NeuroImage 216 (2020) 116870

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Visual target detection in a distracting background relies on neural encoding of both visual targets and background

Cheng Luo [a], Nai Ding [a,b,*]

[a] *Key Laboratory for Biomedical Engineering of Ministry of Education, College of Biomedical Engineering and Instrument Sciences, Zhejiang University, Hangzhou, 310027, China*
[b] *Research Center for Advanced Artificial Intelligence Theory, Zhejiang Lab, Hangzhou, 311121, China*

ARTICLE INFO

Keywords:
Visual attention
EEG
Natural stimulus
Individual difference

ABSTRACT

The ability to detect visual targets in complex background varies across individuals and are affected by factors such as stimulus saliency and top-down attention. Here, we investigated how the saliency of visual background (naturalistic cartoon video vs. blank screen) and top-down attention (single vs. dual tasks) separately affect individual ability to detect visual targets. Behaviorally, we found that target detection accuracy decreased and reaction time elongated when the background was salient or during dual tasking. The EEG response to visual background was recorded using a novel stimulus tagging technique. This response was strongest in occipital electrodes and was sensitive to background saliency but not dual tasking. In contrast, the event-related potential (ERP) evoked by the visual target was strongest in central electrodes, and was affected by both background saliency and dual tasking. With a cartoon background, the EEG responses to visual targets, presented in the central visual field, and the EEG responses to peripheral visual background could both predict individual target detection performance. When these two responses were combined, better prediction was achieved. These results suggest that neural processing of visual targets and background jointly contribute to individual visual target detection performance.

1. Introduction

The natural visual environment is highly dynamic and complex, containing an overwhelming amount of information. Effective processing of the natural visual environment requires selective attention, which is modulated by both top-down goal-driven factors and bottom-up stimulus-driven factors (Desimone and Duncan, 1995; Egeth and Yantis, 1997; Yantis, 2000). On the one hand, attention can be voluntarily oriented to spatial locations (Moran and Desimone, 1985) or visual features (Treue and Trujillo, 1999) to speed up behavioral responses (Posner, 1980; Rossi and Paradiso, 1995; Saenz et al., 2003; Theeuwes and Van der Burg, 2007). Top-down visual attention engages broad cortical networks, including both early visual areas and frontoparietal attention areas (Buschman and Miller, 2007; Gregoriou et al., 2009). In terms of the neural dynamics, attention can modulate both early and late neural responses. For example, human event-related potential (ERP) studies have shown that top-down attention modulates both early visual processing indexed by the N1 and P1 components (Hillyard and Anllo-Vento, 1998; Hillyard et al., 1973; Rugg et al., 1987) and late processing indexed by the P3 component (Hopfinger and West, 2006).

On the other hand, attention is often inevitably distracted by salient task-irrelevant interference (Theeuwes, 1991, 2004), a mechanism that is crucial to monitor unexpected dangers. In general, top-down and bottom-up attention jointly create a biased representation of the external visual world according to both salience and behavioral relevance (Beste et al., 2011; Desimone and Duncan, 1995; Knudsen, 2007). In visual search experiments, for example, salient task-irrelevant stimuli affect eye movements (Theeuwes et al., 1998) and slow down visual search (Theeuwes, 2004). Similarly, neural responses to task-irrelevant peripheral stimuli that share features with the target stimuli are partially enhanced (Andersen et al., 2011; Painter et al., 2014; Serences et al., 2005). Recent evidence also suggests that attention spontaneously

---

Abbreviations: EEG, electroencephalography; ERP, event-related potential; SSVEP, steady-state visual evoked potential; TRF, temporal response function; EOG, electrooculogram; PCA, principal component analysis; FDR, false discovery rate; SD, standard deviation.

* Corresponding author. Key Laboratory for Biomedical Engineering of Ministry of Education, College of Biomedical Engineering and Instrument Sciences, Zhejiang University, 310027, China.

E-mail address: ding_nai@zju.edu.cn (N. Ding).

https://doi.org/10.1016/j.neuroimage.2020.116870
Received 12 November 2019; Received in revised form 11 April 2020; Accepted 17 April 2020

switches between the attended and unattended visual objects (Jia et al., 2017). Task-irrelevant stimuli, however, do not always capture attention (Bettencourt and Xu, 2016; Forster and Lavie, 2008; Lavie, 1995; Rademaker et al., 2019; Sprague et al., 2018). The sensitivity to task-irrelevant also varies across individuals (Fukuda and Vogel, 2009; Gaspar et al., 2016; Sawaki et al., 2017), while the underlying neural correlates remain unclear.

Most previous experiments investigate top-down and bottom-up attention using artificial visual stimuli and focus on the group-level response. Here, we studied how a distracting naturalistic video played in the background interfered with a visual detection task (Fig. 1) and explored the neural correlates of individual visual target detection performance. The background video was chosen to be a cartoon to simulate a common visual distractor in daily life. How the brain processes natural scenes is a fundamental question (Peelen et al., 2009; Vinje and Gallant, 2000), but it is difficult to quantify the neural encoding of natural visual stimulus. Here, we extracted the neural responses to natural videos using a novel stimulus tagging technique (Fig. 1C) that built on the steady-state visual evoked potential (SSVEP) (Regan, 1966) and temporal response function (TRF) (Lalor et al., 2006).

In the experiment, to manipulate bottom-up saliency of the visual background, we compared the condition that presented a cartoon video with a control condition that presented a blank screen. To manipulate top-down attention, we compared a condition in which participants fully focused on the visual target detection task with a condition in which attention was split between the target detection task and a cartoon watching task. The visual target to detect was defined based on both its spatial position and its color, and therefore the task engaged both spatial and feature-based attention. We recorded neural activity using electroencephalogram (EEG). With the stimulus tagging technique, we recorded neural responses to the visual background simultaneously with the ERP to visual targets and other distractors, which allowed us to analyze how each response was modulated by top-down and bottom-up factors and which response could better explain individual behavior.

2. Materials and methods

2.1. Participants

Fifty adults participated in the experiment (20–26 years old, mean age, 22.1 years; 19 females). All participants had normal or corrected-to-normal vision and had no history of psychiatric or neurological disorders. One participant was removed because of excessive noise in EEG data. Consequently, the data from a total of 49 participants were analyzed. The experimental procedures were approved by the Research Ethics Committee of the College of Medicine, Zhejiang University (2019-047). All participants provided written informed consent prior to the start of the experiment and obtained reward contingent on their task performance.

2.2. Visual stimulus

2.2.1. Configuration

The configuration of the visual stimulus and its size were shown in Fig. 1A. A green cross (RGB: 0, 155, 0) was displayed at the center. The size of the visual stimulus was chosen so that participants could clearly see the whole visual stimulus when fixating at the cross. The circular region for visual presentation was further divided into a central region (1.4° visual angle) and a peripheral region (2.25°–9.5° visual angle), which were separated by a 0.85° annular region.

2.2.2. Background

The whole circular region for visual presentation was filled with either a blank screen, i.e., a static gray screen (RGB: 109, 109, 109) or a grayscale cartoon video. The cartoon video was a clip from the famous cartoon "Tom and Jerry" (60 Hz refresh rate). The grayscale of the static gray screen was the same as the averaged grayscale of the cartoon video (Fig. 1B).

We applied a stimulus tagging technique to separately characterize the neural encoding of the central and peripheral parts of the

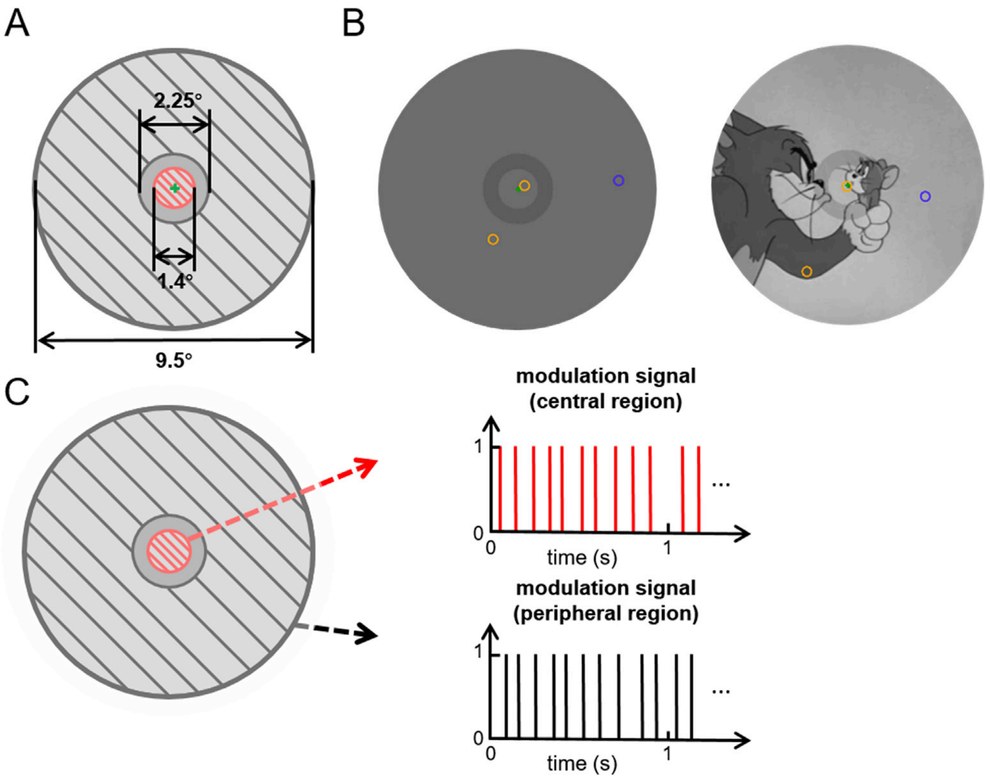

Fig. 1. Visual stimuli. (A) Stimulus configuration. The visual stimulus is presented in a circular region that divides into a central (red hatched lines) and a peripheral region (black hatched lines) that are separated by an annular region. A green fixation cross is shown at the center. (B) Background, distractors, and targets. The background is either a static gray screen (left) or a cartoon video (right). Colored circles are displayed on top of the background. Up to 3 circles can be displayed simultaneously, at most 1 in the central region and 2 in the peripheral region. Circles are either yellow or purple. Yellow circles appear with 30% probability, and only yellow circles in the central region are the targets to detect. (C) Response tagging paradigm. The background is separately tagged in the central and peripheral regions. To tag the neural response, the luminance of the background is modulated by a binary signal. When the modulation signal is 1, the background is replaced by a black frame (RGB: 0, 0, 0). When the modulation signal is 0, the background is not changed. Neural encoding of the background is characterized by the response to the black frames.

background. To achieve this, the luminance of the central and peripheral background was modulated by a binary modulation signal (Fig. 1C). When the modulation signal was 1, the background was replaced by a black frame (the whole region was set to RGB: 0, 0, 0). When the modulation signal was 0, the background was not changed. Specifically, given the refresh rate was 60 Hz, for the central background, in every 60 frames, 10 random nonadjacent frames were replaced by black frames. The same operation was applied to the peripheral background with the only constraint that the central background and the peripheral background would not be replaced by black frames at the same time. Therefore, the central background and the peripheral background flickered separately. Since the background modulation reduced the luminance of the central and peripheral region, the grayscale value of the annular region, which did not flicker, was adjusted to be 5/6 of its original level to achieve similar luminance to the central and peripheral regions.

2.2.3. *Target and distractor*

As shown in Fig. 1B, up to 3 circles could be displayed simultaneously, at most 1 in the central region and 2 in the peripheral region. In the central region, the colored circle was occasionally displayed at a random location. The interval between its appearances was randomized between 500 ms and 1000 ms (uniform distribution) and each circle was displayed for 200 ms each time. Each circle had two colors. The yellow circle appeared with a 30% probability and would not appear in succession. The purple one appeared with a 70% probability. In the peripheral region, two colored circles might be simultaneously presented at random locations. The two circles were independent from each other and the appearance time and color followed the same statistical properties as the circles in the central region. Only the yellow circles in the central region were defined as the target (10% of all circles) while all other circles were distractors including the yellow circles in the peripheral region.

2.3. *Experimental procedures and tasks*

The experiment was divided into three conditions. In the first condition, referred to as the blank(detection) condition, the participants performed a visual target detection task and the visual background was a blank screen. In the second condition, referred to as the cartoon(detection) condition, participants performed the same task while the visual background was a cartoon video. In the third condition, referred to as the cartoon(dual) condition, the participants split attention between a visual target detection task and a cartoon watching task. In the cartoon(detection) and cartoon(dual) condition, different cartoon videos were displayed in the visual background. Each stimulus in each condition lasted 12 min.

In each experimental condition, participants were requested to fixate at the central cross. They had to covertly monitor the colored circles in the central region while ignoring the colored circles in the peripheral region. Only the yellow circle appeared at the central region was the target, and the targets totally appeared 288 times in each experimental condition. Participants had to press a button as soon as possible when a target was spotted.

In the blank(detection) condition and cartoon(detection) condition, participants only performed the target detection task, and were asked to ignore the visual background. In the cartoon(dual) condition, however, participants were asked to split attention to the background cartoon while performing the target detection task. After the cartoon(dual) condition, participants had to answer 12 questions about the cartoon video. After the cartoon(detection) condition, participants also had to answer 12 questions about the cartoon video they saw in that condition, which they did not expect. Before each condition, participants were informed which task they should perform, e.g., the detection task or the dual task. In the blank(detection) and cartoon(detection) conditions, participants were rewarded only based on their target detection accuracy. In the cartoon(dual) condition, however, they were rewarded based on the average of their target detection accuracy and their question answering accuracy (both in %). The order of the three conditions were counterbalanced across participants.

A training section, which lasted 30 s, was given before each experimental condition. For the blank(detection) and cartoon(detection) condition, participants passed the training section if their target detection accuracy was higher than 80%. For the cartoon(dual) condition, they also had to correctly answer 3 questions about the cartoon video. The training section was repeated until the participant passed it.

In sum, in the blank(detection) and cartoon(detection) conditions, participants performed a detection task and we manipulated the saliency of background, i.e., a blank screen or a cartoon video. In the cartoon(detection) and cartoon(dual) conditions, cartoon videos were displayed in the background and we manipulated the task from a detection task to dual tasks.

2.4. *Data recording*

Electroencephalogram (EEG) and Electrooculogram (EOG) were recorded continuously using a Biosemi ActiveTwo system. Sixty-four EEG electrodes were recorded. Two electrodes were placed at the left and right temples and another two electrodes were placed above and below the right eye for recording horizontal EOG (right minus left) and vertical EOG (upper minus lower). Two additional electrodes were placed at the left and right mastoids and their average was the reference for EEG. The EEG/EOG recordings were sampled at 2048 Hz and low-pass filtered below 400 Hz. The EEG recordings were referenced to the average mastoid recording off-line and band-pass filtered between 0.5 Hz and 30 Hz using a linear-phase finite impulse response (FIR) filter, and the latency of the filter was compensated. To remove EOG artifacts in EEG, the horizontal and vertical EOG were regressed out using the least-squares method.

Eyetracking data was recorded using a combined pupil and corneal reflection eye tracker at 500-Hz sampling rate (Eyelink Portable Duo, SR Research, Mississauga, Ontario, Canada). The participants were seated 88 cm from a monitor with their chin resting on a chin rest. At the beginning of each experimental condition, a 9-point (3 × 3 square) calibration and validation was applied and the experimenter had to confirm that the fixation of participants was detected properly. Eyetracking data was only recorded during the task. The eyetracking data showed that all participants maintained good fixation on the central cross (within 1.22° visual angle) throughout the experiment.

2.5. *EEG data analysis*

The EEG recordings were downsampled to 120 Hz. The neural responses to the colored circles were characterized by ERP evoked by the colored circles. Thus, the continuously recorded data was divided into epochs of 1500-ms length starting 200 ms before the onset of the yellow circles or the purple circles in the central region and peripheral region, and then grand averaged. As we applied the stimulus tagging technique, i.e., embedding the black frames as separate markers to tag the background, the neural responses to the central background and the peripheral background were separately characterized by ERP evoked by the black frames in the central region and the peripheral region. Thus, the continuously recorded data was also divided into epochs of 1500-ms length starting 200 ms before the onset of black frames in the central region and peripheral region, and then grand averaged. For each EEG electrode, the mean amplitude of a 200 ms baseline interval was subtracted at all time-points. Based on the response topography (Fig. 3), two clusters of electrodes were selected to analyze the responses to colored circles and visual background: Five central EEG electrodes (Cz, CPz, Pz, CP1, and CP2) were employed to analyze the ERP evoked by colored circles. Five occipital EEG electrodes (O1, Oz, O2, Iz, and POz) were employed to analyze the ERP evoked by central and peripheral background. The choice of electrodes was also consistent with the ERP analysis of previous studies (Kutas et al., 1977; Müller et al., 2007).

2.6. Regression analysis

A linear regression model was used to predict the individual behavioral performance based on different ERP responses, e.g., the response to the targets and the response to peripheral background. For each kind of ERP and for each participant, three electrodes with the highest peak amplitude were chosen within the five electrodes mentioned in section 2.5. For each electrode, a 650-ms response (i.e., 78 samples) starting from time 0 was used in the regression analysis. Data from the three electrodes (234 samples in total) was concatenated and reduced in dimension using the principal component analysis (PCA). The first 10 principal components were retained for the regression analysis. A ridge regression was used to predict detection performance, i.e., reaction time or detection accuracy, based on the 10 principal components. The prediction accuracy of the regression model was measured by the Pearson correlation with boxplot rule (Pernet et al., 2013) between the predicted behavioral results and the actual behavioral results. The regression procedure was separately applied to each condition. The number of EEG electrodes, the number of principal components, and the regularization parameter of the ridge regression were optimized to have a highest prediction accuracy averaged across conditions and participants using 7-fold cross validation.

Additionally, we also decoded behavioral performance using combined neural features. In this analysis, the EEG responses to targets and peripheral background were concatenated to predict individual performance. The concatenated data (468 samples in total) was reduced in dimension by PCA. The first 10 principal components were retained for regression analysis mentioned above to compute the prediction accuracy.

The regression analyses described in previous paragraphs integrated neural responses over channels and time to predict behavioral performance. In an additional time-resolved regression analysis, the analysis window was restricted to narrow time windows. The window length was 100 ms (i.e., 12 samples), and the interval between adjacent windows was 50 ms. Furthermore, since different electrodes might be active at different latency, in this analysis we did not restrict the channels to the 5 channels described in section 2.5. Three channels with the highest amplitude were selected in each narrow time window for the regression analysis. Other procedures were the same as the regression analysis for a single long time interval.

2.7. Statistical tests

The difference between response latency was examined using a Jackknife approach. We first calculated the peak latency for the EEG responses of each participant and then calculated the standard deviation using the following Jackknife procedure: We iteratively removed one participant and computed the latency based on the remaining participants. A t-value and the corresponding significance level was calculated based on the latency averaged over all participants and the standard deviation of the latency calculated using the Jackknife procedure (Miller et al., 1998).

In the regression analysis, whether the prediction accuracy was better than chance level was tested using a permutation test. Specifically, the chance level of prediction accuracy was estimated by shuffling the behavioral results across the forty-nine participants and applying the regression analysis to predict the shuffled behavioral results, which do not match the EEG data. The shuffling was applied 1000 times. If the actual prediction accuracy exceeded the 95% percentile of the chance-level prediction accuracy, it was considered statistically significant ($P < 0.05$).

The difference between the prediction accuracy of different regressor, e.g., the regressor based on a single ERP or the regressor based on combined ERPs, was evaluated using bootstrap (Efron and Tibshirani, 1994). Specifically, the EEG data and behavioral results of all participants were resampled 1000 times with replacement. In each time, the EEG data sampled was used to predict the corresponding behavioral results, and then computed the prediction accuracy. Therefore, a total of 1000 prediction accuracy were produced. The significance level of difference between the prediction accuracy using the combined responses and the respective response was determined as follows: if Ns out of the 1000 difference values were greater (or smaller) than 0, the significance level (p-value) was $Ns/1000$. When multiple comparisons were performed, the p-value was further adjusted using the false discovery rate (FDR) correction (Benjamini and Hochberg, 1995).

3. Results

3.1. Behavioral results

The reaction time, target detection accuracy, and the correction rate for cartoon question answering were shown in Fig. 2. When the cartoon video was played in the background, the reaction time significantly elongated ($P < 0.001$, paired *t*-test, FDR corrected) and the detection accuracy significantly decreased ($P < 0.001$, paired *t*-test, FDR corrected). These results suggested that the naturalistic cartoon background affected the detection of visual targets. Furthermore, the mean correction rate for answering questions was about 46.5% in the cartoon(detection) condition, indicating involuntarily spread of attention to the cartoon background.

When the dual tasking was engaged, the reaction time significantly elongated ($P < 0.001$, paired *t*-test, FDR corrected), the detection accuracy significantly decreased ($P = 0.001$, paired *t*-test, FDR corrected), and the correction rate for answering questions significantly increased ($P < 0.001$, paired *t*-test, FDR corrected). These differences suggested that the division of attention affected visual target detection. Therefore, both the bottom-up saliency of the cartoon background and the top-down dual tasking modulated attention, and had significant effects on the detection performance at group level.

3.2. EEG responses to colored circles

The grand-averaged ERP to colored circles were shown in Fig. 3A–D. The ERP evoked by yellow circles in the central region, i.e., targets, had a prominent P300 component in each condition. In terms of the topography, the P300 component was strongest in central electrodes. The P300 latency did not differ between conditions (blank vs. cartoon: $t_{(48)} = 1.44$, $P = 0.15$; detection vs. dual: $t_{(48)} = 0.841$, $P = 0.40$, Jackknife test). However, the amplitude of P300 component was reduced by both the cartoon video (comparing green and blue curves in Fig. 3A) and dual tasking (blue and orange curves in Fig. 3A). Furthermore, in an earlier time window around 280 ms, a significant ERP difference was observed between conditions with different visual background (comparing green and blue curves in Fig. 3A).

The ERP evoked by peripheral distractors that had the same color as targets showed a negative response when the background was a blank screen (Fig. 3C). The ERP response was strongest in central electrodes. The amplitude of this negativity diminished when cartoon video was presented. The ERP evoked by distractors that have different color from the targets barely showed deviated from the baseline (Fig. 3CD), and had no significant difference between conditions ($P > 0.327$, paired *t*-test, FDR corrected). These results suggested that color was used a cue to allocate attention in the current experiment.

3.3. EEG responses to background

In the experiment, we applied a stimulus tagging technique and embedded black frames as markers to separately tag the central and peripheral parts of the background. The neural responses to the central and peripheral black frames were presented in Fig. 3EF, which was strongest in occipital electrodes. The ERP to central background differed between conditions in a broad window between 180-ms to 400-ms latency (Fig. 3E). The ERP to peripheral background differed between the

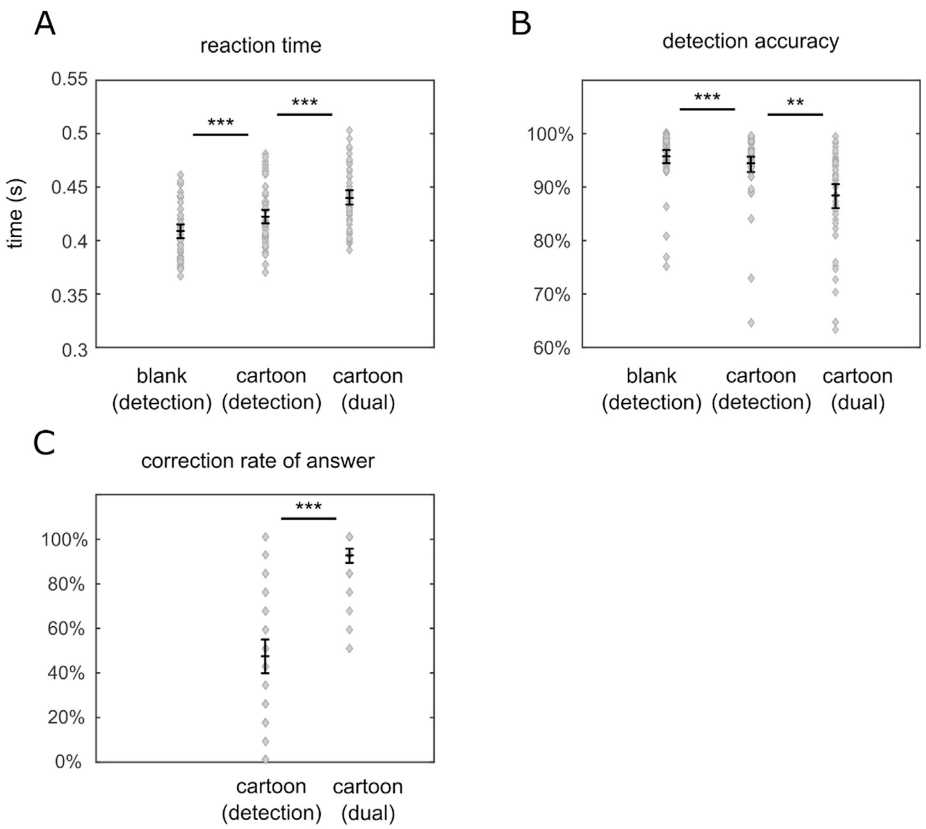

Fig. 2. Behavioral results. Center lines indicate the mean, and error bars indicate 95% confidence interval on each side. The gray points denote results from individual participants. Statistical differences between the blank(detection) condition and the cartoon(detection) condition, and differences between the cartoon(detection) condition and the cartoon(dual) condition are tested. (A) Reaction time. The reaction time is significantly longer when the cartoon video is presented or during dual tasking. (B) Target detection accuracy. The detection accuracy is significantly lower when the cartoon video is presented or during dual tasking. (C) Correction rate for cartoon question answering. The correction rate is significantly higher when attention is partly directed to the cartoon watching task. Significant differences are indicated by black stars. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ (paired *t*-test, FDR corrected).

blank(detection) and cartoon(detection) conditions (Fig. 3F) and was largely insensitive to dual tasking.

3.4. Predicting individual performance using EEG responses

The neural correlates of individual behavioral performance were characterized using a linear regression analysis. Linear regression was applied to predict the individual reaction time and detection accuracy based on different EEG responses. The prediction accuracy, i.e., the Pearson correlation between the predicted behavioral results and the actual behavioral results, was shown in Fig. 4A. The ERP evoked by targets, which had a prominent P300 component, could predict the individual reaction time (blank(detection): $P = 0.013$, cartoon(detection): $P = 0.001$, and cartoon(dual): $P = 0.003$, permutation test) and detection accuracy (blank(detection): $P = 0.020$, cartoon(detection): $P = 0.016$, and cartoon(dual): $P < 0.001$, permutation test) in all three conditions.

The ERP evoked by peripheral background could also predict the individual reaction time (cartoon(detection): $P = 0.016$, and cartoon(dual): $P = 0.002$, permutation test) and detection accuracy (cartoon(detection): $P = 0.008$, and cartoon(dual): $P = 0.016$, permutation test) when the cartoon video was played as the background. Other neural responses, however, could not predict individual performance in any condition ($P > 0.121$, permutation test). For the two responses that could significantly predict behavior, the results from individual participants were shown in Fig. 4BC.

We further tested whether the behavioral difference between conditions could be predicted by the neural responses in single conditions. The results were shown in Fig. 4D. The results showed that the ERP evoked by peripheral background could predict how the background influenced detection accuracy ($P = 0.001$, permutation test), i.e., the difference between the blank(detection) and cartoon(detection) conditions. The ERP evoked by targets could predict how the background influenced reaction time ($P = 0.002$, permutation test) and how the task influences behavioral reaction time ($P = 0.017$, permutation test) and detection accuracy ($P = 0.001$, permutation test), i.e., the difference between the cartoon(detection) and cartoon(dual) conditions.

3.5. Predicting individual performance in different ERP windows

The previous regression analysis integrated the ERP in a long time window spanning from 0 to 650 ms latency. In the following, the regression analysis was restricted to narrow time windows to resolve which time moments contained more information about the behavioral performance (Fig. 5). In this analysis, 3 electrodes with the highest amplitude were selected for each participant in each narrow time window. The distribution of these channels was shown in Fig. 5AC. The early ERP response to targets (up to ~200 ms) was strongest in the occipital electrodes, while the late neural response (~300–650 ms) was strongest in the post-central electrodes (shown in Fig. 5A). The ERP response to peripheral background (up to ~300 ms) was strongest in occipital electrodes (shown in Fig. 5C).

The early response to targets (~150–200 ms) could predict individual reaction time in conditions in which cartoon video was displayed. The late response to targets (~300–500 ms) could significantly predict individual reaction time and detection accuracy in all conditions (Fig. 5B). The early response to peripheral cartoon video (between 100 and 300 ms) could predict individual reaction time and detection accuracy in conditions in which cartoon video was displayed.

3.6. ERP for participants with good or bad performance

To investigate how the ERP evoked by visual targets and the ERP evoked by peripheral background related to individual behavioral performance, we separately averaged the EEG responses over the participants with good and bad performance (10 participants in each group), and analyzed the waveform differences. The participants were divided into groups based on either their reaction time or detection accuracy (Fig. 6 and Supplementary Fig. 2). Ten best/worst performers were

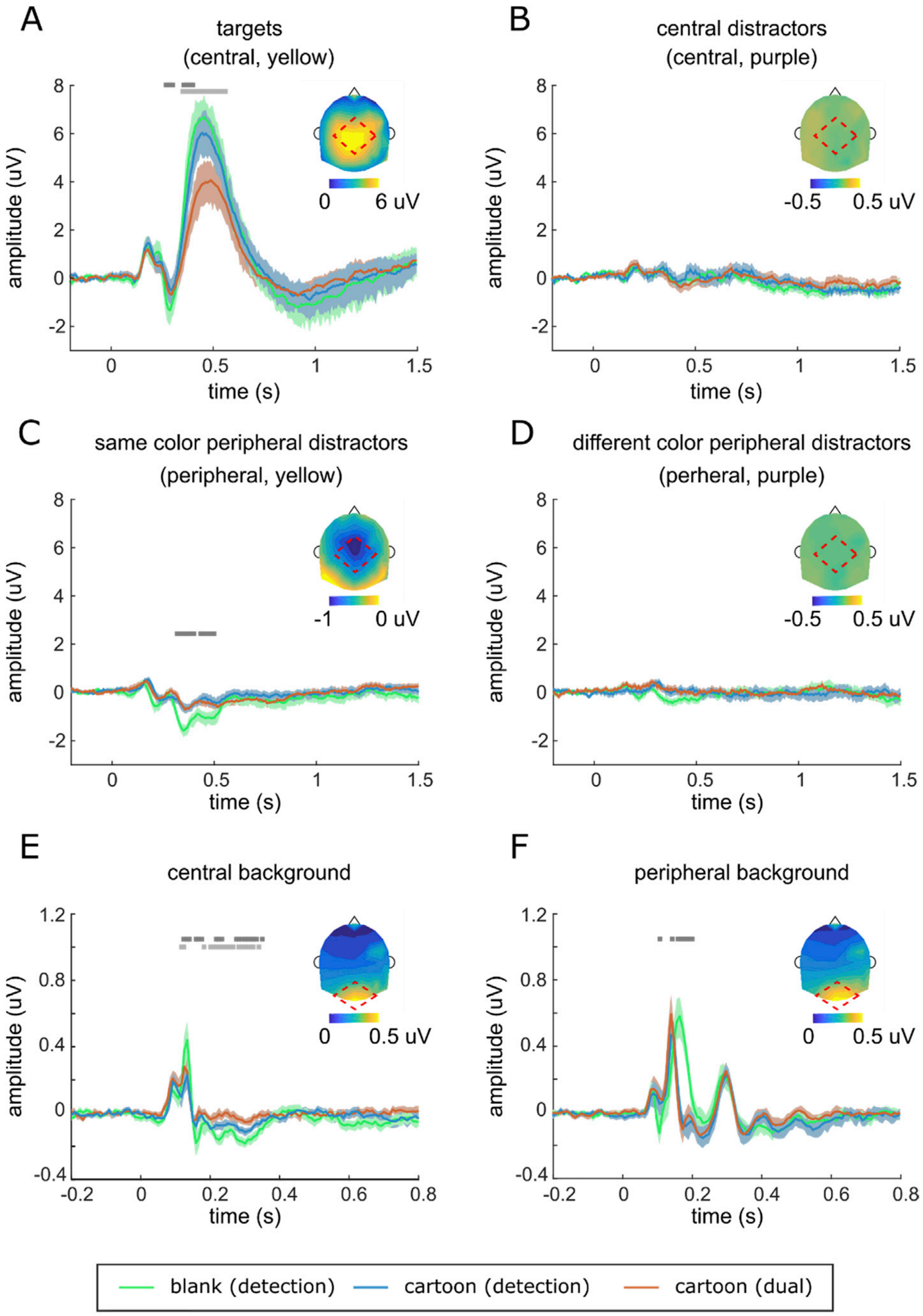

Fig. 3. ERP responses averaged over participants. The channel-averaged ERP waveforms in different conditions are color coded and the shaded area denotes 95% confidence interval. Panels A–D show the ERP responses to the colored circles. The 4 panels separately show ERP responses to targets (A), central distractors (B), peripheral distractors of the same color as the target (C), and peripheral distractors that have a different color from the targets (D). Panels E–F show the ERP responses to the background. The 2 panels separately show the ERP responses to central background (E) and peripheral background (F). The topography of the response peak is shown in each panel. The red dotted boxes on each topography show the positions for central electrodes (Cz, CPz, Pz, CP1, and CP2) or occipital electrodes (O1, Oz, O2, Iz, and POz). The dark gray lines on the top denote the time intervals when the differences between the blank (detection) condition and the cartoon (detection) condition are significant ($P < 0.05$, paired *t*-test, FDR corrected) and the light gray lines on the top denote the time intervals when the differences between the cartoon(detection) condition and the cartoon(dual) condition are significant ($P < 0.05$, paired *t*-test, FDR corrected).

chosen since the rest twenty-nine participants showed relatively little individual difference (Supplementary Fig. 1) and the results did not clearly change if we chose fifteen best/worst performers.

For the two groups divided by reaction time, the ERP evoked by targets differed in an early (~200–300 ms) and a late time window (~300–650 ms) (Fig. 6A). The P300 latency differed between groups in all three conditions. In the blank(detection) condition, the latency difference was 34.5 ± 29.4 ms ($t_{(18)} = 2.46$, $P = 0.02$, two-sample jackknife procedure). In the cartoon(detection) condition, the latency difference was 83.1 ± 46.3 ms ($t_{(18)} = 3.76$, $P < 0.001$, two-sample jackknife procedure). In the cartoon(dual) condition, the latency difference was 136 ± 46.3 ms ($t_{(18)} = 6.92$, $P < 0.001$; two-sample jackknife procedure). For the two groups divided by detection accuracy, significant difference of ERP evoked by targets was only found in a late time window, between about 300 and 650 ms (Supplementary Fig. 2A). Significant P300 latency difference between the two groups was found in conditions that only engaged the detection task. In the blank(detection) condition, the latency difference was 84.8 ± 77.7 ms ($t_{(18)} = 2.29$, $P = 0.030$, two-sample jackknife procedure). In the cartoon(detection) condition, the latency difference was 64.1 ± 39.7 ms ($t_{(18)} = 3.39$, $P = 0.002$, two-sample jackknife procedure). In the cartoon(dual) condition, the latency difference was not significant ($t_{(18)} = 1.22$, $P = 0.234$, two-sample jackknife procedure). For the groups divided by reaction time, the ERP evoked by peripheral background differed between ~100 and 250 ms, when cartoon was presented (Fig. 6B). For the groups divided by detection accuracy, the ERP evoked by peripheral background differed between ~250 and 300 ms, when the cartoon video was presented (Supplementary Fig. 2B).

Given that the target detection task required participants to make a response as soon as possible, the participants must balance between the reaction time and the detection accuracy. Therefore, an additional performance index (PI), i.e., detection accuracy divided by reaction time,

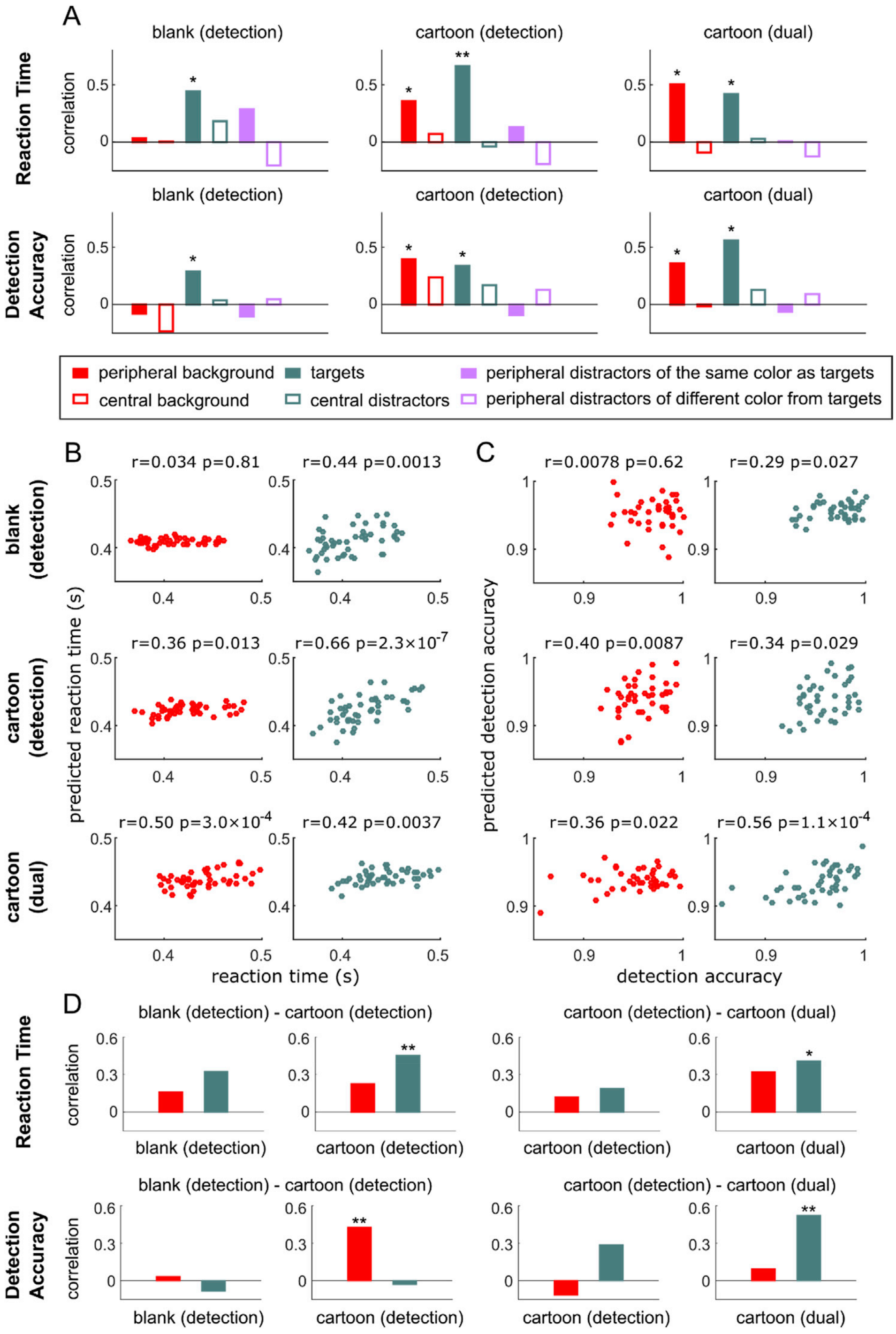

Fig. 4. Predicting individual target detection performance. (A) Individual reaction time (top) and detection accuracy (bottom) are predicted based on the ERP response to different stimulus features. Different color bars indicate different ERP responses. Black stars indicate significant prediction accuracy. $^*P < 0.05$, $^{**}P < 0.01$ (permutation test). ERP evoked by the targets can significantly predict the individual reaction time and detection accuracy in all three conditions (green solid bars). ERP evoked by the peripheral background can predict the individual reaction time and detection accuracy when the cartoon video is played as the background (red solid bars). The scatter plots show the correlation between predicted performance and behavioral performance for individual participants: (B) reaction time, and (C) detection accuracy. Each dot denotes data from one participant. The r-value and the associated p-value are shown in the title of each panel. (D) predicting Individual reaction time difference and detection accuracy difference between conditions based on ERP evoked by the targets (green solid bars) and ERP evoked by the peripheral background (red solid bars). Black stars indicate significant prediction accuracy. $^*P < 0.05$, $^{**}P < 0.01$ (permutation test).

was also calculated and used to divide participants into groups (Supplementary Fig. 1C). EEG responses of fifteen best/worst performers were chosen based on the PI results, and then averaged separately (Supplementary Fig. 3). The results were similar to the results in Fig. 6 and Supplementary Fig. 2.

3.7. EEG responses captured different neural processes

Two ERPs could predict individual behavior, i.e., the ERP evoked by targets and ERP evoked by peripheral background. These two responses had different waveforms and different spatial distributions. To further investigate whether they captured different neural processes, the two responses were combined to predict the individual reaction time and individual detection accuracy (Fig. 7).

When the two ERP responses were combined, the prediction accuracy for individual reaction time was significantly higher in the cartoon(detection) condition (combined responses vs. response to visual targets: $P = 0.010$, and combined responses vs. response to peripheral background: $P = 0.004$, bootstrap, FDR corrected) and the cartoon(dual) condition (combined responses vs. response to visual targets: $P = 0.018$, and combined responses vs. ERP response to peripheral background: $P = 0.004$, bootstrap, FDR corrected) (Fig. 7A).

When predicting individual detection accuracy, in the cartoon(detection) condition, combined responses marginally outperform single

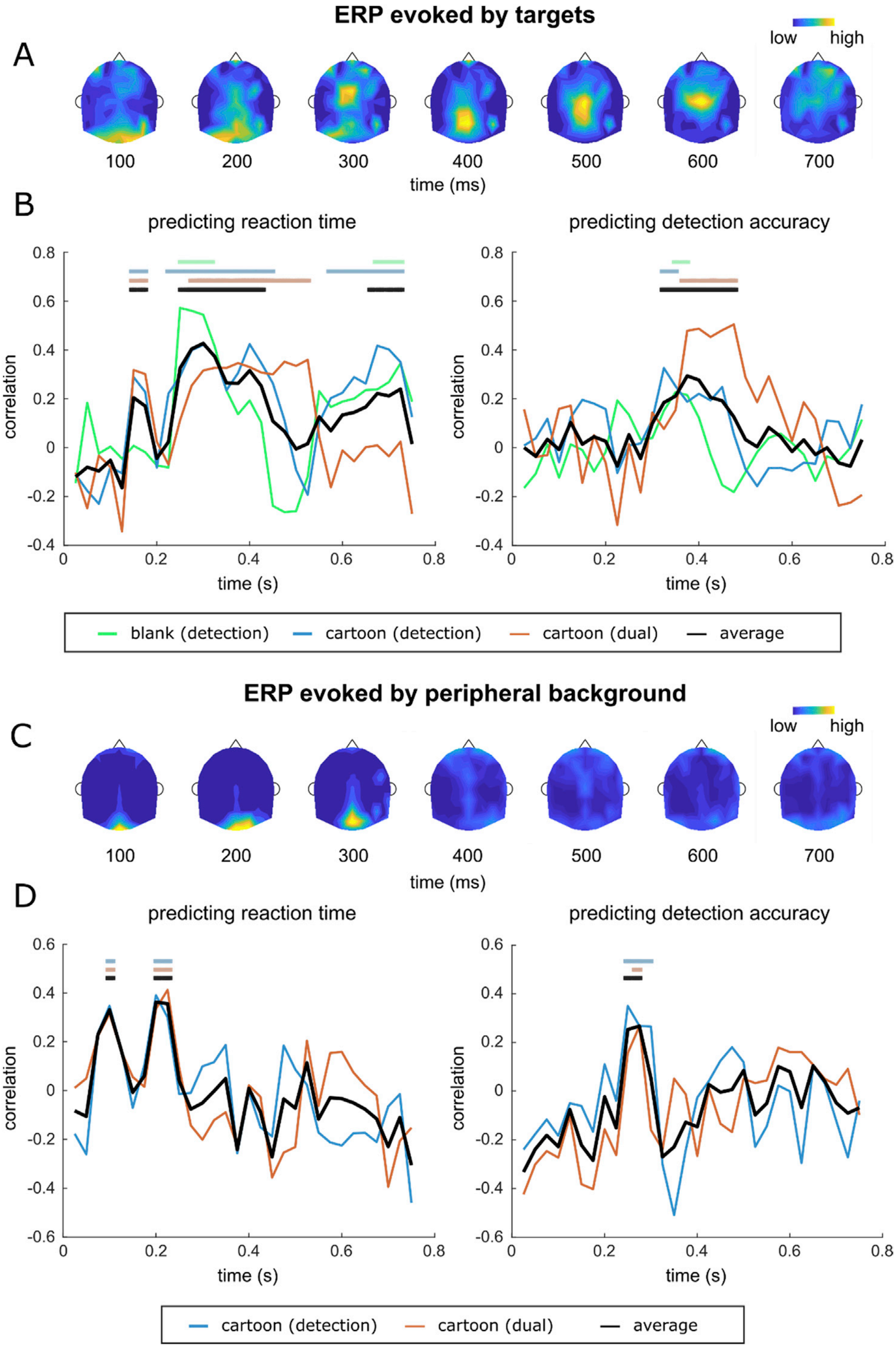

Fig. 5. predicting individual performance using EEG responses in narrow time windows. (AC) topography of the channels used to predict behavior in each time window. In each time window and for each participant, the 3 channels with the highest ERP amplitude are used to predict behavior. The topography of the selected channels is shown in panels A and C for the ERP evoked by targets and peripheral background, respectively. (BD) Neural prediction of individual reaction time (left) and detection accuracy (right). Reaction time and detection accuracy are predicted based on the ERP response to visual target (B) and peripheral background (D), respectively. The colored lines on the top denote the time intervals when the prediction accuracy is significantly above chance ($P < 0.05$, permutation test, FDR corrected).

responses (combined responses vs. response to visual targets: $P = 0.075$, and combined responses vs. response to peripheral background: $P = 0.075$, bootstrap, FDR corrected) (Fig. 7B). In the cartoon(dual) condition, when the two ERP responses were combined, the prediction accuracy for individual detection accuracy was significantly higher (combined responses vs. response to visual targets: $P = 0.035$, and combined responses vs. response to peripheral background: $P = 0.012$, bootstrap, FDR corrected) (Fig. 7B).

4. Discussion

Here, we investigated how the saliency of visual background and top-down attention separately contributed to individual ability to detect visual targets. It was shown that the presence of a salient background, i.e., a cartoon video, slowed down the detection of visual targets and reduced detection accuracy, even when the participants were asked to ignore the background. When the participants split their attention between the visual target detection task and the cartoon watching task, their reaction time to visual target further elongated and their detection accuracy further reduced. In terms of the neural responses, the ERP to visual targets, strongest in the central electrodes, was modulated by both the visual background and dual tasking. In contrast, the ERP response to peripheral background, strongest in occipital electrodes, was largely insensitive to dual tasking but affected by cartoon background.

In the presence of a salient cartoon background, both the responses to targets and background could predict individual performance.

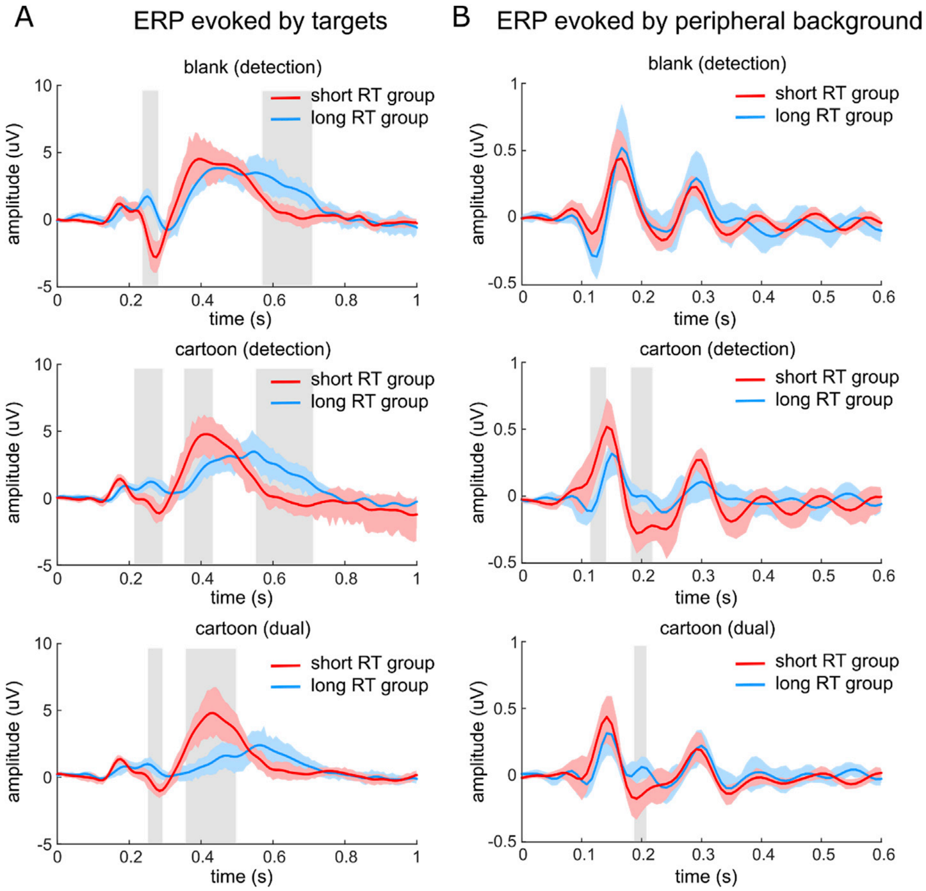

Fig. 6. ERP difference between participants with fast or slow reaction time. Ten participants with the shorted reaction time and ten participants with the longest reaction time are separately averaged for the ERP evoked by targets (A) and peripheral background (B). The shaded area indicates 95% confidence interval. Shaded gray regions denote the time intervals when there are significant differences between the two groups ($P < 0.05$, two-sample *t*-test, FDR corrected).

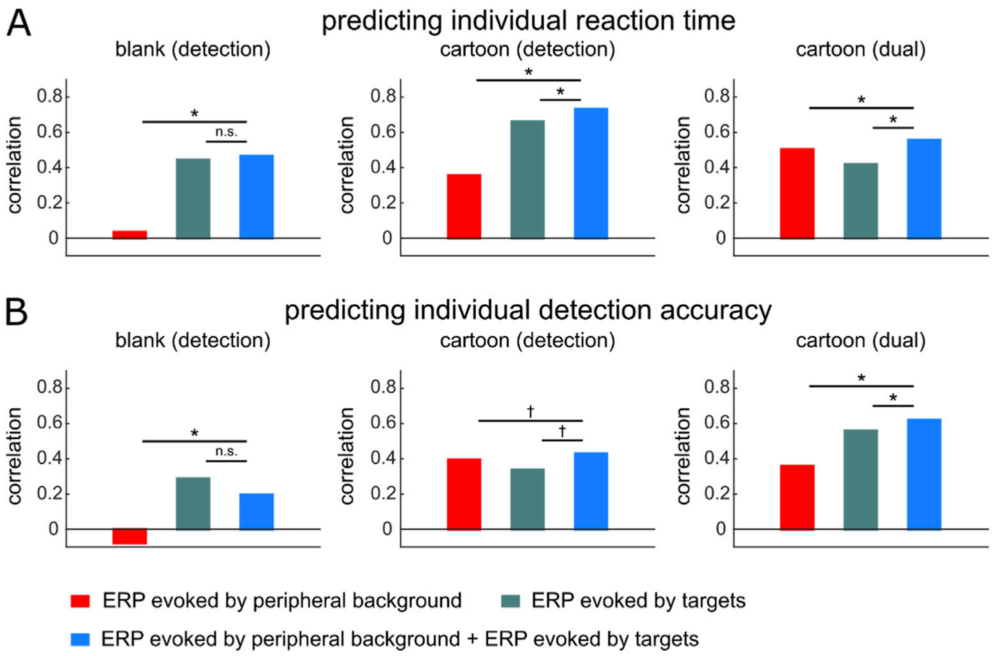

Fig. 7. Predicting individual performance based on combined ERP features. (A) Prediction accuracy for individual reaction time. (B) Prediction accuracy for individual detection accuracy. The red and green bars are same as in Fig. 4. The blue bars show the prediction accuracy when ERP evoked by peripheral background and ERP evoked by targets are combined. Significant differences are indicated by black stars and marginally significant differences are indicated by black crosses. †$P < 0.1$, *$P < 0.05$ (bootstrap, FDR corrected).

Furthermore, combining these two measures could predict individual performance more accurately, demonstrating that the two measures represent distinct processing stages. These results suggest that neural processing of visual targets and background jointly contribute to individual visual target detection performance.

4.1. P300 and late processing of targets

The neural processing of a visual target engages broad cortical networks, and the P300 is a classic neural marker of late processing of visual targets. Previous studies have suggested that the P300 amplitude reflects the subjective probability of a stimulus and its task relevance, while the P300 latency reflects the time required to detect and evaluate a target (Donchin and Coles, 1988). Previous studies have also shown single-trial-level correlation between P300 component and reaction time (Kutas et al., 1977; McCarthy and Donchin, 1981; Polich, 2007; Verleger et al., 2016). Here, we found that participants with good and bad performance differed in both amplitude and latency of the P300 component and good performers showed larger amplitude and shorter latency (Fig. 5). Therefore, the P300 could reflect individual different in visual detection tasks, on top of the within-participant trial-to-trial fluctuation of performance.

In the current study, a large P300 response was observed in the response to visual targets, while no clear P300 response was observed in the responses to distractors (Fig. 3A vs. Fig. 3BCD). The dramatic difference in P300 amplitude might be attributable to two factors. First, the amplitude of the P300 is affected by the stimulus probability, and larger amplitude is observed by less frequent stimuli (Courchesne et al., 1975; Donchin and Coles, 1988; Goldstein et al., 2002). Here, there occurrence frequency of targets was relatively low, only 10% of colored circles were the targets and even within the central area only 30% circles were targets. The response amplitude observed here was comparable to previous studies in which the target stimulus appeared with similar probability (Comerchero and Polich, 1998; Goldstein et al., 2002). Second, in the current study, the target never appeared in succession. As a result, distractors frequently appeared in succession, leading to a large adaptation effect (Morgan et al., 1996).

It has long been hypothesized that late ERP components such as the P300, or the cento-parietal positivity (CPP) in different literature, was linked to perceptual decision making, i.e., the neural transformation of sensory information into adaptive action (Rohrbaugh et al., 1974; Smith et al., 1970). Nevertheless, late ERP components may also reflect other neural processes, such as motor activation (Gratton et al., 1992; Kok, 1986) and affective evaluation (Keil et al., 2002; Liu et al., 2012). Neural correlates of detection performance are often confounded by that movement-related potentials (MRP), which are time locked with overt responding, e.g. a button press. The MRP may temporally overlap with ERP related to stimulus-evaluation, e.g., P300 (Gratton et al., 1992; Kok, 1986; Smith et al., 2008). In the current study, since the participants pressed a button whenever a target appeared, it was theoretically possible that the late ERP component observed here was dominated by motor-related responses such as the lateralized readiness potential (LRP) components instead of the P300. This interpretation, however, is incompatible with our results for the following reasons.

First, the neural response to visual targets were strongest at the central EEG electrodes without obvious lateralization (topography in Fig. 3A) while motor-relative response, such as the LRP, should be lateralized in the current study since participants always used their right forefinger to press the button. Additionally, typical electrodes for the LRP, i.e., C3 and C4 that located above the primary motor cortex (Gratton et al., 1992; Verleger et al., 2006), were not included in the five electrodes (Cz, CPz, Pz, CP1, and CP2) chosen in our analysis. Second, during dual tasking, the reaction time of participants elongated which would increase the latency of MRP. However, we observed significant modulation in the ERP amplitude rather than latency (Fig. 3A). Third, the detection accuracy was compromised during dual tasking and the increase in omission errors might have reduced the MRP amplitude. To test this possibility, we reanalyzed the ERP response to visual targets after removing the epochs of EEG data when participants missed the targets. This reanalysis, however, did not qualitatively change the results: The ERP amplitude remained significantly decreased during dual tasking, which suggested that the attenuate of the ERP amplitude was not caused by an increase in the omission error.

4.2. Attention and early visual processing

Animal studies have showed that attention can modulate early visual processing in cortical visual areas V1, V2, and V4 (Moran and Desimone, 1985; Motter, 1993). In humans, studies on visual ERP found that attention typically enhances the P1 and N1 components (Hillyard and Anllo-Vento, 1998; Mangun and Hillyard, 1990, 1991bib_Mangun_and_Hillyard_1990bib_Mangun_and_Hillyard_1991). Furthermore, previous neurophysiological studies have showed that attention increase the perceived contrast of visual stimuli by boosting early sensory processing in the visual cortex (Störmer et al., 2009), and that neural response gain can account for the behavioral improvements (Itthipuripat et al., 2014). A recent study also showed a link between behavioral gain and enhancement of P1 during behavioral training (Itthipuripat et al., 2017).

Consistent with previous studies, the present study found that the ERP to visual targets was significantly affected by visual background in an early time window before 300 ms (Fig. 3A). A larger negativity, probably the N1 component, was observed in the condition without a distracting cartoon background. Importantly, this early response could predict individual reaction time (Fig. 5B), and was stronger for participants with shorter reaction time (Fig. 6A). These results suggest that the N1 is weakened by distracting visual background and participants with shorter reaction time may better suppress the distracting visual background and maintain a large N1 response. The current study did not observe a clear P1 component (Fig. 3A), which might be attributable to the randomized location of visual targets. Since stimuli presented to the upper and lower visual fields elicit scalp potentials of opposite polarity (Di Russo et al., 2003; Jeffreys and Axford, 1972), their P1 responses might cancel out in the average.

It has been suggested that attention was mediated by biasing the competition between targets and irrelevant stimuli (Duncan et al., 1997). Biasing the competition could be achieved by both enhancing the target and suppressing irrelevant stimuli, e.g., the salient cartoon background in the current study (Hopf et al., 2006; Mazza et al., 2009; Moran and Desimone, 1985; Treue and Trujillo, 1999). Here, we used ERP response evoked by aperiodic markers, i.e., black frames, to characterize neural responses to the visual background in the central and peripheral region, which was related to the steady-state visual evoked potential (SSVEP) approach. The SSVEP is elicited by a stimulus flickering at a constant frequency (Regan, 1966). Previous studies demonstrate that amplitude of SSVEP is substantially increased by both spatial attention (Müller et al., 2003) and feature-selective attention (Andersen et al., 2008; Müller et al., 2006). Additionally, recent studies also show that SSVEP amplitude is stronger for non-target stimuli that share features with the target stimulus (Andersen et al., 2012), even when the non-targets are unattended and presented in peripheral areas (Painter et al., 2014). Here, using an aperiodic approach, we further demonstrate that non-targets, e.g., a salient video background, which are totally irrelevant to the attended targets, can also modulate neural response in early visual processing (Fig. 3EF).

Since visual flickering at a constant rate is likely to induce excessive visual fatigue and sometimes even photosensitive epileptic seizures (Fisher et al., 2005), we employed an aperiodic tagging strategy. Another advantage of the aperiodic tagging strategy was that it preserved temporal information. In the SSVEP, the target response is only characterized by the amplitude and phase at the target frequency and usually only the amplitude is analyzed. The aperiodic tagging technique, however, resulted in an ERP, which allowed distinguishing early and late neural processing. The aperiodic tagging technique was similar to the TRF approach (Jia et al., 2019; Lalor et al., 2006; Liu et al., 2017). However, by using randomly inserted black frames (i.e., spectrally white stimuli), the response could be extracted as a regular ERP, saving sophisticated TRF calculations. More importantly, we extended the stimulus tagging approach to tag natural videos and demonstrated its feasibility.

4.3. Interactions between top-down and bottom-up attention

Previous studies have extensively demonstrated that top-down feature-based or spatial attention both enhances visual processing (Andersen and Müller, 2010; Desimone and Duncan, 1995; Morgan et al., 1996). Here, the targets were defined as yellow circles in the central region, and therefore required both feature-based and spatial attention. A simple feature-based attention mechanism would enhance all yellow objects. Consistent with this hypothesis, in the peripheral region, the ERP to yellow distractors had larger amplitude than the ERP to purple distractors when the background was a gray screen (Fig. 3CD). Nevertheless, this difference largely disappeared when the background was a cartoon video (Fig. 3CD). More importantly, the ERP to peripheral yellow distractors did not predict behavior (Fig. 4A). These results suggest that color-based attention is not effective in peripheral regions in the current experiment.

Similarly, a simple spatial attention mechanism would enhance all visual stimuli in the central region. Consistent with this hypothesis, the response to targets indeed showed a large P300 component. Nevertheless, the ERP to purple distractors in the central region was rather weak (Fig. 3B). More importantly, neither the ERP to central distractors nor the ERP to central background could predict behavior (Fig. 4A). Therefore, the current results showed a complex pattern of attentional modulation that could not be easily explained by simple feature-based or spatial attention mechanisms, especially in the presence of a salient cartoon background. The fact that the neural responses to visual objects of the attended color are selectively enhanced in the central but not peripheral region might be caused by qualitative differences in neural processing between central and peripheral vision (DeValois and DeValois, 1990; Kitterle, 1986). In the central region, different visual objects were well resolved and the color was the only cue to define the target. In the peripheral region, however, the power of feature-based attention was compromised due to the poor spatial resolution of peripheral vision (Carrasco and Chang, 1995; Carrasco et al., 1995).

If top-down attention selectively enhances neural processing of visual targets but not the processing of other visual objects that shared spatial or color features, how could the ERP to peripheral background be predictive of behavior? It was surprising since the peripheral background was outside the spatial area where the participants should attend to and did not have the same color as the target. More importantly, the background was completely irrelevant to the task. It was likely, however, that the cartoon background interfered behavior through bottom up mechanisms. It has been shown that participants can recognize natural images in peripheral regions, in the absence of top-down attention (Li et al., 2002). Neural evidence has also been provided for rapid visual recognition of objects in naturalistic scenes, even if these objects are task-irrelevant and presented outside the focus of spatial attention (Peelen et al., 2009). Therefore, the cartoon in the peripheral background could probably be well recognized even when the participants barely payed attention to them. Since the cartoon video was much larger and easy to recognize than the distractor circles, it likely to be the perceptually dominant stimulus in the periphery that attracted bottom-up attention. Therefore, whether the brain could actively suppress the salient peripheral background was related to target detection performance.

It should be noted that, however, salient visual distractors do not always modulate behavior, especially in some visual working memory (VWM) tasks (Bettencourt and Xu, 2016; Rademaker et al., 2019; Sprague et al., 2018). The load theory of attention and cognitive control have been proposed to account for the inconsistent distractors effects (Lavie et al., 2004). According to this theory, the difficulty of a perceptual task, e.g., a detection task, determines how easily it can be influenced by distractors: Distractors can more easily influence behavior when the perceptual task is easy (Lavie, 2010). In the present study, the simple target detection task might have caused evident attention capture by salient cartoon background.

Finally, it has been hypothesized that the P300, or sometimes called the cento-parietal positivity (CPP), is an index of perceptual decision making, i.e., the neural transformation of sensory information into adaptive action (Rohrbaugh et al., 1974; Smith et al., 1970). Recent studies have demonstrated that, compared with early visual responses, the P300/CPP is more closely related to perceptual decision and better explains behavior (Itthipuripat et al., 2015; Itthipuripat et al., 2014; Kelly and O'Connell, 2013; O'connell et al., 2012; Twomey et al., 2015). If the P300/CPP is already the perceptual decision, it should solely determine behavior. The current results, however, showed that individual behavior was better explained by jointly considering the responses to targets and peripheral background. It was possible that jointly considering two ERPs simply improved the signal-to-noise-ratio of neural responses: Each ERP was noisy and integrating two separate ERPs could suppress noise. It was also possible, however, that the generation of the P300 and the generation of behavior might not be the same process, just like the perceptual detection of targets was not the same as responding by a button press.

In sum, the current study showed behaviorally and neurally that visual processing in a central region could be affected by natural videos in the periphery, and demonstrated complex interactions between top-down feature-based attention, top-down spatial attention, and bottom-up attention.

Declaration of competing interest

None.

CRediT authorship contribution statement

Cheng Luo: Methodology, Investigation, Software, Writing - original draft. **Nai Ding:** Conceptualization, Methodology, Software, Resources, Writing - review & editing.

Acknowledgements

We thank Dr. Michael Breakspear, Dr. Sirawaj Itthipuripat, and the two anonymous reviewers for their constructive comments. We thank Jiajie Zou and Jianfeng Zhang for thoughtful comments on previous versions of the manuscript. The research is supported by National Natural Science Foundation of China 31771248 (ND), Major Scientific Research Project of Zhejiang Lab 2019KB0AC02 (ND), and Fundamental Research Funds for the Central Universities (ND).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2020.116870.

References

Andersen, S., Müller, M., 2010. Behavioral performance follows the time course of neural facilitation and suppression during cued shifts of feature-selective attention. Proc. Natl. Acad. Sci. Unit. States Am. 107, 13878–13882.

Andersen, S.K., Fuchs, S., Müller, M.M., 2011. Effects of feature-selective and spatial attention at different stages of visual processing. J. Cognit. Neurosci. 23, 238–246.

Andersen, S.K., Hillyard, S.A., Müller, M.M., 2008. Attention facilitates multiple stimulus features in parallel in human visual cortex. Curr. Biol. 18, 1006–1009.

Andersen, S.K., Müller, M.M., Martinovic, J., 2012. Bottom-up biases in feature-selective attention. J. Neurosci. 32, 16953–16958.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Stat. Soc. B 57, 289–300.

Beste, C., Wascher, E., Güntürkün, O., Dinse, H.R., 2011. Improvement and impairment of visually guided behavior through LTP-and LTD-like exposure-based visual learning. Curr. Biol. 21, 876–882.

Bettencourt, K.C., Xu, Y., 2016. Decoding the content of visual short-term memory under distraction in occipital and parietal areas. Nat. Neurosci. 19, 150.

Buschman, T.J., Miller, E.K., 2007. Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. Science 315, 1860–1862.

Carrasco, M., Chang, I., 1995. The interaction of objective and subjective organizations in a localization search task. Percept. Psychophys. 57, 1134–1150.

Carrasco, M., Evert, D.L., Chang, I., Katz, S.M., 1995. The eccentricity effect: target eccentricity affects performance on conjunction searches. Percept. Psychophys. 57, 1241–1261.

Comerchero, M.D., Polich, J., 1998. P3a, perceptual distinctiveness, and stimulus modality. Cognit. Brain Res. 7, 41–48.

Courchesne, E., Hillyard, S.A., Galambos, R., 1975. Stimulus novelty, task relevance and the visual evoked potential in man. Electroencephalogr. Clin. Neurophysiol. 39, 131–143.

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

DeValois, R.L., DeValois, K.K., 1990. Spatial Vision. Oxford university press.

Di Russo, F., Martínez, A., Hillyard, S.A., 2003. Source analysis of event-related cortical activity during visuo-spatial attention. Cerebr. Cortex 13, 486–499.

Donchin, E., Coles, M.G., 1988. Is the P300 component a manifestation of context updating? Behav. Brain Sci. 11, 357–374.

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

Efron, B., Tibshirani, R.J., 1994. An Introduction to the Bootstrap. CRC press.

Egeth, H.E., Yantis, S., 1997. Visual attention: control, representation, and time course. Annu. Rev. Psychol. 48, 269–297.

Fisher, R.S., Harding, G., Erba, G., Barkley, G.L., Wilkins, A., 2005. Photic-and pattern-induced seizures: a review for the epilepsy foundation of America working group. Epilepsia 46, 1426–1441.

Forster, S., Lavie, N., 2008. Failures to ignore entirely irrelevant distractors: the role of load. J. Exp. Psychol. Appl. 14, 73.

Fukuda, K., Vogel, E.K., 2009. Human variation in overriding attentional capture. J. Neurosci. 29, 8726–8733.

Gaspar, J.M., Christie, G.J., Prime, D.J., Jolicœur, P., McDonald, J.J., 2016. Inability to suppress salient distractors predicts low visual working memory capacity. Proc. Natl. Acad. Sci. Unit. States Am. 113, 3693–3698.

Goldstein, A., Spencer, K.M., Donchin, E., 2002. The influence of stimulus deviance and novelty on the P300 and novelty P3. Psychophysiology 39, 781–790.

Gratton, G., Coles, M.G., Donchin, E., 1992. Optimizing the use of information: strategic control of activation of responses. J. Exp. Psychol. Gen. 121, 480.

Gregoriou, G.G., Gotts, S.J., Zhou, H., Desimone, R., 2009. High-frequency, long-range coupling between prefrontal and visual cortex during attention. Science 324, 1207–1210.

Hillyard, S.A., Anllo-Vento, L., 1998. Event-related brain potentials in the study of visual selective attention. Proc. Natl. Acad. Sci. Unit. States Am. 95, 781–787.

Hillyard, S.A., Hink, R.F., Schwent, V.L., Picton, T.W., 1973. Electrical signs of selective attention in the human brain. Science 182, 177–180.

Hopf, J.-M., Boehler, C., Luck, S., Tsotsos, J., Heinze, H.-J., Schoenfeld, M., 2006. Direct neurophysiological evidence for spatial suppression surrounding the focus of attention in vision. Proc. Natl. Acad. Sci. Unit. States Am. 103, 1053–1058.

Hopfinger, J.B., West, V.M., 2006. Interactions between endogenous and exogenous attention on cortical visual processing. Neuroimage 31, 774–789.

Itthipuripat, S., Cha, K., Byers, A., Serences, J.T., 2017. Two different mechanisms support selective attention at different phases of training. PLoS Biol. 15.

Itthipuripat, S., Cha, K., Rangsipat, N., Serences, J.T., 2015. Value-based attentional capture influences context-dependent decision-making. J. Neurophysiol. 114, 560–569.

Itthipuripat, S., Ester, E.F., Deering, S., Serences, J.T., 2014. Sensory gain outperforms efficient readout mechanisms in predicting attention-related improvements in behavior. J. Neurosci. 34, 13384–13398.

Jeffreys, D., Axford, J., 1972. Source locations of pattern-specific components of human visual evoked potentials. I. Component of striate cortical origin. Exp. Brain Res. 16, 1–21.

Jia, J., Fang, F., Luo, H., 2019. Selective spatial attention involves two alpha-band components associated with distinct spatiotemporal and functional characteristics. Neuroimage.

Jia, J., Liu, L., Fang, F., Luo, H., 2017. Sequential sampling of visual objects during sustained attention. PLoS Biol. 15, e2001903.

Keil, A., Bradley, M.M., Hauk, O., Rockstroh, B., Elbert, T., Lang, P.J., 2002. Large-scale neural correlates of affective picture processing. Psychophysiology 39, 641–649.

Kelly, S.P., O'Connell, R.G., 2013. Internal and external influences on the rate of sensory evidence accumulation in the human brain. J. Neurosci. 33, 19434–19441.

Kitterle, F.L., 1986. Psychophysics of lateral tachistoscopic presentation. Brain Cognit. 5, 131–162.

Knudsen, E.I., 2007. Fundamental components of attention. Annu. Rev. Neurosci. 30, 57–78.

Kok, A., 1986. Effects of degradation of visual stimuli on components of the event-related potential (ERP) in go/nogo reaction tasks. Biol. Psychol. 23, 21–38.

Kutas, M., McCarthy, G., Donchin, E., 1977. Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time. Science 197, 792–795.

Lalor, E.C., Pearlmutter, B.A., Reilly, R.B., McDarby, G., Foxe, J.J., 2006. The VESPA: a method for the rapid estimation of a visual evoked potential. Neuroimage 32, 1549–1561.

Lavie, N., 1995. Perceptual load as a necessary condition for selective attention. J. Exp. Psychol. Hum. Percept. Perform. 21, 451.

Lavie, N., 2010. Attention, distraction, and cognitive control under load. Curr. Dir. Psychol. Sci. 19, 143–148.

Lavie, N., Hirst, A., De Fockert, J.W., Viding, E., 2004. Load theory of selective attention and cognitive control. J. Exp. Psychol. Gen. 133, 339.

Li, F.F., VanRullen, R., Koch, C., Perona, P., 2002. Rapid natural scene categorization in the near absence of attention. Proc. Natl. Acad. Sci. Unit. States Am. 99, 9596–9601.

Liu, L., Wang, F., Zhou, K., Ding, N., Luo, H., 2017. Perceptual integration rapidly activates dorsal visual pathway to guide local processing in early visual areas. PLoS Biol. 15, e2003646.

Liu, Y., Huang, H., McGinnis-Deweese, M., Keil, A., Ding, M., 2012. Neural substrate of the late positive potential in emotional processing. J. Neurosci. 32, 14563–14572.

Müller, M., Andersen, S., Trujillo, N., Valdes-Sosa, P., Malinowski, P., Hillyard, S., 2006. Feature-selective attention enhances color signals in early visual areas of the human brain. Proc. Natl. Acad. Sci. Unit. States Am. 103, 14250–14254.

Müller, M.M., Andersen, S.K., Keil, A., 2007. Time course of competition for visual processing resources between emotional pictures and foreground task. Cerebr. Cortex 18, 1892–1899.

Müller, M.M., Malinowski, P., Gruber, T., Hillyard, S., 2003. Sustained division of the attentional spotlight. Nature 424, 309.

Mangun, G.R., Hillyard, S., 1990. Allocation of visual attention to spatial locations: tradeoff functions for event-related brain potentials and detection performance. Percept. Psychophys. 47, 532–550.

Mangun, G.R., Hillyard, S.A., 1991. Modulations of sensory-evoked brain potentials indicate changes in perceptual processing during visual-spatial priming. J. Exp. Psychol. Hum. Percept. Perform. 17, 1057.

Mazza, V., Turatto, M., Caramazza, A., 2009. Attention selection, distractor suppression and N2pc. Cortex 45, 879–890.

McCarthy, G., Donchin, E., 1981. A metric for thought: a comparison of P300 latency and reaction time. Science 211, 77–80.

Miller, J., Patterson, T., Ulrich, R., 1998. Jackknife-based method for measuring LRP onset latency differences. Psychophysiology 35, 99–115.

Moran, J., Desimone, R., 1985. Selective attention gates visual processing in the extrastriate cortex. Science 229, 782–784.

Morgan, S., Hansen, J., Hillyard, S., 1996. Selective attention to stimulus location modulates the steady-state visual evoked potential. Proc. Natl. Acad. Sci. Unit. States Am. 93, 4770–4774.

Motter, B.C., 1993. Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. J. Neurophysiol. 70, 909–919.

O'connell, R.G., Dockree, P.M., Kelly, S.P., 2012. A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. Nat. Neurosci. 15, 1729.

Painter, D.R., Dux, P.E., Travis, S.L., Mattingley, J.B., 2014. Neural responses to target features outside a search array are enhanced during conjunction but not unique-feature search. J. Neurosci. 34, 3390–3401.

Peelen, M.V., Fei-Fei, L., Kastner, S., 2009. Neural mechanisms of rapid natural scene categorization in human visual cortex. Nature 460, 94.

Pernet, C.R., Wilcox, R.R., Rousselet, G.A., 2013. Robust correlation analyses: false positive and power validation using a new open source Matlab toolbox. Front. Psychol. 3, 606.

Polich, J., 2007. Updating P300: an integrative theory of P3a and P3b. Clin. Neurophysiol. 118, 2128–2148.

Posner, M.I., 1980. Orienting of attention. Q. J. Exp. Psychol. 32, 3–25.

Rademaker, R.L., Chunharas, C., Serences, J.T., 2019. Coexisting representations of sensory and mnemonic information in human visual cortex. Nat. Neurosci. 22, 1336–1344.

Regan, D., 1966. Some characteristics of average steady-state and transient responses evoked by modulated light. Electroencephalogr. Clin. Neurophysiol. 20, 238–248.

Rohrbaugh, J.W., Donchin, E., Eriksen, C.W., 1974. Decision making and the P300 component of the cortical evoked response. Percept. Psychophys. 15, 368–374.

Rossi, A.F., Paradiso, M.A., 1995. Feature-specific effects of selective visual attention. Vis. Res. 35, 621–634.

Rugg, M., Milner, A., Lines, C., Phalp, R., 1987. Modulation of visual event-related potentials by spatial and non-spatial visual selective attention. Neuropsychologia 25, 85–96.

Saenz, M., Buraĉas, G.T., Boynton, G.M., 2003. Global feature-based attention for motion and color. Vis. Res. 43, 629–637.

Sawaki, R., Kreither, J., Leonard, C.J., Kaiser, S.T., Hahn, B., Gold, J.M., Luck, S.J., 2017. Hyperfocusing of attention on goal-related information in schizophrenia: evidence from electrophysiology. J. Abnorm. Psychol. 126, 106.

Serences, J.T., Shomstein, S., Leber, A.B., Golay, X., Egeth, H.E., Yantis, S., 2005. Coordination of voluntary and stimulus-driven attentional control in human cortex. Psychol. Sci. 16, 114–122.

Smith, D., Donchin, E., Cohen, L., Starr, A., 1970. Auditory averaged evoked potentials in man during selective binaural listening. Electroencephalogr. Clin. Neurophysiol. 28, 146–152.

Smith, J.L., Johnstone, S.J., Barry, R.J., 2008. Movement-related potentials in the Go/NoGo task: the P3 reflects both cognitive and motor inhibition. Clin. Neurophysiol. 119, 704–714.

Sprague, T.C., Itthipuripat, S., Vo, V.A., Serences, J.T., 2018. Dissociable signatures of visual salience and behavioral relevance across attentional priority maps in human cortex. J. Neurophysiol. 119, 2153–2165.

Störmer, V.S., McDonald, J.J., Hillyard, S.A., 2009. Cross-modal cueing of attention alters appearance and early cortical processing of visual stimuli. Proc. Natl. Acad. Sci. Unit. States Am. 106, 22456–22461.

Theeuwes, J., 1991. Exogenous and endogenous control of attention: the effect of visual onsets and offsets. Percept. Psychophys. 49, 83–90.

Theeuwes, J., 2004. Top-down search strategies cannot override attentional capture. Psychon. Bull. Rev. 11, 65–70.

Theeuwes, J., Kramer, A.F., Hahn, S., Irwin, D.E., 1998. Our eyes do not always go where we want them to go: capture of the eyes by new objects. Psychol. Sci. 9, 379–385.

Theeuwes, J., Van der Burg, E., 2007. The role of spatial and nonspatial information in visual selection. J. Exp. Psychol. Hum. Percept. Perform. 33, 1335.

Treue, S., Trujillo, J.C.M., 1999. Feature-based attention influences motion processing gain in macaque visual cortex. Nature 399, 575.

Twomey, D.M., Murphy, P.R., Kelly, S.P., O'Connell, R.G., 2015. The classic P300 encodes a build-to-threshold decision variable. Eur. J. Neurosci. 42, 1636–1643.

Verleger, R., Grauhan, N., Śmigasiewicz, K., 2016. Is P3 a strategic or a tactical component? Relationships of P3 sub-components to response times in oddball tasks with go, no-go and choice responses. Neuroimage 143, 223–234.

Verleger, R., Paehge, T., Kolev, V., Yordanova, J., Jaśkowski, P., 2006. On the relation of movement-related potentials to the go/no-go effect on P3. Biol. Psychol. 73, 298–313.

Vinje, W.E., Gallant, J.L., 2000. Sparse coding and decorrelation in primary visual cortex during natural vision. Science 287, 1273–1276.

Yantis, S., 2000. Goal-directed and stimulus-driven determinants of attentional control. Atten. Perform. 18, 73–103.