Behavioral/Cognitive

# Multisensory Competition Is Modulated by Sensory Pathway Interactions with Fronto-Sensorimotor and Default-Mode Network Regions

Sai Huang,[1]* You Li,[1]* Wei Zhang,[2] Bao Zhang,[3] Xingzhou Liu,[1,2] Lei Mo,[1] and Qi Chen[1]

[1]Center for Studies of Psychological Application and School of Psychology, South China Normal University, Guangzhou 510631, China, [2]Epilepsy Center, Guangdong 999 Brain Hospital, Guangzhou 510631, China, and [3]Department of Psychology and the Center for Mind and Brain, Guangzhou University, Guangzhou 510631, China

Multisensory information competes for preferential access to consciousness. It remains unknown what neural processes cause one particular modality to win multisensory competition and eventually dominate behavior. Thus, in a paradigm in which human participants sought to make simultaneous auditory and visual detection responses, we sought to identify prestimulus and poststimulus neural signals that were associated with auditory and visual dominance on each trial. Behaviorally, visual detection responses preceded auditory responses more frequently than vice versa. Even when visual responses were preceded by auditory responses, they recovered more quickly from previous responses, indicating the dominance of vision over audition. Neurally, visual precedence was associated with increased prestimulus activity in the prefrontal cortex and reduced prestimulus activity in the default-mode network, and increased poststimulus connectivity between the prefrontal cortex and the visual system. Moreover, the dorsal visual stream showed not only increased activity in post-perceptual phases but also enhanced connectivity with the sensorimotor cortex, indicating the functional role of the dorsal visual stream in prioritizing the flow of visual information into the motor system. In contrast, auditory precedence was associated with increased prestimulus activity in the auditory cortex and increased poststimulus neural coupling between the auditory and the sensorimotor cortex. Finally, whenever one modality lost multisensory competition, the corresponding sensory cortex showed enhanced connectivity with the default-mode network. Overall, the outcome of audiovisual competition depended on dynamic interactions between sensory systems and both the fronto-sensorimotor and the default-mode network. Together, these results revealed both the neural causes and the neural consequences of visual and auditory dominance during multisensory competition.

Key words: default mode network; multisensory competition; prefrontal cortex; sensorimotor representations; sensory dominance; visual and auditory systems

## Introduction

Although inundated concurrently by streams of information from multiple sensory modalities, our brain does not give equal weight to different modalities. Rather, visual information more frequently receives preferential processing and dominates the other sensory modalities. One intriguing example of the dominance of vision over audition is the Colavita effect, which refers to the phenomenon that participants often fail in responding to the auditory component of bimodal audiovisual targets (Colavita, 1974; Colavita et al., 1976; Colavita and Weisberg, 1979). The striking pattern of the dominance of vision over audition is almost as if the simultaneous presentation of the visual stimulus leads to the "extinction" of the participants' conscious awareness of the auditory stimulus (Egeth and Sager, 1977; Koppen and Spence, 2007a; Koppen et al., 2009; Hartcher-O'Brien and Alais, 2011). The neural causes that drive one particular sensory modality to receive preferential processing during multisensory competition and eventually dominate awareness and behavior are unknown.

Facing simultaneously presented bimodal audiovisual targets, although human participants are able to make explicit behavioral responses to both the visual and the auditory components, they cannot always respond to them strictly simultaneously. Either the visual response preceded the auditory response or vice versa. In the two situations, both visual and auditory information are consciously perceived, and the critical difference is the differential speed for sensory information to access the corresponding sensorimotor representations. When multisensory information reaches the human brain, neural representations in various sensory systems compete for preferential access to the motor system.

A gain in the neural activation of one object/event representation always occurs at a cost to the others (Desimone and Duncan, 1995; Duncan et al., 1997; Spence et al., 2012). In terms of multisensory competition, neural representations in the dominant sensory modality may suppress neural representations in the dominated modalities. By directly contrasting unimodal visual and auditory targets, we could localize the sensory systems representing the visual and auditory stimuli, respectively. We predicted that neural activity in the localized auditory cortex might be weakened when vision dominated audition, and neural activity in the localized visual cortex might be weakened when audition gained priority.

More critically, to investigate the neural causes of multisensory competition, we calculated trial-to-trial relationships between prestimulus neural activity and response time (RT) difference between the responses to the visual and auditory components of the bimodal audiovisual stimuli. Because enhanced prestimulus activity in the prefrontal cortex and decreased prestimulus activity in the default-mode network (DMN) predicted better task performance (Weissman et al., 2006), we hypothesized that variance of prestimulus activity in the prefrontal cortex and the DMN might predict the extent of sensory dominance as well. Furthermore, by directly comparing the two behavioral conditions in which visual responses preceded auditory responses or not and by testing variations of functional connectivity between the sensory systems, the prefrontal cortex, the sensorimotor cortex, and the DMN, we could not only investigate the neural consequences of sensory dominance but also clarify whether sensory dominance occurred at the early sensory representation stages or the post-perceptual response selection/execution stages via fMRI and event-related potential (ERP) techniques.

## Materials and Methods

### Participants
Three different groups of healthy adult participants volunteered to take part in the present study: 20 in the fMRI experiment (nine females, 21–23 years old), 24 (13 females, 18–26 years old) in the ERP experiment, and 17 (eight females, 19–28 years old) in the behavioral control experiment. The participants were all right-handed, with normal hearing and normal or corrected-to-normal visual acuity. None of them had a history of neurological or psychiatric disorders. All the participants gave their informed consent before the study in accordance with the Declaration of Helsinki. This study was approved by the Ethics Committee of the School of Psychology, South China Normal University.

### Stimuli and experimental design
The auditory target was a 4000 Hz pure tone with the length of 50 ms, and the visual target was a white sphere with a radius of 1.5° visual angle and a luminance of 1.9 cd/m². The default visual display was a white cross that measured 1° × 1° of visual angle on a gray background (red–green–blue value, 128, 128, and 128). To avoid overlap in the spectral content of the target sound and the background echoplanar imaging (EPI) noise in the fMRI experiment (Scarff et al., 2004; Langers et al., 2005), we choose the frequency of the target sound (4000 Hz) to be distinctive from the background EPI noise (~1500 Hz; Ravicz et al., 2000). The same set of visual and acoustic parameters was adopted for the behavioral control, the ERP, and the fMRI experiments. Because the different types of trials were jittered adequately and mixed randomly for each of the participants, the effect of the background scanner noise on the prestimulus, the target presentation, and the poststimulus phases should be counterbalanced and equivalent between critical behavioral conditions in the fMRI experiment. Moreover, according to the psychophysiological pilots, at the frequency band of our auditory stimuli (4000 HZ) and at loudness levels of >60 dB (measured via a sound level meter), with a stimulus duration <100 ms (50 ms in the present study), the auditory stimuli were explicitly

suprathreshold, and the human participants reported that the sound stimuli could be perceived clearly. In this case, the detection time to the unimodal auditory stimuli was comparable with the detection time to the unimodal visual stimuli. However, for the fMRI experiment, the exact loudness level of the auditory stimuli cannot be measured precisely in the presence of the EPI noise. Therefore, for each of the participants, before the formal scanning, we simultaneously switched on the behavioral paradigm and the EPI sequence, communicated with the participants about the loudness of the auditory targets, and accordingly modulated the loudness level until the participants reported that they could clearly and comfortably hear the auditory targets despite the background noise. Behavioral data in the fMRI experiment accordingly showed that the participants were able to detect the unimodal auditory and visual stimuli equally quickly. Because it has been revealed that the Colavita visual dominance effect occurred regardless of whether the visual and the auditory targets were presented from the same or different spatial locations (Koppen and Spence, 2007b), we presented the visual and the auditory targets from the same central location in the present study.

There were three types of trials: (1) unimodal auditory trials in which only the auditory target was presented for 50 ms (i.e., the Auditory_Single condition); (2) unimodal visual trials in which only the visual target was presented for 50 ms at the center of the screen (i.e., the Visual_Single condition); and (3) bimodal trials in which the auditory and the visual targets were presented simultaneously for 50 ms. The three types of trials were presented randomly. In the three experiments, participants were instructed to press one button on the response pad with the thumb of one hand if the auditory target appeared, press the other button with the thumb of the other hand if the visual target appeared, and press both buttons as simultaneously as possible if the auditory and the visual targets both appeared. The mapping between the two response buttons and the visual and auditory targets was counterbalanced across participants. Participants were pushed to press down the two buttons as simultaneously as possible in the bimodal trials via strict instructions before the formal experiments: (1) participants were informed explicitly of the existence of the bimodal trials and (2) were instructed to press the visual key and the auditory key as simultaneously as possible on the bimodal trials.

Most critically, the bimodal trials were *post hoc* categorized into the following six types of behavioral conditions based on participants' online performance (Fig. 1B): (1) the Visual_Auditory (VA) responses, in which participants first responded to the visual component and then to the auditory component; thus, the absolute difference between the RTs to the visual and the auditory components indicated how much the visual response preceded the auditory response in the VA trials; (2) the Auditory_Visual (AV) responses, in which participants first responded to the auditory component and then to the visual component; thus, the absolute difference between the RTs to the visual and the auditory components indicated how much the auditory response preceded the visual response in the AV trials. (3) the "Simultaneous" responses, in which participants responded simultaneously to the auditory and the visual components by pressing down the two response buttons at the same time; based on the uncertainty errors (2–5 ms) recorded by the stimulus delivery system (Presentation Software package; Neurobehavioral Systems), the bimodal trials, in which the absolute RT difference between the responses to the visual and the auditory components was <5 ms (|Visual_RT > Auditory_RT| < 5 ms), were categorized as the simultaneous trials as well; (4) the Visual_Only responses, in which participants responded only to the visual component but not to the auditory component; (5) the Auditory_Only responses, in which participants responded only to the auditory component but not to the visual component. and (6) the "Missed" trials, in which no responses were recorded.

### Statistical analysis of behavioral data
For the behavioral data in the three experiments, the outlier trials, in which the RTs exceeded 3 SDs larger/smaller than the mean RT in each condition, were excluded from additional analysis (1.1% of the overall data points were excluded as outliers in the fMRI experiment, 0.9% in the ERP experiment, and 0% in the behavioral control experiment). Based on the online responses in the bimodal trials, we differentiated between

the six behavioral conditions in the bimodal trials. The ratio of each behavioral condition was calculated as the proportion between the number of bimodal trials in each behavioral condition and the overall number of bimodal trials. Note that the smaller the proportion of the Auditory_Visual responses, the larger the visual dominance effect at the behavioral level. However, the extremely small proportions of the Auditory_Visual responses ($<10\%$) could not give us enough statistical power to calculate the underlying neural substrates in the ERP and the fMRI experiments. Thus, three participants in the fMRI experiment were excluded from additional analysis because of the extremely small proportions of the Auditory_Visual responses ($<10\%$), leaving 17 participants in total in the fMRI experiment. For the ERP experiment, in addition to the data in the two participants who were discarded because of low ratios of the Auditory_Visual responses ($<10\%$), another two participants were discarded because of excessive EEG artifacts, leaving 20 participants in total in the ERP experiment.

For RTs, we focused our analysis on RTs to the visual and the auditory components in the two critical behavioral conditions in the bimodal trials, i.e., the Visual_Auditory and the Auditory_Visual trials. Omissions, incorrect responses, and trials with RTs exceeding 3 SDs away from the mean RT for each condition were first excluded from additional analysis. Mean RTs of the remaining trials were then calculated for each condition and submitted to a 2 (type of response: responses to the visual components vs responses to the auditory components) × 2 (response order: the first response vs the second response) repeated-measures ANOVA.

In addition, participants in the three experiments were all right-handed, and the correspondence between the responding hand (left vs right hand) and the sensory stimuli (auditory vs visual target) was counterbalanced across participants. Half of the participants used the left hand to respond to the auditory stimuli and the right hand to the visual stimuli, i.e., the "LHA_RHV" group, and the reversed assignment for the other half of participants, i.e., the "LHV_RHA" group. To explore whether the assignment of responding hand to sensory stimuli altered the nature of audiovisual competition in the present paradigm, we collapsed the behavioral data from the three experiments and split the participants into two groups according to the mapping between the responding hand and sensory stimuli (26 participants in the LHA_RHV group and 28 participants in the LHV_RHA group). First, to test whether the assignment of responding hand affected the response speed to unimodal stimuli, the unimodal RTs were submitted to a 2 (the between-group factor: LHA_RHV vs LHV_RHA) × 2 (type of unimodal trials: Visual_Single vs Auditory_Single) repeated-measures ANOVA. Second, to test whether the assignment of responding hand affected the pattern of visual dominance in terms of the
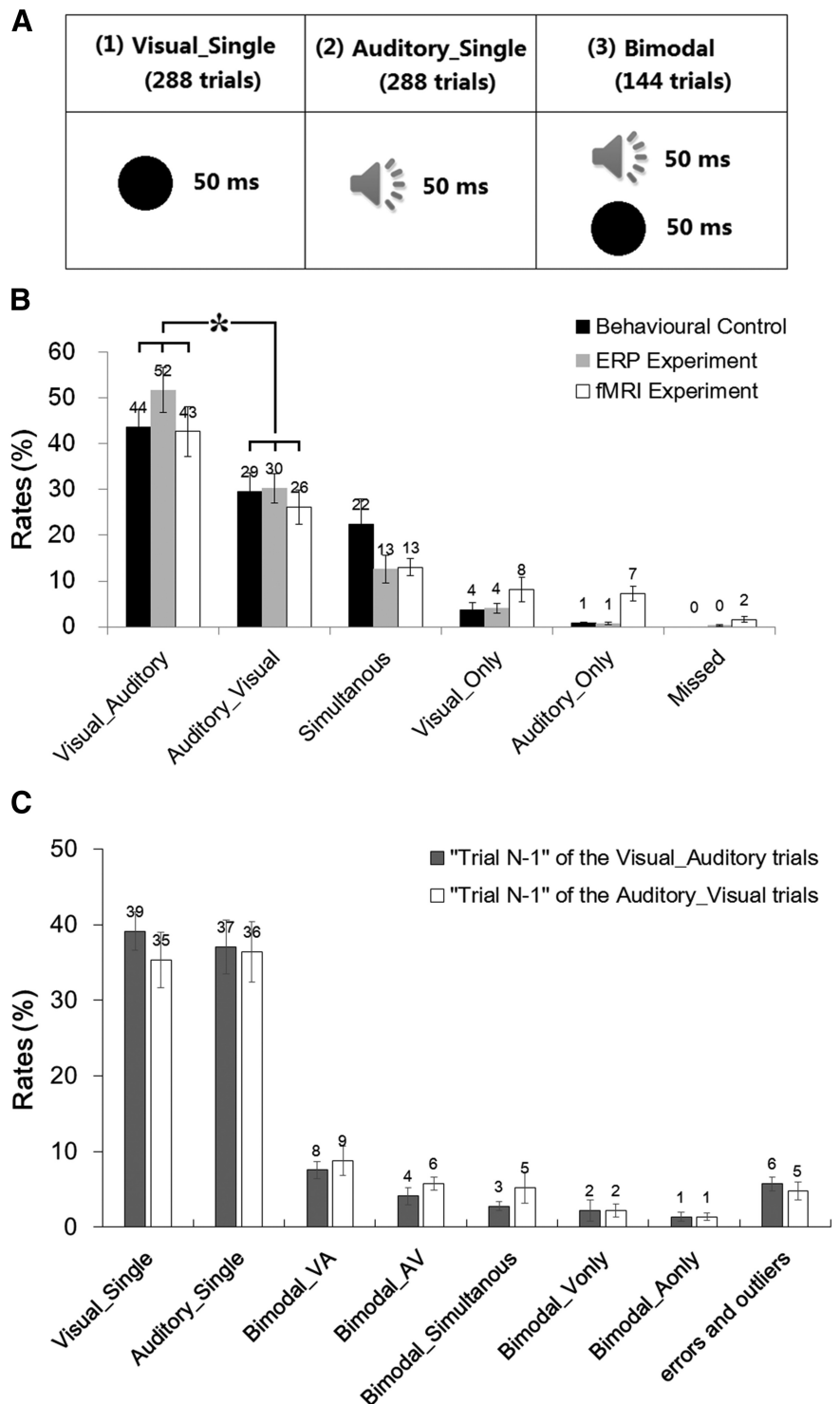


**Figure 1.** Stimuli and behavior data in the three experiments. **A**, Stimuli in the present study. Three types of stimuli were used: (1) unimodal visual target; (2) unimodal auditory target; and (3) bimodal audiovisual target. Participants were instructed to press one prespecified button when the visual target was presented, press another button when the auditory target was presented, and simultaneously press both buttons if the audiovisual target was presented. **B**, Proportions of the six different types of behavioral conditions in the bimodal trials of the three experiments. Bimodal trials were categorized into six behavioral conditions based on participants' online performance: (1) the Visual_Auditory trials in which participants first responded to the visual component and then to the auditory component; (2) the Auditory_Visual trials in which participants first responded to the auditory component and then to the visual component; (3) the Simultaneous trials in which participants responded simultaneously to the auditory and the visual components; (4) the Visual_Only trials in which participants responded only to the visual component; (5) the Auditory_Only trials in which participants responded only to the auditory component; and (6) the Missed trials in which no responses were made. **C**, Proportions of the different types of trials among the Trials N − 1 to the Visual_Auditory and the Auditory_Visual trials, respectively. *$p < 0.05$. The error bars indicate SEs.

proportion of trials, proportions of the six different types of bimodal behavioral conditions were also calculated for the two groups of participants, respectively, and were submitted to a 2 (the between-group factor: LHA_RHV vs LHV_RHA) × 2 (type of both responded but asynchronous bimodal trials: VA vs AV) repeated-measures ANOVA. Third, to test whether the assignment of responding hand altered the pattern of visual dominance in terms of the RTs in the bimodal trials, RTs in the bimodal trials were submitted to a 2 (the between-group factor: LHA_RHV vs LHV_RHA) × 2 (response type: visual vs auditory) × 2 (response order: first vs second) repeated-measures ANOVA.

### fMRI experiment

*Experimental procedures.* The auditory target was delivered binaurally to the participants via MR-compatible stereo headphones. The visual target was presented through an LCD projector onto a rear projection screen located behind the participants' head. Participants viewed the screen through an angled mirror on the head coil of the MRI setup. Participants were instructed to fixate at the central fixation cross throughout the experiment without moving their eyes and to detect the appearance of the target stimuli by pressing the prespecified buttons. The mapping between the two response buttons and the auditory and visual targets was counterbalanced across participants.

The fMRI experiment consisted of 900 trials in total, including 288 Auditory_Single trials, 288 Visual_Single trials, 144 bimodal trials, and 180 null trials. The 180 null trials, in which only the central fixation cross was presented without an onset/offset of the fixation cross, were used as the implicit baseline. The intertrial intervals were jittered from 2000 to 3000 ms (2000, 2250, 2500, 2750, and 3000 ms). Because the formal scanning lasted relatively long (~30 min), the participants were asked to rest for a short period of time [8.8 s, i.e., four repetition times (TRs)] after every 10 min task performance, which made two short periods of rest in total. During the two short rest periods, the scanner kept running and a visual instruction "rest" was presented on the screen throughout. One TR (2.2 s) after the disappearance of the "rest" visual instruction, the behavioral task reassumed. The temporal order of trials was randomized for each participant individually to avoid potential problems of unbalanced transition probabilities. Before the fMRI experiment, all participants were familiarized with the tasks and the experimental setup by a training session of 10 min.

*Data acquisition and preprocessing.* A Siemens 3T Trio system with a standard head coil was used to obtain T2-weighted echoplanar images with blood oxygenation level-dependent contrast. The matrix size was $64 \times 64$, and the voxel size was $3.4 \times 3.4 \times 3.5$ mm$^3$. Thirty-seven transversal slices of 3.5 mm thickness that covered the whole brain were acquired sequentially with a 0.4 mm gap (TR, 2.2 s; echo time, 30 ms; field of view, 220 mm; flip angle, 90°). There was one run of functional scanning (820 EPI volumes). 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, UK; http://www.fil.ion.ucl.ac.uk). Images were realigned to the first volume to correct for interscan head movements. Then the mean echo planar image for each participant was 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 applied subsequently to individual EPI volumes. All images were thus transformed into standard MNI space and resampled to $2 \times 2 \times 2$ 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 imaging data.* Data were high-pass-filtered at $\frac{1}{128}$ Hz and analyzed with a general linear model (GLM) as implemented in SPM8. Temporal autocorrelation was modeled using an AR(1) process. At the first level, the GLM was used to construct a multiple regression design matrix. Both the unimodal trials and the critical bimodal trials were modeled as regressors of interest: the Auditory_Single trials, the Visual_Single trials, the Auditory_Visual trials, the Visual_Auditory trials, the Auditory_Only trials, the Visual_Only trials, and the Simultaneous trials. In addition, the two short rest periods (four TRs/each) were modeled as confounds, and the behaviorally missed trials, error trials, and trials in which RTs were outside of the mean RT ± 3 SDs were included as another regressor of no interest. The null trials were not modeled and treated as the implicit baseline in the GLM model. The six head movement parameters derived from the realignment procedure were also included as confounds. All the trials were time locked to the onset of the target stimuli by a canonical synthetic hemodynamic response function and its first-order time derivative with event duration of 0 s. 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.

The unimodal Auditory_Single and Visual_Single trials and the bimodal Auditory_Visual and Visual_Auditory trials were the four conditions of interest. Therefore, for each participant, simple main effects for each of the four critical types of events were computed by putting 1 on the regressor of interest and 0 on all the other regressors, respectively, that is, the experimental trials versus the baseline mean. The four first-level individual contrast images were then fed into a within-participants ANOVA at the second group level using a random-effects model (flexible factorial design in SPM8 including an additional factor modeling the subject means). In the modeling of variance components, we allowed for violations of sphericity by modeling non-independence across parameter estimates from the same subject and allowing unequal variances both between conditions and participants using the standard implementation in SPM8. We were especially interested in the differential neural activity between the two types of unimodal trials (Visual_Single vs Auditory_ Single) and between the two bimodal behavioral conditions (Visual_ Auditory vs Auditory_Visual). Areas of activation were identified as significant only if they passed a conservative threshold of $p < 0.05$, family-wise error (FWE) corrected for multiple comparisons at the cluster level, with an underlying voxel level of $p < 0.005$, uncorrected (Poline et al., 1997).

*Psychophysiological interaction analysis.* To further investigate how the primary/secondary visual and auditory cortices (see Fig. 3) and the dorsal visual stream (see Fig. 4A) were involved in preferentially selecting visual versus auditory information during audiovisual competition, psychophysiological interaction (PPI) analysis was used to examine the context-specific functional modulation of neural activity across the brain by the neural activity in the primary/secondary visual and auditory cortices and the dorsal visual stream, respectively. 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). Neural activity in the primary/secondary visual and auditory cortices and in the dorsal visual stream was used as the physiological factor, respectively, and the contrast "Visual_Auditory > Auditory_Visual" as the psychological factor.

For each participant, the contrasts "Visual_Single > Auditory_ Single," "Auditory_Single > Visual_Single," and "Visual_Auditory > Auditory_Visual" were first calculated at the individual level, respectively. Subsequently, for neural activations in each of the above three neural contrasts, participant's individual peak voxels were 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 from the second-level group analysis (Table 1; see Figs. 3, 4A). Individual peak voxels from every participant were located in the same anatomical structure [left middle occipital gyrus (MOG), $x = -22 \pm 6$, $y = -100 \pm 3$, $z = 1 \pm 5$; right MOG, $x = 29 \pm 5$, $y = -93 \pm 5$, $z = 1 \pm 6$; left superior temporal gyrus (STG), $x = -63 \pm 4$, $y = -30 \pm 7$, $z = 9 \pm 4$; right STG, $x = 69 \pm 2$, $y = -20 \pm 6$, $z = 4 \pm 4$; precuneus, $x = -11 \pm 4$, $y = -65 \pm 6$, $z = 58 \pm 4$]. Next, time series were extracted as the first principal component from a sphere of 4 mm radius (twice the voxel size) around the individual peak voxels. PPI analysis at the first individual level used

**Table 1. Brain regions showing significant relative increases of BOLD response associated with the unimodal Visual_Single and Auditory_Single trials, and the bimodal Visual_Auditory and Auditory_Visual trials**

| Anatomical region | Side | Cluster peak (mm) | $t$ score | $k_E$ (voxels) |
|---|---|---|---|---|
| **Unimodal trials** | | | | |
| *A*, Visual_Single > Auditory_Single | | | | |
| MOG | R | 30, −90, 0 | 9.76 | 486 |
| | L | −18, −102, −2 | 6.79 | 415 |
| *B*, Auditory_Single > Visual_Single | | | | |
| STG | R | 66, −18, 4 | 11.88 | 1430 |
| | L | −62, −30, 10 | 11.18 | 2787 |
| Cerebellar vermis | M | 4, −62, −2 | 6.72 | 758 |
| **Bimodal trials** | | | | |
| *C*, Visual_Auditory > Auditory_Visual | | | | |
| Precuneus | L | −10, −66, 58 | 5.73 | 7767 |
| Human V6A | L | *−6, −82, 50* | *5.41* | |

The coordinates ($x$, $y$, $z$) correspond to MNI coordinates. Displayed are the coordinates of the maximally activated voxel within a significant cluster and the coordinates of relevant local maxima within the cluster (in italics). L, Left; M, middle; R, right.

one regressor representing the extracted time series in the given ROI (i.e., the physiological variable), one regressor representing the psychological variable of interest (i.e., Visual_Auditory > Auditory_Visual), and a third regressor representing the cross product of the previous two (the PPI term). At the individual level, an SPM was calculated to reveal brain areas in which the neural activation was predicted by the PPI term, with the physiological and the psychological regressors being treated as confound variables, i.e., by putting 1 on the PPI regressor and 0 on the physiological and the psychological regressors, respectively. At the group level, random-effects analysis was adopted: the individual SPMs corresponding to the PPI term in each participant were subsequently entered into a one-sample $t$ test ($p < 0.05$, FWE correction for multiple comparisons at cluster level with an underlying voxel threshold at $p < 0.005$, uncorrected).

*Statistical analysis of prestimulus neural activity.* To further investigate how variations of neural activity before the actual appearance of sensory stimuli (i.e., the prestimulus preparation phases) predicted the direction and the extent of sensory dominance on the appearance of audiovisual stimuli, a new GLM was estimated. In the new model, different types of events were locked to the time points when participants made their final responses in the previous trials ("Trials N − 1"), i.e., the prestimulus preparation phase of the current trial ("Trials N"). All the outliers, errors, and missed trials and trials preceded by outliers and errors were modeled separately as a regressor of no interest. Therefore, by directly contrasting the VA and the AV bimodal behavioral conditions, we could test whether prestimulus variations of neural activity in a brain region predicted the direction of sensory dominance (visual vs auditory dominance).

More critically, behavioral performance on Trials N was included as a parametric regressor for four critical types of events (Auditory_Single, Visual_Single, Auditory_Visual, and Visual_Auditory, respectively). The parametric regressor modeled the trial-to-trial variance in the average prestimulus BOLD signal that varied linearly with trial-to-trial variance in task performance within each of four types of events. Because all the Trials N belonged to the same type of behavioral condition with the same bottom-up stimuli, by calculating the parametric modulation effect of task performance in Trials N on the prestimulus neural activity in Trials N − 1, we could investigate how the variations of neural activity before the actual appearance of bottom-up stimuli predicted the subsequent task performance.

For the unimodal trials (Visual_Single and Auditory_Single), because only one RT was obtained in each trial, we included the RT on each unimodal trial as a parametric regressor for the Visual_Single and the Auditory_Single trials, respectively. The relative RT for each trial was measured as the mean corrected score: RT for that trial minus the mean RT of all correct trials within each type of unimodal trials. In contrast to the unimodal trials, in which only one RT was obtained in each trial, two RTs were obtained on each bimodal trial (one for the visual and one for

the auditory component). By subtracting the visual and the auditory RTs in the bimodal trials, we could quantify how much the visual/auditory response was preceded the other one: "Auditory RT > Visual RT" for the size of visual dominance in the Visual_Auditory behavioral condition; "Visual RT > Auditory RT" for the size of auditory dominance in the Auditory_Visual behavioral condition. Subsequently, the mean corrected size of the visual and the auditory dominance in the current VA and AV trials (Trials N) was included as a parametric regressor, respectively. Because the neural events were time locked to the prestimulus preparation phases of the VA and the AV trials (Trials N − 1), the parametric regressors of the size of sensory dominance in Trials N modeled the trial-to-trial variance of the prestimulus BOLD signal that varied linearly with the trial-to-trial variance in the size of sensory dominance in the VA and the AV trials, respectively. Therefore, we could test how much the variations of neural activity in a brain region before the actual appearance of audiovisual stimuli predicted the size of sensory dominance in the subsequent trial. The parametric regressor for a single bimodal trial was coded only when the responses in both the Trial N and the Trial N − 1 were neither missed nor outliers. Unless illustrated otherwise, brain regions activated by the parametric modulation effects were identified as significant only if they passed a conservative threshold of $p < 0.05$, FWE correction for multiple comparisons at the cluster level, with an underlying voxel level of $p < 0.005$, uncorrected.

It has been well documented that increased prestimulus neural activity in the parietofrontal attention control regions predicted better task performance, but increased (less deactivated) prestimulus neural activity in the DMN predicted worse task performance (Weissman et al., 2006). Therefore, we predicted that the prefrontal attention control regions should be significantly involved in the positive parametric modulation effect of sensory dominance, i.e., higher prestimulus activity predicted higher sensory dominance. Moreover, the DMN should be significantly involved in the negative parametric modulation effect of sensory dominance, i.e., higher (less deactivated) prestimulus activity predicted lower sensory dominance.

To more clearly demonstrate how the size of sensory dominance changed as a function of the height of prestimulus neural activity, we split Trials N − 1 according to the size of sensory dominance on Trials N. Specifically, the trials before the VA and the AV trials (Trials N − 1) were split into the higher and the lower halves, respectively, according to the size of the visual and the auditory dominance on the current VA and AV trials (Trials N). Subsequently, time courses for the BOLD responses in the preceding trials (Trials N − 1) of the high versus low dominance VA and AV trials were extracted from the brain regions activated significantly by the parametric modulation effects, respectively. A finite impulse response (FIR) model was used to estimate the mean event-related BOLD responses during Trials N − 1 in the activated brain regions for each participant. The FIR model uses a linear model to provide unbiased estimates of the average signal intensity at each time point of the Trials N − 1 rather than making a priori assumptions about the shape of the BOLD response (Burock and Dale, 2000). We used seven 2.2 s time bins (corresponding to the TR), starting from the beginning of Trials N − 1. The dependent measure in the time course plots was in units of percentage signal change from the means over the whole session measured within the activated clusters (see Fig. 5).

In addition, because different types of trials were mixed randomly for each participant, the composition of the different types of events among Trials N − 1 should not vary and accordingly should not contribute to the potential difference between the critical behavioral conditions. To prove this point, we further calculated the proportions of the different types of trials among the Trials N − 1 to the VA and the AV trials, respectively (for details, see the behavioral results; Fig. 1C).

### ERP experiment
*Experimental procedures.* The experiment was conducted in a dimly lighted and soundproof room. The stimuli and procedures were similar to those in the fMRI experiment, except for the following: (1) the visual target was present on an LCD monitor; (2) the auditory target was delivered via a loudspeaker that was positioned directly behind the LCD monitor to ensure that the auditory tone sounded like it was coming from the

same central spatial position as the visual target; and (3) the ERP experiment consisted of 10 blocks in total, and each block included 80 Auditory_Single trials, 80 Visual_Single trials, and 40 bimodal trials, which were mixed randomly. Each trial was followed by a time interval that was selected randomly among 1350, 1450, 1550, 1650, and 1750 ms. The temporal order of trials was randomized for every participant.

*ERP recording and analysis.* EEGs were recorded continuously from 64 Ag/AgCl electrodes (10–20 system) with BrainAmp DC amplifiers (low-pass, 100 Hz; high-pass, 0.01 Hz; sampling frequency, 500 Hz). Signals were referenced online to the left mastoid and re-referenced offline to the two mastoids average. Electrooculograms (EOGs) were recorded using three facial bipolar electrodes, with two placed on the outer canthi of each eye to record the horizontal EOG and one positioned in the inferior areas of the left orbit to record the vertical EOG. All the electrode impedances were kept below 5 k$\Omega$.

During offline data analysis, as we did in the fMRI experiment, we *post hoc* classified the bimodal trials into six types based on the participants' online responses and segmented the EEGs of the bimodal trials according to the six types of bimodal behavioral conditions and the two types of unimodal trials. Each segment was 800 ms, including a 100 ms pretarget interval for baseline correction. All the segments with EEG exceeding ±100 $\mu$v relative to baseline and with EOG exceeding ±80 $\mu$v relative to baseline were excluded.

*Statistical analysis of prestimulus neural activity.* In addition to the poststimulus evoked neural activity, we further examined whether there existed any prestimulus difference between the VA and AV bimodal behavioral conditions and when the potential difference began. Stepwise paired-sample *t* tests were performed to compare the differences in EEG activity before the appearance of audiovisual targets between the VA and the AV trials. The prestimulus intervals of −1000 to −800, −800 to −600, −600 to −400, and −400 to −200 ms were used as the baseline, respectively.

### Behavioral control experiment

There were two major differences between the ERP and the fMRI experiments: (1) loud scanner noise existed only in the fMRI experiment, not in the ERP experiment; and (2) participants were lying in the scanner in the fMRI experiment but were sitting in a chair in the ERP experiment. To examine whether the scanner noise and the body position altered the nature of audiovisual competition in the present paradigm, we ran a behavioral control experiment in the scanner without switching on the EPI sequence. A new group of 17 participants was instructed to perform the same behavioral task as that in the fMRI experiment while lying in the scanner without the background EPI noise. We predicted that, because the scanner noise was present only in the fMRI experiment but not in the behavioral control, with the body position being further controlled for, any potential difference between the fMRI and the behavioral control experiments should be attributed to the effect of scanner noise. In addition, because the body position changed between the behavioral control and the ERP experiment, with the background noise being removed, any potential difference between them should be attributed to the effect of body position.

## Results

### Behavioral data

To test whether the scanner noise and the body position of the participants altered the nature of audiovisual competition, behavioral data from the behavioral control (no noise and lying), the ERP (no noise and sitting), and the fMRI (noise and lying) experiments were submitted to the same repeated-measures ANOVA, with the three experiments being treated as a between-group factor.

#### Proportions of different types of behavioral conditions

Proportions of the six different types of bimodal behavioral conditions in the three experiments were illustrated in Figure 1B. To examine whether there existed the visual dominance effect in the three experiments, a 3 (the between-group factor: behavioral control, ERP, and fMRI experiments) × 2 (type of both responded but asynchronous bimodal trials: VA vs AV) repeated-measures ANOVA was performed on the proportions. The main effect of the two bimodal behavioral conditions was the only significant effect ($F_{(1,51)} = 16.14$, $p < 0.001$). Neither the main effect of the between-group factor nor the interaction was significant (both *p* values >0.05). This pattern of results suggested that the proportion of the VA responses was significantly higher than the proportion of the AV responses, i.e., a visual dominance effect, in all the three experiments (all *p* values <0.05).

In addition, to ensure that the composition of the different types of trials among the Trials N − 1 to the AV and the VA trials did not differ and accordingly did not contribute to the potential prestimulus difference, we calculated the proportions of the different types of trials among the Trials N − 1 to the AV and VA trials, respectively (Fig. 1C). Planned paired-sample *t* tests suggested no significant difference between the AV and the VA behavioral conditions for all types of trials (all *p* values >0.1), indicating that the AV and VA behavioral conditions were preceded by comparable proportions of the different types of trials.

#### RTs in the unimodal trials

There was no significant difference between unimodal RTs in the Auditory_Single and the Visual_Single trials in all the three experiments (Fig. 2A). RTs in the unimodal trials were submitted to a 3 (the between-group factor: behavioral control, ERP, and fMRI experiment) × 2 (the type of unimodal trials: Visual_Single vs Auditory_Single) repeated-measures ANOVA (Fig. 2A). The only significant effect was the main effect of the between-group factor ($F_{(2,51)} = 9.65$, $p < 0.001$), indicating that unimodal RTs in the fMRI experiment (578 ms) were significantly slower than unimodal RTs in the behavioral control (466 ms) and the ERP (457 ms) experiments (all *p* values <0.05, Bonferroni's correction). Neither the main effect of the type of unimodal trials nor the two-way interaction was significant (both *F* values <1).

#### RTs in the bimodal trials

To test how the pattern of RTs during audiovisual competition changed as a function of the three experiments, RTs in the bimodal trials were submitted to a 3 (the between-group factor: behavioral control, ERP, and fMRI experiments) × 2 (type of response: auditory response vs visual response) × 2 (response order: first vs second) repeated-measures ANOVA (Fig. 2B). The main effect of the type of response was significant ($F_{(1,51)} = 7.02$, $p < 0.05$), indicating that responses to the auditory components of the bimodal trials (602 ms) were slower than responses to the visual components (590 ms). The main effect of the response order was significant ($F_{(1,51)} = 229.16$, $p < 0.001$), indicating that the first responses (523 ms) were significantly faster than the second responses (669 ms). The main effect of the between-group factor was marginally significant ($F_{(2,51)} = 2.997$, $p = 0.059$), indicating that there was a trend that the bimodal RTs in the fMRI experiment (649 ms) was slower than the bimodal RTs in the behavioral control (569 ms) and the ERP (570 ms) experiments. The two-way interaction between the type of response and the response order was significant ($F_{(1,51)} = 21.70$, $p < 0.001$), but the three-way interaction was not ($F < 1$). This pattern of results suggested that, in all the three experiments, it was faster for the visual responses to recover from the previous auditory responses than for the auditory responses to recover from the previous visual responses (Fig. 2B): 149 versus 210 ms in the behavioral control experiment, 143 versus 199 ms in the ERP experiment, and 57 versus 120 ms in the fMRI experiment (planned paired-
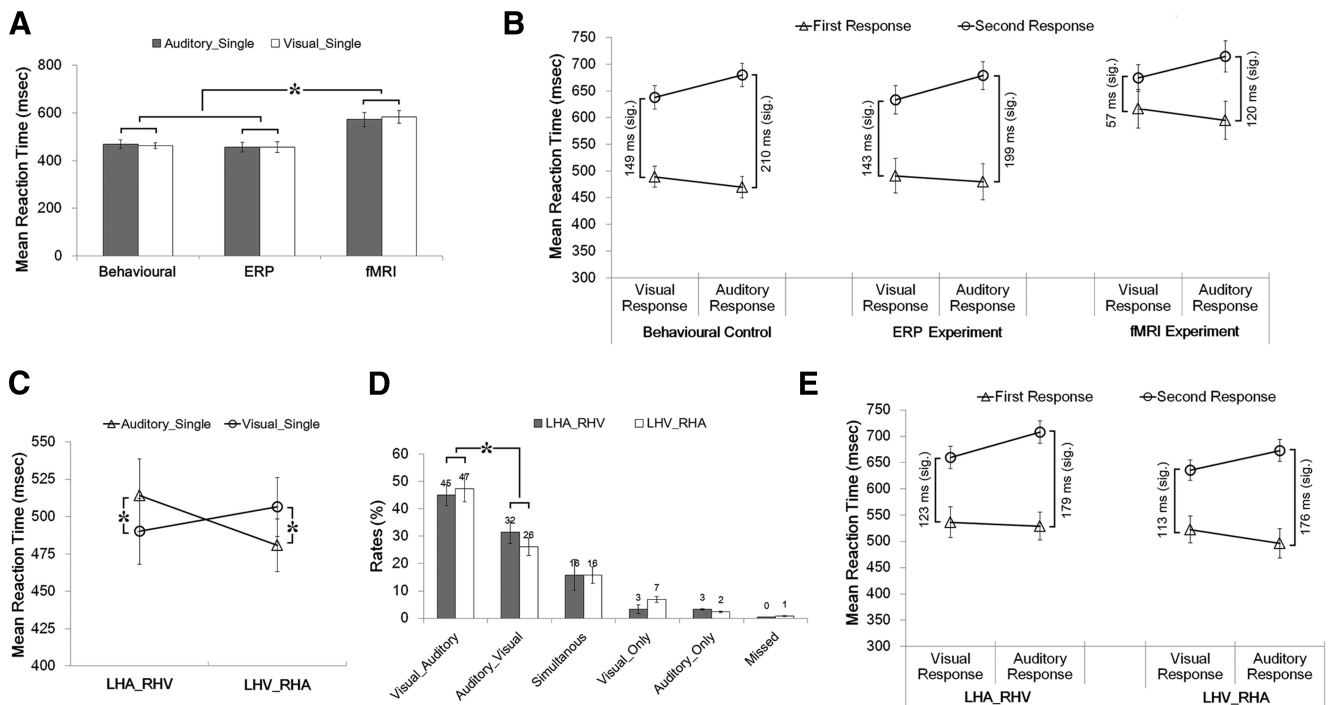
**Figure 2.** Behavioral results in the three experiments. **A**, Mean RTs in the unimodal trials of the three experiments. **B**, RTs to the visual and the auditory components of the Visual_Auditory and the Auditory_Visual trials shown as a function of response order across the three experiments. **C**, RTs in the unimodal trials shown as a function of the correspondence between the responding hand and the sensory stimuli (collapsed across the three experiments). LHA_RHV, The group of participants who used the left hand for auditory stimuli and the right hand for visual stimuli; LHV_RHA, the group of participants who used the left hand for visual stimuli and the right hand for auditory stimuli. **D**, Proportions of the six different types of behavioral conditions in the bimodal trials shown as a function of the LHA_RHV versus LHV_RHA between-group factor. **E**, RTs to the visual and the auditory components of the Visual_Auditory and Auditory_Visual trials shown as a function of the response order for the LHA_RHV and the LHV_RHA groups, respectively. *$p < 0.05$. The error bars indicate SEs.

sample $t$ tests were significant in all three experiments, all $p$ values <0.05, Bonferroni's correction). The two-way interaction between the type of response and the response order was significant in all the three experiments (all $p$ values <0.05). In addition, the two-way interaction between the between-group factor and the response order was significant ($F_{(2,51)} = 8.68$, $p < 0.001$), indicating that the RT difference between the first and the second response was significantly smaller in the fMRI experiment (89 ms) than in the behavioral (180 ms) and the ERP (171 ms) experiments. No other significant effect was found.

Together, the scanner noise in the fMRI experiment (1) generally slowed the behavioral performance in both the unimodal and the bimodal trials (Fig. 2A,B) and (2) shrunk the difference between the first and the second response in the bimodal trials (Fig. 2B). However, the scanner noise did not alter the critical pattern of visual dominance across the three experiments. The proportion of the VA responses was significantly higher than the proportion of the AV responses (Fig. 1B), and the visual responses recovered more quickly from the previous auditory responses than vice versa in all the three experiments (Fig. 2B). In addition, although the body position of the participants changed between the behavioral control (lying) and the ERP experiment (sitting), we did not find significant differences between the behavioral control and the ERP experiments (Figs. 1B, 2B), indicating that the body position did not change the nature of audiovisual competition in the present paradigm as well. However, note that, although the scanner noise did not alter the nature of visual dominance at the behavioral level, we cannot conclusively rule out the possibility that the scanner noise affects some of the neural data measured during the fMRI experiment.

*Effect of responding hand on sensory dominance*
In responding to the unimodal stimuli, the response speed of the right hand was significantly faster than the response speed of the left hand, regardless of which sensory modality was assigned to the right hand (Fig. 2C). Although the assignment of responding hand to sensory stimuli influenced the response speed to the unimodal stimuli, it did not alter the nature of audiovisual competition in the bimodal trials (Fig. 2D,E). Specifically, the proportion of the VA responses was significantly higher than the proportion of the AV responses in both the LHA_RHV and the LHV_RHA group (both $p$ values <0.05; Fig. 2D). Furthermore, the delay in responding to the visual components of bimodal stimuli was significantly smaller than the delay in responding to the auditory components, regardless of the correspondence between the responding hand and the sensory stimuli (123 vs 179 ms in the LHA_RHV group; 113 vs 176 ms in the LHV_RHA group; both $p$ values <0.05; Fig. 2E).

**fMRI data**
*Sensory systems involved in processing the unimodal visual and auditory stimuli*
By contrasting the unimodal visual and auditory trials, we first identified the sensory cortices that were involved in selectively representing the visual and the auditory targets, respectively. To ensure that the localized sensory cortices showed enhanced neural activity relative to the implicit baseline (i.e., the null trials in which only the central fixation was presented, and neither the visual nor the auditory targets were presented), the Visual_Single > Auditory_Single and the Auditory_Single > Visual_Single contrasts were inclusively masked by the simple main
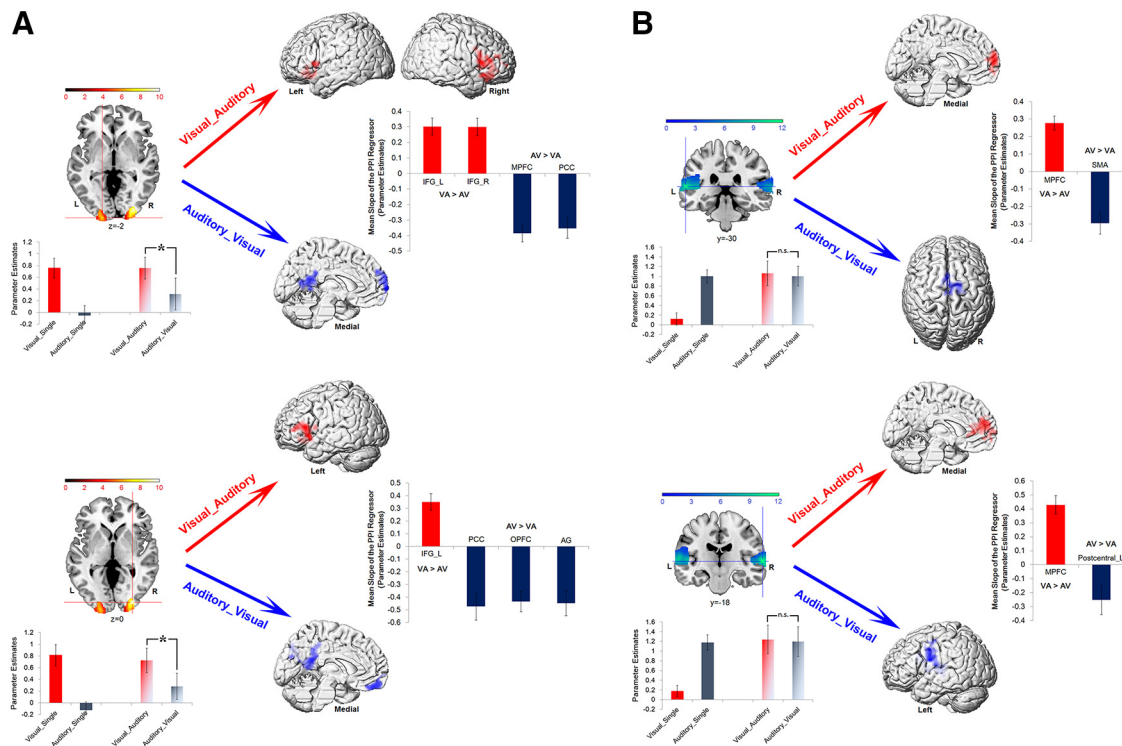
**Figure 3.** Neural activity in the unimodal trials. **A**, Visual_Single > Auditory_Single. The bilateral MOG in the primary/secondary visual cortex was significantly activated. **B**, Auditory_Single > Visual_Single. The bilateral STG in the primary/secondary auditory cortex was significantly activated. The color scale bars indicate *t* values. Left panels of (**A**) and (**B**): for the localized primary/secondary auditory and visual cortices, mean parameter estimates in the unimodal visual and auditory trials and in the bimodal Visual_Auditory and Auditory_Visual trials were extracted from the activated clusters, respectively. *$p < 0.05$, significant difference between bimodal behavioral conditions. The error bars are standard errors. **A**, **B**, Right columns, PPI analysis based on neural activity in the localized primary/secondary visual and auditory cortices, with the contrast Visual_Auditory > Auditory_Visual as the psychological factor. The visual and auditory cortices showed enhanced neural coupling with the DMN whenever the corresponding sensory modality lost multisensory competition in the Auditory_Visual or the Visual_Auditory trials, respectively. The visual and auditory cortices showed enhanced functional connectivity with the IFG and the sensorimotor areas, respectively, when the corresponding sensory modality won multisensory competition in the Visual_Auditory or the Auditory_Visual trials. Mean parameter estimates of the PPI regressor were extracted from the activated areas. The error bars indicate SEs.

effects of the Visual_Single and the Auditory_Single conditions (i.e., the "1 0" baseline contrasts), respectively, at the threshold of $p < 0.001$, uncorrected at the voxel level. In this way, only those voxels that reached a level of significance at $p < 0.001$ (uncorrected) in the mask contrasts [i.e., in the experimental conditions vs implicit baseline (null trials) contrasts] were included in the analysis.

First, bilateral MOG in the primary/secondary visual cortex (BA 17/18) showed significantly enhanced neural activity to the Visual_Single trials compared with the Auditory_Single trials (Fig. 3A, left; Table 1A). Second, bilateral primary/secondary auditory cortex in the STG and the cerebellar vermis showed significantly enhanced neural activity in the Auditory_Single trials compared with the Visual_Single trials (Fig. 3B, left; Table 1B). Mean parameter estimates in the four critical types of events (Visual_Single, Auditory_Single, Visual_Auditory, and Auditory_Visual) were further extracted from the activated clusters in the localized visual and auditory processing systems, respectively. For the bilateral MOG, neural activity was significantly weakened in the AV trials compared with the VA trials ($t_{left(16)} = 2.13$, $p < 0.05$; $t_{right(16)} = 2.50$, $p < 0.05$), although the bottom-up audiovisual inputs were identical in the two types of bimodal behavioral conditions (Fig. 3A, left). In addition, neural activity was equally high in the Visual_Single trials and the Visual_Auditory trials (*t* values <1 for both the left and the right MOG). Therefore, the visual processing system in the bilateral MOG showed specific neural selectivity to the visual components in the VA trials but significantly weakened neural selectivity to the visual

components in the AV trials. However, for the auditory processing system in the bilateral STG, neural selectivity to the auditory components was equally high in the VA and the AV trials (*p* values >0.1; Fig. 3B, left). In addition, neural activity was comparable between the Auditory_Single, the Visual_Auditory, and the Auditory_Visual trials (all *t* values <1). Therefore, the auditory processing system showed specific neural selectivity to the auditory components of both the VA and the AV trials.

*Neural correlates underlying visual dominance in bimodal trials*
We further investigated the neural correlates underlying the visual dominance effect by directly calculating the neural contrast "VA vs AV." An extended cluster along the medial dorsal visual stream, including the anterior and dorsal bank of parieto-occipital sulcus extending to the precuneus, showed significantly higher neural activity in the VA than in the AV bimodal behavioral conditions (Fig. 4A; Table 1C). Mean parameter estimates in the four critical types of events (Visual_Single, Auditory_Single, Visual_Auditory, and Auditory_Visual) were further extracted from the activated cluster (Fig. 4C, left). Neural activity in the dorsal visual stream did not show specific selectivity toward the unimodal visual targets compared with the unimodal auditory targets ($t_{(16)} = 1.14$, $p = 0.27$). In contrast, neural activity was enhanced significantly in the VA trials compared with the AV trials ($t_{(16)} = 2.33$, $p < 0.05$). No significant activation was found in the reverse contrast, i.e., "AV > VA."

*Functional connectivity of the primary/secondary visual cortex*

PPI analysis was performed with neural activity in the left MOG as the physiological factor and with the contrast "VA > AV" as the psychological factor. The left MOG showed significantly enhanced functional connectivity with bilateral inferior frontal gyrus (IFG) in the VA trials than in the AV trials and showed significantly increased functional connectivity with the medial prefrontal cortex (MPFC) and the posterior cingulate cortex (PCC) of the DMN in the AV trials than in the VA trials (Fig. 3A, top right; Table 2A). Similarly, the right MOG showed significantly enhanced neural coupling with the left IFG in the VA trials than in the AV trials and enhanced neural coupling with the orbital prefrontal cortex (OPFC), the PCC, and the left angular gyrus (AG) of the DMN in the AV trials than in the VA trials (Fig. 3A, bottom right; Table 2A).

*Functional connectivity of the early auditory system*

PPI analysis with neural activity in the left STG as the physiological factor and with the contrast VA > AV as the psychological factor revealed that the left STG showed significantly enhanced functional connectivity with the MPFC in the DMN in the VA trials than in the AV trials and enhanced functional connectivity with the supplementary motor area (SMA) in the AV trials than in the VA trials (Fig. 3B, top right; Table 2B). Similarly, the right STG showed significantly enhanced neural coupling with the MPFC in the DMN in the VA trials than in the AV trials and enhanced neural coupling with the left postcentral gyrus in the AV trials than in the VA trials (Fig. 3B, bottom right; Table 2B).

←

over the 300–370 ms period was overlaid on the three-concentric head model. The red and blue shadows indicated the SE of ERPs in the two bimodal behavioral conditions. *C*, PPI analysis based on neural activity in the precuneus, with the contrast Visual_Auditory > Auditory_Visual as the psychological factor. Left, Mean parameter estimates in the unimodal visual and auditory trials and in the bimodal Visual_Auditory and Auditory_Visual trials were extracted from the precuneus. Right top, The precuneus showed significantly higher functional connectivity with the bilateral IFG and the bilateral postcentral gyrus in the Visual_Auditory trials than in the Auditory_Visual trials. Right bottom, The precuneus showed significantly higher functional connectivity with the OPFC and the PCC in the Auditory_Visual trials than in the Visual_Auditory trials. Mean parameter estimates of the PPI regressor were extracted from the activated areas. *p < 0.05. The error bars indicate SEs.
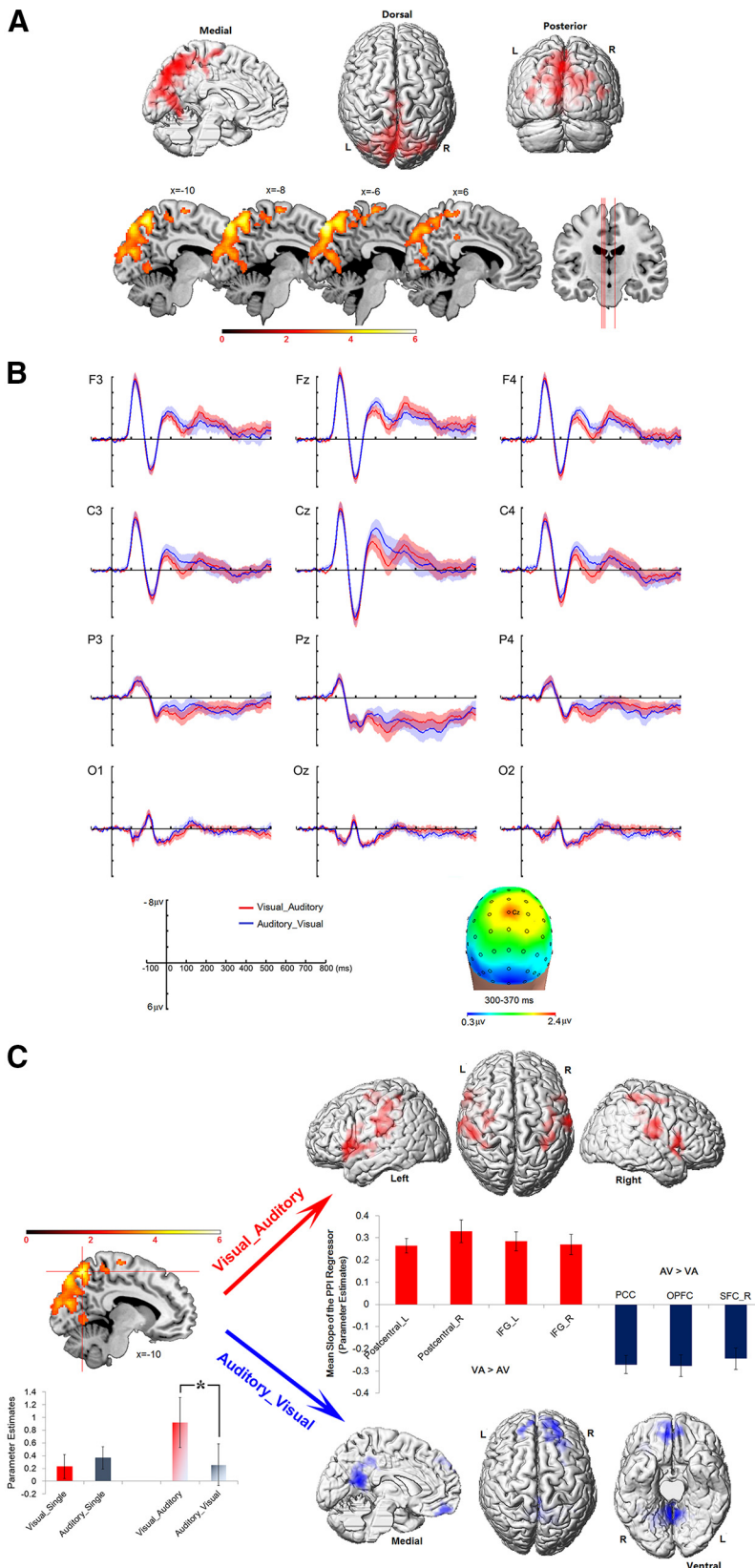
**Figure 4.** Neural activity in the bimodal trials. *A*, Visual_Auditory > Auditory_Visual. The medial dorsal visual stream, including the precuneus and the human V6A, showed significantly higher neural activity in the Visual_Auditory trials than in the Auditory_Visual trials. *B*, Grand-average ERP waveforms in the 800 ms poststimulus interval at representative analyzed electrodes and topography for the Visual_Auditory and the Auditory_Visual trials. In all the ERP graphs, a low-pass filter (30 Hz at 24 dB/octave) was used for visual clarity. The divergence between the two bimodal behavioral conditions started to reach significance at 250 ms after the onset of the target. The topography of the Visual_Auditory versus Auditory_Visual difference voltage maps

**Table 2. Brain regions that showed higher functional connectivity with the primary/secondary visual and auditory cortices and with the dorsal visual stream in the bimodal Visual_Auditory versus Auditory_Visual trials**

| Anatomical region | Hemisphere | Cluster peak (mm) | t score | $k_E$ (voxels) |
|---|---|---|---|---|
| A, Primary/secondary visual cortex in bilateral MOG | | | | |
| Left MOG | | | | |
| Visual_Auditory > Auditory_Visual | | | | |
| IFG | R | 40, 28, 12 | 6.07 | 1382 |
| | L | −34, −24, −2 | 5.96 | 645 |
| Auditory_Visual > Visual_Auditory | | | | |
| PCC | M | 4, −64, 14 | 7.03 | 1029 |
| MPFC | M | 12, 64, 16 | 6.92 | 1533 |
| Right MOG | | | | |
| Visual_Auditory > Auditory_Visual | | | | |
| IFG | L | −48, 16, 14 | 7.47 | 1132 |
| Auditory_Visual > Visual_Auditory | | | | |
| PCC | M | −2, −48, 24 | 6.27 | 1628 |
| OPFC | M | −2, 44, −16 | 6.23 | 428 |
| AG | L | −52, −76, 28 | 4.77 | 480 |
| B, Primary/secondary auditory cortex in bilateral STG | | | | |
| Left STG | | | | |
| Visual_Auditory > Auditory_Visual | | | | |
| MPFC | M | 2, 56, 12 | 8.65 | 748 |
| Auditory_Visual > Visual_Auditory | | | | |
| SMA | R | 18, −12, 66 | 7.54 | 845 |
| Right STG | | | | |
| Visual_Auditory > Auditory_Visual | | | | |
| MPFC | M | 4, 50, 24 | 5.68 | 873 |
| Auditory_Visual > Visual_Auditory | | | | |
| Postcentral gyrus | L | −50, −16, 18 | 7.22 | 1534 |
| C, Dorsal visual stream | | | | |
| Left precuneus | | | | |
| Visual_Auditory > Auditory_Visual | | | | |
| IFG | L | −60, 14, 4 | 11.62 | 1898 |
| | R | 54, 8, 12 | 6.5 | 707 |
| Postcentral gyrus | L | −48, −34, 36 | 7.56 | 1887 |
| | R | 62, −20, 36 | 6.65 | 1446 |
| Auditory_Visual > Visual_Auditory | | | | |
| PCC | R | 20, −60, 18 | 7.83 | 2035 |
| OPFC | M | 8, 40, −16 | 6.6 | 692 |
| Superior frontal cortex | R | 28, 42, 48 | 6.48 | 1254 |

The coordinates (x, y, z) correspond to MNI coordinates. Displayed are the coordinates of the maximally activated voxel within a significant cluster. L, Left; M, middle; R, right.

### Functional connectivity of the precuneus

PPI analysis with neural activity in the precuneus as the physiological factor and with the contrast VA > AV as the psychological factor revealed that the precuneus showed significantly higher functional connectivity with the bilateral IFG and the bilateral postcentral gyrus in the VA trials than in the AV trials (Fig. 4C, top right; Table 2C) and significantly higher functional connectivity with the OPFC and PCC of the DMN in the AV trials than in the VA trials (Fig. 4C, bottom right; Table 2C).

### Variations of prestimulus neural activity in bimodal trials

Direct comparisons of the prestimulus neural activity between the bimodal VA and AV behavioral conditions did not reveal any significant activation. Subsequently, parametric modulation effects of the size of sensory dominance on prestimulus neural activity in the bimodal trials were calculated. For the VA (visual dominance) trials, prestimulus neural activity in the bilateral IFG and the anterior cingulate cortex (ACC) was correlated positively with the size of visual dominance in the subsequent VA trial: the higher the neural activity in the bilateral IFG and ACC before the appearance of the audiovisual stimuli, the higher the visual dominance (Fig. 5A, top; Table 3A). Accordingly, prestimulus neural activity in the bilateral IFG and the ACC was significantly higher for the high visual dominance VA trials than for the low visual dominance VA trials (Fig. 5A, bottom left, taking the left IFG as an example) but was comparable between the high and the low auditory dominance AV trials (Fig. 5A, bottom right). For the negative parametric modulation effect of visual dominance in the VA trials, no significant activation was revealed at the conservative threshold of $p < 0.05$, FWE correction for multiple comparisons at the cluster level with an underlying voxel threshold of $p < 0.005$, uncorrected. However, we had clear a priori hypothesis that the DMN should be involved in the negative parametric modulation effect of sensory dominance, i.e., higher (less deactivated) prestimulus neural activity in the DMN should predict lower sensory dominance. Therefore, at a less conservative threshold of $p < 0.05$, FWE correction at the cluster level with an underlying voxel level threshold of $p < 0.01$, uncorrected, we indeed found significant activations in the DMN (Fig. 5B; Table 3B). Prestimulus neural activity in the DMN was correlated negatively with the size of visual dominance in the subsequent VA trial: the less deactivated the DMN before the appearance of the audiovisual stimuli, the lower the visual dominance in the upcoming VA trials (Fig. 5B, top; Table 3B). Accordingly, prestimulus neural activity in the DMN was significantly more deactivated in the high visual dominance VA trials than in the low visual dominance VA trials (Fig. 5B, bottom left, taking the PCC as an example) but was comparable between the high and low auditory dominance AV trials (Fig. 5B, bottom right).

For the AV (auditory dominance) trials, prestimulus neural activity in the bilateral STG of the primary/secondary auditory cortex and SMA was correlated positively with the size of auditory dominance in the upcoming AV trials: the higher the prestimulus neural activity, the higher the auditory dominance (Fig. 5C, top; Table 3C). Accordingly, in the bilateral STG and the SMA, prestimulus neural activity was significantly higher in the high auditory dominance AV trials than in the low auditory dominance AV trials (Fig. 5C, bottom right, taking the left STG as an example) but was comparable between the high and low visual dominance VA trials (Fig. 5C, bottom left). No significant effect was found in the negative parametric modulation effect of auditory dominance even at a less conservative threshold of $p < 0.05$, FWE correction at the cluster level with an underlying voxel level threshold of $p < 0.05$, uncorrected.

### ERP experiment
#### ERP data
We first compared the waveforms elicited by the VA and the AV bimodal behavioral conditions at the electrodes of F3, F4, Fz, C3, C4, Cz, P3, P4, Pz, O1, O2, and Oz. The grand-average ERPs to the two types of bimodal behavioral conditions were shown in Figure 4B. The early ERP components for the two types of bimodal behavioral conditions overlapped perfectly at all analyzed electrodes: stepwise paired-sample t tests failed to reveal any significant difference between the VA and the AV trials during the early 0–250 ms interval. The waveforms of the early perceptual processes (from the onset of the target until to 250 ms) were statistically identical between the VA and the AV trials at all the analyzed electrodes, indicating that there was no significant difference in the early perceptual processes between the VA and the AV trials.

The divergence between the VA and the AV trials reached significance after 250 ms (Fig. 4B). The stepwise paired-sample t
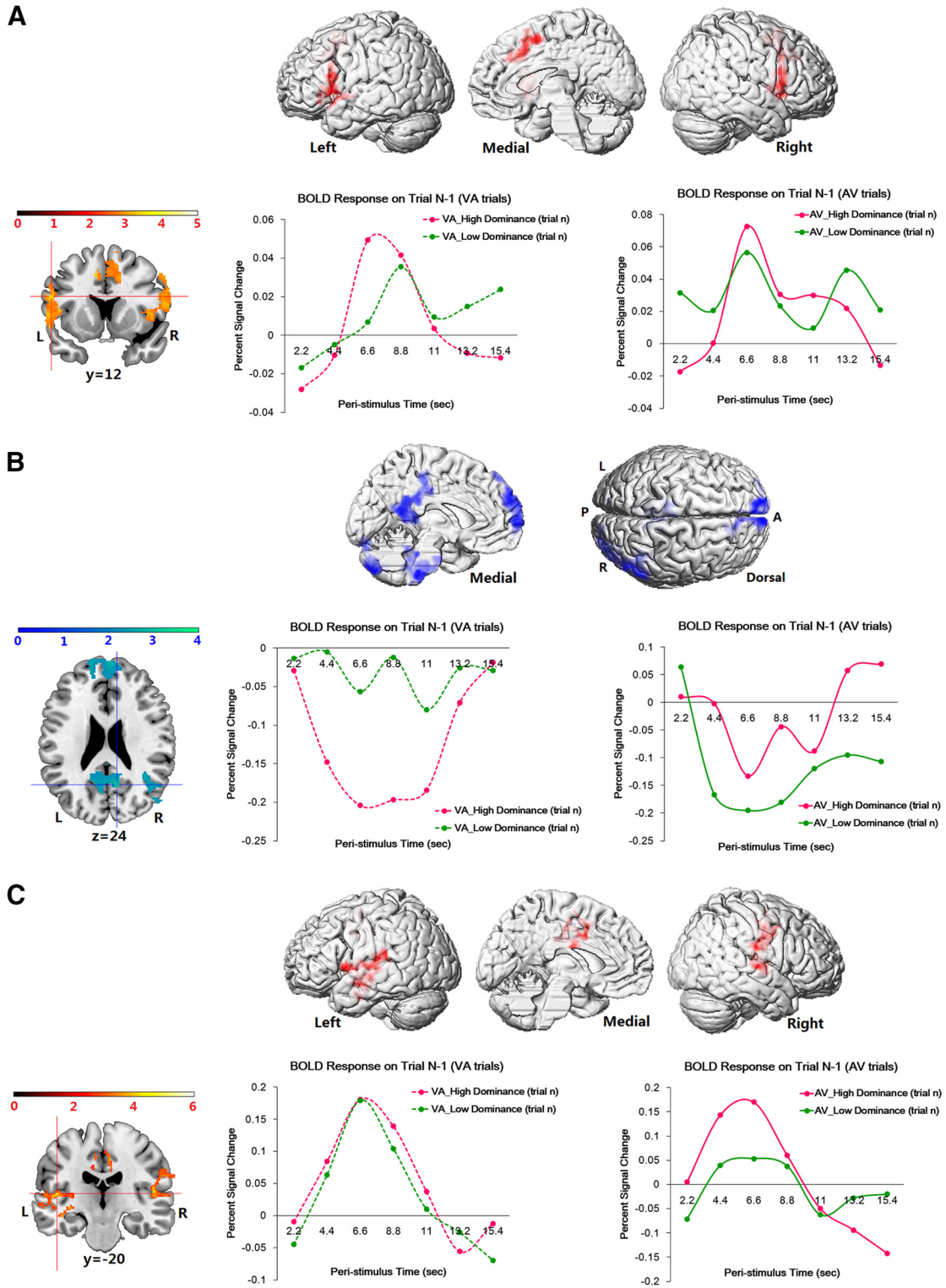
**Figure 5.** Variations of prestimulus neural activity as a function of the size of sensory dominance. ***A***, Top, Positive parametric modulation effect of visual dominance on prestimulus neural activity in the VA trials. Bottom, For example, in the left IFG, the height of BOLD responses in the trials before the high visual dominance VA trials was significantly higher than in the trials before the low visual dominance VA trials (left). In contrast, the height of BOLD responses was comparable between the Trials N − 1 of the high versus low auditory dominance AV trials (right). The other activated regions showed a similar pattern of prestimulus neural activity as the left IFG. ***B***, Top, Negative parametric modulation effect of visual dominance on prestimulus neural activity in the VA trials. Bottom, For example, in the PCC, the BOLD responses in the trials before the high visual dominance VA trials were significantly more deactivated than in the trials before the low visual dominance VA trials (left). In contrast, the height of BOLD responses was comparable between the Trials N − 1 of the high versus low auditory dominance AV trials (right). The other activated regions in the DMN showed a similar pattern of prestimulus neural activity as the PCC. ***C***, Top, Positive parametric modulation effect of auditory dominance on prestimulus neural activity in the AV trials. Bottom, For example, in the left STG, the height of BOLD responses in the trials before the high auditory dominance AV trials was significantly higher than in the trials before the low auditory dominance AV trials (right). In contrast, the height of BOLD responses was comparable between the Trials N − 1 of the high versus low visual dominance VA trials (left). The other activated regions showed similar pattern of prestimulus neural activity as the left STG.

**Table 3. Brain regions activated by the parametric modulation effects of the size of sensory dominance on prestimulus neural activity in the Visual_Auditory and Auditory_Visual trials**

| Anatomical region | Side | Cluster peak (mm) | t Score | $k_E$ (voxels) |
|---|---|---|---|---|
| **Visual_Auditory trials** | | | | |
| *A,* Positive parametric modulation effect of visual dominance | | | | |
| IFG | L | −54, 12, 22 | 3.58 | 596 |
| | R | 60, 8, 10 | 3.39 | 495 |
| ACC | R | 12, 28, 38 | 3.58 | 729 |
| *B,* Negative parametric modulation effect of visual dominance | | | | |
| PCC | L | 10, −58, 24 | 3.11 | 1549 |
| MPFC | M | 2, 52, 30 | 3.09 | 1769 |
| AG | R | 38, −48, 26 | 2.72 | 864 |
| Inferior occipital gyrus | R | 32, −84, −2 | 3.56 | 1488 |
| Parahippocampus | R | 22, −6, −28 | 3.78 | 1836 |
| **Auditory_Visual trials** | | | | |
| *C,* Positive parametric modulation effect of auditory dominance | | | | |
| STG | L | −46, −20, 10 | 4.57 | 1015 |
| | R | 50, −20, 10 | 4.09 | 438 |
| SMA | R | 10, −16, 44 | 3.86 | 606 |

The coordinates (*x, y, z*) correspond to MNI coordinates. Displayed are the coordinates of the maximally activated voxel within a significant cluster. L, Left; M, middle; R, right.

tests with significant *t* values included the time intervals of 300–380 ms at F3, 260–390 ms at F4, 280–390 ms and 400–490 ms at Fz, 250–400 ms at C3, C4, and Cz, 270–420 ms at P3 and P4, 280–390 ms at Pz, 320–420 ms at O1, and 500–590 ms at O2 and Oz. The topography of the VA > AV difference voltage maps over the 300–370 ms period was shown in Figure 4*B* (bottom). The voltage map showed that neural activity evoked by the VA responses was stronger than neural activity evoked by the AV responses over the centroparietal regions.

*Variations of prestimulus neural activity in bimodal trials*
In contrast to the fMRI results, the comparisons of the prestimulus EEG activity between the VA and the AV behavioral conditions did not reveal any significant prestimulus difference at the electrodes of F3, F4, Fz, C3, C4, Cz, P3, P4, Pz, O1, O2, and Oz, with the four prestimulus intervals (−1000 to −800, −800 to −600, −600 to −400, and −400 to −200 ms) being used as the baseline, respectively.

Note that, although significant prestimulus differences between the VA and the AV behavioral conditions were revealed in the fMRI experiment (Fig. 5), no significant prestimulus difference was observed in the EEG experiment. The statistical null effect of the EEG data could be caused by various sources of noises, and one cannot put elaborate discussions on a statistical null effect. The tension, in terms of prestimulus neural activity, between the EEG and the fMRI data may be attributable to differences in timescale of the effects or relative sensitivity of EEG and fMRI.

## Discussion

When audiovisual information was presented simultaneously, responses to the visual targets preceded responses to the auditory targets more frequently than vice versa (Fig. 1*B*). Even when visual responses were preceded by auditory responses, they recovered more quickly from previous auditory responses than vice versa (Fig. 2*B*), indicating the dominance of vision over audition in terms of both frequency of occurrence and response speed. This visual dominance effect existed regardless of background

noise, body position, and responding hand (Figs. 1*B*, 2). When individuals perform two sensorimotor tasks in immediate succession, response to the second task is often postponed, which is termed as the psychological refractory period (PRP; Telford, 1931; Smith, 1967; Pashler, 1994). PRP has been suggested to occur even when stimuli are chosen between different sensory modalities (Pashler, 1994). Our RT data, for the first time, suggested that there existed asymmetries in cross-modal PRP: visual responses and the corresponding neural motor preparations recovered from the PRP caused by previous auditory responses more quickly than vice versa.

We then localized the visual and auditory cortices that selectively processed the present visual and auditory targets, respectively, and then tested variations of neural activity in the localized sensory cortex during multisensory competition in the bimodal trials (Fig. 3*A*,*B*, left). During multisensory selective attention, neural representations in the dominant modality win multisensory competition probably at the cost of neural representations in the other modalities, i.e., the modality-based biased competition (Spence et al., 2012). Our results suggested that the dominant sensory cortices showed specific neural selectivity to its corresponding sensory information, and the critical difference occurred in the dominated sensory cortices. When vision was dominated by audition in the AV trials, the primary/secondary visual cortex showed significantly weakened activity toward the visual components (Fig. 3*A*, left). However, when audition was dominated by vision in the VA trials, the primary/secondary auditory cortex still showed specific neural selectivity toward the auditory components (Fig. 3*B*, left). Therefore, compared with the auditory cortex, the primary/secondary visual cortex showed more flexibly changing neural activity depending on the outcome of multisensory competition. The decreased activity in the primary/secondary visual cortex during auditory dominance was also consistent with previous evidence showing cross-modal deactivations in the visual cortex during selective attention to the auditory modality (Laurienti et al., 2002; Mozolic et al., 2008).

Connectivity analysis based on neural activity in the visual and auditory cortices suggested that the sensory systems dynamically interacted with the prefrontal cortex, the sensorimotor cortex, and the DMN during multisensory competition. On the one hand, whenever one sensory modality lost the multisensory competition, it showed enhanced connectivity with the DMN (Figs. 3*A*,*B*, right, 4*C*, right). The DMN of the human brain has been suggested to show task-induced deactivations during the performance of various goal-directed tasks (Gusnard and Raichle, 2001; Raichle et al., 2001), and momentary lapses of selective attention are characterized by less deactivation of the DMN (Weissman et al., 2006). Recent evidence further showed how visual areas are dynamically coupled with the frontoparietal network and the DMN depending on current behavioral goals. The fusiform face area was coupled with the frontoparietal network when faces were task relevant, whereas the parahippocampal place area was coupled with the DMN when houses were task irrelevant (Chadick and Gazzaley, 2011). In the present study, whenever neural activity in one sensory system was synchronized functionally with the DMN, momentary lapses in attention to the corresponding sensory information might be evoked, and this exact sensory system accordingly lost the multisensory competition.

On the other hand, when vision dominated audition, the primary/secondary visual cortex showed enhanced connectivity with the IFG (Fig. 3*A*). The prefrontal cortex biases top-down selective attention toward task-relevant visual information and causes the visual cortex to code high-quality perceptual represen-

tations that can be fed forward into the sensorimotor cortex to determine behavior (Desimone, 1998; Kastner et al., 1998, 1999; Kastner and Ungerleider, 2000; Pessoa et al., 2003). In the present study, although the visual and auditory targets were both behaviorally relevant, if the visual targets received enhanced attention via increased neural coupling between the IFG and the visual cortex, they would eventually dominate audition. When audition dominated vision, the auditory cortex showed enhanced connectivity directly with the sensorimotor areas (Fig. 3B, right). Previous anatomical and functional data consistently suggested that there exists an inherent link between the auditory and the motor systems. Anatomically, the auditory cortex is connected directly with the premotor cortex (Zatorre et al., 2007), and functionally, passive listening to purely perceptual rhythmic sounds can engage the premotor cortex, indicating that the motor system may be sensitive to the basic physical properties of auditory stimuli (Chen et al., 2008). Our results further suggested that the enhanced functional connectivity between the auditory and the motor system may expedite the access of auditory information to its sensorimotor representations.

Furthermore, by directly contrasting the VA trials with the AV trials, we found that the dorsal visual stream showed significantly increased neural activity when vision dominated audition (Fig. 4A). The ERP data further suggested that, during the later postperceptual phases, rather than the earlier perceptual phases, the VA trials elicited a more positive neural response than the AV trials over dorsal frontoparietal areas (Fig. 4B). In contrast to the primary/secondary visual cortex in the bilateral MOG (Fig. 3A), the dorsal visual stream did not show specific neural selectivity toward the unimodal visual targets. Rather, it showed specifically enhanced neural activity during visual dominance in the VA trials (Fig. 4C, left). Moreover, the dorsal visual stream showed significantly increased neural coupling not only with the bilateral IFG, as the bilateral MOG did (Figs. 3A, 4C), but also with the bilateral sensorimotor areas when vision dominated audition (Fig. 4C, top right). Anatomically, the maximally activated regions in the dorsal visual stream (Fig. 4A, C; Table 1C) correspond to the human homolog of macaque area V6A, which is a visuomotor area located to the anterior dorsal bank of parieto-occipital sulcus (Pitzalis et al., 2013). Functionally, the V6A complex contains both visual and sensorimotor cells (Breveglieri et al., 2006; Gamberini et al., 2011), which makes it functionally suitable to feed visual representations forward into the sensorimotor system (Chen et al., 2012). In humans, the V6A area responds to finger pointing and reaching movements (Pitzalis et al., 2013). Consistent with its functional role, the dorsal visual stream governs the visual control of online movements in a primarily fast and automatic manner (Goodale et al., 1991; Goodale and Milner, 1992; Shmuelof and Zohary, 2005; Milner and Goodale, 2008; Milner, 2012). The lack of specific neural selectivity to unimodal visual stimuli in the dorsal visual stream may be attributable to the visuomotor, rather than pure perceptual, nature of this area.

Although the poststimulus variations of neural activity were revealed (Figs. 3, 4), the neural causes and consequences of sensory dominance remain to be elucidated. By testing trial-to-trial relationships between prestimulus neural activity and the size of sensory dominance, we further investigated whether sensory dominance was caused by altered neural activity in the sensory systems per se or was driven by top-down modulations from prefrontal attention control regions. Our results suggested that increased prestimulus activity in the prefrontal cortex and decreased prestimulus activity in the DMN predicted enhanced visual dominance (Fig. 5A, B), whereas increased prestimulus

activity in the auditory cortex predicted enhanced auditory dominance (Fig. 5C). Therefore, an integrated view of visual dominance started with prestimulus variations of neural activity in the prefrontal cortex and DMN. Increased prefrontal and decreased DMN activity before the appearance of audiovisual stimuli would enhance the connectivity from the IFG to both the primary/secondary visual cortex and the dorsal visual stream in terms of facilitatory computational roles of the IFG (Miller and Cohen, 2001; O'Reilly et al., 2010). The top-down modulation from IFG would facilitate visual processing in the bilateral MOG to code high-quality perceptual representations on the one hand (Fig. 3A) and expedite the transformation of perceptual visual representations to sensorimotor representations via the dorsal visual stream on the other hand (Fig. 4C), which eventually resulted in enhanced visual dominance (Fig. 5A). However, the auditory dominance started with prestimulus variations of neural activity in the auditory cortex. Increased activity in the auditory cortex before the appearance of audiovisual stimuli could increase the connectivity between the auditory cortex and the sensorimotor areas (Fig. 3B), expedite the access of auditory information to the corresponding sensorimotor representations, and eventually result in enhanced auditory dominance (Fig. 5C).

To summarize, multisensory competition implicates both the low-level sensory systems and the high-level fronto-sensorimotor networks. However, it remained unclear whether the outcome of multisensory competition was caused by top-down control from the prefrontal cortex or by variations of bottom-up processing in the sensory systems. The present results revealed that visual dominance originated from top-down control, while auditory dominance originated from altered sensory processing in the auditory cortex. Moreover, the dynamic neural coupling between the sensory systems and both the fronto-sensorimotor and the DMN prioritized the flow of information in one particular sensory modality into the motor system, making this modality eventually win multisensory competition.

## References

Breveglieri R, Galletti C, Gamberini M, Passarelli L, Fattori P (2006) Somatosensory cells in area PEc of macaque posterior parietal cortex. J Neurosci 26:3679–3684. CrossRef Medline

Burock MA, Dale AM (2000) Estimation and detection of event-related fMRI signals with temporally correlated noise: a statistically efficient and unbiased approach. Hum Brain Mapp 11:249–260. CrossRef Medline

Chadick JZ, Gazzaley A (2011) Differential coupling of visual cortex with default or frontal-parietal network based on goals. Nat Neurosci 14:830–832. 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

Chen Q, Weidner R, Weiss PH, Marshall JC, Fink GR (2012) Neural interaction between spatial domain and spatial reference frame in parietal–occipital junction. J Cogn Neurosci 24:2223–2236. CrossRef Medline

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

Colavita FB, Weisberg D (1979) A further investigation of visual dominance. Percept Psychophys 25:345–347. CrossRef Medline

Colavita FB, Tomko R, Weisberg D (1976) Visual prepotency and eye orientation. Bull Psychon Soc 8:25–26. CrossRef

Desimone R (1998) Visual attention mediated by biased competition in extrastriate visual cortex. Philos Trans R Soc Lond B Biol Sci 353:1245–1255. CrossRef Medline

Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. Annu Rev Neurosci 18:193–222. CrossRef Medline

Duncan J, Humphreys G, Ward R (1997) Competitive brain activity in visual attention. Curr Opin Neurobiol 7:255–261. CrossRef Medline

Egeth HE, Sager LC (1977) On the locus of visual dominance. Percept Psychophys 22:77–86. CrossRef

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

Gamberini M, Galletti C, Bosco A, Breveglieri R, Fattori P (2011) Is the medial posterior parietal area V6a a single functional area? J Neurosci 31:5145–5157. CrossRef Medline

Goodale MA, Milner AD (1992) Separate visual pathways for perception and action. Trends Neurosci 15:20–25. CrossRef Medline

Goodale MA, Milner AD, Jakobson LS, Carey DP (1991) Perceiving the world and grasping it. A neurological dissociation. Nature 349:154–156. 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

Hartcher-O'Brien J, Alais D (2011) Temporal ventriloquism in a purely temporal context. J Exp Psychol Hum Percept Perform 37:1383–1395. CrossRef Medline

Kastner S, Ungerleider LG (2000) Mechanisms of visual attention in the human cortex. Annu Rev Neurosci 23:315–341. CrossRef Medline

Kastner S, De Weerd P, Desimone R, Ungerleider LG (1998) Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. Science 282:108–111. CrossRef Medline

Kastner S, Pinsk MA, De Weerd P, Desimone R, Ungerleider LG (1999) Increased activity in human visual cortex during directed attention in the absence of visual stimulation. Neuron 22:751–761. CrossRef Medline

Koppen C, Spence C (2007a) Seeing the light: exploring the Colavita visual dominance effect. Exp Brain Res 180:737–754. CrossRef Medline

Koppen C, Spence C (2007b) Spatial coincidence modulates the Colavita visual dominance effect. Neurosci Lett 417:107–111. 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

Langers DR, Van Dijk P, Backes WH (2005) Interactions between hemodynamic responses to scanner acoustic noise and auditory stimuli in functional magnetic resonance imaging. Magn Reson Med 53:49–60. CrossRef Medline

Laurienti PJ, Burdette JH, Wallace MT, Yen YF, Field AS, Stein BE (2002) Deactivation of sensory-specific cortex by cross-modal stimuli. J Cogn Neurosci 14:420–429. CrossRef Medline

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

Milner AD (2012) Is visual processing in the dorsal stream accessible to consciousness? Proc R Soc B Biol 279:2289–2298. CrossRef

Milner AD, Goodale MA (2008) Two visual systems re-viewed. Neuropsychologia 46:774–785. CrossRef Medline

Mozolic JL, Joyner D, Hugenschmidt CE, Peiffer AM, Kraft RA, Maldjian JA, Laurienti PJ (2008) Cross-modal deactivations during modality-specific selective attention. BMC Neurol 8:35. CrossRef Medline

O'Reilly RC, Herd SA, Pauli WM (2010) Computational models of cognitive control. Curr Opin Neurobiol 20:257–261. CrossRef Medline

Pashler H (1994) Dual-task interference in simple tasks: data and theory. Psychol Bull 116:220–244. CrossRef Medline

Pessoa L, Kastner S, Ungerleider LG (2003) Neuroimaging studies of attention: from modulation of sensory processing to top-down control. J Neurosci 23:3990–3998. Medline

Pitzalis S, Sereno MI, Committeri G, Fattori P, Galati G, Tosoni A, Galletti C (2013) The human homologue of macaque area V6A. Neuroimage 82:517–530. CrossRef Medline

Poline JB, Worsley KJ, Evans AC, Friston KJ (1997) Combining spatial extent and peak intensity to test for activations in functional imaging. Neuroimage 5:83–96. 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

Ravicz ME, Melcher JR, Kiang NY (2000) Acoustic noise during functional magnetic resonance imaging. J Acoust Soc Am 108:1683–1696. CrossRef Medline

Scarff CJ, Dort JC, Eggermont JJ, Goodyear BG (2004) The effect of MR scanner noise on auditory cortex activity using fMRI. Hum Brain Mapp 22:341–349. CrossRef Medline

Shmuelof L, Zohary E (2005) Dissociation between ventral and dorsal fMRI activation during object and action recognition. Neuron 47:457–470. CrossRef Medline

Smith MC (1967) Theories of the psychological refractory period. Psychol Bull 67:202–213. CrossRef Medline

Spence C, Parise C, Chen YC (2012) The Colavita visual dominance effect. In: The neural bases of multisensory processes (Murray MM, Wallace MT, eds). Boca Raton, FL: CRC.

Telford CW (1931) The refractory phase of voluntary and associative responses. J Exp Psychol 14:1–36. CrossRef

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