Frontiers | Frontiers in Neuroscience

The effects of attention in auditory–visual integration revealed by time-varying networks

Yuhao Jiang[1,2,3], Rui Qiao[1,2], Yupan Shi[1,2], Yi Tang[1,2], Zhengjun Hou[1,2]* and Yin Tian[1,2]*

[1]Institute for Advanced Sciences, Chongqing University of Posts and Telecommunications, Chongqing, China, [2]Guangyang Bay Laboratory, Chongqing Institute for Brain and Intelligence, Chongqing, China, [3]Central Nervous System Drug Key Laboratory of Sichuan Province, Luzhou, China

Attention and audiovisual integration are crucial subjects in the field of brain information processing. A large number of previous studies have sought to determine the relationship between them through specific experiments, but failed to reach a unified conclusion. The reported studies explored the relationship through the frameworks of early, late, and parallel integration, though network analysis has been employed sparingly. In this study, we employed time-varying network analysis, which offers a comprehensive and dynamic insight into cognitive processing, to explore the relationship between attention and auditory-visual integration. The combination of high spatial resolution functional magnetic resonance imaging (fMRI) and high temporal resolution electroencephalography (EEG) was used. Firstly, a generalized linear model (GLM) was employed to find the task-related fMRI activations, which was selected as regions of interesting (ROIs) for nodes of time-varying network. Then the electrical activity of the auditory-visual cortex was estimated via the normalized minimum norm estimation (MNE) source localization method. Finally, the time-varying network was constructed using the adaptive directed transfer function (ADTF) technology. Notably, Task-related fMRI activations were mainly observed in the bilateral temporoparietal junction (TPJ), superior temporal gyrus (STG), primary visual and auditory areas. And the time-varying network analysis revealed that V1/A1↔STG occurred before TPJ↔STG. Therefore, the results supported the theory that auditory-visual integration occurred before attention, aligning with the early integration framework.

1. Introduction

Individuals are constantly exposed to a plethora of sensory information that they unconsciously integrate in order to comprehend their environment. Visual and auditory information constitutes the majority (over 90%) of the information that is perceived (Treichler, 1967; Ristic and Capozzi, 2023). Auditory–visual integration occurs when auditory and visual stimuli coincide temporally and spatially, and when two stimuli are presented within a close time interval and similar spatial arrangement (Stein and Meredith, 1990; Frassinetti et al., 2002; Stevenson et al., 2012; Spence, 2013; Tang et al., 2016; Ľuboš et al., 2021). Attention plays a crucial role in selectively processing external information and improving information processing performance through focusing on target locations (Posner and Rothbart, 2006; Zhang T. et al., 2022). Attention is instrumental in processing dynamic stimuli efficiently and enhancing

perception, as it directs limited cognitive resources toward information relevant to the current task (Tian et al., 2014; Li et al., 2015). In addition, the researches on the attention mechanism may help to improve deep neural networks for visual processing tasks (Zhang et al., 2019; Wang et al., 2020).

There is ongoing debate regarding the role of attention in multisensory integration, particularly in the case of auditory–visual integration. Three mainstream theories about the relationship between auditory–visual integration and attention were proposed in previous studies (Koelewijn et al., 2010; Xu et al., 2020). The first, the early integration framework, asserts that integration occurs prior to attention and can even drive it (Vroomen et al., 2001; Rachel et al., 2022). Evidence for this is seen in the "pip-pop effect," where the addition of auditory stimulation to a visual search task led to faster results (Erik et al., 2008). Then non-spatial auditory stimulation was added to the spatial visual experiment. The second theory, the late integration framework, demonstrates that multisensory integration appears behind attention. In other words, two unimodal (i.e., auditory and visual) events are attended to separately before they are integrated. This model indicates that attention is necessary for multisensory integration (Laura et al., 2005; Sébastien et al., 2022). A later study used a cross-modal attention preference task to prove that cross-modal interactions are influenced by attention (Romei et al., 2013; Wen et al., 2021). Furthermore, late integration suggests that late cross-modal effects are mediated by attentional mechanisms. The third theory is the parallel integration framework; here, the stage at which multisensory integration takes place is uncertain. Multisensory integration can be early or late, and it depends on experimental or external conditions (Calvert and Thesen, 2004; Sébastien et al., 2022). Some studies extended the seminal methods of the parallel integration framework (Talsma et al., 2010; Stoep et al., 2015). This may produce different results as a result of several factors, including task type (detection or identification), stimulus properties (simple or complex), and attention resources (exogenous or endogenous).

In the study of the relationship between attention and auditory–visual integration, various methods have been employed. Early research utilized behavioral data and discovered that an auditory stimulus influences the reaction time (RT) of a synchronous or nearly synchronous visual stimulus (Mcdonald et al., 2000; Shams et al., 2000; Laura et al., 2005; Zhang X. et al., 2022) and the reverse is also true (Platt and Warren, 1972; Bertelson, 1999). These results indicate that a simultaneous or near-simultaneous bimodal stimulus reduces stimulation uncertainty (Calvert et al., 2000), potentially supporting the early integration framework or enhancing stimulation response for the late framework (Stein et al., 1989; Zhang et al., 2021). However, external factors, such as the state of the experimental subjects, may be overlooked.

With the advancement of brain imaging technology, increasing numbers of researchers have turned to brain imaging to investigate the relationship between attention and auditory–visual integration. By utilizing an event-related potential component (ERP) of an auditory–visual streaming design and a rapid serial visual presentation paradigm, they explored the interactions between multisensory integration and attention (Durk and Woldorff, 2005; Kang-jia and Xu, 2022). The results indicated that activity associated with multisensory integration processes is heightened when they are attended to, suggesting that attention plays a critical role in auditory–visual integration and aligning with the late integration criteria. The

improvement of the spatial resolution of scalp EEG has long been a subject of interest for researchers.

Studies using functional magnetic resonance imaging (fMRI) with high spatial resolution have reported the accurate location of many areas involved in auditory–visual integration and attention; these mainly include the prefrontal, parietal, and temporal cortices (Calvert et al., 2001; Macaluso et al., 2004; Tedersäläjärvi et al., 2005; Noesselt et al., 2007; Cappe et al., 2010; Chen et al., 2015). The superior temporal gyrus (STG) and sulcus (STS) both participate in speech auditory–visual integration (Klemen and Chambers, 2012; Rupp et al., 2022) and non-speech auditory–visual stimuli (Yan et al., 2015). In the past, STG was considered an area of pure sound input (Mesgarani et al., 2014). The temporoparietal junction (TPJ), which is close to the STG, is an important area of the ventral attention network (VAN) that is located mostly in the right hemisphere, and is recruited at the moment a behaviorally relevant stimulus is detected (Corbetta et al., 2008; Tian et al., 2014; Branden et al., 2022). The TPJ is activated during detection of salient stimuli in a sensory environment for a visual (Corbetta et al., 2002, 2008), auditory (Alho et al., 2015), and auditory–visual task (Mastroberardino et al., 2015). However, as many studies have mentioned, it is difficult to determine accurately the timing characteristics when using fMRI with poor temporal resolution.

For the reason that EEG and fMRI are two prominent noninvasive functional neuroimaging modalities, and they demonstrate highly complementary attributes, there has been a considerable drive toward integrating these modalities in a multimodal manner (Abreu et al., 2018). The combination of scalp EEG's exceptional temporal resolution and fMRI's remarkable spatial resolution enables a more comprehensive exploration of brain activity, surpassing the limitations inherent to individual techniques (Bullock et al., 2021). Previous investigations have examined the functional aspects of the brain in various pathological conditions, such as schizophrenia (Baenninger et al., 2016; Ford et al., 2016). Multiple researchers have employed combination of EEG and fMRI to explore cognitive mechanisms (Jorge et al., 2014; Shams et al., 2015; Wang et al., 2018). Some other studies have investigated brain dynamics in relation to complex cognitive processes like decision-making and the onset of sleep (Bagshaw et al., 2017; Pisauro et al., 2017; Hsiao et al., 2018; Muraskin et al., 2018). In this study, we used these two neuroimaging technologies to investigate the appearance order of auditory–visual integration and attention. Previous studies have tended to apply a specific experimental paradigm to investigate this relationship, but few have used network analyses to resolve this conundrum. We employed time-varying network analysis based on the adaptive directed transfer function (ADTF) method to uncover dynamic information processing. This method can uncover the dynamic information processing with a multivariate adaptive autoregressive mode (Li et al., 2016; Tian et al., 2018b; Nazir et al., 2020). This approach may offer new insights into the temporal order of multisensory integration and attention in a stimulated EEG network.

2. Materials and methods

2.1. Participants

The data for this study was obtained through separate EEG and fMRI recordings, conducted on 15 right-handed, healthy adult males

(mean ± standard deviation (SD) = 21.4 ± 2.8 years). Participants provided informed consent and were free from visual or auditory impairments and any mental health conditions. Upon completion of the experiments, participants were compensated for their time. The study was approved by the Ethics Committee of the University of Electronic Science and Technology of China.

2.2. Experimental design

Throughout the experiment, a white fixation cross of dimensions (0.5° × 0.5°) was presented at the center of a black monitor. The visual stimuli consisted of rectangular boxes that randomly appeared in either the left or right visual field (LVF or RVF, respectively). The box was 2° × 2° and its width was 0.2ϒ. The boxes remained on the screen for 50 ms and were followed by an auditory stimulus, a 1,000 Hz pure tone that also randomly appeared in the left or right auditory field (LAF and RAF, respectively) after a 50 or 750 ms interval. Participants were instructed to respond by pressing the 'Z' key with their left hand if the tone appeared in the LAF, and the '/' key with their right hand if it appeared in the RAF. Participants were required to react as soon as they heard the pure tone, which lasted for 200 ms. The fixation cross remained on the monitor for an additional 800 ms to ensure participants had sufficient time to respond correctly. The experimental procedure is illustrated in Figure 1.

2.3. Behavioral data and analysis

The behavioral data was obtained via EEG and fMRI. We analyzed RT using repeated measures analysis of variance (ANOVA) with the following factors: stimulus visual field (LVF vs. RVF), cue validity (valid vs. invalid), stimulus-onset asynchrony (SOA), and the interval between the cue and target stimulus (long vs. short). Data consistency was ensured by excluding RTs greater than 900 ms and less than 200 ms, as well as any instances of missed or incorrect key presses.

2.4. EEG and fMRI data recording

In the study, EEG and fMRI data were collected separately. We used a Geodesic Sensor Net (GSN) with 129-scalp electrodes located according to the International 10–20 system (Tucker, 1993) to record the EEG at a rate of 250 Hz. The Oz, Pz, CPz, Cz, FCz, and Fz electrodes were placed in the middle of the skull, and the remaining electrodes were distributed along both sides of the midline. The central top electrode (Cz) was used as the reference electrode and all electrodes had impedances lower than 40 kΩ (Tucker, 1993).

fMRI data was collected using the fast T2*-weighted gradient echo EPI sequence on a 3-T GE MRI scanner (TR = 2000 ms, TE = 30 ms, FOV = 24 cm × 24 cm, flip angle = 90°, matrix = 64 × 64, 30 slices) at the University of Electronic Science and Technology of

FIGURE 1
Illustration of stimulus sequence in the experiment.
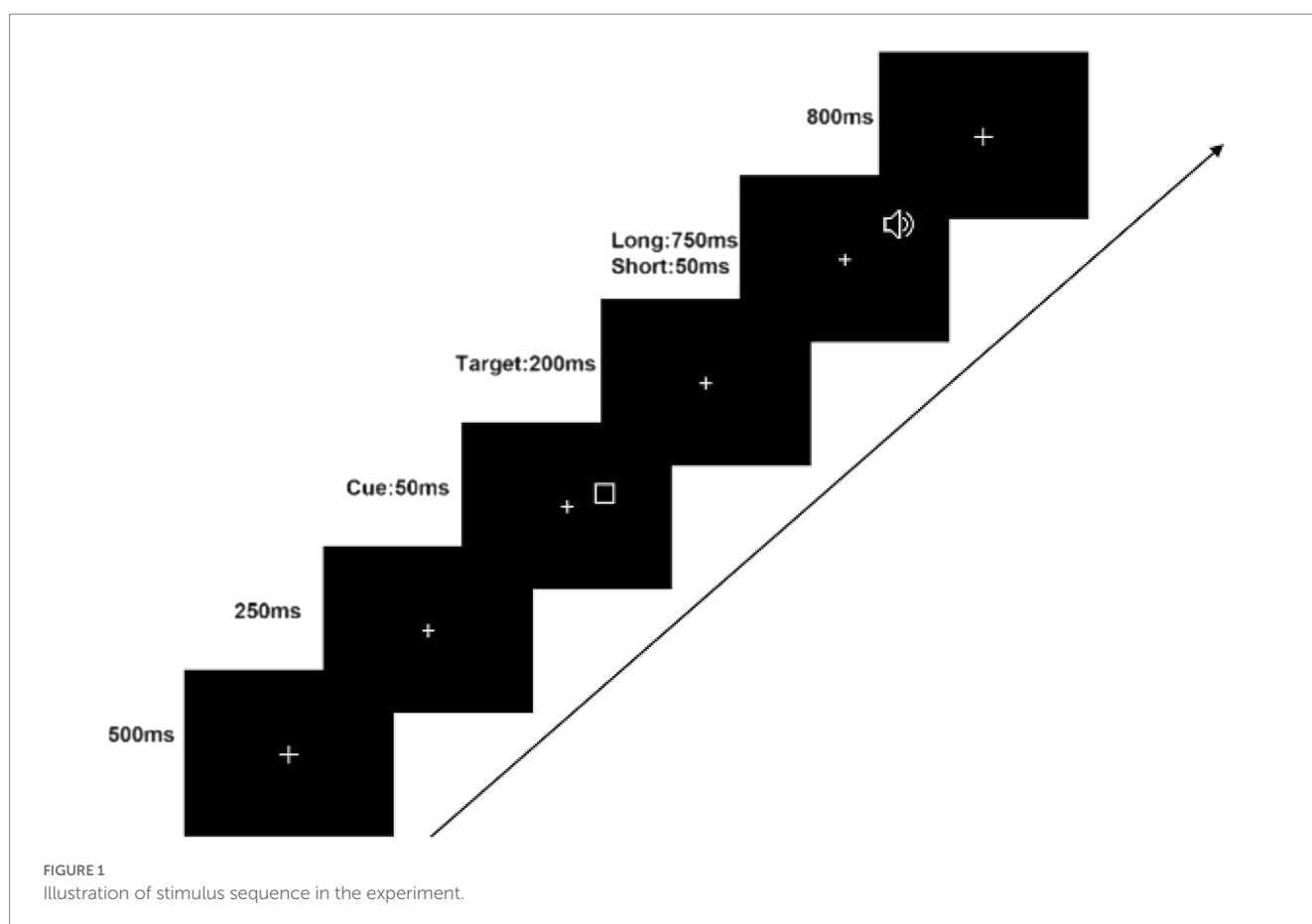
China. This method obtained 198 volumes for each session. Because the machine at was unstable at the beginning of the data collection, we discarded the first five image volumes of each run for preprocessing.

2.5. The processing framework for time-varying networks

The processing framework for calculating time-varying networks consisted of three stages, as illustrated in Figure 2.

1. ROI selection based on task-related fMRI activations, as shown in Figure 2A.

The fMRI data was preprocessed and constructed by a generalized linear model (GLM). The results of the GLM were then subjected to a statistical test. Reply on the statistical results, 4 activations for the left cue and 4 activations for the right cue in the fMRI experiment were selected as ROIs (nodes) in the cerebral cortex, providing relatively accurate MNI coordinates for the construction of the time-varying network in the following steps.

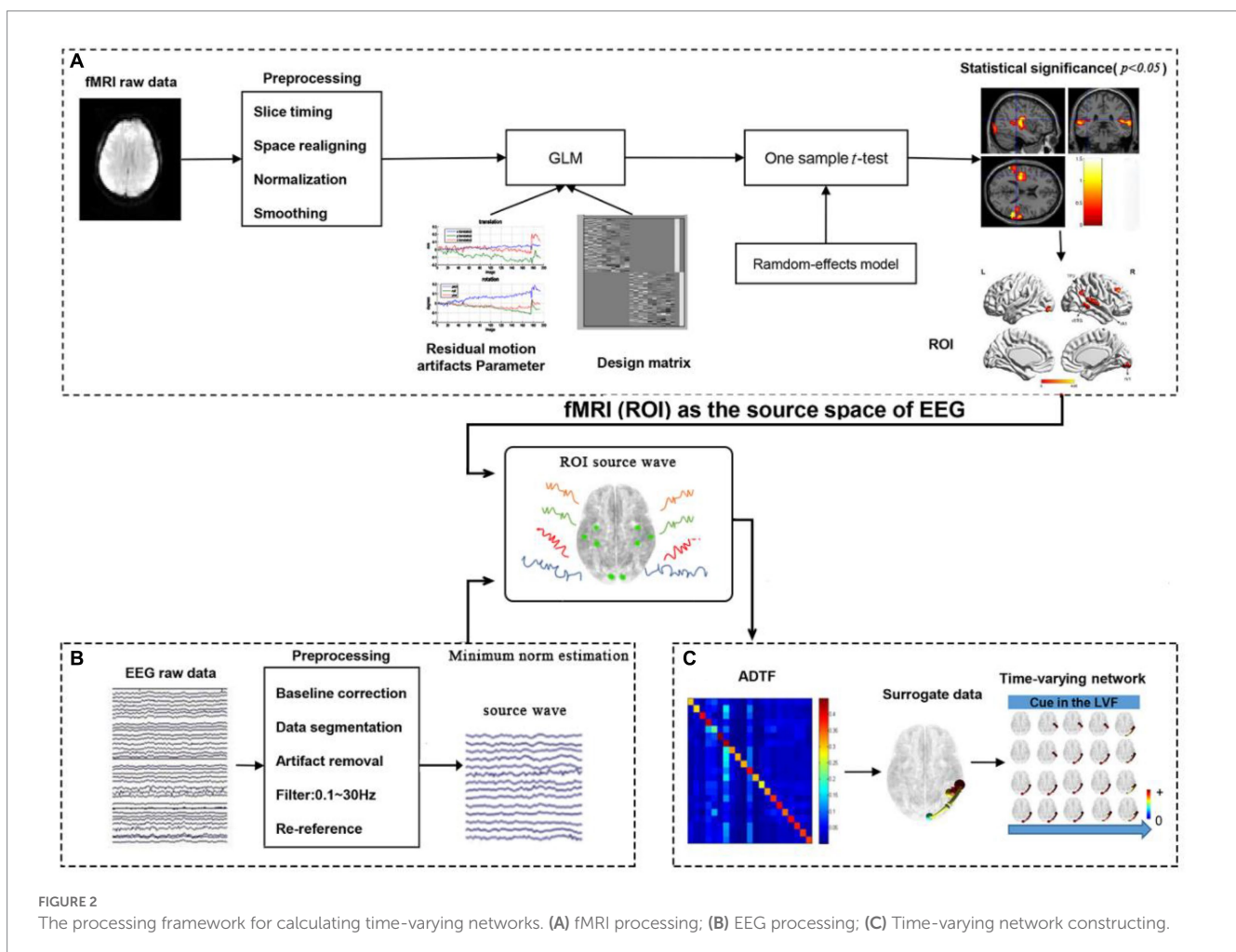2. Source wave extraction (Figure 2B).

The EEG data was preprocessed, and the scalp electrical signals are mapped to the cerebral cortex by MNE source localization method. Then, the MNI coordinates provided by fMRI were converted to the corresponding positions of the head model and the corresponding time series of the cortical electrical signals are extracted.

3. Time-varying network construction (Figure 2C).

In the third stage, the time-varying network was constructed using the ADTF technology, based on the results from steps 1 and 2.

2.6. fMRI data processing

The remaining volumes underwent preprocessing using Statistical Parametric Mapping version 8 (SPM8) software. Four preprocessing pipelines were applied in this study. Firstly, slice timing correction was implemented to address temporal differences among the slices. Secondly, spatial realignment was performed to eliminate head movement, whereby all volumes were aligned with the first volume. Participants whose head movement exceeded 2 mm or 2 degrees were excluded (Bonte et al., 2014). Thirdly, normalization was carried out to standardize each participant's original fMRI image to the standard Montreal Neurological Institute (MNI) space using EPI templates.

FIGURE 2
The processing framework for calculating time-varying networks. **(A)** fMRI processing; **(B)** EEG processing; **(C)** Time-varying network constructing.

Voxel resampling to $3 \times 3 \times 3 \, mm^3$ was performed to overcome head size inconsistencies. Lastly, spatial smoothing was implemented to ensure high signal-to-noise ratio (SNR) by smoothing the functional images with a Gaussian kernel of full width half maximum (FWHM) of $6 \times 6 \times 6 \, mm^3$.

After data preprocessing, the time series of all voxels underwent a high-pass filter at 1/128 Hz and were then analyzed with a general linear model (GLM; Friston et al., 1995) using SPM8 software. Temporal autocorrelation was modeled using a first-order autoregressive process. At the individual level, a multiple regression design matrix was constructed using the GLM, that included two experimental events based on the cue location (left visual field or right visual field). The two events were time-locked to the target of each trial by a canonical synthetic hemodynamic response function (HRF) and its temporal and dispersion derivatives. By including dispersion derivatives, the analysis accounted for variations in the duration of neural processes induced by the cue location. Nuisance covariates, such as realignment parameters, were included to account for residual motion artifacts. Parameter estimates were obtained for each voxel using weighted least-squares, which provided maximum likelihood estimators based on the temporal autocorrelation of the data (Wang et al., 2013).

In this study, to compute simple main effects for each participant, baseline contrasts were applied to the experimental conditions. Subsequently, the resulting individual contrast images were entered into a second-level one sample *t*-test using a random-effects model. In order to identify areas of significant activation, a threshold of $p < 0.05$ (false discovery rate [FDR] corrected) and a minimum cluster size of 10 voxels were utilized. These stringent criteria were employed to ensure robust and reliable identification of neural activation patterns.

2.7. EEG data preprocessing

The EEG data underwent five preprocessing steps. Firstly, the EEG epochs were set to a time range of −200 to 1,000 ms. Secondly, we used the average of 200 ms pre-stimulus data as a baseline to correct the epochs. Thirdly, we performed artifact rejection, excluding epochs contaminated by eye blinks, eye movements, amplifier clipping or muscle potentials that exceeded ±75 μv. Fourthly, we filtered the EEG recordings using a band-pass filter of 0.1-30 Hz. Finally, we re-referenced the data using the reference electrode standardization technique (REST) (Yao et al., 2005; Tian and Yao, 2013; Tian et al., 2018a). We excluded trials with incorrect behavioral responses and bad channel replacements, and averaged the ERPs from the stimulus onset time point based on the validity of the cue, visual field, and SOA length.

2.8. Minimum norm estimation

The volume conductor effect may lead to the generation of pseudo-connections during brain network construction using scalp brain electricity. And invasive methods for directly collecting brain electricity in the cerebral cortex are challenging to use. To overcome this problem, we employed source localization technology to transfer scalp brain electrical signal to the cortex, enabling estimation of cortical electrical signals (Tian et al., 2018a; Tian and Ma, 2020), and

we converted 129 scalp electrodes into 19 electrodes covering the whole brain.

In this study, we used the normalized minimum norm estimation (MNE) source localization method to estimate the electrical activity of the auditory–visual cortex. Compared to other methods, the normalized MNE offers higher dipole positioning accuracy, especially in depth source analysis. Our head model consisted of a three-layer realistic representation of the cortex, skull, and scalp. The formula for MNE calculation is expressed as follows:

$$\varphi(t) = \omega x(t) \tag{1}$$

Where $x(t)$ is the EEG collected by the scalp, $\varphi(t)$ is the corresponding cortical EEG, and ω is the field matrix, which can be obtained from the following formula:

$$\omega = C_s A^T \left(A C_s A^T + \mu^2 C_n \right)^{-1} \tag{2}$$

where C_s is the signal covariance, C_n is the noisy covariance, and A is the transfer matrix obtained by the boundary element theory. μ is a regularization parameter and is obtained by the following formula:

$$\mu = \frac{trace\left(A C_s A^T \right)}{trace\left(C_n \right) \times snr^2} \tag{3}$$

where *snr* is the signal to noise ratio.

2.9. Cortical time-varying network

The MNE source localization method was employed to transfer scalp electrical signals to the cerebral cortex. Next, MNI coordinates obtained from fMRI were mapped to the corresponding positions on the head model, and the cortical electrical signal time series at these positions were extracted. Subsequently, we designated these positions as nodes of the network and constructed the time-varying network using the relationship between these time series as the network edges.

To calculate the ADTF, we computed the multivariate adaptive autoregressive (MVAAR) model for all conditions. The model was normalized and expressed by following equation:

$$X(t) = \sum_{j}^{p} \omega(j,t) X(t-j) + \eta(t) \tag{4}$$

where $X(t)$ is the EEG data vector over the entire time window, $\omega(j,t)$ is the coefficient matrix of the time-varying model, which can be calculated by the Kalman filter algorithm, and $\eta(t)$ represents the multivariate independent white noise. The symbol p denotes the MVAAR model order selected by Schwarz Bayesian Criterion (Schwarz, 1978; Wilke et al., 2008; Tian et al., 2018b).

After obtaining the coefficients of the MVAAR model, we calculated the ADTF by applying Equation (5) to convert the model coefficient $\omega(j,t)$ to the frequency domain. The H_{ij} element of

$H(f,t)$ describes the directional information flow between the jth and the ith element at each time point t as:

$$\omega(f,t) * X(f,t) = \eta(f,t) \tag{5}$$

$$X(f,t) = \omega^{-1}(f,t) * \eta(f,t) = H(f,t) * \eta(f,t) \tag{6}$$

where $\omega(f,t) = \sum_{k=0}^{p} \omega_k(t) e^{-j2\pi f \cdot tk} \omega_k$ is the matrix of the time-varying coefficients. $\acute{E}(f,t)$ and $\eta(f,t)$ are transformed into the frequency domain as $X(t)$ and $\eta(f,t)$, respectively.

Defining the directed causal interrelation from the jth to the ith element, the normalized ADTF is described between (0,1) as follows:

$${}^{1}2_{ij}(f,t) = \frac{\left|H_{ij}(f,t)\right|^2}{\sum_{k}^{n}\left|H_{ik}(f,t)\right|^2} \tag{7}$$

To obtain total information flow from a single node, the integrated ADTF is calculated as the ratio of summed ADTF values divided by the interested frequency bands (f_1, f_2):

$$\upsilon_{ij}^2(t) = \frac{\sum_{f1}^{f2} \acute{e}_{ij}^2(k,t)}{f2 - f1} \tag{8}$$

Surrogate data were used to establish the empirical ADTF value distribution under the connectionless zero assumption since the ADTF function has a highly non-linear correlation with the time series it derives, making it impossible to determine the distribution of the ADTF estimator under zero assumption without causality. The shuffling procedure independently and randomly iterated Fourier coefficient phases to produce new surrogate data while preserving the spectral structure of the time series (Wilke et al., 2008). To establish a statistical network, the nonparametric signed rank test was used to select statistically significant edges. The shuffling procedure was repeated 200 times for each model-derived time series from each participant to obtain the significance threshold of $p < 0.05$ with Bonferroni correction (Tian et al., 2018b).

2.10. Correlation analysis

The relationship between the information flow and the corresponding average response time (RT) was calculated using Pearson correlation based on the results of time-varying network analysis.

3. Results

3.1. Behavioral data analysis

Significant effects were observed for SOA ($F[1,14] = 9.85$, $p < 0.01$) and validity ($F[1,14] = 8.74$, $p < 0.05$), as well as their interaction ($F[1,14] = 27.54$, $p < 0.001$). However, no significant visual field effect

($F[1,14] = 3.60$, $p > 0.05$) or interactions between visual field and SOA or validity were found.

Because SOA, validity, and their interaction were significant, we conducted paired t-tests for the effects of SOA and validity (Figure 3). The results showed that participants reacted significantly faster in long SOA-invalid trials (268.94 ± 19.33 ms) than in long SOA-valid trials (277.79 ± 17.91 ms). In short SOA-invalid trials (291.91 ± 20.76 ms), participants took significantly more time to react than in short SOA-valid trials (273.80 ± 20.87 ms). There were also significant differences between long and short SOA-invalid trials. Although the RTs of long SOA-valid trials were slower than those of short SOA-valid trials, the difference was not significant.

3.2. fMRI results

A single sample t-test was performed to analyze fMRI data, revealing areas related to visual (V1), auditory (A1), multisensory integration (STG), and attention (angular, middle frontal cortex [MFG]) in both the left and right visual field (LVF and RVF). In the LVF, the main activated areas ($p < 0.05$, FDR correction) included the right angular gyrus (BA39), which is part of the right temporoparietal junction (rTPJ), right STG (BA21), right Heschl's gyrus as A1 (BA48), right lingual gyrus as rV1 (BA18), and right MFG (BA46), as shown in Figure 4A. In the RVF, more activated areas ($p < 0.05$, FDR correction) included left STG (BA21), left A1 (BA48), left V1 (BA17), bilateral MFG (BAs44/45/46), right TPJ (BA39), and so on, as shown in Figure 4B.

We selected four ROIs based on the task-related fMRI activations depicted in Figure 4 for both LVF and RVF cues. Specifically, when the cue appeared in either the LVF or RVF, the right TPJ (rTPJ) was selected for both LVF and RVF cues. When the cue appeared in the LVF, we chose the right STG (rSTG), right A1 (rA1), and right V1 (rV1). When the cue appeared in the RVF, we chose the contralateral STG, A1, and V1. The coordinates and sizes of the ROIs are presented in Table 1.
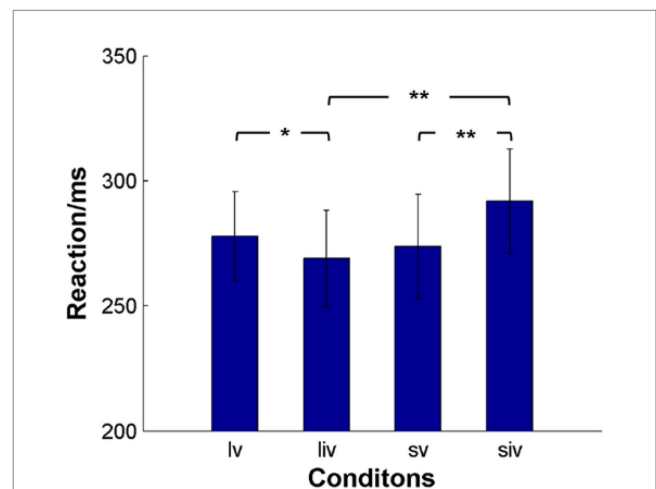

FIGURE 3
The average response time (RT) of subjects for the four conditions. lv denotes long SOA-valid condition, liv denotes long SOA-invalid condition, sv denotes short SOA-valid condition, siv denotes short SOA-valid condition. **$p < 0.001$, *$p < 0.05$.
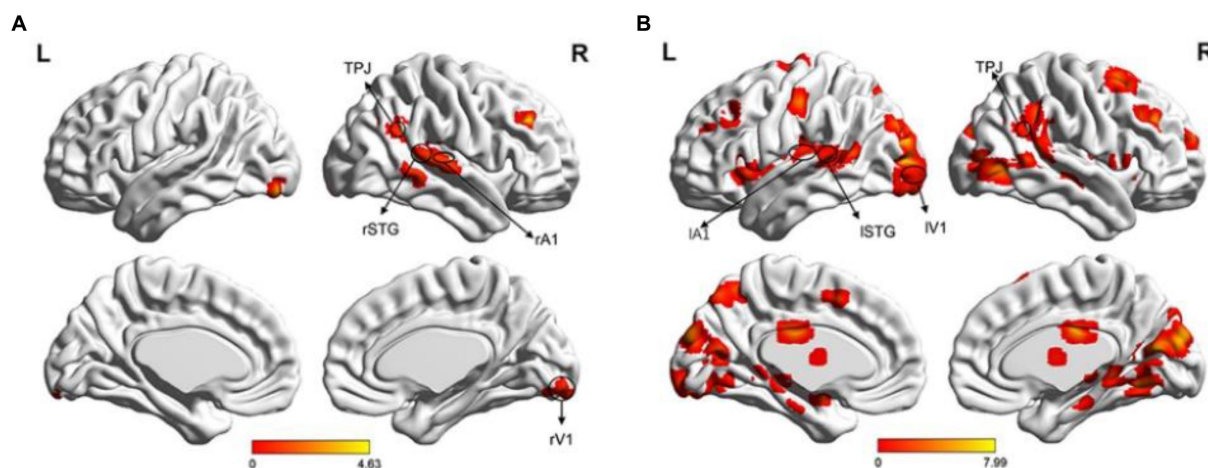
FIGURE 4
Brain activation maps. **(A)** Illustration of the activation map when the cue stimulus appeared in the LVF. The main activated areas ($p < 0.05$, FDR) included the right TPJ, right STG, right A1, right V1, and right MFG; **(B)** Illustration of the activation map when the cue stimulus appeared in the RVF ($p < 0.05$, FDR correction). The activated areas included the left STG, left A1, left V1, bilateral MFG, right TPJ, and right MT.

TABLE 1 The four selected regions of interest (ROIs) in each visual field.

ROI	MNI coordinates (mm)			The size of ROI (mm)
	x	y	z	
Cue in the LVF				
rTPJ	45	−54	30	6
rSTG	63	−39	18	6
rA1	45	−21	12	10
rV1	9	96	−3	10
Cue in the RVF				
rTPJ	45	−54	33	6
lSTG	−64	−46	18	6
lA1	−40	−26	14	10
lV1	−9	−96	−9	10

3.3. Time-varying network

We computed the time-varying network at time points ranging from 200 ms to 900 ms and displayed the connection time points only when it changed in the four conditions. When the cue appeared in the LVF or RVF, the changes in cue conditions were illustrated in Figure 5A and Figure 5B, respectively. Figure 5C summarizes the results of the time-varying network analysis. The first step for the long SOA condition was A1↔STG, whereas for the short SOA condition, it was V1↔STG. The last step for both long and short SOA conditions was V1↔STG and STG↔TPJ. Notably, in the long SOA-valid condition, V1↔STG was the middle step.

3.4. Correlation analysis

Our analysis revealed significant correlations between reaction and information flow (such as STG → TPJ and TPJ → STG) for all conditions, as shown in Figure 6. For the long SOA, distinct differences were observed for each condition. Negative correlations were evident

when the cue was invalid and appeared in the LVF (STG → TPJ: $r = −0.54$, $p < 0.05$; TPJ → STG: $r = −0.52$, $p < 0.05$) or RVF (STG → TPJ: $r = −0.51$, $p < 0.05$; TPJ → STG: $r = −0.58$, $p < 0.05$). Conversely, positive correlations were evident when the cue was valid and appeared in either the LVF (STG → TPJ: r = 0.65, $p < 0.01$; TPJ → STG: r = 0.52, $p < 0.05$) or RVF (STG → TPJ: r = 0.53, $p < 0.05$; TPJ → STG: r = 0.57, $p < 0.05$). Similar trends were noted for all conditions for the short SOA, as shown in Figure 6. Positive correlations between mean RT and information flow were observed when the cue was invalid and appeared in the LVF (STG → TPJ: $r = 0.59$, $p < 0.05$; TPJ → STG: $r = 0.56$, $p < 0.05$) or RVF (STG → TPJ: $r = 0.59$, $p < 0.05$; TPJ → STG: $r = 0.56$, $p < 0.05$). Similarly, positive correlations were observed when the cue was valid, regardless of whether it appeared in the LVF (STG → TPJ: r = 0.72, $p < 0.005$ TPJ → STG: r = 0.55, $p < 0.05$) or RVF (STG → TPJ: r = 0.52, $p < 0.05$; TPJ → STG: r = 0.53, $p < 0.05$)

4. Discussions

In this study, the behavioral results showed that the RT for a valid cue was significantly shorter than an invalid cue in the short SOA condition, while the opposite was opposite for the long SOA, which was similar to the unimodal task. In both long and short SOA conditions, we observed STG activation, a critical auditory–visual integration region (Klemen and Chambers, 2012). Additionally, we observed activation in TPJ and MFG, which are important VAN areas (Corbetta et al., 2008), indicating that attention plays a role in auditory–visual integration. Our time-varying network analysis revealed that V1/A1↔STG occurred before TPJ↔STG, as shown in Figure 5, indicating that pre-attention in auditory–visual integration.

4.1. Similar results observed between the bimodal and unimodal cue-target paradigms

Previous researches have reported that there is a significant cue effect for short SOAs in the visual cue-target paradigm. On the
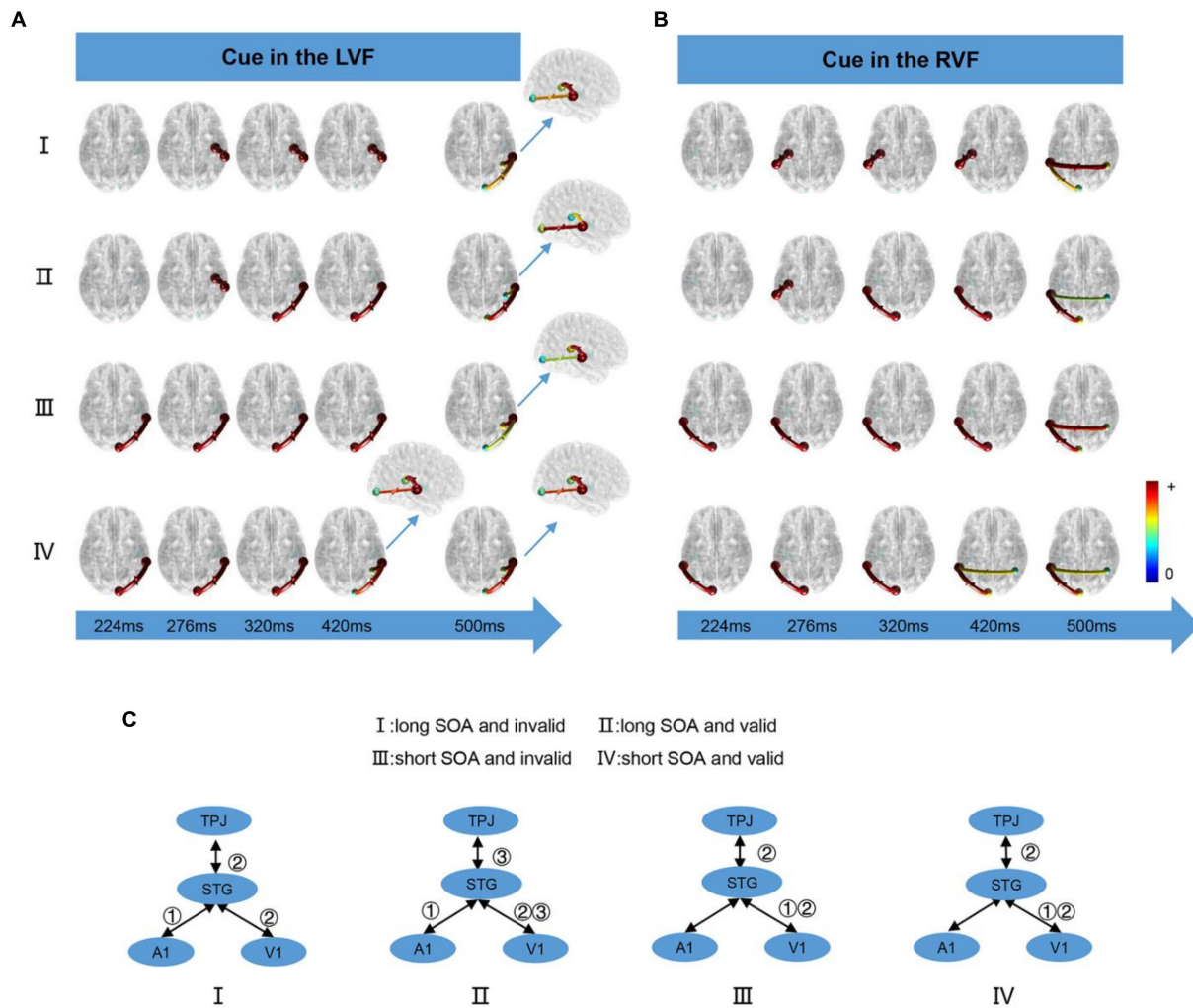
FIGURE 5
The time-varying networks when the cue stimulus appeared in the **(A)** LVF and **(B)** RVF. **(C)** Summary of time-varying networks. ①②③ denote the order in which the connections appear.

condition of the time interval of the cue and target stimulus is shorter than 300 ms, the subjects exhibited faster responses when the cue was valid as compared to when it was invalid. However, the subjects showed slower responses when the cue was valid rather than invalid for long SOA (more than 300 ms). These findings were consistent with previous studies (Lepsien and Pollmann, 2002; Mayer et al., 2004a,b; Tian and Yao, 2008; Tian et al., 2011) and suggested that stimulus-driven attention effects are faster and more transient than goal-directed attention effects (Jonides and Irwin, 1981; Shepherd and Müller, 1989; Corbetta et al., 2002; Busse et al., 2008; Macaluso et al., 2016; Tang et al., 2016). Similar outcomes have been observed in the auditory paradigm (Alho et al., 2015; Hanlon et al., 2017). Our behavioral analysis aligns with previous research on the unimodal paradigm and suggests that there is no difference between unimodal and bimodal paradigms in the cue-target paradigm.

4.2. Integration and attention exist in the bimodal cue-target paradigm

Previous studies have emphasized that auditory–visual integration in the cue-target paradigm occurs when the cue with one modal

stimulus and the target with a different modal stimulus are presented from around the same spatial position (Stein and Meredith, 1990; Spence, 2013; Wu et al., 2020) and at approximately the same time (Stein and Meredith, 1990; Frassinetti et al., 2002; Bolognini et al., 2005; Spence and Santangelo, 2010; Stevenson et al., 2012; Tang et al., 2016). However, it will not appear if the cue precedes the target by more than 300 ms (Spence, 2010). In our paradigm, the time intervals between cue and target stimulus were divided into 100 and 800 ms, which cannot be directly compared to previous studies. Our fMRI results, where the STG appeared in all conditions, indicate that auditory and visual integration occurs even when these two stimuli are not aligned in space or time (i.e., more than 300 ms interval).

The role of attention in multisensory integration is still under debate. Some studies proposed that multisensory integration is an automatic process (Vroomen et al., 2001; José et al., 2020), while others suggested that attention played an important role in multisensory integration (Talsma et al., 2007, 2010; Fairhall and Macaluso, 2009; Tang et al., 2016). Since rTPJ and MFG important parts of the ventral attention network (Corbetta et al., 2008; Klein et al., 2021), the experimental activation of these rTPJ and MFG suggested that attention may also be involved in integration. We used time-varying networks to determine the temporal order between
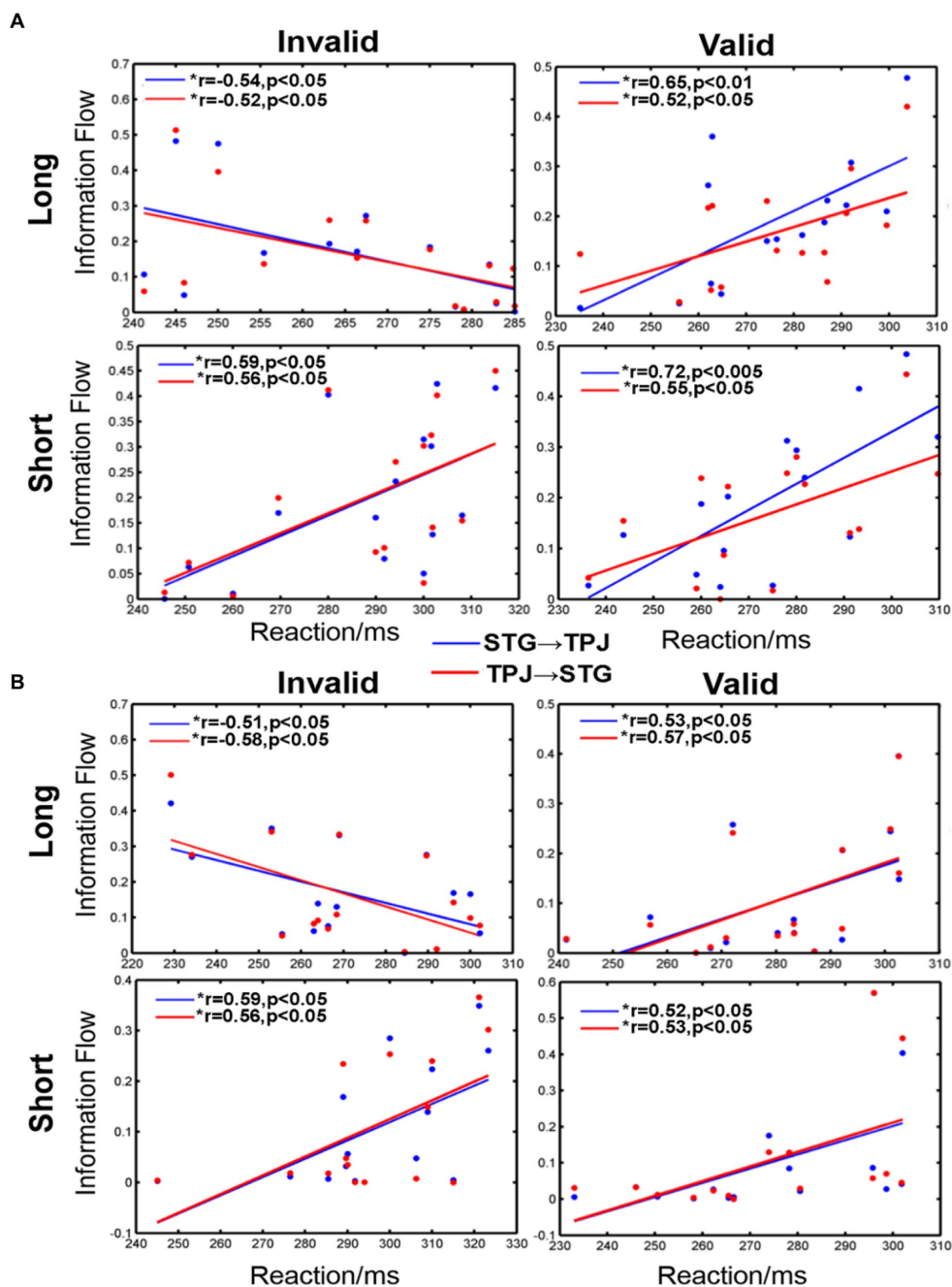
FIGURE 6
The correlation between information flow and average reaction time (RT) when the cue stimulus appeared in the **(A)** LVF and **(B)** RVF. *a significant difference between the two regions. STG → TPJ denotes the causal flow from STG to TPJ. TPJ → STG denotes the causal flow from TPJ to STG.

multisensory integration and attention using combination of fMRI and EEG data, which allowed for greater precision than EEG data alone. Additionally, the fMRI data provided a more precise spatial resolution for the time-varying networks.

4.3. Auditory–visual integration prior to attention

In this research, we constructed a time-varying network using task-related fMRI activations as nodes, including TPJ as the core of the VAN (Corbetta et al., 2008), and STG as an important area for

integration (Yan et al., 2015). Our aim was to investigate the relationship between multisensory integration and attention. As depicted in Figure 5C, the V1/A1↔STG connection was always the first order, followed by STG↔TPJ, regardless of the conditions. This finding supports the notion that pre-attention is involved in auditory and visual integration, which is consistent with previous studies (Erik et al., 2008). However, we observed some differences under different conditions, such as the SOA length. For short SOA, the first connection was V1↔STG, as visual stimuli are dominant in processing spatial characteristics, while auditory events dominate temporal characteristic processing (Bertelson et al., 2000; Stekelenburg et al., 2004; Bonath et al., 2007; Navarra et al., 2010). Conversely, for long SOA,

information flowed from the A1 to STG. This result might be due to the auditory–visual stimuli being temporally unsynchronized in our data collection. As the SOA increased, the dominant role of the visual stimulus diminished, and the auditory effect became stronger, leading to a significant A1↔STG flow. Interestingly, TPJ↔STG and STG↔V1 were the last step in all conditions, indicating that the TPJ modulates the primary cortex by using integration areas as a transfer node in all cases.

4.4. Relationship between information flow and RT

Numerous studies have investigated how attention affects a subject's reaction time, but there is disagreement on whether attention boosts or limits the reflection (Senkowski et al., 2005; Karns and Knight, 2009; Macaluso et al., 2016). Some studies have suggested that attention accelerates reaction speed (Mcdonald et al., 2005, 2009; Van der Stoep et al., 2017), while others have proposed that attention may actually inhibit reaction (Tian and Yao, 2008). A recent study has demonstrated that both stimulus-driven attention and multisensory integration can accelerate responses (Van der Stoep et al., 2017; Motomura and Amimoto, 2022).

In this study, we compared the correlation between mean RT and the information flow of STG↔TPJ under different circumstances. Our findings suggest that attention has a direct influence on multisensory integration, as the extent of information flow reflects the mutual influence of the two brain regions. Specifically, we observed a negative correlation between the two regions in the long SOA-invalid condition, indicating that larger information flow led to faster reflection times. We inferred that this phenomenon is due to bottom-up attention, where increased information flow leads to greater information exchange between the STG and TPJ and, thus, faster reactions. However, in other conditions, we observed positive correlations, which we attribute to the modulation of attention. Specifically, greater attention modulation results in inhibited reactions.

5. Conclusion

In this paper, our analysis of the behavioral data showed no discernible difference between the multisensory and unisensory cue-target paradigms. We also employed fMRI data analysis to demonstrate the existence of auditory–visual integration in the long SOA condition and the necessity of attention for such integration. The constructed time-varying networks based on fMRI coordinates revealed that multisensory integration occurs prior to attention and pre-attention is involved in auditory–visual integration. Furthermore, our findings suggest that attention can impact the subject's reaction time, but the effect depends on the situation, and greater attention modulation results in inhibited reactions.

References

Abreu, R., Leal, A., and Figueiredo, P. (2018). EEG-informed fMRI: a review of data analysis methods. *Front. Hum. Neurosci.* 12:29. doi: 10.3389/fnhum.2018.00029

Alho, K., Salmi, J., Koistinen, S., Salonen, O., and Rinne, T. (2015). Top-down controlled and bottom-up triggered orienting of auditory attention to pitch activate

Data availability statement

Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee of the University of Electronic Science and Technology of China. The patients/participants provided their written informed consent to participate in this study.

Author contributions

YJ performed experiments and data analysis. RQ and YS contributed to data analysis. YTa wrote the draft of the manuscript. ZH and YTi contributed to experiments design and conception. All the authors edited the manuscript. All authors contributed to the article and approved the submitted version.

Funding

Conflict of interest

Publisher's note

overlapping brain networks. *Brain Res.* 1626, 136–145. doi: 10.1016/j.brainres.2014.12.050

Baenninger, A., Diaz Hernandez, L., Rieger, K., Ford, J. M., Kottlow, M., and Koenig, T. (2016). Inefficient preparatory fMRI-BOLD network activations predict working

memory dysfunctions in patients with schizophrenia. *Front. Psych.* 7:29. doi: 10.3389/fpsyt.2016.00029

Bagshaw, A. P., Hale, J. R., Campos, B. M., Rollings, D. T., Wilson, R. S., Alvim, M. K. M., et al. (2017). Sleep onset uncovers thalamic abnormalities in patients with idiopathic generalised epilepsy. *Neuroimage Clin.* 16, 52–57. doi: 10.1016/j.nicl.2017.07.008

Bertelson, P. (1999). Ventriloquism: a case of crossmodal perceptual grouping. *Adv. Psychol.* 129, 347–362. doi: 10.1016/S0166-4115(99)80034-X

Bertelson, P., Vroomen, J., De, G. B., and Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Percept. Psychophys.* 62, 321–332. doi: 10.3758/BF03205552

Bolognini, N., Frassinetti, F., Serino, A., and Làdavas, E. (2005). "Acoustical vision" of below threshold stimuli: interaction among spatially converging audiovisual inputs. *Exp. Brain Res.* 160, 273–282. doi: 10.1007/s00221-004-2005-z

Bonath, B., Noesselt, T., Martinez, A., Mishra, J., Schwiecker, K., Heinze, H. J., et al. (2007). Neural basis of the ventriloquist illusion. *Curr. Biol.* 17, 1697–1703. doi: 10.1016/j.cub.2007.08.050

Bonte, M., Hausfeld, L., Scharke, W., Valente, G., and Formisano, E. (2014). Task-dependent decoding of speaker and vowel identity from auditory cortical response patterns. *J. Neurosci.* 34, 4548–4557. doi: 10.1523/JNEUROSCI.4339-13.2014

Branden, J. B., Arvid, G., Mark, P., and Andrew, I. W. (2022). Michael SAG right temporoparietal junction encodes inferred visual knowledge of others. *Neuropsychologia* 171:108243. doi: 10.1016/j.neuropsychologia.2022.108243

Bullock, M., Jackson, G. D., and Abbott, D. F. (2021). Artifact reduction in simultaneous EEG-fMRI: a systematic review of methods and contemporary usage. *Front. Neurol.* 12:622719. doi: 10.3389/fneur.2021.622719

Busse, L., Katzner, S., and Treue, S. (2008). Temporal dynamics of neuronal modulation during exogenous and endogenous shifts of visual attention in macaque area MT. *Proc. Natl. Acad. Sci.* 105, 16380–16385. doi: 10.1073/pnas.0707369105

Calvert, G. A., Campbell, R., and Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr. Biol.* 10, 649–657. doi: 10.1016/s0960-9822(00)00513-3

Calvert, G. A., Hansen, P. C., Iversen, S. D., and Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *NeuroImage* 14, 427–438. doi: 10.1006/nimg.2001.0812

Calvert, G. A., and Thesen, T. (2004). Multisensory integration: methodological approaches and emerging principles in the human brain. *J. Physiol. Paris* 98, 191–205. doi: 10.1016/j.jphysparis.2004.03.018

Cappe, C., Thut, G., Romei, V., and Murray, M. M. (2010). Auditory-visual multisensory interactions in humans: timing, topography, directionality, and sources. *J. Neurosci.* 30, 12572–12580. doi: 10.1523/JNEUROSCI.1099-10.2010

Chen, T., Michels, L., Supekar, K., Kochalka, J., Ryali, S., and Menon, V. (2015). Role of the anterior insular cortex in integrative causal signaling during multisensory auditory-visual attention. *Eur. J. Neurosci.* 41, 264–274. doi: 10.1111/ejn.12764

Corbetta, M., Kincade, J. M., and Shulman, G. L. (2002). Neural systems for visual orienting and their relationships to spatial working memory. *J. Cogn. Neurosci.* 14, 508–523. doi: 10.1162/089892902317362029

Corbetta, M., Patel, G., and Shulman, G. L. (2008). The reorienting system of the human brain: from environment to theory of mind. *Neuron* 58, 306–324. doi: 10.1016/j.neuron.2008.04.017

Durk, T., and Woldorff, M. G. (2005). Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. *J. Cogn. Neurosci.* 17, 1098–1114. doi: 10.1162/0898929054475172

Erik, V. D. B., Olivers, C. N. L., Bronkhorst, A. W., and Theeuwes, J. (2008). Pip and pop: nonspatial auditory signals improve spatial visual search. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 1053–1065. doi: 10.1037/0096-1523.34.5.1053

Fairhall, S. L., and Macaluso, E. (2009). Spatial attention can modulate audiovisual integration at multiple cortical and subcortical sites. *Eur. J. Neurosci.* 29, 1247–1257. doi: 10.1111/j.1460-9568.2009.06688.x

Ford, J. M., Roach, B. J., Palzes, V. A., and Mathalon, D. H. (2016). Using concurrent EEG and fMRI to probe the state of the brain in schizophrenia. *NeuroImage Clin.* 12, 429–441. doi: 10.1016/j.nicl.2016.08.009

Frassinetti, F., Bolognini, N., and Làdavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Exp. Brain Res.* 147, 332–343. doi: 10.1007/s00221-002-1262-y

Friston, K. J., Holmes, A. P., Poline, J. B., Grasby, P. J., Williams, S. C. R., Frackowiak, R. S. J., et al. (1995). Analysis of fMRI time-series revisited. *NeuroImage* 2, 45–53. doi: 10.1006/nimg.1995.1007

Hanlon, F. M., Dodd, A. B., Ling, J. M., Bustillo, J. R., Abbott, C. C., and Mayer, A. R. (2017). From behavioral facilitation to inhibition: the neuronal correlates of the orienting and reorienting of auditory attention. *Front. Hum. Neurosci.* 11:293. doi: 10.3389/fnhum.2017.00293

Hsiao, F. C., Tsai, P. J., Wu, C. W., Yang, C. M., Lane, T. J., Lee, H. C., et al. (2018). The neurophysiological basis of the discrepancy between objective and subjective sleep during the sleep onset period: an EEG-fMRI study. *Sleep* 41:zsy056. doi: 10.1093/sleep/zsy056

Jonides, J., and Irwin, D. E. (1981). Capturing attention. *Cognition* 10, 145–150. doi: 10.1016/0010-0277(81)90038-X

Jorge, J., Zwaag, W. V. D., and Figueiredo, P. (2014). EEG-fMRI integration for the study of human brain function. *NeuroImage* 102, 24–34. doi: 10.1016/j.neuroimage.2013.05.114

José, P., Ossandón, K. P., and Heed, T. (2020). No evidence for a role of spatially modulated alpha-band activity in tactile remapping and short-latency, overt orienting behavior. *J. Neurosci.* 40, 9088–9102. doi: 10.1523/JNEUROSCI.0581-19.2020

Kang-jia, J., and Xu, W. (2022). Research and Prospect of visual event-related potential in traumatic brain injury and visual function evaluation. *J. Forensic Med.* 40, 9088–9102. doi: 10.1523/JNEUROSCI.0581-19.2020

Karns, C. M., and Knight, R. T. (2009). Intermodal auditory, visual, and tactile attention modulates early stages of neural processing. *J. Cogn. Neurosci.* 21, 669–683. doi: 10.1162/jocn.2009.21037

Klein, H. S., Vanneste, S., and Pinkham, A. E. (2021). The limited effect of neural stimulation on visual attention and social cognition in individuals with schizophrenia. *Neuropsychologia* 157:107880. doi: 10.1016/j.neuropsychologia.2021.107880

Klemen, J., and Chambers, C. D. (2012). Current perspectives and methods in studying neural mechanisms of multisensory interactions. *Neurosci. Biobehav. Rev.* 36, 111–133. doi: 10.1016/j.neubiorev.2011.04.015

Koelewijn, T., Bronkhorst, A., and Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: a review of audiovisual studies. *Acta Psychol.* 134, 372–384. doi: 10.1016/j.actpsy.2010.03.010

Laura, B., Roberts, K. C., Crist, R. E., Weissman, D. H., and Woldorff, M. G. (2005). The spread of attention across modalities and space in a multisensory object. *Proc. Natl. Acad. Sci.* 102, 18751–18756. doi: 10.1073/pnas.0507704102

Lepsien, J., and Pollmann, S. (2002). Covert reorienting and inhibition of return: an event-related fMRI study. *J. Cogn. Neurosci.* 14, 127–144. doi: 10.1162/089892902317236795

Li, F., Chen, B., Li, H., Zhang, T., Wang, F., Jiang, Y., et al. (2016). The time-varying networks in P300: a task-evoked EEG study. *IEEE Trans. Neural Syst. Rehabil. Eng.* 24, 725–733. doi: 10.1109/TNSRE.2016.2523678

Li, F., Liu, T., Wang, F., Li, H., Gong, D., Zhang, R., et al. (2015). Relationships between the resting-state network and the P3: evidence from a scalp EEG study. *Sci. Rep.* 5:15129. doi: 10.1038/srep15129

Ľuboš, H., Aaron, R., Seitz, , and Norbert, K. (2021). Auditory-visual interactions in egocentric distance perception: ventriloquism effect and aftereffecta. *J. Acoust. Soc. Am.* 150, 3593–3607. doi: 10.1121/10.0007066

Macaluso, E., George, N., Dolan, R., Spence, C., and Driver, J. (2004). Spatial and temporal factors during processing of audiovisual speech: a PET study. *NeuroImage* 21, 725–732. doi: 10.1016/j.neuroimage.2003.09.049

Macaluso, E., Noppeney, U., Talsma, D., Vercillo, T., Hartcher-O'Brien, J., and Adam, R. (2016). The curious incident of attention in multisensory integration: bottom-up vs Top-down. *Multisens. Res.* 29, 557–583. doi: 10.1163/22134808-00002528

Mastroberardino, S., Santangelo, V., and Macaluso, E. (2015). Crossmodal semantic congruence can affect visuo-spatial processing and activity of the fronto-parietal attention networks. *Front. Integr. Neurosci.* 9:45. doi: 10.3389/fnint.2015.00045

Mayer, A. R., Dorflinger, J. M., Rao, S. M., and Seidenberg, M. (2004a). Neural networks underlying endogenous and exogenous visual-spatial orienting. *NeuroImage* 23, 534–541. doi: 10.1016/j.neuroimage.2004.06.027

Mayer, A., Seidenberg, M., Dorflinger, J., and Rao, S. (2004b). An event-related fMRI study of exogenous orienting: supporting evidence for the cortical basis of inhibition of return? *J. Cogn. Neurosci.* 16, 1262–1271. doi: 10.1162/0898929041920531

Mcdonald, J. J., Hickey, C., Green, J. J., and Whitman, J. C. (2009). Inhibition of return in the covert deployment of attention: evidence from human electrophysiology. *J. Cogn. Neurosci.* 21, 725–733. doi: 10.1162/jocn.2009.21042

Mcdonald, J. J., Teder-Sälejärvi, W. A., Di, R. F., and Hillyard, S. A. (2005). Neural basis of auditory-induced shifts in visual time-order perception. *Nat. Neurosci.* 8, 1197–1202. doi: 10.1038/nn1512

Mcdonald, J. J., Teder-sälejärvi, W. A., and Hillyard, S. A. (2000). Involuntary orienting to sound improves visual perception. *Nature* 407, 906–908. doi: 10.1038/35038085

Mesgarani, N., Cheung, C., Johnson, K., and Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science* 343, 1006–1010. doi: 10.1126/science.1245994

Motomura, K., and Amimoto, K. (2022). Development of stimulus-driven attention test for unilateral spatial neglect – accuracy, reliability, and validity. *Neurosci. Lett.* 772:136461. doi: 10.1016/j.neulet.2022.136461

Muraskin, J., Brown, T. R., Walz, J. M., Tu, T., Conroy, B., Goldman, R. I., et al. (2018). A multimodal encoding model applied to imaging decision-related neural cascades in the human brain. *NeuroImage* 180, 211–222. doi: 10.1016/j.neuroimage.2017.06.059

Navarra, J., Soto-Faraco, S., and Spence, C. (2010). Assessing the role of attention in the audiovisual integration of speech. *Inf. Fusion* 11, 4–11. doi: 10.1016/j.inffus.2009.04.001

Nazir, H. M., Hussain, I., Faisal, M., Mohamd Shoukry, A., Abdel Wahab Sharkawy, M., Fawzi Al-Deek, F., et al. (2020). Dependence structure analysis of multisite river inflow

data using vine copula-CEEMDAN based hybrid model. *PeerJ* 8:e10285. doi: 10.7717/peerj.10285

Noesselt, T., Rieger, J. W., Schoenfeld, M. A., Kanowski, M., Hinrichs, H., Heinze, H. J., et al. (2007). Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. *J. Neurosci.* 27, 11431–11441. doi: 10.1523/JNEUROSCI.2252-07.2007

Pisauro, M. A., Fouragnan, E., Retzler, C., and Philiastides, M. G. (2017). Neural correlates of evidence accumulation during value-based decisions revealed via simultaneous EEG-fMRI. *Nat. Commun.* 8:15808. doi: 10.1038/ncomms15808

Platt, B. B., and Warren, D. H. (1972). Auditory localization: the importance of eye movements and a textured visual environment. *Percept. Psychophys.* 12, 245–248. doi: 10.3758/BF03212884

Posner, M. I., and Rothbart, M. K. (2006). Research on attention networks as a model for the integration of psychological science. *Annu. Rev. Psychol.* 58, 1–23. doi: 10.1146/annurev.psych.58.110405.085516

Rachel, W., Kathleen, B., and Mark, A. (2022). Evaluating the co-design of an age-friendly, rural, multidisciplinary primary care model: a study protocol. *Methods Protoc.* 5:23. doi: 10.3390/mps5020023

Ristic, J., and Capozzi, F. (2023). The role of visual and auditory information in social event segmentation. *Q. J. Exp. Psychol.* 1:174702182311764. doi: 10.1177/17470218231176471

Romei, V., Murray, M. M., Cappe, C., and Thut, G. (2013). The contributions of sensory dominance and attentional Bias to cross-modal enhancement of visual cortex excitability. *J. Cogn. Neurosci.* 25, 1122–1135. doi: 10.1162/jocn_a_00367

Rupp, K., Hect, J. L., Remick, M., Ghuman, A., Chandrasekaran, B., Holt, L. L., et al. (2022). Neural responses in human superior temporal cortex support coding of voice representations. *PLoS Biol.* 20:e3001675. doi: 10.1371/journal.pbio.3001675

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Stat.* 6, 461–464. doi: 10.1214/aos/1176344136

Sébastien, A. L., Arin, E. A., Kristina, C., Blake, E. B., and Ryan, A. S. (2022). The relationship between multisensory associative learning and multisensory integration. *Neuropsychologia* 174:108336. doi: 10.1016/j.neuropsychologia.2022.108336

Senkowski, D., Talsma, D., Herrmann, C. S., and Woldorff, M. G. (2005). Multisensory processing and oscillatory gamma responses: effects of spatial selective attention. *Exp. Brain Res.* 166, 411–426. doi: 10.1007/s00221-005-2381-z

Shams, L., Kamitani, Y., and Shimojo, S. (2000). What you see is what you hear. *Nature* 408, 788–2671. doi: 10.1038/35048669

Shams, N., Alain, C., and Strother, S. (2015). Comparison of BCG artifact removal methods for evoked responses in simultaneous EEG-fMRI. *J. Neurosci. Methods* 245, 137–146. doi: 10.1016/j.jneumeth.2015.02.018

Shepherd, M., and Müller, H. J. (1989). Movement versus focusing of visual attention. *Percept. Psychophys.* 46, 146–154. doi: 10.3758/BF03204974

Spence, C. (2010). Crossmodal spatial attention. *Ann. N. Y. Acad. Sci.* 1191, 182–200. doi: 10.1111/j.1749-6632.2010.05440.x

Spence, C. (2013). Just how important is spatial coincidence to multisensory integration? Evaluating the spatial rule. *Ann. N. Y. Acad. Sci.* 1296, 31–49. doi: 10.1111/nyas.12121

Spence, C., and Santangelo, V. (2010). Auditory development and learning. *Oxf. Handb. Audit. Sci.* 3:249. doi: 10.1093/oxfordhb/9780199233557.013.0013

Stein, B. E., and Meredith, M. A. (1990). Multisensory integration. Neural and behavioral solutions for dealing with stimuli from different sensory modalities. *Ann. N. Y. Acad. Sci.* 608, 51–70. doi: 10.1111/j.1749-6632.1990.tb48891.x

Stein, B. E., Meredith, M. A., Huneycutt, W. S., and Mcdade, L. (1989). Behavioral indices of multisensory integration: orientation to visual cues is affected by auditory stimuli. *J. Cogn. Neurosci.* 1, 12–24. doi: 10.1162/jocn.1989.1.1.12

Stekelenburg, J. J., Vroomen, J., and De, G. B. (2004). Illusory sound shifts induced by the ventriloquist illusion evoke the mismatch negativity. *Neurosci. Lett.* 357, 163–166. doi: 10.1016/j.neulet.2003.12.085

Stevenson, R. A., Fister, J. K., Barnett, Z. P., Nidiffer, A. R., and Wallace, M. T. (2012). Interactions between the spatial and temporal stimulus factors that influence multisensory integration in human performance. *Exp. Brain Res.* 219, 121–137. doi: 10.1007/s00221-012-3072-1

Stoep, N., Der, V., Stigchel, S., Der, V., and Nijboer, T. C. W. (2015). Exogenous spatial attention decreases audiovisual integration. *Atten. Percept. Psychophys.* 77, 464–482. doi: 10.3758/s13414-014-0785-1

Talsma, D., Doty, T. J., and Woldorff, M. G. (2007). Selective attention and audiovisual integration: is attending to both modalities a prerequisite for optimal early integration? *Cereb. Cortex* 17, 691–701. doi: 10.1093/cercor/bhk020

Talsma, D., Senkowski, D. F. S., and Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends Cogn. Sci.* 14, 400–410. doi: 10.1016/j.tics.2010.06.008

Tang, X., Wu, J., and Shen, Y. (2016). The interactions of multisensory integration with endogenous andexogenous attention. *Neurosci. Biobehav. Rev.* 61, 208–224. doi: 10.1016/j.neubiorev.2015.11.002

Tedersäläjärvi, W. A., Di, R. F., Mcdonald, J. J., and Hillyard, S. A. (2005). Effects of spatial congruity on audio-visual multimodal integration. *J. Cogn. Neurosci.* 17, 1396–1409. doi: 10.1162/0898929054985383

Tian, Y., Klein, R. M., Satel, J., Xu, P., and Yao, D. (2011). Electrophysiological explorations of the cause and effect of inhibition of return in a Cue-target paradigm. *Brain Topogr.* 24, 164–182. doi: 10.1007/s10548-011-0172-3

Tian, Y., Liang, S., and Yao, D. (2014). Attentional orienting and response inhibition: insights from spatial-temporal neuroimaging. *Neurosci. Bull.* 30, 141–152. doi: 10.1007/s12264-013-1372-5

Tian, Y., and Ma, L. (2020). Auditory attention tracking states in a cocktail party environment can be decoded by deep convolutional neural networks. *J. Neural Eng.* 17:036013. doi: 10.1088/1741-2552/ab92b2

Tian, Y., Xu, W., and Yang, L. (2018a). Cortical classification with rhythm entropy for error processing in cocktail party environment based on scalp EEG recording. *Sci. Rep.* 8:6070. doi: 10.1038/s41598-018-24535-4

Tian, Y., Xu, W., Zhang, H., Tam, K. Y., Zhang, H., Yang, L., et al. (2018b). The scalp time-varying networks of N170: reference, latency, and information flow. *Front. Neurosci.* 12:250. doi: 10.3389/fnins.2018.00250

Tian, Y., and Yao, D. (2008). A study on the neural mechanism of inhibition of return by the event-related potential in the go/Nogo task. *Biol. Psychol.* 79, 171–178. doi: 10.1016/j.biopsycho.2008.04.006

Tian, Y., and Yao, D. (2013). Why do we need to use a zero reference? Reference influences on the ERPs of audiovisual effects. *Psychophysiology* 50, 1282–1290. doi: 10.1111/psyp.12130

Treichler, D. (1967). Are you missing the boat in training aids. *Film AV Commun.* 1, 14–16.

Tucker, D. M. (1993). Spatial sampling of head electrical fields: the geodesic sensor net. *Electroencephalogr. Clin. Neurophysiol.* 87, 154–163. doi: 10.1016/0013-4694(93)90121-B

Van der Stoep, N., Van der Stigchel, S., Nijboer, T. C., and Spence, C. (2017). Visually induced inhibition of return affects the integration of auditory and visual information. *Perception* 46, 6–17. doi: 10.1177/0301006616661934

Vroomen, J., Bertelson, P., and Gelder, B. D. (2001). Directing spatial attention towards the illusory location of a ventriloquized sound. *Acta Psychol.* 108, 21–33. doi: 10.1016/S0001-6918(00)00068-8

Wang, K., Li, W., and Dong, L. (2018). Clustering-constrained ICA for Ballistocardiogram artifacts removal in simultaneous EEG-fMRI. *Front. Neurosci.* 12:59. doi: 10.3389/fnins.2018.00059

Wang, P., Fuentes, L. J., Vivas, A. B., and Chen, Q. (2013). Behavioral and neural interaction between spatial inhibition of return and the Simon effect. *Front. Hum. Neurosci.* 7:572. doi: 10.3389/fnhum.2013.00572

Wang, W., Wang, Y., Sun, J., Liu, Q., Liang, J., and Li, T. (2020). Speech driven talking head generation via attentional landmarks based representation. *Proc. Interspeech* 2020, 1326–1330. doi: 10.21437/Interspeech.2020-2304

Wen, H., You, S., and Fu, Y. (2021). Cross-modal dynamic convolution for multi-modal emotion recognition. *J. Vis. Commun. Image Represent.* 78:103178. doi: 10.1016/j.jvcir.2021.103178

Wilke, C., Ding, L., and He, B. (2008). Estimation of time-varying connectivity patterns through the use of an adaptive directed transfer function. *IEEE Trans. Biomed. Eng.* 55, 2557–2564. doi: 10.1109/TBME.2008.919885

Wu, X., Wang, A., and Zhang, M. (2020). Cross-modal nonspatial repetition inhibition: an ERP study, neuroscience letters. *ISSN* 734:135096. doi: 10.1016/j.neulet.2020.135096

Xu, Z., Yang, W., and Zhou, Z. (2020). Cue–target onset asynchrony modulates interaction between exogenous attention and audiovisual integration. *Cogn. Process.* 21, 261–270. doi: 10.1007/s10339-020-00950-2

Yan, T., Geng, Y., Wu, J., and Li, C. (2015). Interactions between multisensory inputs with voluntary spatial attention: an fMRI study. *Neuroreport* 26, 605–612. doi: 10.1097/WNR.0000000000000368

Yao, D., Li, W., Oostenveld, R., Nielsen, K. D., Arendt-Nielsen, L., and Chen, C. A. N. (2005). A comparative study of different references for EEG spectral mapping: the issue of the neutral reference and the use of the infinity reference. *Physiol. Meas.* 26, 173–184. doi: 10.1088/0967-3334/26/3/003

Zhang, J., Li, Y., Li, T., Xun, L., and Shan, C. (2019). License plate localization in unconstrained scenes using a two-stage CNN-RNN. *IEEE Sensors J.* 19, 5256–5265. doi: 10.1109/JSEN.2019.2900257

Zhang, T., Gao, Y., and Hu, S. (2022). Focused attention: its key role in gaze and arrow cues for determining where attention is directed. *Psychol. Res.* 2022, 1–15. doi: 10.1007/s00426-022-01781-w

Zhang, X., Wang, Z., and Liu, T. (2022). Anti-disturbance integrated position synchronous control of a dual permanent magnet synchronous motor system. *Energies* 15:6697. doi: 10.3390/en15186697

Zhang, X., Zhang, J., and Liu, G. (2021). Comprehensive framework for the integration and analysis of geo-environmental data for urban geohazards. *Earth Sci. Inf.* 14, 2387–2399. doi: 10.1007/s12145-021-00642-1