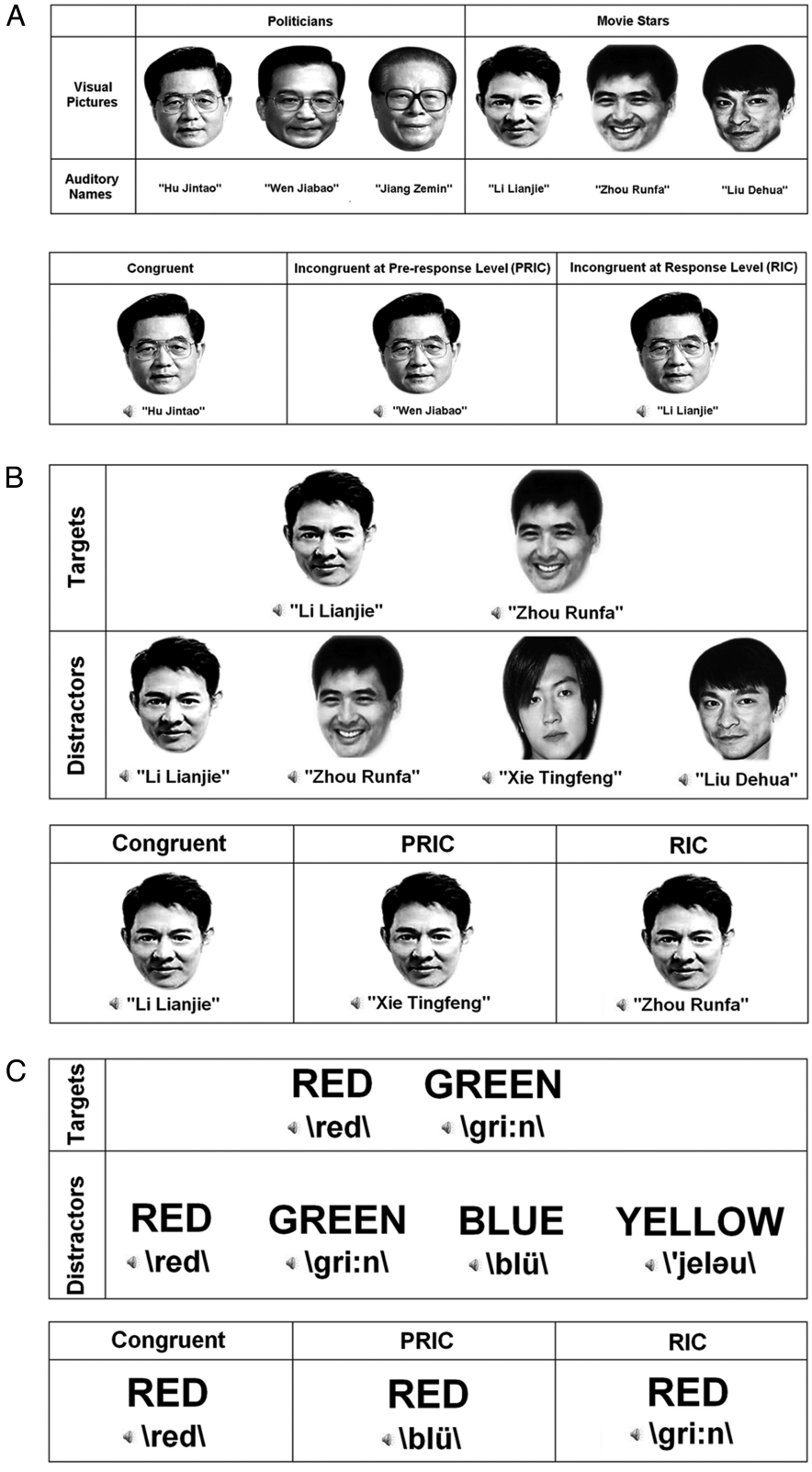

Figure 1. ***A***, Design and exemplar stimuli in Experiment 1. Top: Three faces of politicians and three faces of movie stars were used as the visual stimuli and the spoken names of the six persons were used as the auditory stimuli. Bottom: Three levels of congruency were created. In the C condition, the auditory name and the visual face refer to the same person. In the PRIC condition, the auditory and visual stimuli refer to two different persons, either both politicians or both movie stars. In the RIC condition, the auditory and visual stimuli refer to a politician and a movie star or vice versa. ***B***, Design and exemplar stimuli in Experiment 2. Top: Faces of two movie stars and their spoken names served as targets. Another two movie stars and the two target movie stars served as distractors. Examples of the manipulation of the three levels of congruency are given for the situation in which the visual modality was attended. Bottom: Examples of the manipulation of the three levels of congruency are also given for the situation in which the visual modality was attended. ***C***, Design and exemplar stimuli in Experiment 3. Two written color words and their verbal sounds served as targets. Another two color words and the two target color words served as distractors. Examples of the manipulation of the three levels of congruency are also given for the situation in which the visual modality was attended.

the same identity requiring the same response (congruent, C), different identities requiring the same response (incongruent at the preresponse level, PRIC), or different identities requiring different responses (incongruent at both the preresponse and response levels, RIC; Fig. 1). In this way, we could disentangle the cross-modal conflict at the preresponse (PRIC > C) and response (RIC > PRIC) levels. In the classical paradigm, targets and distractors are from the same sensory modality (Eriksen and Schultz, 1979). However, in the present study, distractors were delivered via a modality different from targets such that preresponse and response conflicts between different sensory modalities could be disentangled explicitly. We predicted that if vision dominates at the preresponse level while audition dominates at the response level, then visual distractors should cause larger preresponse interference (PRIC > C) to auditory processing than vice versa, while auditory distractors should cause larger response level interference (RIC > PRIC) to visual processing than vice versa. Prefrontal areas may be involved neurally in resolving the potentially larger response level interference caused by auditory compared with visual distractors (MacDonald et al., 2000; van Veen and Carter, 2005). We did not have specific predictions concerning neural correlates underlying the potentially larger preresponse interference caused by visual compared with auditory distractors.

Materials and Methods

Experiment 1

Participants. Eighteen healthy university students (nine female and nine male, 21–23 years of age) participated in the experiment. Participants were all right handed and had normal or corrected-to-normal visual acuity. None had a history of neurological or psychiatric disorders. All participants gave informed written consent before the experiments in accordance with the Declaration of Helsinki. This experiment was approved by the ethics committee of the Department of Psychology, Peking University.

Experimental design and stimuli. Visual stimuli consisted of photos of six faces, three of politicians and three of movie stars; auditory stimuli consisted of six spoken names corresponding to the six persons (Fig. 1*A*, top). The

six celebrities were all male and their names all consisted of three syllables (i.e., three Chinese characters). Visual stimuli were presented through an LCD projector onto a rear screen located behind the participant's head, and the participant viewed the screen via an angled mirror mounted on the head-coil of the MRI setup. Auditory stimuli were voice recordings of a male speaker delivered binaurally via MR-compatible stereo headphones, with the length of each name being 450 ms. Headphone volume was adjusted for each participant so that the auditory stimuli could be heard clearly over the background scanner noise. All of the visual pictures measured 5° (horizontally) × 6° (vertically) in visual angle, with the presentation duration of each picture being 450 ms. The default visual display was a cross (1.5° × 1.5°) at the center of the screen.

The experimental task was to judge whether the attended visual face or auditory name referred to a politician or a movie star. For a participant and across the experiment, all of the faces and spoken names were potential targets that required responses. Participants used the index and middle fingers of their right hands to respond by pressing one key on the response box for politicians or another key for movie stars. The mapping between the two response keys and politicians versus movie stars was counterbalanced across participants. In the C condition, the visual face and the auditory name were the same person (Fig. 1*A*, bottom). In the PRIC condition, the visual face and the auditory name referred to two different people, but these two people were either both politicians or both movie stars. Therefore, the face and the name in a PRIC trial potentially required *the same response* (i.e., mapped onto the same response key). The crucial difference between the C and the PRIC conditions was that the face and the name referred to the same semantic entity (the same person) in the former but to different semantic entities (different persons) in the latter. In the RIC condition, the face was that of either a politician or a movie star, and the auditory name referred to the unused profession in the pairing (i.e., a politician's face was paired with the auditory stimulus of a movie star's name and vice versa). In this way, the visual and auditory stimuli referred to different semantic entities and were mapped to different response keys. Therefore, the experimental design was a 2 (modality attended: Attend_Auditory vs Attend_Visual) × 3 (congruency: C, PRIC, and RIC) factorial design. In this method for differentiating conflicts at different levels of cognitive processing, the response level conflict (RIC > PRIC) was mapped explicitly onto the response selection stage. The preresponse level conflict (PRIC > C), however, could be mapped to any stage before response selection, including both perceptual and postperceptual (e.g., semantic) representation stages. Although the present method does not clearly differentiate the stages of processing at the preresponse level, given that the initial sensory/perceptual differences between the two input modalities were present in all three (C, PRIC, and RIC) conditions of the present *cross-modal* paradigm, the preresponse conflict (PRIC > C) was most likely to take place at the postperceptual semantic stage.

Procedures. The presentation of stimuli used a hybrid fMRI design in which the attended modality was blocked, and the C, PRIC, and RIC trials were mixed randomly within each block. Participants were asked, in each block, to pay attention to either the visual or the auditory stimuli while ignoring stimuli from the other modality. Participants were instructed to fixate on the central cross throughout the experiment without moving their eyes. In each trial, a face and an auditory name were simultaneously presented for 450 ms. Experiment 1 had two sessions of functional scanning. Within each session, there were 48 trials for each of the six experimental conditions, resulting in a total of 384 trials (288 experimental trials and 96 null trials). In a null trial, only the central fixation cross was displayed. For the Attend_Visual and the Attend_Auditory conditions, respectively, null trials and C, PRIC, and RIC trials were randomly mixed and then divided into 24 test blocks. Each block had 8 trials and lasted for 20 s. Attend_Visual and Attend_Auditory blocks alternated with each other. Each block started with a 2 s visual instruction telling participants which modality to attend. Event-related procedures were embedded within the Attend_Visual and Attend_Auditory blocks, and the time interval between the eight trials in a block was jittered. The intertrial intervals were jittered from 2000–3000 ms (2000, 2250, 2500, 2750, and 3000 ms). The temporal order of trials was randomized for each participant individually to avoid potential problems of unbalanced transition probabilities. All participants completed a training section of 10 min before the scanning.

Data acquisition and preprocessing. A Siemens 3T Trio system with a standard head coil at Beijing MRI Center for Brain Research was used to obtain T2*-weighted echo-planar images (EPIs) with blood oxygenation level-dependent contrast. The matrix size was 64 × 64 mm^3 and the voxel size was 3.4 × 3.4 × 5 mm^3. Twenty-four transversal slices of 4 mm thickness that covered the whole brain were acquired sequentially with a 1 mm gap (TR = 1.5 s, TE = 30 ms, FOV = 220 mm, flip angle = 90°). There were two runs of functional scanning, each of which included 760 EPI volumes. For each run, the first five volumes were discarded to allow for T1 equilibration effects.

Data were preprocessed with Statistical Parametric Mapping software SPM8 (Wellcome Department of Imaging Neuroscience, London, http://www.fil.ion.ucl.ac.uk). Images were realigned to the first volume to correct for interscan head movements. The mean EPI image of each participant was then computed and spatially normalized to the MNI single participant template using the "unified segmentation" function in SPM8. This algorithm is based on a probabilistic framework that enables image registration, tissue classification, and bias correction to be combined within the same generative model. The resulting parameters of a discrete cosine transform, which define the deformation field necessary to move individual data into the space of the MNI tissue probability maps, were then combined with the deformation field transforming between the latter and the MNI single participant template. The ensuing deformation was subsequently applied to individual EPI volumes. All images were thus transformed into standard MNI space and resampled to 2 × 2 × 2 mm^3 voxel size. The data were then smoothed with a Gaussian kernel of 8 mm full-width half-maximum to accommodate interparticipant anatomical variability.

Statistical analysis of behavioral data. For each of the six experimental conditions, omissions, incorrect responses, and trials with reaction times (RTs) 3 SDs away from the mean RT were excluded from further analysis (1.1% of the overall data points). Mean RTs of the remaining trials were then calculated. Error rates in each of the six experimental conditions were calculated as the proportion between the sum of omissions and incorrect trials and the overall number of trials. Mean RTs and error rates were submitted to a 2 (modality attended: Attend_Auditory vs Attend_Visual) × 3 (congruency: C, PRIC and RIC) repeated-measures ANOVA.

Statistical analysis of imaging data. Data were high-pass-filtered at 1/128 Hz and analyzed with a general linear model as implemented in SPM8. Temporal autocorrelation was modeled using an AR(1) process. At the individual level, the general linear model was used to construct a multiple regression design matrix. For each of the two sessions, six experimental conditions were modeled: Auditory_C, Auditory_PRIC, Auditory_RIC, Visual_C, Visual_PRIC, and Visual_RIC. The six event types were time locked to the onset of stimuli by a canonical synthetic hemodynamic response function and its first-order time derivative with an event duration of 0 s. In addition, all of the instructions were included as confounds. All of the error trials and trials in which RTs were outside of the mean RT ± 3 SD were modeled separately as another regressor of no interest. The six head movement parameters derived from the realignment procedure were also included as confounds for each session. Parameter estimates were calculated subsequently for each voxel using weighted least-squares analysis to provide maximum likelihood estimators based on the temporal autocorrelation of the data. No global scaling was applied.

For each participant, simple main effects for each of the six experimental conditions were computed by applying appropriate "1 0" baseline contrasts; that is, the experimental conditions versus implicit baseline (null trials) contrasts. The 6 first-level individual contrast images were then fed to a 2 × 3 within-subject ANOVA at the second group level using a random-effects model (i.e., the flexible factorial design in SPM8 including an additional factor modeling the subject means). In the modeling of variance components, violations of sphericity were allowed for by modeling nonindependence across parameter estimates from the same participant and allowed for unequal variances between conditions and between participants using the standard implementation in SPM8. Areas of activation were identified as significant only if they passed the

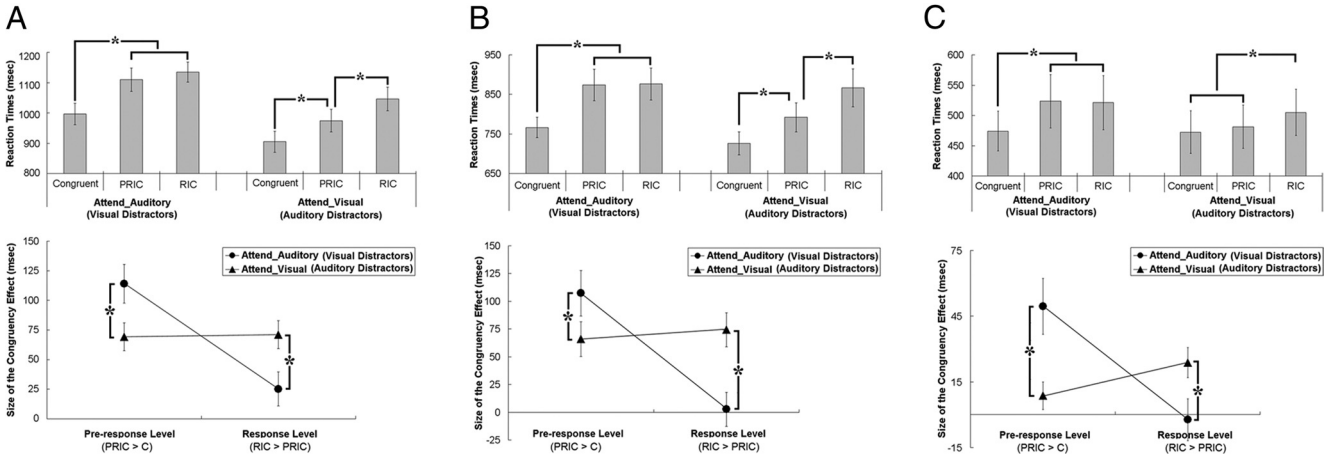

Figure 2. ***A***, Behavioral results of Experiment 1. Top: Mean RTs are shown as a function of the six experimental conditions. Bottom: Sizes of cross-modal conflict at the preresponse (PRIC > C) and response (RIC > PRIC) levels are shown as a function of the attended modality. ***B***, Behavioral results of Experiment 2. Top: Mean RTs are shown as a function of the six experimental conditions. Bottom: Sizes of cross-modal conflict at the preresponse (PRIC > C) and response (RIC > PRIC) levels are shown as a function of the attended modality. ***C***, Behavioral results of Experiment 3. Top: Mean RTs are shown as a function of the six experimental conditions. Bottom: Sizes of cross-modal conflict at the preresponse (PRIC > C) and response (RIC > PRIC) levels are shown as a function of the attended modality. Conditions denoted by an asterisk indicate a significant difference between them ($p < 0.05$, Bonferroni corrected).

threshold of $p < 0.001$ family-wise error (FWE) corrected for multiple comparisons at the cluster level with an underlying voxel level of $p < 0.001$ uncorrected (Poline et al., 1997).

Psychophysiological interaction analysis. To investigate further how the default mode network was involved in the visual dominance at the preresponse level, psychophysiological interaction (PPI) analysis was used to examine the context-specific functional modulation of neural activity across the brain by the neural activity in orbital prefrontal cortex (OPFC; Fig. 4*A*, Table 2*a*). PPI analysis allows for detecting regionally specific responses in one brain area in terms of the interaction between input from another brain region and a cognitive/sensory process (Friston et al., 1997). The neural activity in the OPFC region was used as the physiological factor and the Attend_Auditory versus Attend_Visual contrast as the psychological factor. For each participant, the preresponse interaction contrast Attend_Auditory (PRIC > C) > Attend_Visual (PRIC > C) inclusively masked by Attend_Auditory (PRIC > C) was first calculated at the individual level. Subsequently, each participant's individual peak voxel was determined as the maximally activated voxel within a sphere of 16 mm radius (i.e., twice the smoothing kernel) around the coordinates of the peak voxel within OPFC from the second-level group analysis (MNI: 0, 42, −2, Fig. 4*A*, Table 2*a*). Individual peak voxels from every participant ($x = -3 \pm 5, y = 46 \pm 6, z = 0 \pm 7$) were located in the same anatomical structure. Next, OPFC time series were extracted from a sphere of 4 mm radius (twice the voxel size) around the individual peak voxels for the two sessions, respectively (without deconvolution because of the *blocked* Attend_Auditory and Attend_Visual factors). For each of the two sessions at the first individual level, PPI analysis used one regressor representing the extracted time series in the given region of interest in OPFC (i.e.,the physiological variable), one regressor representing the psychological variable of interest (i.e., Attend_Auditory > Attend_Visual), and a third regressor representing the cross product of the previous two (the PPI term). An SPM was calculated to reveal areas for which activation was predicted by the PPI term, with the physiological and the psychological regressors being treated as confound variables. The PPI analysis was performed for each participant and then entered into a random-effects group analysis ($p < 0.001$ with FWE correction for multiple comparisons at the cluster level with an underlying voxel threshold at $p < 0.005$ uncorrected).

The preresponse conflict (PRIC > C) caused by visual distractors when the auditory modality was attended was larger than that of the auditory distractors when the visual modality was attended (Fig. 2*A*), which corresponded to enhanced neural activity (reduced deactivation) in OPFC in the PRIC condition compared with the C condition when the auditory modality was attended (Fig. 4*A*, left, shading). Therefore, it may be hypothesized that in a multimodal environment, OPFC may be attracted selectively by irrelevant visual, rather than auditory, stimuli and thus pass them into visual awareness. Correspondingly, compared with the irrelevant auditory stimuli, the perceptual and/or the semantic representations of the irrelevant PRIC visual stimuli may cause larger conflicts and the preresponse representations of the irrelevant C visual stimuli may cause larger facilitations, resulting in larger preresponse conflicts (PRIC > C). If this hypothesis is accurate, OPFC in the Attend_Auditory condition using visual distractors should show higher neural coupling with the visual perceptual processing and semantic representations compared with the Attend_Visual condition using auditory distractors.

Experiment 2

In Experiment 1, the PRIC condition induced cross-modal conflict between two semantic entities within the same semantic category, while the RIC condition induced cross-modal conflict both between different semantic categories and between different response codes (Fig. 1*A*). One may argue that although PRIC > C gave a "pure" preresponse conflict, RIC > PRIC may reveal effects of conflicts not only between different responses but also between different semantic categories. To rule out this possibility, a person-identification task was used in Experiment 2 in which participants were asked to make decisions regarding the identity of the target person. For a given participant, the target stimuli were two of the four movie stars (Fig. 1*B*); the experimental task involved discriminating which one of the two target stars was presented in the attended modality while ignoring the input from the other modality. The two target persons corresponded to two response keys. In the C condition, the visual face and the auditory name referred to the same person in a trial. In the PRIC condition, the distractor in the unattended modality was one of the two nontarget movie stars who were not assigned to any response keys. In the RIC condition, the distractor in the unattended modality was a target person different from the one in the attended modality. In this way, the PRIC condition induced conflicts between two semantic entities within the same semantic category but did not induce conflicts at the response level. Moreover, the RIC condition induced conflicts both between two semantic entities within the same semantic category and between two different response codes. Therefore, the contrast RIC > PRIC could reveal cross-modal conflict at the "pure" response level.

Participants. Seventeen healthy university students (9 female and 8 male, 19–23 years of age) participated in the experiment. The recruitment conditions were the same as those in Experiment 1. All participants gave informed consent before the experiment in accordance with the Declaration of Helsinki and were paid for their participation. This exper-

iment was approved by the ethics committee of the Department of Psychology, South China Normal University.

Experimental design and stimuli. The experimental design was essentially the same as that in Experiment 1, except that the target stimuli belonged to the same semantic category. Visual stimuli were photos of four movie stars, two of which served as targets and two as distractors; the four names of the movie stars served as the auditory stimuli, with each name corresponding to three syllables and three characters (Fig. 1*B*, top). Participants were asked to identify which one of the two target movie stars was presented in the attended modality. They were instructed to use the index or middle finger of the right hand to press one of the two keys that corresponded to the two targeted movie stars. The mapping between the two keys and the two stars was counterbalanced across participants. For a given participant, the potential response targets were two names and faces, whereas the remaining two names and faces served as distractors. The selection of the targeted two movie stars was counterbalanced across participants.

Procedures. Experiment 2 had only one session of scanning; all of the settings were the same as those in Experiment 1.

Data acquisition and preprocessing. The matrix size was 64×64 mm^3 and the voxel size was $3.1 \times 3.1 \times 3.0$ mm^3. Thirty-six transversal slices of 3 mm thickness that covered the whole brain were acquired sequentially with a 0.3 mm gap (TR = 2.2 s, TE = 30 ms, FOV = 220 mm, flip angle = 90°). The one-run functional scanning had 468 EPI volumes.

Statistical analysis of behavioral data. The same procedure used in Experiment 1 was used, with 1.5% of the overall data points being excluded as outliers.

Statistical analysis of imaging data. Steps of statistical analysis of imaging data were essentially the same as those in Experiment 1. Areas of activation were identified as significant only if they passed the threshold of $p < 0.005$ with FWE correction for multiple comparisons at the cluster level with an underlying voxel level of $p < 0.005$ uncorrected. A less conservative criterion for activation was used in Experiment 2 because this experiment had only half of the trial numbers in each condition compared with Experiment 1.

PPI analysis. Similar to Experiment 1, OPFC ($x = -12 \pm 5$, $y = 53 \pm 6$, $z = -3 \pm 8$) was treated as the physiological factor (source region) and the contrast Attend_Auditory > Attend_Visual was treated as the psychological factor. Areas of activation were identified as significant only if they passed the threshold of $p < 0.001$ with FWE corrected for multiple comparisons at the cluster level with an underlying voxel level of $p < 0.005$ uncorrected (Poline et al., 1997).

Experiment 3

Because three-syllable auditory words were used in Experiments 1 and 2, one may argue that the reason for auditory information causing more interference at the response level could be due to the relatively slower auditory processing. One might also argue that the relatively faster visual processing could be responsible for visual information causing more interference at the earlier preresponse level. Moreover, the complexity of visual faces and auditory words may also contribute to these modality differences. To rule out this possibility, we performed a behavioral experiment using only *one-syllable*, *one-character* color words (Fig. 1*C*, top).

Participants. Fifteen university students (8 female, 7 male, 20–24 years of age) participated in the experiment. They were recruited using the same criteria as in Experiments 1 and 2. In addition, they had no color blindness/weakness. All participants gave informed consent before the experiment in accordance with the Declaration of Helsinki and were paid for their participation. This study was approved by the ethics committee of the Department of Psychology, South China Normal University.

Experimental design and stimuli. The experimental design was essentially the same as that in Experiment 2, with the exception being the stimuli. The targets were two colors (e.g., red and green) and their corresponding written words and verbal pronunciations; the distractors were the written forms and pronunciations of two other color words (e.g., blue and yellow) (Fig. 1*C*, top). The duration of each of the visual or auditory stimuli was 300 ms. Participants were asked to judge which one of the two target colors was presented in the attended modality while ignoring the distractor in the unattended modality. The two target color words were assigned to two response buttons. The correspondence between the target words and the response buttons was counterbalanced across participants. In the C condition, the bimodal stimuli referred to the same color. In the PRIC condition, the distractor was one of the two other color words to which no response codes were assigned. In the RIC condition, the distractor was the other target color word not used in the attended modality.

Procedures. The experiment was run in a soundproof and dimly lit room. Participants sat in frontal of a monitor screen, with the eye-to-monitor distance being kept at 65 cm. Auditory stimuli were delivered binaurally via stereo headphones. Other aspects of the experimental procedure were essentially the same as those in Experiment 2.

Statistical analysis of behavioral data. The same procedure used in Experiment 1 was used, with 0.9% of the overall data points being excluded as outliers.

Table 1. Mean RTs (ms) and error rates (%) with SEs in Experiments 1–3

	Modality attended	Auditory			Visual		
	Congruency	C	PRIC	RIC	C	PRIC	RIC
a. Experiment 1	RT (SE)	996 (36)	1110 (38)	1135 (33)	905 (35)	975 (38)	1046 (39)
	Error rates (SE)	5.6 (2.0)	5.7 (1.3)	12.6 (1.9)	4.4 (1.2)	5.2 (1.1)	10.5 (2.2)
b. Experiment 2	RT (SE)	766 (26)	874 (40)	876 (40)	726 (29)	792 (37)	866 (48)
	Error rates (SE)	2.6 (1.0)	6.6 (2.4)	10.0 (2.8)	3.8 (1.9)	5.0 (2.2)	11.3 (3.1)
c. Experiment 3	RT (SE)	474 (33)	523 (44)	521 (45)	472 (35)	481 (36)	505 (38)
	Error rates (SE)	1.4 (0.4)	3.1 (0.8)	3.6 (0.8)	3.5 (0.7)	3.1 (0.5)	7.4 (1.3)

Results

Experiment 1

Behavioral data

For RTs, the main effect of attended modality was significant ($F_{(1,17)} = 49.19$, $p < 0.001$), indicating that RTs to the visual targets (975 ms) were significantly faster than RTs to the auditory targets (1080 ms) (Fig. 2*A*, top, Table 1*a*). The main effect of congruency was also significant ($F_{(2,34)} = 72.78$, $p < 0.001$). Further pairwise comparisons with Bonferroni correction indicated that RTs in the PRIC condition (1042 ms) were significantly slower than RTs in the C condition (951 ms), and RTs in the RIC condition (1090 ms) were significantly slower than RTs in the PRIC condition ($p < 0.05$), indicating significant cross-modal conflicts at the preresponse and response levels. The interaction between attended modality and congruency was also significant ($F_{(2,34)} = 4.95$, $p = 0.01$). Planned *t* tests on simple effects indicated that the preresponse level conflict was significant both in the Attend_Auditory condition (114 ms; $t_{(17)} = 6.97$, $p < 0.001$) and in the Attend_Visual condition (69 ms; $t_{(17)} = 5.3$, $p < 0.001$) (Fig. 2*A*, top). Conversely, the response level conflict was significant only in the Attend_Visual condition (71 ms; $t_{(17)} = 6.05$, $p < 0.001$), not in the Attend_Auditory condition (25 ms; $t_{(17)} = 1.77$, $p = 0.095$) (Fig. 2*A*, top). The size of the preresponse level conflict was significantly larger when auditory modality was attended (114 ms) than when visual modality was attended (69 ms; $t_{(17)} = 2.30$, $p < 0.05$), whereas the size of the response level conflict was significantly larger when visual modality was attended (71 ms) than when auditory modality was attended (25 ms; $t_{(17)} = 2.70$, $p < 0.05$) (Fig. 2*A*, bottom).

Analysis of error rates revealed a significant main effect of attended modality ($F_{(1,17)} = 4.70$, $p < 0.05$), with more errors in responding to auditory stimuli (8.0%) than to visual stimuli (6.7%) ($p < 0.05$. The main effect of congruency was also significant ($F_{(2,34)} = 23.4$, $p < 0.001$), with more errors in the RIC condition (11.5%) than in the PRIC (5.4%) and the C (5%) conditions ($p < 0.05$, Bonferroni corrected). The interaction be-

tween attended modality and congruency was not significant ($F < 1$).

Imaging data

We first calculated neural activations associated with cross-modal conflict at different levels. The main effect contrast PRIC > C, collapsing over attended modalities, revealed activations in left precentral gyrus (PCG) extending to left inferior frontal gyrus (IFG), right middle frontal gyrus extending to right IFG, and left supplemental motor area (SMA) (Fig. 3*A*, green). The contrast RIC > C activated a neural network similar to the contrast PRIC > C (Fig. 3*A*, red). In previous studies on cross-modal conflicts, the preresponse and response level conflicts were not disentangled and were combined in the incongruent condition (Weissman et al., 2004; Macaluso and Driver, 2005; Mayer et al., 2009; Orr and Weissman, 2009), which was identical to the RIC condition in the present experiment. Accordingly, the bilateral prefrontal, premotor, and SMA activations in the main effect contrast RIC > C (collapsed over attended modalities) in the present experiment was consistent with the prefrontal activations in the Incongruent > C contrasts in the previous studies. Moreover, in other studies in which the preresponse and response level conflicts *were* dissociated for the same-modality (visual) targets and distractors (van Veen et al., 2001; e.g., Milham et al., 2001; Bunge et al., 2002; van Veen and Carter, 2005; Kim et al., 2011), similar prefrontal activations were also revealed for the RIC conditions.

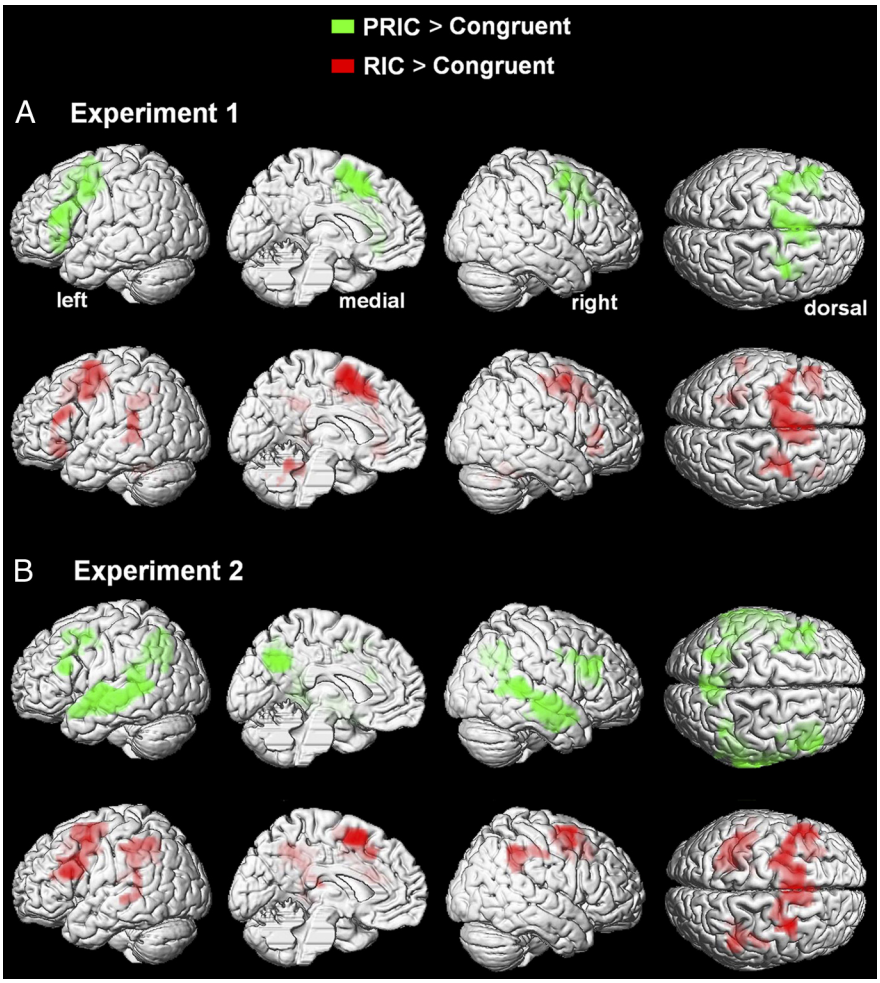

Figure 3. Main effects of cross-modal congruency in Experiment 1 (***A***) and Experiment 2 (***B***), collapsing over attended modalities.

The contrast RIC > PRIC, however, did not reveal any significant activations. Because our behavioral data suggested that the response level conflict (RIC > PRIC) was induced only by auditory distractors when the visual modality was attended, and not by visual distractors when the auditory modality was attended (Fig. 2*A*), neural substrates of the response level conflict specific to the Attend_Visual condition may be localized by the neural interaction contrast, rather than by the main effect.

Visual dominance at the preresponse level. Modality-specific activation at the preresponse/response level was defined as those regions with larger PRIC > C/RIC > PRIC differences for one modality than the other (Schumacher et al., 2011). The behavioral data suggested that visual distractors caused larger preresponse conflicts (PRIC > C) to auditory target processing than auditory distractors did to visual target processing (i.e., visual dominance at the preresponse level; Fig. 2*A*); therefore, the following contrast was used to identify regions specific to the visual dominance at the preresponse level: the interaction contrast Attend_Auditory (PRIC > C) > Attend_Visual (PRIC > C) was inclusively masked by the mask contrast Attend_Auditory (PRIC > C) at the threshold of $p < 0.05$, uncorrected for multiple comparisons at the voxel level. In this way, only those voxels specifically involved in the preresponse conflict caused by visual distractors were included in the statistical analysis for neural interaction.

Activations in OPFC and posterior cingulate cortex (PCC) were identified in this analysis (Fig. 4*A*, Table 2*a*). Mean parameter estimates extracted from the activated clusters are shown in Figure 4*A* as a function of experimental condition. The two areas showed deactivations in all six experimental conditions relative to null trials, representing the typical response pattern of the "default mode network" in the human brain (Gusnard and Raichle, 2001; Raichle et al., 2001). Neural interaction in these two areas was due to less deactivation in the PRIC condition relative to the C condition when the auditory modality was attended (OPFC: $t_{(17)} = 2.69$, $p < 0.05$; PCC: $t_{(17)} = 2.31$, $p < 0.05$) and was due to less deactivation in the C condition relative to the PRIC condition when the visual modality was attended (OPFC: $t_{(17)} = 2.60$, $p < 0.05$; PCC: $t_{(17)} = 2.28$, $p < 0.05$).

Auditory dominance at the response level. Because the behavioral data suggested that auditory distractors caused significantly larger response conflicts (RIC > PRIC) to visual target processing than visual distractors did to auditory targets processing (i.e., auditory dominance at the response level), the following contrast was used to identify the regions specific to the auditory dominance at the response level: the interaction contrast Attend_Visual (RIC > PRIC) > Attend_Auditory (RIC > PRIC) was inclusively

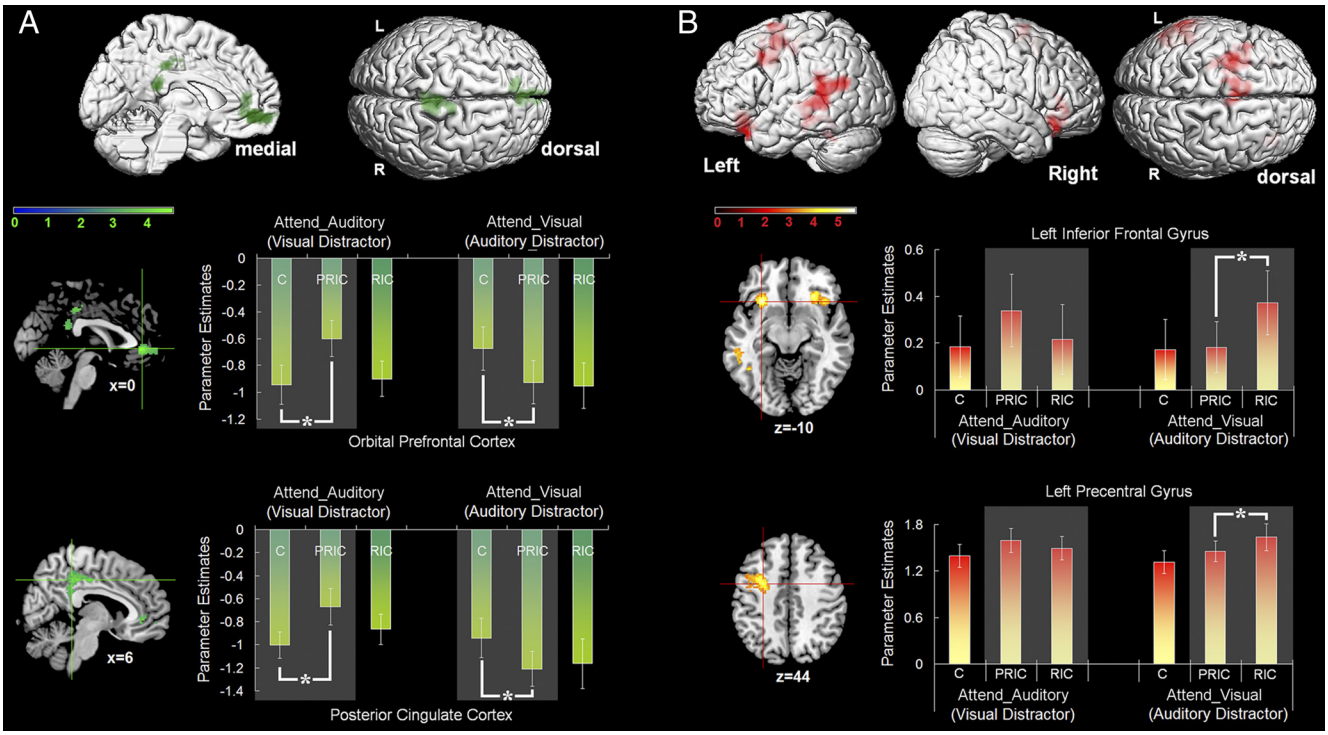

Figure 4. Neural correlates underlying the visual and auditory dominance at different levels in Experiment 1. ***A***, Visual dominance at the preresponse level. OPFC and PCC were significantly activated by the neural interaction contrast Attend_Auditory (PRIC > C) > Attend_Visual (PRIC > C), inclusively masked by Attend_Auditory (PRIC > C). Mean parameter estimates in the two activated clusters are shown as a function of the six experimental conditions. ***B***, Auditory dominance at the response level. PCG, bilateral IFG, left superior temporal gyrus, and left inferior occipital cortex were significantly activated by the interaction contrast Attend_Visual (RIC > PRIC) > Attend_Auditory (RIC > PRIC), inclusively masked by Attend_Visual (RIC > PRIC). Mean parameter estimates in left IFG and PCG are shown as a function of the six experimental conditions. The pattern of neural activity in other regions was similar to that in the above two regions. The four conditions shaded were the conditions involved in the interaction contrasts. Error bars represent SEs. Conditions denoted by an asterisk indicate a significant difference between them ($p < 0.05$).

Table 2. Brain regions activated by the neural interactions between attended modality and cross-modal conflict at the preresponse and the response levels, respectively

Anatomical region	Hemisphere	Cluster peak (mm)	*t* score	k_E (voxels)
Experiment 1				
a. Visual dominance at the preresponse level: Attend_Auditory (PRIC > C) > Attend_Visual (PRIC > C) masked by Attend_Auditory (PRIC > C)				
OPFC	M	0, 42, −2	4.83	403
PCC	R	6, −38, 42	4.23	404
b. Auditory dominance at the response level: Attend_Visual (RIC > PRIC) > Attend_Auditory (RIC > PRIC) masked by Attend_Visual (RIC > PRIC)				
PCG	L	−28, 0, 44	5.40	1390
SMA	*L*	*−10, 8, 52*	*4.75*	
IFG	L	−30, 22, −10	5.71	545
	R	26, 28, −10	5.25	519
Posterior superior temporal sulcus	L	−50, −44, 24	5.63	1124
Superior temporal gyrus	*L*	*−62, −46, 14*	*5.32*	
Middle temporal gyrus	*L*	*−66, −38, 6*	*4.77*	
Inferior occipital cortex	L	−44, −52, −10	4.29	418
Experiment 2				
c. Visual dominance at the preresponse level: Attend_Auditory (PRIC > C) > Attend_Visual (PRIC > C) masked by Attend_Auditory (PRIC > C)				
OPFC	L	−12, 50, −4	4.13	1261
d. Auditory dominance at the response level: Attend_Visual (RIC > PRIC) > Attend_Auditory (RIC > PRIC) masked by Attend_Visual (RIC > PRIC)				
IFG	L	−32, 20, −16	3.58	48
	R	32, 30, −10	3.08	15
Superior frontal cortex	L	−14, 12, 52	3.19	107

The coordinates (*x*, *y*, *z*) correspond to MNI coordinates. Displayed are the coordinates of the maximally activated voxel within a significant cluster as well as the coordinates of relevant local maxima within the cluster (in italics).

masked by the mask contrast Attend_Visual (RIC > PRIC) at the threshold of $p < 0.05$ uncorrected for multiple comparisons at the voxel level. In this way, only those voxels specifically involved in the response conflict caused by auditory distractors would be included in the statistical analysis for neural interaction.

The bilateral IFG, PCG extending to left SMA, left posterior superior temporal sulcus, and left inferior occipital cortex were

significantly activated (Fig. 4*B*, Table 2*b*). Mean parameter estimates extracted from the activated clusters in the left IFG and left PCG are shown in Figure 4*B* as a function of experimental condition, although the other areas showed similar patterns of neural activity. In all of these areas, neural interaction was due to significantly higher activity in the RIC than in the PRIC condition when the visual modality was attended (all $p < 0.05$), whereas there were no significant differences between the RIC and PRIC conditions when the auditory modality was attended (Fig. 4*B*).

PPI analysis with OPFC as the source region. PPI analysis with OPFC as the source region and with the contrast Attend_Auditory versus Attend_Visual as the psychological factor revealed that OPFC showed significantly higher neural coupling with bilateral occipital and temporal cortex extending to bilateral fusiform gyrus and bilateral hippocampus (Fig. 5*A*, Table 3*a*) in the Attend_Auditory blocks compared with the Attend_Visual blocks. There was no significant modulation of neural coupling in the reverse contrast (i.e., Attend_Visual > Attend_Auditory).

Experiment 2

Behavioral data

Analysis of RTs revealed a pattern almost identical to that in Experiment 1. The main effect of attended modality was significant ($F_{(1,16)} = 6.25, p < 0.05$), suggesting that responses to auditory targets (839 ms) were significantly slower than responses to visual targets (795 ms) (Fig. 2*B*, top, Table 1*b*). The main effect of congruency was also significant ($F_{(2,32)} = 33.11$, $p < 0.001$). RTs were increasingly slower over the C (746 ms), the PRIC (833 ms), and the RIC (871 ms) conditions, and the differences between conditions were all significant ($p < 0.005$, Bonferroni corrected). Moreover, the interaction between attended modality and congruency was significant ($F_{(2,32)} = 5.59, p < 0.01$). Further planned *t* tests on simple effects suggested that the preresponse level conflict was significant both when the auditory modality was attended (107 ms; $t_{(16)} = 5.22, p < 0.001$) and when the visual modality was attended (66 ms; $t_{(16)} = 4.21$, $p < 0.001$) (Fig. 2*B*, top). However, the response level conflict was significant only when the visual modality was attended (74 ms; $t_{(16)} = 4.83, p < 0.005$), not when the auditory modality was attended (3 ms; $t_{(16)} < 1$). The size of the preresponse level

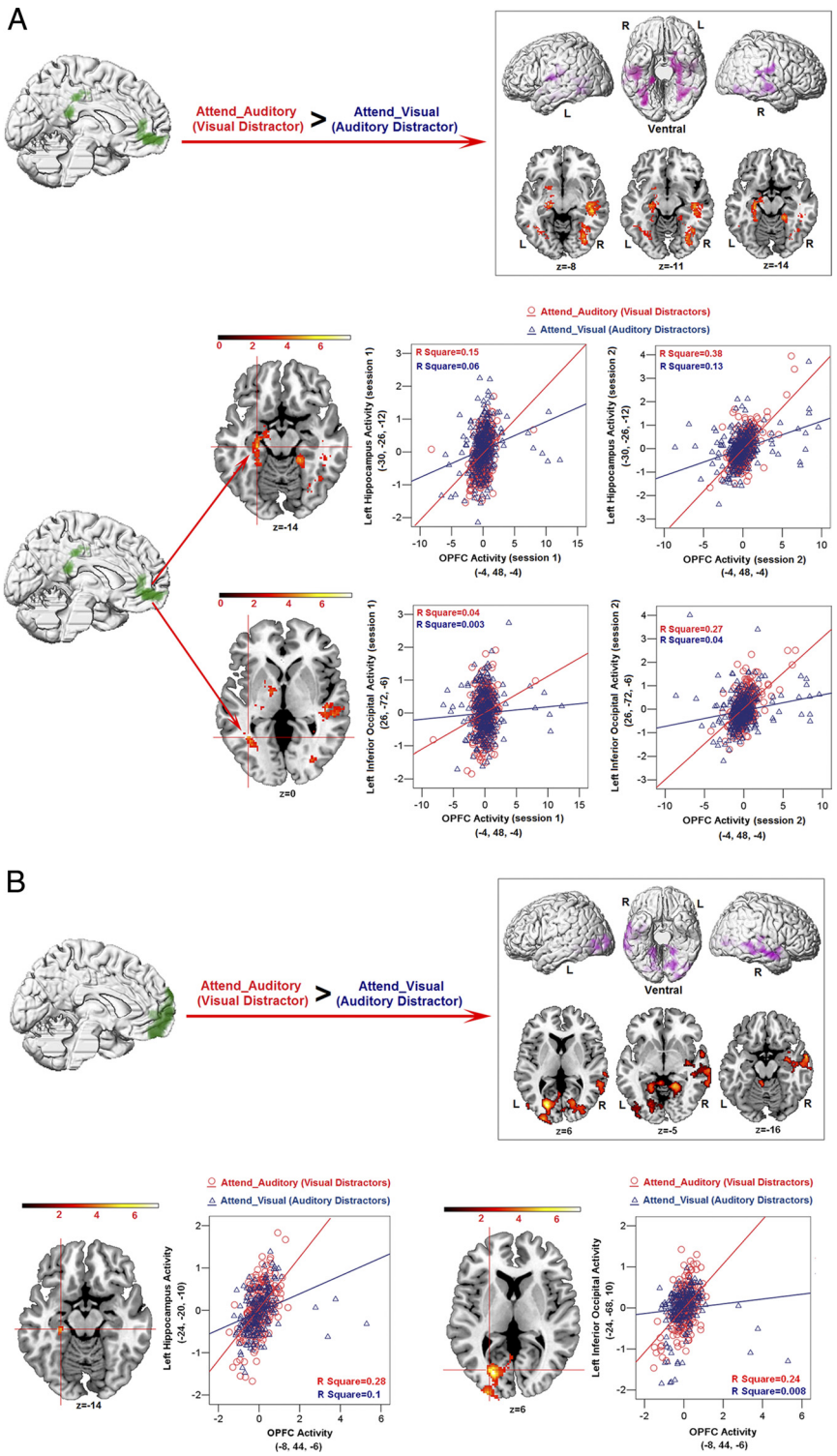

Figure 5. PPI analysis based on neural activity in OPFC with the contrast Attend_Auditory > Attend_Visual as the psychological factor. The source region in OPFC is marked in green. ***A***, Experiment 1. Top: Bilateral temporal and occipital cortex extending to bilateral fusiform and bilateral hippocampus showed significant context-dependent covariations with the neural activity in OPFC. The coupling was stronger in the Attend_Auditory condition than in the Attend_Visual condition. To give a clear view of ventral cortical structures, the cerebellum is removed in the display. Bottom: PPI analysis based on the neural activity in OPFC (green) for a representative participant. Mean corrected neural activity in left hippocampus and left inferior occipital cortex is displayed as a function of mean corrected activity in OPFC (i.e., the first principal component from a sphere of 4 mm radius) in the Attend_Auditory blocks (red dots and lines) and the Attend_Visual blocks (blue triangles and lines) for the two sessions, respectively. ***B***, Experiment 2. Top: Bilateral occipital cortex and right middle temporal gyrus extending to right hippocampus showed stronger functional connectivity with OPFC in the Attend_Auditory than in the Attend_Visual blocks. Bottom: For a representative participant, mean corrected neural activity in left hippocampus and left inferior occipital cortex (from the SVC analysis based on the activations in the PPI analysis in Experiment 1) is displayed as a function of mean corrected neural activity in OPFC in the Attend_Auditory blocks compared with the Attend_Visual blocks.

Table 3. Brain regions that showed higher functional connectivity with OPFC in the Attend_Auditory blocks (with visual distractors) than in the Attend_Visual blocks (with auditory distractors)

Anatomical region	Hemisphere	Cluster peak (mm)	*t* score	k_E (voxels)
a. Experiment 1				
Superior temporal gyrus	L	−44, −26, 22	7.76	1054
Hippocampus	*L*	*−26, −24, −14*	*5.22*	
Fusiform gyrus	*L*	*−26, −42, −12*	*4.02*	
Middle occipital gyrus	L	−28, −52, 10	5.66	608
Middle temporal gyrus	R	46, −30, −6	6.12	871
Hippocampus	*R*	*36, −22, −8*	*4.05*	
Inferior occipital cortex	R	32, −76, −4	5.94	469
Fusiform gyrus	*R*	*20, −38, −14*	*5.10*	
b. Experiment 2				
Middle occipital gyrus	L	−22, −78, 6	7.23	1367
Lingual gyrus	R	16, −46, −4	5.55	690
Middle temporal gyrus	R	62, 0, −16	5.47	1301
Hippocampus	*R*	*42, −20, −14*	*4.95*	

The coordinates (*x*, *y*, *z*) correspond to MNI coordinates. Displayed are the coordinates of the maximally activated voxel within a significant cluster as well as the coordinates of relevant local maxima within the cluster (in italics).

conflict was significantly larger when auditory modality was attended (107 ms) than when visual modality was attended (66 ms; $t_{(16)} = 2.66$, $p < 0.05$), whereas the size of the response level conflict was significantly larger when visual modality was attended (74 ms) than when auditory modality was attended (3 ms; $t_{(16)} = 3.0$, $p < 0.01$; Fig. 2*B*, bottom).

For error rates (Table 1*b*), the only significant effect was the main effect of congruency ($F_{(2,32)} = 15.9, p < 0.001$). Participants made increasingly more errors over the C (3.2%), the PRIC (5.8%), and the RIC (10.7%) conditions, and the differences between conditions were all significant (all $p < 0.05$, Bonferroni corrected), indicating significant cross-modal conflicts at both the preresponse (PRIC > C) and the response (RIC > PRIC) level.

Imaging data

Compared with the C condition, the PRIC condition significantly activated bilateral IFG, bilateral superior temporal gyrus, and precuneus (Fig. 3*B*, green). The bilateral IFG activations were similar to those in Experiment 1 (Fig. 3*A*, green). However, the bilateral superior temporal gyrus and precuneus activations in Experiment 2 and the SMA activations in Experiment 1 were different (Fig. 3, green). Note that a key difference between Experiments 1 and 2 was in the PRIC condition. In Experiment 1, the target and the distractor in the PRIC condition corresponded to the same response key (Fig. 1*A*). In Experiment 2, however, the distractor in the PRIC condition was not assigned to a response key (Fig. 1*B*). Therefore, the contrast PRIC > C resulted in two C response keys versus two C response keys in Experiment 1, while it resulted in one response key versus two C response keys in Experiment 2. This difference might result in the different patterns of activation for the PRIC > C contrast in the two experiments. Conversely, because the RIC conditions were essentially the same for Experiments 1 and 2, the RIC condition in Experiment 2, compared with the C condition, significantly activated bilateral parietal cortex and left superior frontal gyrus extending to SMA and left IFG (Fig. 3*B*, red), a pattern similar to the one for the main effect contrast RIC > C (collapsed over attended modalities) in Experiment 1 (Fig. 3*A*, red).

No significant activations were found in the contrast RIC > PRIC. Similar to Experiment 1, because the response level conflict (RIC > PRIC) was specific to the Attend_Visual condition, we expected that neural mechanisms underlying the response level cross-modal conflict would be localized by the neural interaction contrast rather than the main effect contrasts.

Visual dominance at the preresponse level. OPFC was significantly involved in the interaction contrast at the preresponse level Attend_Auditory (PRIC > C) > Attend_Visual (PRIC > C), which was inclusively masked by the contrast Attend_Auditory (PRIC > C) at the threshold of $p < 0.05$, uncorrected for multiple comparisons at the voxel level (Fig. 6*A*, Table 2*c*). This interaction (Fig. 6*A*) was due to less deactivation in the PRIC condition relative to the C condition when auditory modality was attended ($t_{(16)} = 2.55$, $p < 0.05$) and was due to less deactivation in the C condition relative to the PRIC condition when visual modality was attended ($t_{(16)} = 2.97$, $p < 0.05$).

Auditory dominance at the response level. For the interaction contrast Attend_Visual (RIC > PRIC) > Attend_Auditory (RIC > PRIC) inclusively masked by Attend_Visual (RIC > PRIC) at the threshold of $p < 0.05$, uncorrected for multiple comparisons at the voxel level, we did not get significant activations at the threshold of $p < 0.005$, FWE correction for multiple comparisons at the cluster level with an underlying voxel level of $p < 0.005$ (uncorrected). However, because Experiments 1 and 2 were two independent studies and because of our clear a priori hypothesis that prefrontal executive areas similar to those in Experiment 1 should be activated by the auditory dominance at the response level, small volume correction (SVC) analysis was performed within the brain areas activated by the auditory dominance at the response level in Experiment 1 (Fig. 4*B*, Table 1*b*). The searching areas were spheres centering at peak voxels in Table 2*b* with a radius of 8 mm (one smooth kernel). Bilateral IFG and left superior frontal gyrus were significantly activated ($p < 0.05$, FWE correction at the voxel level; Fig. 6*B*, Table 1*d*). In the three areas, neural activity in the RIC condition was significantly higher than neural activity in the PRIC condition when the visual modality was attended (all $p < 0.05$), but not when the auditory modality was attended (all $t < 1$; Fig. 6*B*).

PPI analysis with OPFC as the source region. Compared with the Attend_Auditory blocks, OPFC in the Attend_Visual blocks showed significantly higher neural coupling with bilateral inferior occipital cortex and left middle temporal cortex extending to left hippocampus (Fig. 5*B*, top, Table 3*b*). There was no significant modulation of neural coupling for the reverse contrast (i.e., Attend_Visual > Attend_Auditory). To further show the consistency between Experiments 1 and 2, SVC analysis was performed within the brain areas showing higher neural coupling with OPFC in the Attend_Auditory blocks than in the Attend_Visual blocks of Experiment 1 (Fig. 5*A*, Table 3*a*). The searching areas were spheres centering at peak voxels in Table 3*a* with a radius of 8 mm (1 smooth kernel). Left hippocampus (MNI: $x = -26$, $y = -26$, $z = -14$, $t = 4.45$, 64 voxels), left middle occipital gyrus (MNI: $x = -32$, $y = -46$, $z = 6$, $t = 5.21$, 36 voxels), right hippocampus (MNI: $x = 42$, $y = -20$, $z = -14$, $t = 4.95$, 112 voxels), and right lingual gyrus (MNI: $x = 18$, $y = -44$, $z = -4$, $t = 4.9$, 18 voxels) survived the threshold of $p < 0.05$ with FWE corrected at the voxel level (Fig. 5*B*, bottom).

Experiment 3

Analysis of RTs revealed a pattern similar to those of Experiments 1 and 2, although the main effect of attended modality did not reach significance ($F_{(1,14)} = 1.76$, $p = 0.21$), indicating that visual and auditory processing of the targets were equally fast. The main effect of congruency was significant ($F_{(2,28)} = 12.66$, $p < 0.01$), as was the interaction between congruency and attended modality ($F_{(2,28)} = 3.78$, $p < 0.05$) (Fig. 2*C*, top, Table 1*c*). Planned *t* tests

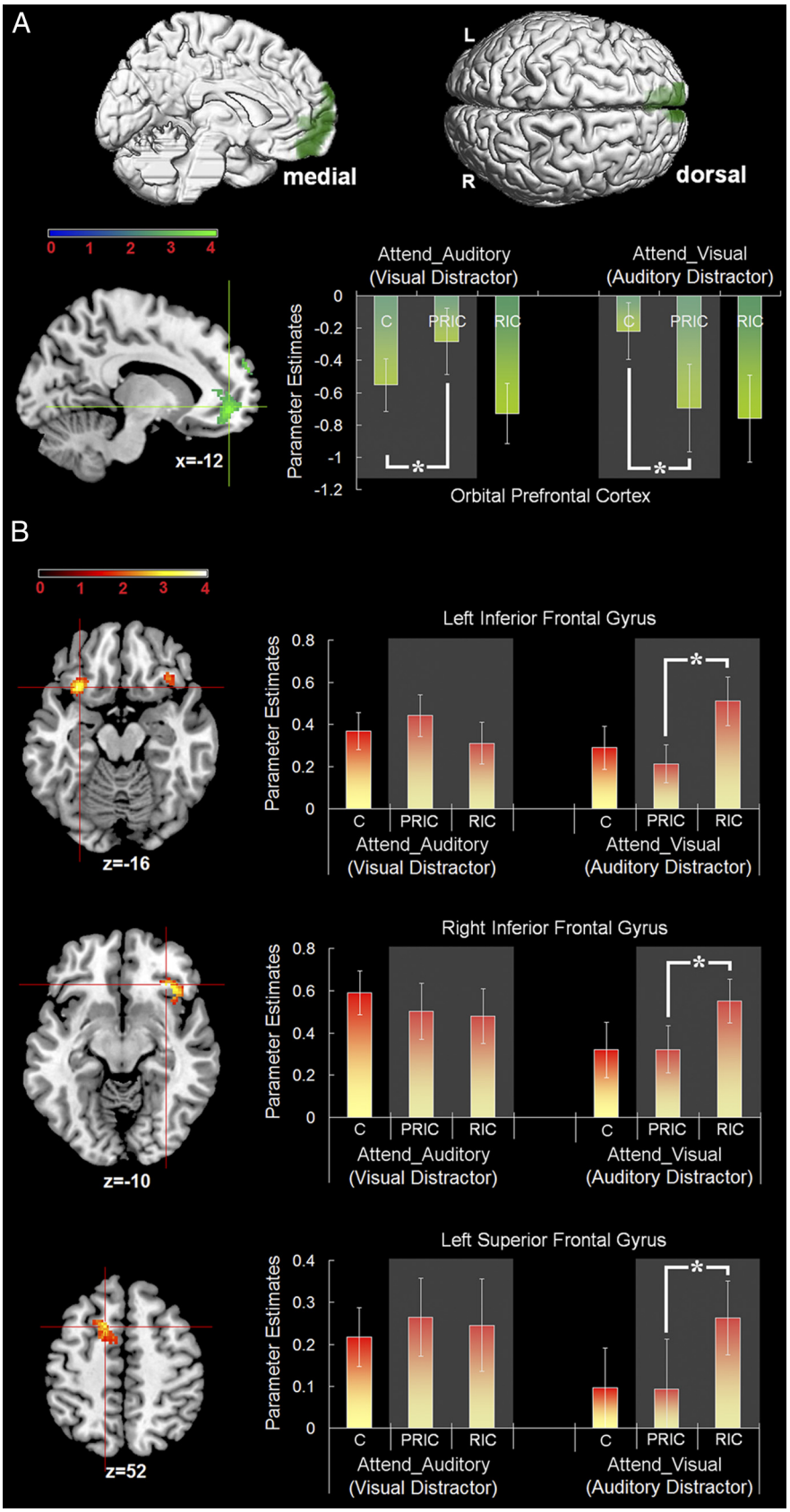

Figure 6. Neural correlates underlying the visual and auditory dominance at different levels in Experiment 2. ***A***, Visual dominance at the preresponse level. OPFC was significantly activated in the interaction contrast Attend_Auditory (PRIC > C) > Attend_Visual (PRIC > C), inclusively masked by Attend_Auditory (PRIC > C). Mean parameter estimates in the activated cluster are shown as a function of the six experimental conditions. ***B***, Auditory dominance at the response level. Bilateral IFG and left superior frontal gyrus were significantly activated in the contrast Attend_Visual (RIC > PRIC) > Attend_Auditory (RIC > PRIC), inclusively masked by Attend_Visual (RIC > PRIC). Mean parameter estimates in the three areas are shown as a function of the six experimental conditions. The four conditions shaded are the conditions involved in the interaction contrasts. Error bars represent SEs. Conditions denoted by an asterisk indicate a significant difference between them ($p < 0.05$).

on simple effects indicated that the preresponse level conflict was significant only when the auditory modality was attended (49 ms; $t_{(14)} = 3.89$, $p < 0.01$), not when the visual modality was attended (9 ms; $t_{(14)} = 1.38$, $p = 0.19$); the response level conflict was significant only when the visual modality was attended (24 ms; $t_{(14)} = 3.44$, $p < 0.01$), not when the auditory modality was attended (−2 ms; $t_{(14)} < 1$). Moreover, the size of preresponse level conflict was significantly larger when auditory modality was attended (49 ms) than when visual modality was attended (9 ms; $t_{(14)} = 2.76$, $p < 0.05$), whereas the size of response level conflict was significantly larger when visual modality was attended (24 ms) than when auditory modality was attended (−2 ms; $t_{(14)} = 2.53$, $p < 0.05$; Fig. 2*C*, bottom).

Analysis of error rates revealed a significant main effect of attended modality ($F_{(1,14)} = 7.21$, $p < 0.05$), with more errors when the visual modality was attended (4.6%) than when the auditory modality was attended (2.7%). The main effect of congruency was significant ($F_{(2,28)} = 7.9$, $p < 0.005$), as was the interaction between congruency and attended modality ($F_{(2,28)} = 4.26$, $p < 0.05$). Planned t tests on simple effects suggested that the size of the preresponse conflict (PRIC > C) was significantly larger when the auditory modality was attended (1.7%) than when the visual modality was attended (−0.4%; $t_{(14)} = 2.49$, $p < 0.05$), and the size of the response conflict (RIC > PRIC) was significantly larger when the visual modality was attended (4.3%) than when the auditory modality was attended (0.6%; $t_{(14)} = 2.58$, $p < 0.05$).

Discussion

In the present study, by disentangling the preresponse and response conflicts between simultaneous visual and auditory inputs, we investigated whether the direction of sensory dominance in cross-modal interaction depends on the level of cognitive processing. Behaviorally, visual distractors caused significantly larger preresponse interference with auditory processing than vice versa, demonstrating visual dominance at the preresponse level.

In contrast, auditory distractors caused significantly larger response interference with visual processing than vice versa, indicating auditory dominance at the response level. This pattern of results was obtained regardless of whether the behavioral task was category-based or identity-based, whether the experimental stimuli were celebrities or colors, and whether visual processing was faster than auditory processing or visual and auditory processing were equally fast (Fig. 2). Neurally, the visual dominance at the preresponse level and the auditory dominance at the response level were associated with neural activity in the default mode network and the prefrontal network, respectively (Fig. 4, Fig. 6). Moreover, the pattern of behavioral results was consistent with the pattern of neural activity in the two neural networks. First, the larger preresponse conflicts (PRIC > C) caused by visual distractors than by auditory distractors (i.e., the visual dominance at the preresponse level; Fig. 2) corresponded to the higher neural activity (less deactivation) in the default mode network in the PRIC condition than in the C condition when the auditory modality was attended (i.e., visual distractors) (Fig. 4*A*, Fig. 6*A*, left, shading). Second, the larger response conflicts (RIC > PRIC) caused by auditory distractors than by visual distractors (i.e., the auditory dominance at the response level; Fig. 2) corresponded to the higher neural activity in the prefrontal executive regions in the RIC condition than in the PRIC condition when the visual modality was attended (i.e., auditory distractors) (Fig. 4*B*, Fig. 6*B*, right, shading).

The neural interaction at the preresponse level was underpinned jointly by the reduced deactivation in the default mode network for the comparison PRIC > C when the auditory modality was attended (Fig. 4*A*, Fig. 6*A*, left, shading) and by the reduced deactivation for the comparison C > PRIC when the visual modality was attended (Fig. 4*A*, Fig. 6*A*, right, shading). The level of deactivation in the default mode network has been known to be negatively related to task difficulty: the easier the task, the shorter the RTs and the less the deactivation (McKiernan et al., 2003). When the visual modality was attended, deactivation in the default mode network was significantly diminished in the C condition (shorter RTs) compared with the PRIC and RIC conditions (longer RTs) (Fig. 4*A*, Fig. 6*A*, right, shading).

Therefore, the visual dominance at the preresponse level, the larger preresponse conflict (PRIC > C) caused by visual distractors (Fig. 2), was indeed associated with the reduced deactivation of the default mode network in the PRIC condition when compared with the C condition (Fig. 4A, Fig. 6A, left, shading). The default mode network has been associated with relative decreases of neural activity during the performance of various goal-directed tasks (Gusnard and Raichle, 2001; Raichle et al., 2001; Fox et al., 2005). In a global/local task, a typical *perceptual* interference task, a momentary reduction of the efficiency of selective attention (i.e., increase of RT when correctly identifying target stimulus) is characterized by less deactivation of the default mode network (Weissman et al., 2006). In this study, if the default mode network was less deactivated in the PRIC conditions, irrelevant visual information was more likely to attract attention and cause preresponse interference with auditory processing. PPI analysis in Experiments 1 and 2 consistently revealed that OPFC in the default mode network showed a more enhanced neural coupling with the ventral visual perceptual and semantic representation regions in response to visual distractors than to auditory distractors (Fig. 5). These ventral visual areas in the fusiform gyrus and hippocampus have been known to be specifically involved in processing human faces (Kanwisher et al., 1997), especially famous faces (Bernard et al., 2004; Elfgren et al., 2006) such as those used in the present study. Anatomically, primate studies indicate that OPFC has cortical connections with the ventral visual stream (Barbas, 1993; Cavada et al., 2000; Price, 2007). Functionally, OPFC is known to be involved in processing visual stimuli (Petrides et al., 2002; Rolls et al., 2005). For example, successful visual object recognition activates not only the ventral visual stream, but also OPFC (Bar et al., 2001). Magnetoencephalography studies further suggest that the recognition-related activity in OPFC precedes the corresponding activity in the ventral visual areas, indicating a top-down active selection function of OPFC (Bar et al., 2006). Moreover, a time-frequency, trial-by-trial covariance analysis suggests that the synchrony between OPFC and the ventral visual stream is significantly stronger for coarse than for fine images. Because details in unattended visual images are less processed than those in attended visual images, the coarse versus fine distinction corresponds very well to the unattended versus attended visual stimuli in the present study. The default mode network seems to be specifically designed to select visual stimuli in a multisensory environment and to pass them to visual awareness via enhanced neural coupling with the ventral visual stream, especially when the visual stimuli are coarse or unattended.

The auditory dominance at the response level, as evidenced by the stronger response conflict caused by auditory distractors during the attendance of the visual modality (Fig. 2), was associated with enhanced neural activity in prefrontal executive regions (Fig. 4B, Fig. 6B, right, shading). Auditory areas in the posterior superior temporal gyrus are known to be anatomically connected with the premotor cortex (for a review see Zatorre et al., 2007). Functionally, passive listening to purely perceptual rhythmic sounds can engage the premotor cortex, indicating an inherent link between the auditory and the motor systems: the motor system may be sensitive to the basic physical properties of auditory stimuli (Chen et al., 2008). In the present multisensory environment, the response codes of the unattended auditory stimuli might be accessed directly by the motor system, causing response level conflicts. Goal-directed behavior depends on the ability to suppress inappropriate actions and to resolve interference between competing responses. Previous neuropsychological and neuroimaging studies have suggested that IFG, premotor cortex, and (pre) SMA are involved in inhibiting inappropriate responses in go/no-go and stop-signal tasks (Aron et al., 2003; Chambers et al., 2007; Leung and Cai, 2007; Swick et al., 2008; Chikazoe et al., 2009). IFG, via anatomical connections, could signal the motor system to override a tendency of making a response inconsistent with the behavioral goal (MacDonald et al., 2000; Miller and Cohen, 2001; Kerns et al., 2004; Nee et al., 2007; Mostofsky and Simmonds, 2008).

Previous research suggests that sensory dominance in cross-modal interaction can be determined by task demands and/or by the reliability of information conveyed by different modalities along specific dimensions (Ernst and Banks, 2002; Heron et al., 2004; Witten and Knudsen, 2005; Yuval-Greenberg and Deouell, 2009). For example, the modality precision hypothesis argues that the modality that exhibits more accurate discrimination for the kind of information required by the task is favored in cross-modal interaction. Similarly, the modality appropriateness hypothesis proposes that because vision is specifically designed to process spatial features of stimuli (e.g., location, orientation, and shape) and audition is designed to process temporal information (e.g., sensorimotor synchronization tasks), the former is given precedence in spatial tasks, whereas the latter is given precedence in temporal tasks (Welch and Warren, 1980). These hypotheses,

however, cannot account for the findings in the present study. Neither the semantic categorization task adopted in Experiment 1 nor the identity discrimination task used in Experiments 2 and 3 was a sensorimotor or a temporal task. Nevertheless, we demonstrated not only visual but also auditory dominance in the same task. Specifically, with task demands held constant and with the reliability of information fixed in each modality, we showed that the sensory dominance varies as a function of the level of cognitive processing. Therefore, our results suggest that, at least under certain circumstances, what determines the direction of sensory dominance is the level of processing rather than simply task demands.

To summarize, by differentiating cross-modal conflict into the preresponse and response levels, we found that visual distractors caused more interference to auditory processing than vice versa at the preresponse level, whereas auditory distractors caused more interference to visual processing than vice versa at the response level, indicating a visual dominance at the preresponse level and an auditory dominance at the response level. Neurally, the default mode network was involved in the visual dominance at the preresponse level and the prefrontal executive network was involved in the auditory dominance at the response level, indicating that the default mode network inclines to select irrelevant visual rather than auditory information at earlier stages of cognitive processing, while the prefrontal executive network resolves response level conflicts caused by irrelevant auditory information at later stages of processing.

References

Aron AR, Fletcher PC, Bullmore ET, Sahakian BJ, Robbins TW (2003) Stop-signal inhibition disrupted by damage to right inferior frontal gyrus in humans. Nat Neurosci 6:115–116. CrossRef Medline

Bar M, Tootell RB, Schacter DL, Greve DN, Fischl B, Mendola JD, Rosen BR, Dale AM (2001) Cortical mechanisms specific to explicit visual object recognition. Neuron 29:529–535. CrossRef Medline

Bar M, Kassam KS, Ghuman AS, Boshyan J, Schmid AM, Dale AM, Hämäläinen MS, Marinkovic K, Schacter DL, Rosen BR, Halgren E (2006) Top-down facilitation of visual recognition. Proc Natl Acad Sci U S A 103:449–454. CrossRef Medline

Barbas H (1993) Organization of cortical afferent input to orbitofrontal areas in the rhesus monkey. Neuroscience 56:841–864. CrossRef Medline

Bernard FA, Bullmore ET, Graham KS, Thompson SA, Hodges JR, Fletcher PC (2004) The hippocampal region is involved in successful recognition of both remote and recent famous faces. Neuroimage 22:1704–1714. CrossRef Medline

Bunge SA, Hazeltine E, Scanlon MD, Rosen AC, Gabrieli JD (2002) Dissociable contributions of prefrontal and parietal cortices to response selection. Neuroimage 17:1562–1571. CrossRef Medline

Cavada C, Compañy T, Tejedor J, Cruz-Rizzolo RJ, Reinoso-Suárez F (2000) The anatomical connections of the macaque monkey orbitofrontal cortex. A review. Cereb Cortex 10:220–242. CrossRef Medline

Chambers CD, Bellgrove MA, Gould IC, English T, Garavan H, McNaught E, Kamke M, Mattingley JB (2007) Dissociable mechanisms of cognitive control in prefrontal and premotor cortex. J Neurophysiol 98:3638–3647. CrossRef Medline

Chen JL, Penhune VB, Zatorre RJ (2008) Listening to musical rhythms recruits motor regions of the brain. Cereb Cortex 18:2844–2854. CrossRef Medline

Chikazoe J, Jimura K, Asari T, Yamashita K, Morimoto H, Hirose S, Miyashita Y, Konishi S (2009) Functional dissociation in right inferior frontal cortex during performance of go/no-go task. Cereb Cortex 19:146–152. CrossRef Medline

Colavita FB (1974) Human sensory dominance. Percept Psychophys 16:409–412. CrossRef

Elfgren C, van Westen D, Passant U, Larsson EM, Mannfolk P, Fransson P (2006) fMRI activity in the medial temporal lobe during famous face processing. Neuroimage 30:609–616. CrossRef Medline

Eriksen CW, Schultz DW (1979) Information processing in visual search: A continuous flow conception and experimental results. Percept Psychophys 25:249–263. CrossRef Medline

Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415:429–433. CrossRef Medline

Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van Essen DC, Raichle ME (2005) The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proc Natl Acad Sci U S A 102:9673–9678. CrossRef Medline

Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ (1997) Psychophysiological and modulatory interactions in neuroimaging. Neuroimage 6:218–229. CrossRef Medline

Gusnard DA, Raichle ME (2001) Searching for a baseline: functional imaging and the resting human brain. Nat Rev Neurosci 2:685–694. CrossRef Medline

Heron J, Whitaker D, McGraw PV (2004) Sensory uncertainty governs the extent of audio-visual interaction. Vision Res 44:2875–2884. CrossRef Medline

Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci 17:4302–4311. Medline

Kato M, Konishi Y (2006) Auditory dominance in the error correction process: a synchronized tapping study. Brain Res 1084:115–122. CrossRef Medline

Kerns JG, Cohen JD, MacDonald AW 3rd, Cho RY, Stenger VA, Carter CS (2004) Anterior cingulate conflict monitoring and adjustments in control. Science 303:1023–1026. CrossRef Medline

Kim C, Kroger JK, Kim J (2011) A functional dissociation of conflict processing within anterior cingulate cortex. Hum Brain Mapp 32:304–312. CrossRef Medline

Koppen C, Levitan CA, Spence C (2009) A signal detection study of the Colavita visual dominance effect. Exp Brain Res 196:353–360. CrossRef Medline

Leung HC, Cai W (2007) Common and differential ventrolateral prefrontal activity during inhibition of hand and eye movements. J Neurosci 27:9893–9900. CrossRef Medline

Macaluso E, Driver J (2005) Multisensory spatial interactions: a window onto functional integration in the human brain. Trends Neurosci 28:264–271. CrossRef Medline

MacDonald AW 3rd, Cohen JD, Stenger VA, Carter CS (2000) Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. Science 288:1835–1838. CrossRef Medline

Marks LE (2004) Cross-modal interactions in speeded classification. In: The handbook of multisensory processes (Calvert GA, Spence C, Stein BE, eds), pp 85–105. Cambridge: MIT.

Mayer AR, Franco AR, Canive J, Harrington DL (2009) The effects of stimulus modality and frequency of stimulus presentation on cross-modal distraction. Cereb Cortex 19:993–1007. CrossRef Medline

McKiernan KA, Kaufman JN, Kucera-Thompson J, Binder JR (2003) A parametric manipulation of factors affecting task-induced deactivation in functional neuroimaging. J Cogn Neurosci 15:394–408. CrossRef Medline

Milham MP, Banich MT, Webb A, Barad V, Cohen NJ, Wszalek T, Kramer AF (2001) The relative involvement of anterior cingulate and prefrontal cortex in attentional control depends on nature of conflict. Brain Res Cogn Brain Res 12:467–473. CrossRef Medline

Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. Annu Rev Neurosci 24:167–202. CrossRef Medline

Mostofsky SH, Simmonds DJ (2008) Response inhibition and response selection: two sides of the same coin. J Cogn Neurosci 20:751–761. CrossRef Medline

Nee DE, Jonides J, Berman MG (2007) Neural mechanisms of proactive interference-resolution. Neuroimage 38:740–751. CrossRef Medline

Orr JM, Weissman DH (2009) Anterior cingulate cortex makes 2 contributions to minimizing distraction. Cereb Cortex 19:703–711. CrossRef Medline

Petrides M, Alivisatos B, Frey S (2002) Differential activation of the human orbital, mid-ventrolateral, and mid-dorsolateral prefrontal cortex during the processing of visual stimuli. Proc Natl Acad Sci U S A 99:5649–5654. CrossRef Medline

Poline JB, Worsley KJ, Evans AC, Friston KJ (1997) Combining spatial ex-

tent and peak intensity to test for activations in functional imaging. Neuroimage 5:83–96. CrossRef Medline

Price JL (2007) Definition of the orbital cortex in relation to specific connections with limbic and visceral structures and other cortical regions. Ann N Y Acad Sci 1121:54–71. CrossRef Medline

Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL (2001) A default mode of brain function. Proc Natl Acad Sci U S A 98:676–682. CrossRef Medline

Repp BH, Penel A (2004) Rhythmic movement is attracted more strongly to auditory than to visual rhythms. Psychol Res 68:252–270. Medline

Rolls ET, Browning AS, Inoue K, Hernadi I (2005) Novel visual stimuli activate a population of neurons in the primate orbitofrontal cortex. Neurobiol Learn Mem 84:111–123. CrossRef Medline

Schumacher EH, Schwarb H, Lightman E, Hazeltine E (2011) Investigating the modality specificity of response selection using a temporal flanker task. Psychol Res 75:499–512. CrossRef Medline

Swick D, Ashley V, Turken AU (2008) Left inferior frontal gyrus is critical for response inhibition. BMC Neurosci 9:102. CrossRef Medline

van Veen V, Carter CS (2005) Separating semantic conflict and response conflict in the Stroop task: a functional MRI study. Neuroimage 27:497–504. CrossRef Medline

van Veen V, Cohen JD, Botvinick MM, Stenger VA, Carter CS (2001) Anterior cingulate cortex, conflict monitoring, and levels of processing. Neuroimage 14:1302–1308. CrossRef Medline

Weissman DH, Warner LM, Woldorff MG (2004) The neural mechanisms for minimizing cross-modal distraction. J Neurosci 24:10941–10949. CrossRef Medline

Weissman DH, Roberts KC, Visscher KM, Woldorff MG (2006) The neural bases of momentary lapses in attention. Nat Neurosci 9:971–978. CrossRef Medline

Welch RB, Warren DH (1980) Immediate perceptual response to intersensory discrepancy. Psychol Bull 88:638–667. CrossRef Medline

Witten IB, Knudsen EI (2005) Why seeing is believing: merging auditory and visual worlds. Neuron 48:489–496. CrossRef Medline

Yuval-Greenberg S, Deouell LY (2007) What you see is not (always) what you hear: induced gamma band responses reflect cross-modal interactions in familiar object recognition. J Neurosci 27:1090–1096. CrossRef Medline

Yuval-Greenberg S, Deouell LY (2009) The dog's meow: asymmetrical interaction in cross-modal object recognition. Exp Brain Res 193:603–614. CrossRef Medline

Zatorre RJ, Chen JL, Penhune VB (2007) When the brain plays music: auditory-motor interactions in music perception and production. Nat Rev Neurosci 8:547–558. CrossRef Medline