Comparing neural correlates of visual target detection in serial visual presentations having different temporal correlations

An Luo and Paul Sajda*

Department of Biomedical Engineering, Columbia University, New York, NY, USA

Most visual stimuli we experience on a day-to-day basis are continuous sequences, with spatial structure highly correlated in time. During rapid serial visual presentation (RSVP), this correlation is absent. Here we study how subjects' target detection responses, both behavioral and electrophysiological, differ between continuous serial visual sequences (CSVP), flashed serial visual presentation (FSVP) and RSVP. Behavioral results show longer reaction times for CSVP compared to the FSVP and RSVP conditions, as well as a difference in miss rate between RSVP and the other two conditions. Using mutual information, we measure electrophysiological differences in the electroencephalography (EEG) for these three conditions. We find two peaks in the mutual information between EEG and stimulus class (target vs. distractor), with the second peak occurring 30–40 ms earlier for the FSVP and RSVP conditions. In addition, we find differences in the persistence of the peak mutual information between FSVP and RSVP conditions. We further investigate these differences using a mutual information based functional connectivity analysis and find significant fronto-parietal functional coupling for RSVP and FSVP but no significant coupling for the CSVP condition. We discuss these findings within the context of attentional engagement, evidence accumulation and short-term visual memory.

Keywords: target detection, visual presentation, electroencephalography, mutual information

INTRODUCTION

We typically observe the world within the context of moving imagery – i.e., the visual scene is temporally continuous and highly correlated in time. This is in contrast to a visual paradigm that is commonly used in the experimental settings called rapid serial visual presentation (RSVP) (Chun and Potter, 1995; Potter and Levy, 1969), where a sudden onset of a target stimulus is followed by non-relevant stimuli (e.g., distractor images). Two attractive elements in the RSVP paradigm are: (1) images are presented so rapidly that eye movements play little if any role in the task (Öquist et al., 2004) and (2) the speed of visual processing and decision making can be driven to the limits of temporal processing (Keysers et al., 2001). However in RSVP, the correlation in space and time is typically discontinuous, i.e., the targets in RSVP only appear for a very brief time period before they are masked by subsequent distracting frames, while targets in a continuous sequence normally are present for a much longer time.

To understand the target detection process in the context of natural vision, one must consider the possible differences between target detection during a continuous sequence and the experimentally-used RSVP stimuli. However, to the best of our knowledge, no one has yet systematically studied this topic. The study of attentional blink (Broadbent and Broadbent, 1987; Chun and Potter, 1995; Raymond et al., 1992) has shed some light on how target detection is affected by the correlations of image frames across time – i.e., spatio-temporal correlations in the sequence. The attentional blink is the phenomenon that when two targets are presented amongst distractors during RSVP, the correct identification of the first (T1) is followed by poor detection of the second target (T2) when the two targets are separated by 200–500 ms and T1 is

followed by distractor images. It has been shown that replacing a distractor image with a uniform gray image at the T1 + 1 position in the RSVP sequence, i.e., T1 is followed by a blank image, results in improved detection accuracy of both T1 and T2 (Bowman and Wyble, 2007; Chun and Potter, 1995; Raymond et al., 1992). It has also been shown that the greater the similarity of T1 + 1 with T1, the longer it takes to identify T1 (Chun and Potter, 1995; Peterson and Juola, 2000). These findings suggest that the detection of a target can be affected by the correlations of image frames across time, and that the effect is likely due to the difference in the interaction between the instantaneous visual input and stored visual information, potentially from persistent activity.

Starting from these findings, we hypothesize that the temporal correlation of image frames will affect target detection in the two following ways: (1) temporally-continuous sequences will result in increased response times (RTs) and lower miss rates (MRs), relative to RSVP stimuli, since there is additional evidence in the incoming stimulus that can be accumulated over time and (2) temporally-discontinuous RSVP will result in masking of target images by non-relevant distractor images, with detectability of the target image reduced relative to that for continuous sequences or flashed presentations with no distractor images.

We expect that the two hypothesized observations would lead to differences in neural activity between serial visual presentations having different temporal correlation. These would include differences in the timing and/or amount of information which can be decoded from the neural activity that could be used to predict the stimulus class (i.e., target vs. distractor). Moreover, it has been suggested by Honey et al. (2002) that increased functional connectivity between prefrontal and parietal cortices is

related to higher demand for maintenance and executive processes during a working memory task. Consistent with this finding, we would expect to see functional connectivity differences between a continuous sequence and a sequence with little temporal correlation, if in fact the latter requires short-term visual memory and visual persistence.

To test these hypotheses and their corresponding neural correlates we consider three stimulus conditions in this paper: (1) continuous serial visual sequences (CSVP), where the visual input is concatenated video clips so the input and the stored representation are highly correlated; (2) flashed serial visual presentation (FSVP), where the visual input only appears for a very brief period of time and there is no masking from subsequent distractor images; and (3) RSVP, where the inputs are independent/random frames and there is no correlation between the input and the stored visual representation.

It is widely accepted that target detection elicits the P300 (Hruby and Marsalek, 2003; Menon et al., 1997; Picton, 1992; Squires et al., 1976; Tanaka et al., 1998), which is a large positivity in the average event-related potential (ERP) difference between target and distractor trials starting at around 300-ms post-stimulus and is maximal over the parietal scalp. Although differences exist in the scalp distribution and time course of the P300, it is believed that the P300 is independent of sensory modality (auditory, visual, and somatosensory; Ji et al., 1999; Naumann et al., 1992). Evidence also has shown that the P300 is closely related to a post-perceptual, capacity-limited stage in target detection (Dell'Acqua et al., 2003; Kranczioch et al., 2003; Nieuwenhuis et al., 2005b; Rolke et al., 2001; Vogel and Luck, 2002; Vogel et al., 1998), and thus is directly linked to processes such as evidence accumulation and short-term memory retrieval.

A traditional way to analyze electroencephalography (EEG) is by averaging across multiple trials to compute the mean ERP difference between target and distractor. However, this method only provides the mean difference between conditions and cannot quantitatively measure the discriminability of the EEG for the two conditions – i.e., it does not incorporate the variance across trials. Mutual information has been widely used in the neuroscience community (Borst and Theunissen, 1999; Chen et al., 2007; Jeong et al., 2001; Rozell and Johnson, 2005) and is seen as an informative measure of the statistical relationship between the stimulus and the response (Rozell and Johnson, 2005). In this paper, we use mutual information as a metric to quantify the target-related information content in the EEG as well as a means to estimate functional connectivity between clusters of electrodes at the peak target-informative times. We focus our analysis on electrophysiological correlates whose timing overlaps with the P300 ERP. We show that the neural correlates of target detection processes, as measured via our mutual information framework, show systematic differences for the three stimulus conditions, suggesting that the interaction between short-term visual memory and the incoming visual stimulus can substantially impact perceptual decision making.

MATERIALS AND METHODS
SUBJECTS
Ten healthy right-handed subjects (one female and nine male, age from 25 to 31, mean age 29 years) participated in the study. All subjects had normal or corrected to normal vision and reported no history of neurological problems. Informed consent was obtained from all participants in accordance with the guidelines and approval of the Columbia University Institutional Review Board.

STIMULI
To generate our stimuli, video clips from movies were manually extracted and inspected to make sure all clips had no scene changes and minimal camera/angle changes (i.e., smooth temporal correlation). If a video clip contained a target (defined as a person somewhere in the image frame), we also made sure that the target appeared through-out the entire clip (from first to last frame) and was clearly visible to subjects. The presentation rate was set to 25 frames per second (i.e., 40 ms per frame). The duration of each video clip was uniformly distributed between 18 and 30 frames (720–1200 ms). The size of each frame was 352×288 pixels. A Dell Precision 530 Workstation (Round Rock, TX, USA) with nVidia Quadro4 900XGL graphics card (Santa Clara, CA, USA) was used for stimulus presentation. A program was specially designed, using Visual C++, to present the stimuli and to ensure the timing of each frame during display. Visual stimuli were presented on a front projection screen using an LCD projector (InFocus LP130, Wilsonville, OR, USA) through an RF-shielded window. Stimuli subtended $19° \times 17°$ of the visual field.
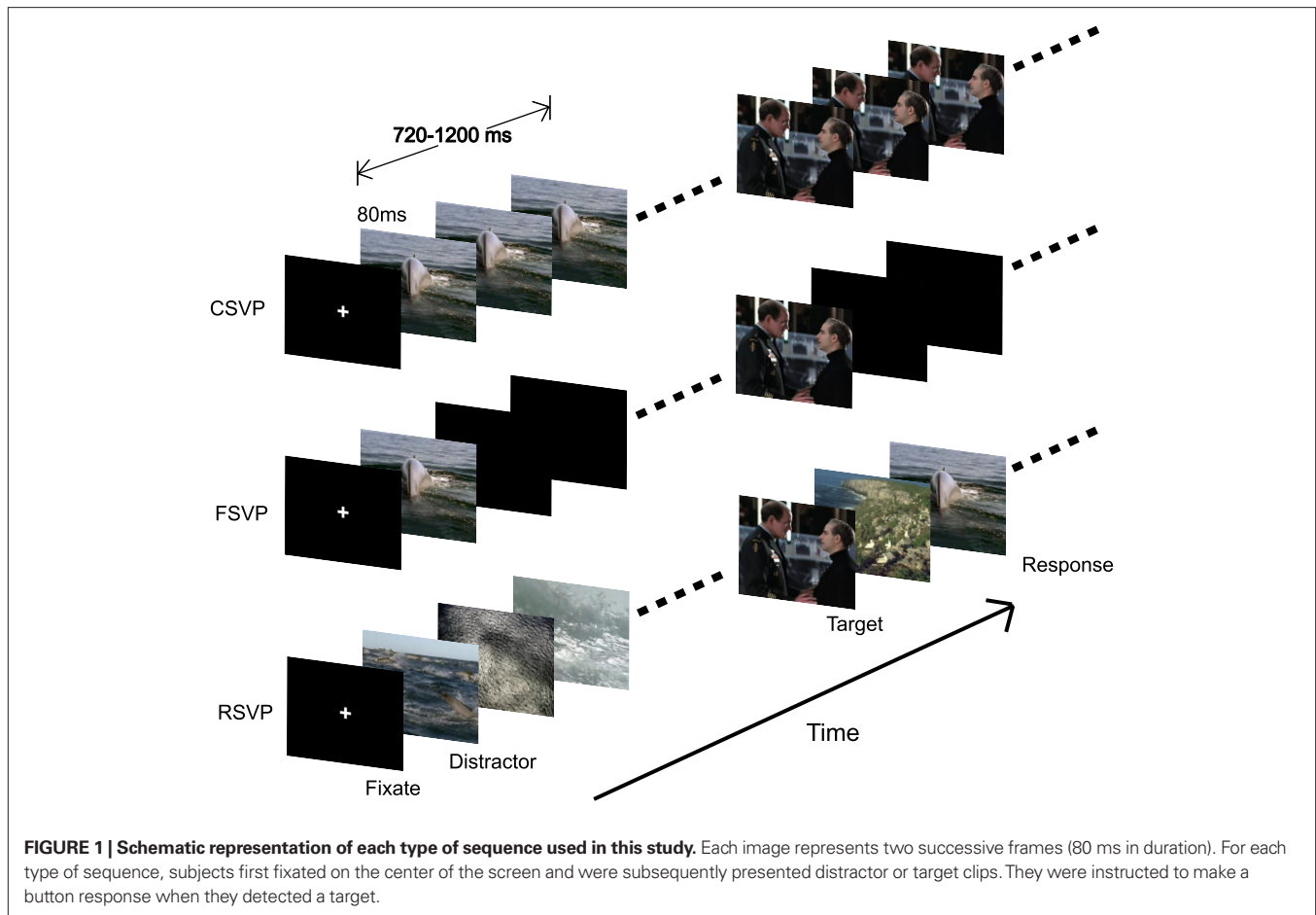
EXPERIMENTAL PARADIGM
The main experimental paradigm was designed to compare the behavioral and neuronal responses to target presentation in CSVP, FSVP, and RSVP sequence. A secondary paradigm (see Supplementary Material) was used to assess the effect of the behavioral response (button press) on target-related EEG signals.

The experiment consisted of three stimulus conditions: CSVP, FSVP, and RSVP. The task for subjects was to detect targets in a sequence of frames by pressing a button with their right index finger as soon as possible. The targets were defined as human(s) somewhere in the frame. Thus, the target could be at any position, scale or pose and low level features were unlikely to be discriminative. Each clip was considered a trial, with trials containing people termed "target trials", and otherwise they were "distractor trials".

Each CSVP sequence began with a fixation cross followed by eight clips. Every CSVP sequence had a corresponding FSVP and an RSVP sequence. For the FSVP, the first two frames of each clip in CSVP were presented at the same relative place in the sequence (i.e., same frame number), with blank (black) frames replacing the remaining frames in the clip. For RSVP, only the first two frames of target clips were kept the same as the other two types of sequences; the rest were non-target frames randomly picked from the same CSVP sequence. **Figure 1** shows an illustration of the three types of sequences used in this paradigm.

There were 70 sequences for each stimulus condition, with 50 of these containing one and only one target clip (trial). The remaining 20 sequences were distractor only trials. Sequences were randomly presented to subjects. For sequences with a target, the target could appear from the second to the last (eighth) clip. All subjects participated in this experiment and EEG was simultaneously recorded for each subject.

FIGURE 1 | Schematic representation of each type of sequence used in this study. Each image represents two successive frames (80 ms in duration). For each type of sequence, subjects first fixated on the center of the screen and were subsequently presented distractor or target clips. They were instructed to make a button response when they detected a target.

DATA ACQUISITION AND PREPROCESSING

EEG data were recorded in an electro-statically shielded room (ETS-Lindgren, Glendale Heights, IL, USA) by a Sensorium EPA-6 Electro-physiological Amplifier (Charlotte, VT, USA). We used an EEG cap having 60 Ag/AgCl electrodes (Electro-Cap, Eaton, OH, USA), with electrodes positioned according to the International 10–20 system. All channels were referenced to the left mastoid and chin ground. Data were sampled at 1 kHz with analog band-pass filtering of 0.01–300 Hz. Raw EEG data were visually-inspected and trials with large eye movements were excluded. One subject was later excluded from data analysis because this subject's EEG data contained frequent eye-blink artifacts. Following data acquisition, a software-based second-order 0.5 Hz Butterworth high-pass filter was used to remove DC drifts. The 60-Hz noise and 120-Hz harmonics were filtered out by two second-order Butterworth band-stop filters. Eye-blink and eye-movement activities were recorded and later removed from EEG recordings using a maximum difference method (Parra et al., 2005).

USING MUTUAL INFORMATION TO QUANTIFY TARGET DETECTION

We use mutual information as a metric to quantify the discriminability of the EEG and identify neural correlates of the target detection process in each of the three stimulus conditions. The approach enables us to, within the same framework, investigate

coupling between electrode activities, thereby providing a means for inferring functional connectivity.

The mutual information between two variables measures the mutual dependence of the two variables, and is defined as (using discrete random variables as an example),

$$I(X,Y) = \sum_{y \in Y} \sum_{x \in X} P(x,y) \log \frac{P(x,y)}{P(x)P(y)}, \tag{1}$$

where $P(x,y)$ is the joint probability distribution function of the variables X and Y, and $P(x)$ and $P(y)$ are the marginal probability distribution functions of X and Y, respectively. Mutual information measures how much knowing one of the two variables reduces the uncertainty of the other. If X and Y are independent random variables, i.e., $P(x,y) = P(x) P(y)$, their mutual information is zero, meaning the variables are not informative about one another; otherwise, the value of mutual information is greater than zero. The widely-used units of mutual information include the "bit" and the "nit", with one bit being the amount of information required to distinguish between two equally likely possibilities. In this paper, we use the natural logarithm to compute the mutual information, so units are in "nits" (one nit of mutual information equals 1/ln2 bits).

We first estimate the mutual information between spatio-temporal EEG signals and their class labels (targets vs. distractors),

across time and on each electrode for each subject, to quantify the information the EEG provides about target detection. If the mutual information is estimated at each sample in time, a large number of trials are needed for estimating the probability density functions from each subject. The number of trials in our experiment is limited, so we compute the mutual information within temporal windows. If $S_{t_\delta, i}$ represents all EEG samples in a temporal window starting at a post-stimulus onset time t_δ with a duration of δ on channel i, and C represents the corresponding class labels ($C \in \{0,1\}$), Eq. 1 can be rewritten as,

$$I(S_{t_{\delta,i}}, C) = \sum_{c \in C} \int_{s \in S_{t_{\delta,i}}} p(s,c) \log \frac{p(s,c)}{p(s)p(c)} ds. \quad (2)$$

Here $c = 1$ indicates a target trial (clip) and zero a non-target trial. In terms of the temporal window length δ, it should be long enough to include sufficient data samples for the mutual information estimation, but not so long such that we are estimating over different processes and/or noise that is not informative about the class – i.e., within the temporal window the signal should be stationary. In our analysis, we fixed the window length to be 50 ms ($\delta = 50$ when fs = 1000 Hz). As an example, if for one subject there are 100 target trials ($c = 1$) and 200 distractor trials ($c = 0$), with window length $\delta = 50$ ms, sampling frequency fs = 1000 Hz, and a post-stimulus window ranging from t to $t + \delta$, then $(100 + 200) \times 50 = 15,000$ EEG samples per channel are used to estimate the mutual information between EEG and class labels for that subject. We shift the 50-ms window every 10 ms, and estimate the mutual information between EEG signal and class across time, for each channel and for each type of sequence (CSVP, FSVP, and RSVP). The result is a spatio-temporal distribution of the mutual information between EEG and class labels. In this way, we focused on the discriminating activity, i.e., the target detection process, irrespective of the perceptual difference introduced by the different types.

Similar to displaying ERP responses on a series of scalp plots over time, the spatial distribution of the mutual information can be displayed in the same way, with different scalp plots representing the mutual information estimated over different temporal windows.

Mutual information can also be used to quantify the mutual dependence of two EEG signals. For example,

$$I(S_1, S_2) = \int_{s_1 \in S_1} \int_{s_2 \in S_2} p(s_1, s_2) \log \frac{p(s_1, s_2)}{p(s_1)p(s_2)} ds_1 ds_2 \quad (3)$$

is the mutual information between EEG signals S_1 and S_2. Here S_1 and S_2 are two sets of EEG samples, for example for some temporal window on a particular set of channels from a subject. In this paper, we used this method to study the functional connectivity of EEG recorded from different clusters of electrodes. In the Supplementary Material, we show the discriminating activity between target and distractor conditions identified by a receiver operating characteristic (ROC) analysis and compare the results with that of this mutual information method. We also compare the utility of linear correlation for studying the functional connectivity between EEG clusters. In sum, we find the mutual information method is more suited for studying the neural correlates of target detection than the ROC analysis and linear correlation.

ESTIMATING MUTUAL INFORMATION

Mutual information, as defined in Eq. 1, requires three probability distribution functions P(x), P(y), and P(x,y) to be estimated. When estimating the one-dimensional distributions P(x) or P(y), we first construct a histogram with the variable x or y binned by either two levels (zeroes and ones, for class labels) or m levels (for EEG signals). To quantize EEG variables, the range from the lower bound to the upper bound of the EEG signal is equally divided into m bins. The probability distribution function is computed as the histogram divided by the total number of samples. Likewise, the two-dimensional histogram with the variable x and y is constructed to estimate the joint distribution P(x,y).

A crucial parameter for estimating mutual information for a continuous variable is m. In previous work (Luo and Sajda, 2006), we used mutual information to identify a set of features for classifying EEGs and obtained good results, where we did the binning within each of the 50-ms windows with a bin size of eight ($m = 8$). Specifically, for each 50-ms temporal window and on each electrode, we combined all the target and distractor trials and found the lower and upper bound of the EEGs, divided the range between these two bounds into eight bins with equal size, and computed the probability that EEG falls into each of these bins. In this paper, we started with this method and compared results with different values of m.
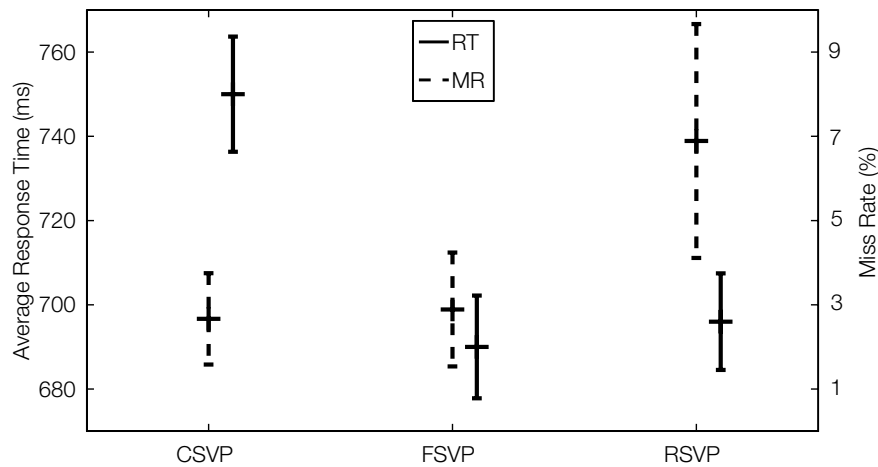
RESULTS
BEHAVIORAL RESPONSES
Statistical analysis of the subjects' RTs and MRs across all trials is shown in **Figure 2**. One can see a significant delay in the behavioral response for the CSVP targets condition. Although for the RSVP condition the mean RT was slightly longer than for the FSVP case, there was no significant difference between their RTs. A Wilcoxon signed-rank test was performed and showed that the average RT to CSVP targets is significantly larger than that of the FSVP and RSVP cases across subjects ($p < 0.05$ for both cases, with 9 degrees of freedom). Also note that the MR in the RSVP was significantly higher than the other two stimuli. Surprisingly, the MR in FSVP was only slightly higher than in the CSVP stimuli, although the duration of the target presentation in FSVP was the same as that of RSVP (80 ms), and was much shorter than in CSVP (lasting from 720 to 1200 ms).

AVERAGE ERP RESULTS
As a first step in analyzing the EEG data, we constructed scalp plots of the group average ERP difference (target minus distractor trials) across all subjects, locked to clip (trial) onset time [see **Figure 3** (top)]. For targets, only correct trials were included in this analysis. **Figure 3** (bottom) shows the average ERP difference on channels FCZ, CZ, and P1 (for the locations of these channels please refer to **Figure 5**). These ERPs showed an earlier deflection in response to FSVP stimuli than to the other two stimuli. Using a Wilcoxon signed-rank test we found that the ERP activity is significantly earlier during the FSVP cases than during the CSVP and RSVP cases across subjects ($p < 0.05$, 9 degrees of freedom). There was no significant difference in the timing of ERPs between CSVP and RSVP trials. This result cannot explain why RTs to FSVP and RSVP stimuli were significantly shorter than that of CSVP stimuli in the

FIGURE 2 | Group results showing miss rates (MRs) and response times (RTs) to targets for the three types of stimuli. The solid lines show the mean RT (labeled on the left *Y*-axis of the figure) to the three types of targets, with the error bars indicating the 95% confidence interval of the mean RT (the standard error is computed on a within-trial basis). The dashed lines represent the miss rate for each type of target stimuli (labeled on the right *Y*-axis), and the error bars show the 95% confidence interval of the average miss rate for within-subject measures.

behavioral results. However as we have stated, the average ERP difference only considers the mean response, it does not consider the variance, which also holds important information about the discriminality of the EEG. Therefore, as a next step, we estimated the mutual information between EEG and class labels to exploit variance, and potentially higher order statistical information, to identify discriminating activity.

QUANTIFYING DISCRIMINATING ACTIVITY BETWEEN TARGET AND DISTRACTOR TRIALS: MUTUAL INFORMATION BETWEEN EEG AND CLASS LABELS

Figure 4 shows the temporal evolution of the mutual information, averaged across subjects, between the EEG (for each electrode) and the corresponding class labels for all three stimuli conditions. At approximately 300–350 ms after stimulus onset, central scalp electrodes were most informative (e.g., EEG is discriminating of class label) and the latency of the peak was earlier for the FSVP condition relative to the other two conditions ($p < 0.05$, Wilcoxon signed-rank test, with 9 degrees of freedom). This is consistent with the average ERP result. After 450 ms, two areas (one frontal and one left-parietal) were informative for target detection for all three types' conditions. Interestingly, at this time the latencies of the peaks of the mutual information were earlier, by approximately 30 ms, in the FSVP and RSVP cases ($p < 0.05$, Wilcoxon signed-rank test with 9 degrees of freedom on left-parietal electrodes; electrodes were selected based on **Figure 5**). Note that this finding differs from what one sees in the average ERP analysis results, where these two conditions differ in their latencies for the ERP peaks around 450–500 ms. The mutual information result is in fact consistent with the behavioral results, where the two conditions have more similar mean RTs than for the CSVP condition, and this can be inferred from the relative time of the peaks in **Figure 4**.

Figure 4 also illustrates a rapid reduction in discrimination in left-parietal electrodes late in the trial. For example for electrode P1 (**Figure 4**, bottom) the mutual information was "suppressed" after

450 ms for the RSVP condition while the discriminating activity in FSVP persisted. To test the significance of this early reduction in discriminability for the RSVP condition, for each subject we measured the duration in which the mutual information between channel P1 and class label was above 75% of the second peak value in the FSVP and RSVP cases. We found that the duration of the second peak in mutual information in the RSVP condition was significantly shorter than that of FSVP ($p < 0.05$, Wilcoxon signed-rank test, 9 degrees of freedom; for this channel).
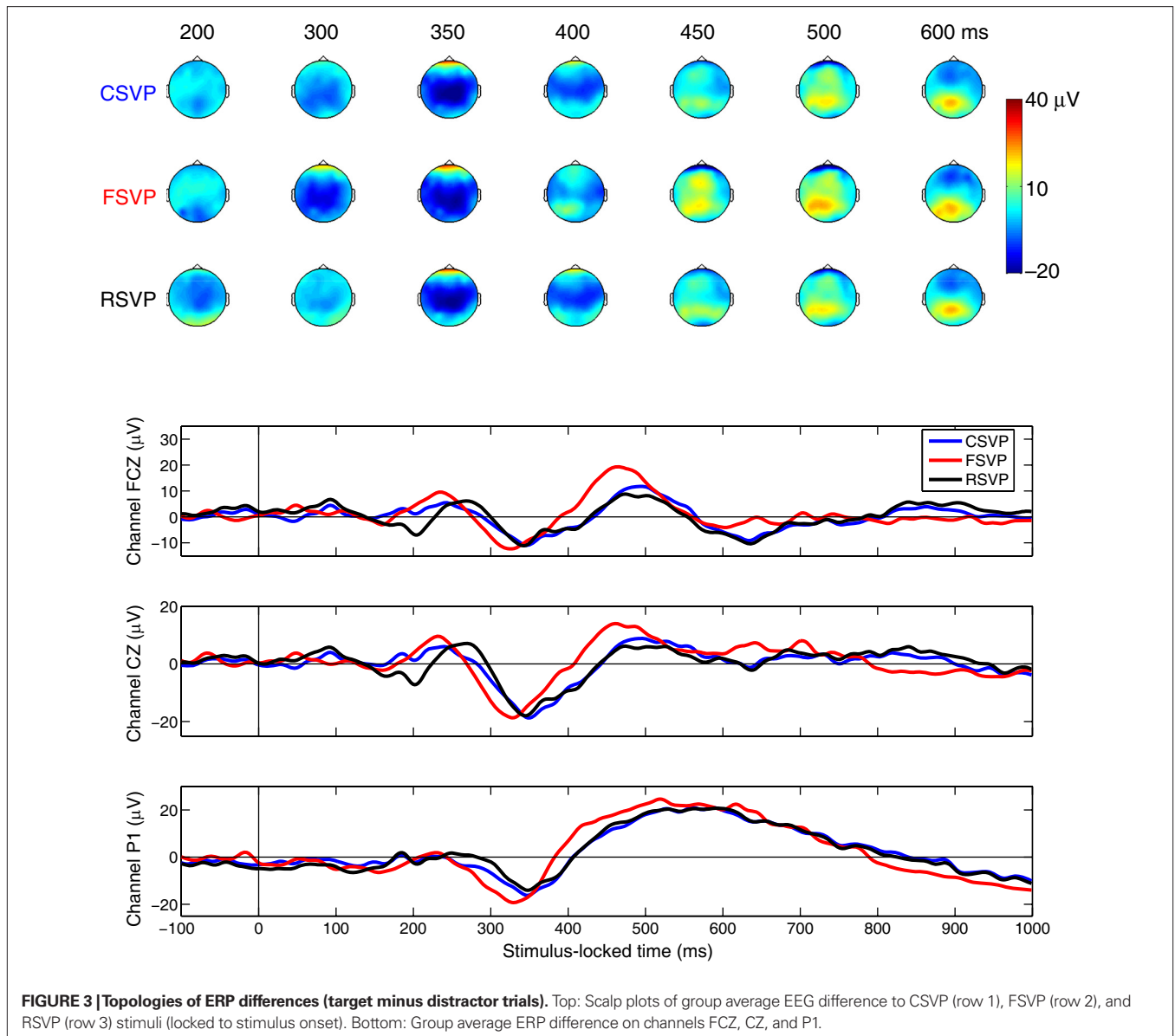
QUANTIFYING THE MUTUAL DEPENDENCE BETWEEN BRAIN SIGNALS: MUTUAL INFORMATION FOR INFERRING FUNCTIONAL CONNECTIVITY

As we have stated in Section "Materials and Methods", we can use mutual information to quantify the mutual dependence of two EEG signals to study the functional connectivity of brain responses at different incidents/areas. In this section, we identified a set of EEG clusters involved in target detection and quantified their mutual dependence, comparing results between the three conditions.

Identify discriminating clusters between target and distractor trials
In the previous section, we identified three EEG clusters that were most discriminating: one with a central scalp topography, occurring approximately 300-ms post-stimulus; one with a frontal topography and the third with a left-parietal topography, both occurring at about 450-ms post-stimulus. To investigate the functional connectivity among these clusters we evaluated their mutual dependence with respect to each cluster's corresponding EEG signals. Gerson et al. (2005) found that the discriminating activity between target and distractor trials was locked to both stimulus onset and response, and as time progressed from stimulus onset to response, discriminating activities became more locked to response. Thus to identify the exact timing of these clusters for each trial, we first quantified their degree of locking relative to stimulus onset and RT.

Figure 5 shows channels and approximate latencies that were used for identification of the three clusters. In our analysis, we used

FIGURE 3 | Topologies of ERP differences (target minus distractor trials). Top: Scalp plots of group average EEG difference to CSVP (row 1), FSVP (row 2), and RSVP (row 3) stimuli (locked to stimulus onset). Bottom: Group average ERP difference on channels FCZ, CZ, and P1.

the sum of the selected channel data as the input to estimate the mutual information. **Figure 6** shows each subject's peak latency of mutual information in each cluster as a function of their average RTs. Similar to Gerson et al. (2005), linear fitting was performed to quantify the degree in which the EEG was locked to the stimulus and response, as Eq. 4 shows,

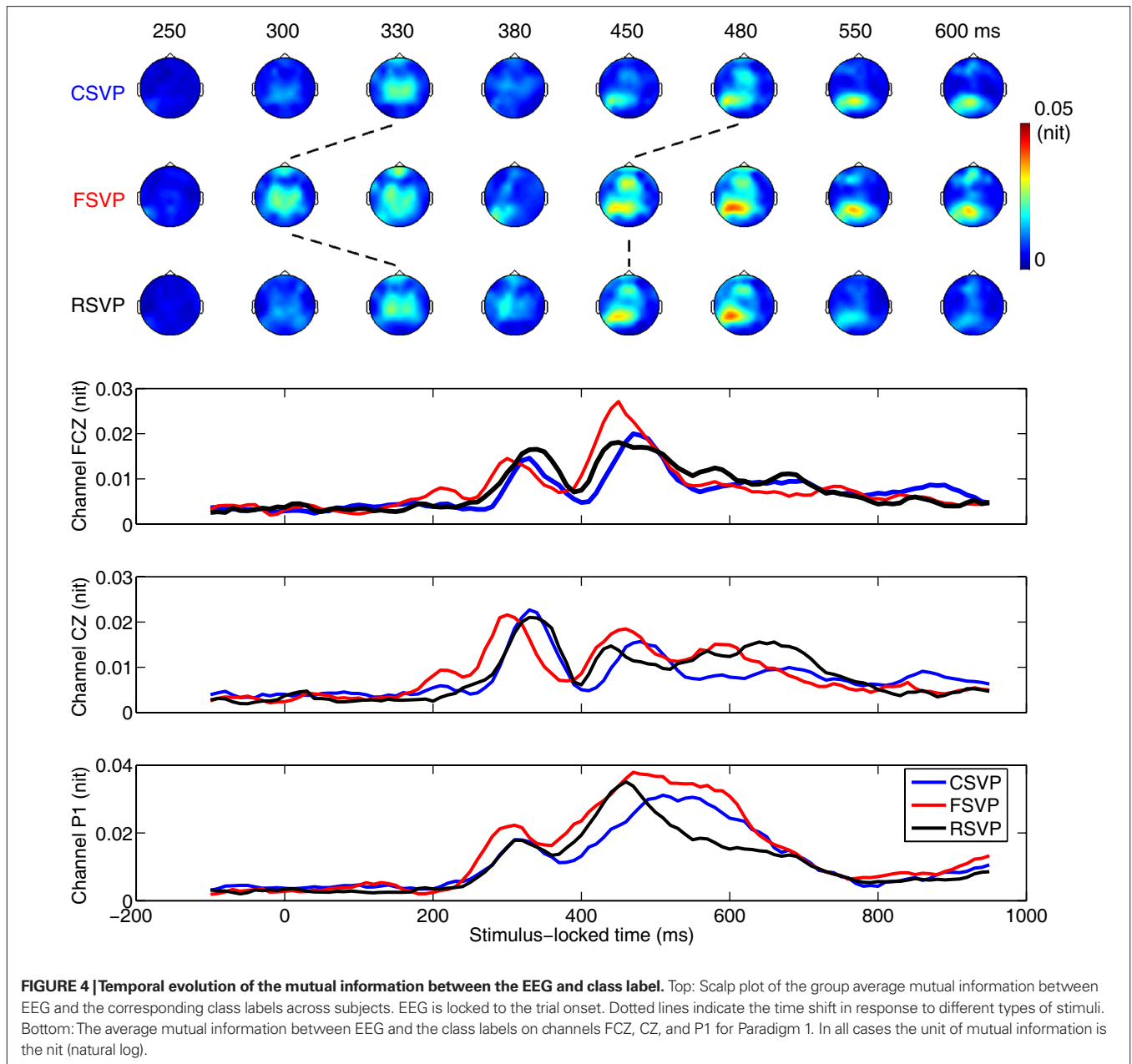$$PL = s \cdot RT \pm b, \tag{4}$$

where PL represents the peak latency of the mutual information between EEG and class labels on each cluster, RT is response time, and b is the intercept on the axis RT = 0. A slope of $s = 0$ indicates the latency is strictly locked to the stimulus while $s = 1$ means it is 100% locked to the response. From **Figure 6**, we see that all peak values of the mutual information co-varied with RT, while the two late clusters (frontal and left-parietal) were more locked to RT [$s_{frontal} = 0.61 \pm 0.09$ (SE),

$s_{parietal} = 0.65 \pm 0.08$ (SE)] and the early cluster was more locked to stimulus [$s = 0.19 \pm 0.06$ (SE)]. This is consistent with findings of Gerson et al. (2005).

Once b and s were found, using linear fitting for each subject, we were able to identify subject-specific cluster latency for each trial and stimulus condition, given a stimulus onset and an RT.

Mutual information between EEG clusters

Next we estimated the mutual information between EEG clusters extracted from all correctly-responded target trials. As we have mentioned, discretizing EEG signals with different bin sizes can produce different results. If the bin size is too small, the detailed structure in the signal distribution might be lost; while if the bin size is too large, the estimation might be biased as there are not enough samples. In a previous analysis, we started with a bin size of eight (Luo and Sajda, 2006). In this paper, when estimating the

FIGURE 4 | Temporal evolution of the mutual information between the EEG and class label. Top: Scalp plot of the group average mutual information between EEG and the corresponding class labels across subjects. EEG is locked to the trial onset. Dotted lines indicate the time shift in response to different types of stimuli. Bottom: The average mutual information between EEG and the class labels on channels FCZ, CZ, and P1 for Paradigm 1. In all cases the unit of mutual information is the nit (natural log).

mutual information between EEG clusters, we varied the number of bins from 6 to 12 (bin size is in µV).

The dots in **Figure 7** (left) show the group average mutual information for central-frontal, central-parietal, and frontal-parietal clusters for the three types of targets. To test the significance of these results, a bootstrap procedure was also performed: (1) for each subject, we sampled *M* trials with replacement from the original data, where *M* is the number of the subject's correctly-responded trials, and then extract their first EEG cluster; (2) we repeat the method in step 1 to extract the second and the third EEG clusters, so that the three clusters generally did not come from the same trial, and computed the mutual information between each pair of these three clusters; (3) we computed the group average mutual information across subjects; and (4) repeated steps 1–3 1000 times, to calculate statistics for each

stimulus condition. The dashed lines in **Figure 7** (left) show the upper bound of the 95% confidence interval of the group average mutual information, computed via this procedure. The solid line shows the mean bootstrap result. For all nine plots (three stimulus types × three pairs of clusters), we see consistent results across different number of bins, i.e., the group average mutual information is either consistently significant [when the dots are above the dashed line in **Figure 7** (left)] or nearly significant, or consistently not significant [when the dots are below the dashed line in **Figure 7** (left)]. This shows the estimation of mutual information was not overly sensitive to the number of bins for the range we considered.

We further summarized these results in **Figure 7** (right), where for each stimulus condition two scalp plots are shown, with the first representing the early central cluster, and the second representing

the late frontal and left-parietal clusters. Clusters that had mutual information significantly above chance are connected with a solid line; otherwise they are connected with a dashed line. For CSVP targets, there was no strong connectivity among the three most discriminating clusters; while for the other two conditions, especially the FSVP case, the connectivity between the areas was significant.

DISCUSSION

In this paper, we study how the temporal correlation of the visual stimulus affects target detection. We estimated the mutual informa-

tion between the EEG and class label and identified three clusters of EEG electrodes that are most informative of target detection. We also demonstrated that these clusters were not identified using a traditional ERP analysis. In the following, we relate our findings to current cognitive theories of target detection and explain the differences in the timing of these neural activities for serial visual presentation having different temporal correlations.

VISUAL TARGET DETECTION, ATTENTIONAL AROUSAL, AND THE P300

Many cognitive models of visual target detection argue that a post-perceptual capacity-limited stage is crucial for target identification (Nieuwenhuis et al., 2005a; Shapiro et al., 1997). For example, Shapiro et al. (1994) "interference theory" assumes that all stimuli are processed to a varying degree and that they then compete, most likely during retrieval from visual short-term memory. Another model by Chun and Potter (1995) proposed that target detection involves a two-stage process, in which a pre-attentive stage of a short-lived visual representation is followed by a capacity-limited target identification/decision-making process (the second stage). The second stage is initiated by a transient attentional arousal response that actively selects the target and is followed by the identification and consolidation of the targets (decision making). Evidence shows that this post-perceptual capacity-limited stage is reflected in the P300 (Nieuwenhuis et al., 2005a). For example, it was suggested by Chun and Potter (1995) that this stage of processing happens between 200 and 500 ms, which of course overlaps with the P300. In attentional blink studies, it was found that the P300 is completely suppressed during the attentional blink period, while P1, N1, and N4 in EEG are not affected. Note that the P1 and N1 are thought to reflect early sensory processing and N4 semantic analysis (Dell'Acqua et al., 2003; Kranczioch et al., 2003; Rolke et al., 2001; Vogel et al., 1998). Further, Vogel and Luck (2002) assumed that the P300 indicates the consolidation of transient perceptual representations into a more durable short-term memory. Menon et al. (1997) suggested that the P300 marks the completion of initial sensory processing and the onset of a process of directed attention
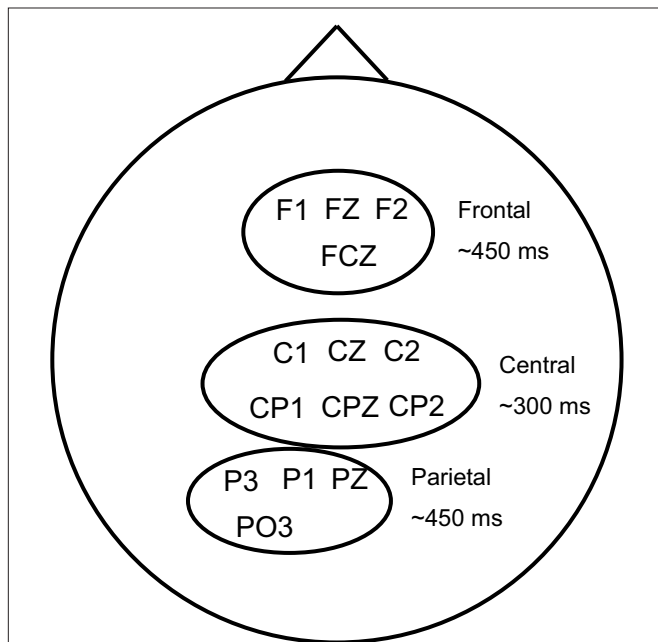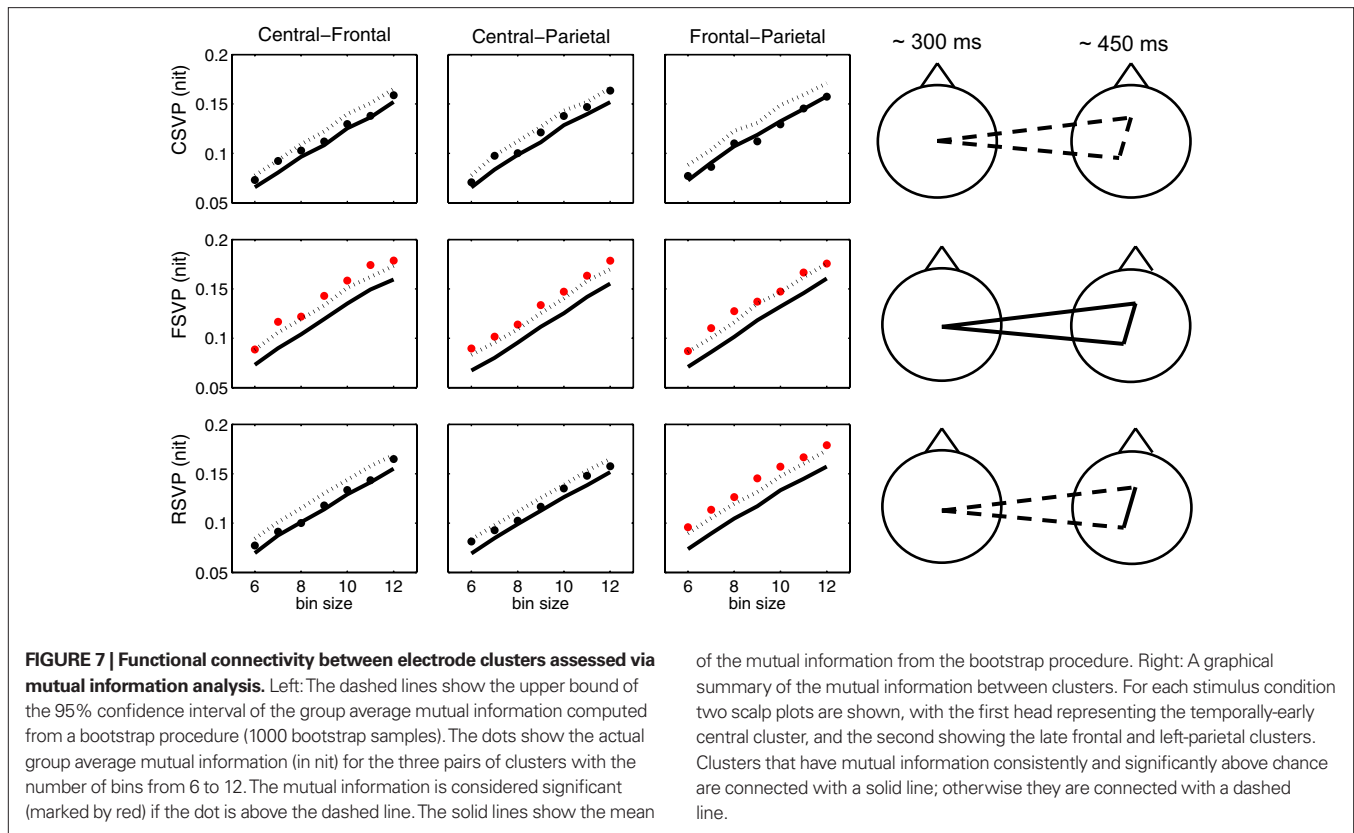

FIGURE 5 | Channels and latencies used for identification of the three electrode clusters. Latencies are approximate; they were ultimately determined by RT (see **Figure 6**).
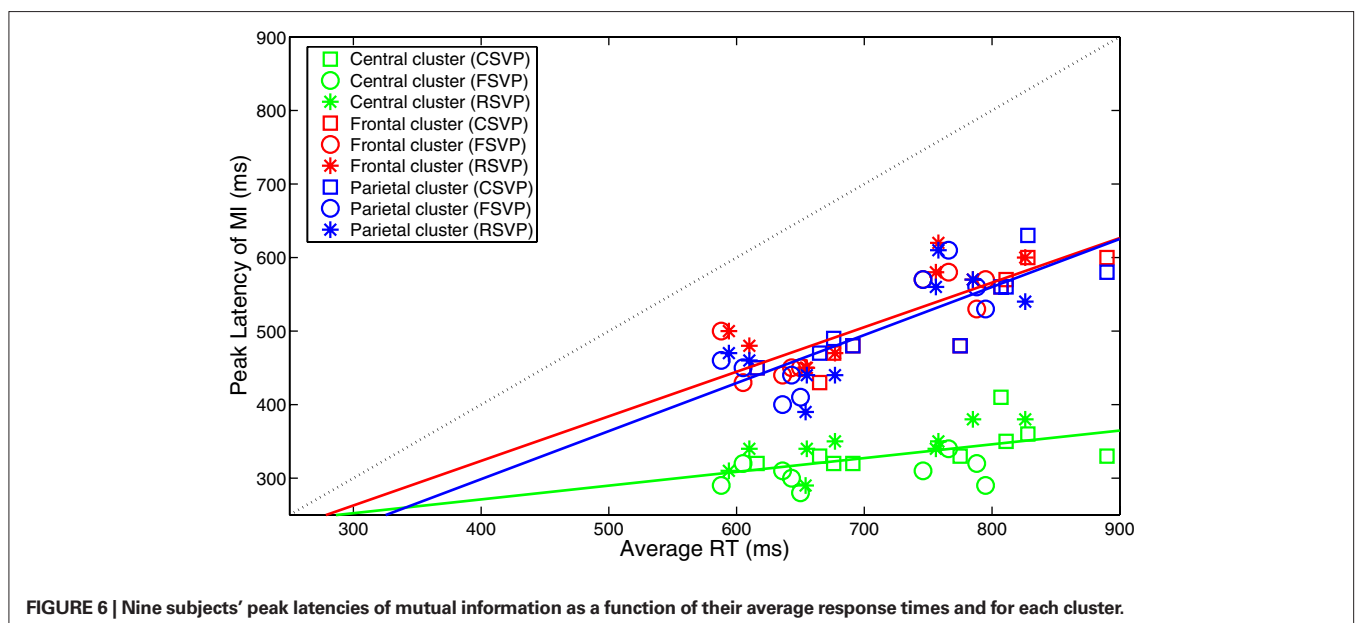
FIGURE 6 | Nine subjects' peak latencies of mutual information as a function of their average response times and for each cluster.

FIGURE 7 | Functional connectivity between electrode clusters assessed via mutual information analysis. Left: The dashed lines show the upper bound of the 95% confidence interval of the group average mutual information computed from a bootstrap procedure (1000 bootstrap samples). The dots show the actual group average mutual information (in nit) for the three pairs of clusters with the number of bins from 6 to 12. The mutual information is considered significant (marked by red) if the dot is above the dashed line. The solid lines show the mean of the mutual information from the bootstrap procedure. Right: A graphical summary of the mutual information between clusters. For each stimulus condition two scalp plots are shown, with the first head representing the temporally-early central cluster, and the second showing the late frontal and left-parietal clusters. Clusters that have mutual information consistently and significantly above chance are connected with a solid line; otherwise they are connected with a dashed line.

leading to conscious awareness of salient stimuli. Together these studies suggest a strong link between the P300 ERP, target detection and attention.

The time course and scalp topography of the discriminating activity between target and non-target conditions in our task are in general analogous to previously reported visual P300 activity (Ji et al., 1999; Sangal and Sangal, 1996), although our results show a relatively stronger frontal activity and a more negative deflection at the onset of the ERP. Most previous studies used simple visual or auditory odd-ball tasks to study the P300, while we used much more complex stimuli in which targets (i.e., people) were defined with variable position, scale and/or pose, and low level features were unlikely to be discriminative. This may account for the difference in the evoked P300 between our results and other visual odd-ball paradigms.

We find three clusters of activity related to target detection and occurring during the timing of the P300. Menon et al. (1997) speculated that a small fronto-central discriminating activity happening at around 300-ms post-stimulus may characterize the initial orienting response, and may be due to the smaller or deeper activation of the anterior cingulate. The scalp topography and timing of this activity are analogous to that of the temporally-early central cluster found in our analysis. The temporally-late clusters identified in our mutual information analysis, which have a frontal-parietal topography and happen at around 350–700 ms post-stimulus, is analogous to the previously reported P3b component [see Picton (1992) for a review] and thus may underlie the post-sensory processing of stimuli for evaluation, identification

and decision making (Menon et al., 1997). Thus, our functional connectivity analyses links the orienting response, stimulus identification, and decision making to the specific temporal correlations in the visual stimulus.

DIFFERENCES IN BOTH BEHAVIORAL RESPONSES AND EEG MEASURED NEURAL CORRELATES FOR CSVP, FSVP, AND RSVP STIMULI

One prominent difference between the stimulus conditions is that we find the RTs for CSVP targets are significantly longer than the RTs for both FSVP and RSVP targets. This finding is consistent with our first hypothesis, namely that the temporally-continuous stimulus would result in an increased RT to the target. Such a behavioral delay is also reflected in the EEG response, with both the early and late cluster activities delayed in the CSVP condition relative to the FSVP and RSVP conditions. The behavioral and electrophysiological delays we find are in agreement with Chun's two-stage model (Chun and Potter, 1995): once an item has been engaged in the consolidation stage, the time to process that item is a function of its discriminability relative to the previous item (Peterson and Juola, 2000). Thus, it takes longer for the limited-capacity second stage to identify the CSVP targets.

We can also view these differences within the context of a decision-making process. In the diffusion model of decision making (Ratcliff, 1978; Ratcliff and Rouder, 1998, 2000), two-choice decisions are described in terms of an evidence accumulation process. In this case, the decision mechanism accumulates information over time, modeled via a drift rate at which information is accumulated, and after which the evidence crosses a threshold representing one

of the two possible response criteria. From the perspective of this model, in the FSVP and RSVP conditions, the characteristics of the stimulus, namely an abrupt change in the stimulus from a target to non-target frame, forces the evidence accumulation to stop and in turn forces a decision. In terms of the diffusion model, this could be representative of an earlier starting point for the accumulation of evidence, a larger drift rate, or lower decision/response threshold, all of which could potentially results in faster RTs for these two stimulus conditions.

In terms of the neural activations differentiating the stimulus conditions, our results show that the early central cluster of activity elicited by FSVP targets is earlier than that seen for the RSVP and CSVP conditions. This is possibly an indication of an earlier orienting response which results when a target appears for a very short period of time after which attentional resources that are needed are reduced. This reconciles with the theory of Bowman and Wyble (2007), namely that targets followed by blanks trigger a stronger and faster attentional gating response.

Another factor that may contribute to the earlier neural component seen in the FSVP compared to the RSVP and CSVP conditions is that for the FSVP condition the stimuli before a target presentation are also blanks frames. Thus at the time of target onset, visual storage is empty and ready for new inputs (Chun and Potter, 1995). Smith et al. (2004) found that inattention delays the entry of stimuli into short-term memory. This may also account for the delayed central component seen in the RSVP and CSVP conditions.

Although the statistics of RTs to RSVP targets is about the same as that of FSVP stimuli, RSVP's MR is significantly higher than that of FSVP. This result is consistent with our second hypothesis that when a short presentation of a target is masked by non-relevant natural image distractors, its detectability will be reduced. The neural correlate of this may be the earlier reduction of the discriminating activity in the parietal cortex at around 450-ms post-stimulus in RSVP case, reflective of a loss of the target representation in visual storage, or a reduction in iconic memory (Coltheart, 1980; Sperling, 1960). The masking (i.e., overwriting of a visual buffer) by distractor images that follow the target results in a mismatch of the incoming input (distracting images) to the internal representation of targets in RSVP condition. This is consistent with Smith et al. (2004) and Kawahara et al. (2001), that visual masks limit the visual persistence of stimuli. Similarly, in an attentional blink study, the authors found that distractor images following T1 resulted in poorer detection of T1, compared to blank images (Chun and Potter, 1995; Raymond et al., 1992). In contrast, it seems that in the FSVP condition the visual input is allowed to persist in iconic memory. In Philiastides et al. (2006) and Philiastides and Sajda (2007), the authors used an FSVP stimulus design and showed that neural activity of a "late neural component" reflective of evidence accumulation within a stored representation of the stimulus was in fact localized to visual areas previously reported to be the site of iconic memory and visual persistence.

FUNCTIONAL CONNECTIVITY DIFFERENCES BETWEEN CSVP AND FSVP/RSVP PROCESSING

One of the attractive aspects of using mutual information as a quantitative metric for analyzing EEG is that it can be used to study the functional connectivity between EEG measured at different scalp locations. For CSVP, we found that the connectivity between the three most discriminating clusters is weaker than that for the FSVP and RSVP cases. This reconciles with Honey et al. (2002), where the authors found that during a working memory task the connectivity between prefrontal cortex and parietal area increases, and speculated that it reflects greater demand for maintenance and executive processes. In the CSVP condition, subjects do not have to recruit memory stores but instead can rely on the instantaneous input to form a decision. For FSVP, there are strong connections among all three clusters, which may also contribute to the improved performance (shorter RTs and lower MRs) in this case.

CONSISTENCY WITH PREVIOUS STUDIES

In a two-letter identification experiment (Ratcliff and Rouder, 2000), the authors varied the stimulus duration from 12 to 84 ms and found as stimulus duration increases, responses are faster and more accurate. This is in contrast to our results that show responses are slower for stimuli having longer duration (such as CSVP targets). However, we believe these two results do not conflict. The first and more obvious explanation is that the span of the stimulus duration for the two experiments does not overlap. The second and more interesting explanation reflects basic properties of simple decision making. As the difficulty of the decision increases, accuracy decreases and RTs get longer. This is reflected in Ratcliff and Rouder (2000): as the signal strength decreases with shorter presentation durations, so too does the rate of information accumulation, producing a decrease in accuracy and an increase in RT. However, subjects can trade-off speed and accuracy and respond rapidly and inaccurately, or slowly and accurately, as the task requires (Smith et al., 2004). This trade-off can also be affected by the very nature of the stimulus presentation, for example the abrupt ending of a stimulus forces a faster RT at the cost of accuracy. This is consistent with our finding that in the CSVP condition subjects are able to accumulate more information to make slow but accurate decisions, while in the RSVP case, the sudden change in the image characteristics (target vs. distractor) forces a decision which is also accompanied by masking of the visual representation of the target image in short-term memory. The result is then subjects make faster responses with more errors.

CONCLUSIONS

In this paper, we study the effects that the temporal correlation in visual stimuli has on one's behavioral and electrophysiological responses for target detection. We report longer reaction times for CSVP compared to the FSVP and RSVP conditions, and a difference in MR between the FSVP and RSVP conditions. Using mutual information, we find neural correlates of these behavioral observations. These neural correlates were not identified by traditional ERP analysis. We also investigate functional connectivity between clusters of electrodes and find significant fronto-parietal functional coupling for the FSVP and RSVP conditions but no significant coupling during the CSVP condition. These findings suggest that the interaction between visual short-term memory and the visual input can impact the target detection process.

SUPPLEMENTARY MATERIAL

Supplementary material can be found online at htpp://www.frontiersin.org/humanneuroscience/paper/10.3389/neuro.09.005.2009/.

REFERENCES

Borst, A., and Theunissen, F. E. (1999). Information theory and neural coding. *Nat. Neurosci.* 2, 947–957.

Bowman, H., and Wyble, B. (2007). The simultaneous type, serial token model of temporal attention and working memory. *Psychol. Rev.* 114, 38–70.

Broadbent, D. E., and Broadbent, M. H. P. (1987). From detection to identification: response to multiple targets in rapid serial visual presentation. *Percept. Psychophys.* 42, 105–113.

Chen, C., Hsieh, J., Wu, Y., Lee, P., Chen, S., Niddam, D. M., Yeh, T., and Wu, Y. (2007). Mutual-information-based approach for neural connectivity during self-paced finger lifting task. *Hum. Brain Mapp.* 29, 265–280.

Chun, M. M., and Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *J. Exp. Psychol. Hum. Percept. Perform.* 21, 109–127.

Coltheart, M. (1980). Iconic memory and visible persistence. *Percept. Psychophys.* 27, 183–228.

Dell'Acqua, R., Jolicoeur, P., Pesciarelli, F., Job, C. R., and Palomba, D. (2003). Electrophysiological evidence of visual encoding deficits in a cross-modal attentional blink paradigm. *Psychophysiology* 40, 629–639.

Gerson, A. D., Parra, L. C., and Sajda, P. (2005). Cortical origins of response time variability during rapid discrimination of visual objects. *Neuroimage* 28, 342–353.

Honey, G. D., Fu, C. H. Y., Kim, J., Brammer, M. J., Croudance, T. J., Suckling, J., Pich, E. M., Williams, S. C. R., and Bullmore, E. T. (2002). Effects of verbal working memory load on corticocortical connectivity modeled by path analysis of functional magnetic resonance imaging data. *Neuroimage* 17, 573–582.

Hruby, T., and Marsalek, P. (2003). Event-related potentials – the P3 wave. *Acta Neurobiol. Exp.* 63, 55–63.

Jeong, J., Gore, J. C., and Peterson, B. S. (2001). Mutual information analysis of the EEG in patients with Alzheimer's disease. *Clin. Neurophysiol.* 112, 827–835.

Ji, J., Porjesz, B., Begleiter, H., and Chorlian, D. (1999). P300: the similarities and differences in the scalp distribution of visual and auditory modality. *Brain Topogr.* 11, 315–327.

Kawahara, J., Di Lollo, V., and Enns, J. T. (2001). Attentional requirements in visual detection and identification: evidence from the attentional blink. *J. Exp. Psychol. Hum. Percept. Perform.* 27, 969–984.

Keysers, C., Xiao, D. K., Földiák, P., and Perrett, D. I. (2001). The speed of sight. *J. Cogn. Neurosci.* 13, 90–101.

Kranczioch, C., Debener, S., and Engel, A. K. (2003). Event-related potential correlates of the attentional blink phenomenon. *Brain Res. Cogn. Brain Res.* 17, 177–187.

Luo, A., and Sajda, P. (2006). Using single-trial EEG to estimate the timing of target onset during rapid serial visual presentation. 28th IEEE EMBS Annual International Conference, New York, pp. 79–82.

Menon, V., Ford, J. M., Lim, K. O., Glover, G. H., and Pfefferbaum, G. H. (1997). Combined event-related fMRI and EEG evidence for temporal-parietal cortex activation during target detection. *Neuroreport* 8, 3029–3037.

Naumann, E., Huber, C., Maier, S., Plihal, W., Wustmans, A., Diedirch, O., and Bartussek, D. (1992). The scalp topography of P300 in the visual and auditory modalities: a comparison of three normalization methods and the control of statistical type II error. *Electroencephalogr. Clin. Neurophysiol.* 83, 254–264.

Nieuwenhuis, S., Gilzenrat, M. S., Holmes, B. D., and Cohen, J. D. (2005a). The role of the locus coeruleus in mediating the attentional blink: a neurocomputational theory. *J. Exp. Psychol. Gen.* 134, 291–307.

Nieuwenhuis, S., Aston-Jones, G., and Cohen, J. D. (2005b). Decision making, the P3, and the locus coeruleus–norepinephrine system. *Psychol. Bul.* 131, 510–532.

Öquist, G., Hein, A. S., Ygge, J., and Goldstein, M. (2004). Eye movement study of reading on a mobile device using the page and RSVP text presentation formats. Proceedings of the 6th International Symposium on Mobile Human–Computer Interaction (MobileHCI), Glasgow, pp. 108–119.

Parra, L., Spence, C., Gerson, A., and Sajda, P. (2005). Recipes for the linear analysis of EEG. *Neuroimage* 28, 326–341.

Peterson, M. S., and Juola, J. F. (2000). Evidence for distinct attentional bottlenecks in attention switching and attentional blink tasks. *J. Gen. Psychol.* 127, 6–26.

Philiastides, M. G., Ratcliff, R., and Sajda, P. (2006). Neural representation of task difficulty and decision-making during perceptual categorization: a timing diagram. *Neuroscience* 26, 8965–8975.

Philiastides, M. G., and Sajda, P. (2007). EEG-informed fMRI reveals spatio-temporal characteristics of perceptual decision making. *J. Neurosci.* 27, 13082–13091.

Picton, T. W. (1992). The P300 wave of the human event-related potential. *J. Clin. Neurophysiol.* 9, 456–479.

Potter, M. C., and Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *J. Exp. Psychol.* 81, 10–15.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychol. Rev.* 85, 59–108.

Ratcliff, R., and Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychol. Sci.* 9, 347–356.

Ratcliff, R., and Rouder, J. N. (2000). A diffusion model account of masking in two-choice letter identification. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 127–140.

Raymond, J. E., Shapiro, K. L., and Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: an attentional blink? *J. Exp. Psychol. Hum. Percept. Perform.* 18, 849–860.

Rolke, B., Heil, M., Streb, J., and Hennighausen, E. (2001). Missed prime words within the attentional blink evoke an N400 semantic priming effect. *Psychophysiology* 38, 165–174.

Rozell, C. J., and Johnson, D. H. (2005). Examining methods for estimating mutual information in spiking neural systems. *Neurocomputing* 65–66, 429–434.

Sangal, B., and Sangal, J. M. (1996). Topography of auditory and visual P300 in normal adults. *Clin. Electroencephalogr.* 27, 145–150.

Shapiro, K. L., Arnell, K. M., and Raymond, J. E. (1997). The attentional blink. *Trends Cogn. Sci.* 1, 291–296.

Shapiro, K. L., Raymond, J. E., and Arnell, K. M. (1994). Attention to visual pattern information produces the attentional blink in RSVP. *J. Exp. Psychol. Hum. Percept. Perform.* 20, 357–371.

Smith, P. L., Ratcliff, R., and Wolfgang, B. J. (2004). Attention orienting and the time course of perceptual decisions: response time distributions with masked and unmasked displays. *Vision Res.* 44, 1297–1320.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied* 74, 1–30.

Squires, K., Wickens, C., Squires, N., and Donchin, E. (1976). The effect of stimulus sequence of the waveform of the cortical event-related potential. *Science* 193, 1142–1146.

Tanaka, F., Kachi, T., Yamada, T., and Sobue, G. (1998). Auditory and visual event-related potentials and flash visual evoked potentials in Alzheimer's disease: correlations with mini-mental state examination and Raven's coloured progressive matrices. *J. Neurolog. Sci.* 156, 83–88.

Vogel, E. K., and Luck, S. J. (2002). Delayed working memory consolidation during the attentional blink. *Psychon. Bull. Rev.* 9, 739–743.

Vogel, E. K., Luck, S. J., and Shapiro, K. L. (1998). Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. *J. Exp. Psychol. Hum. Percept. Perform.* 24, 1656–1674.