Dynamic competition between bottom-up saliency and top-down goals in early visual cortex

Dan Wang[1*], Kabir Arora[1], Jan Theeuwes[2,3,4], Stefan Van der Stigchel[1], Surya Gayet[1†], Samson Chota[1†]

1. Department of Experimental Psychology, Helmholtz Institute, Utrecht University, Utrecht, Netherlands.

2. Department of Experimental and Applied Psychology, Vrije Universiteit Amsterdam, Amsterdam, Netherlands.

3. Institute Brain and Behavior Amsterdam (iBBA), Amsterdam, Netherlands.

4. William James Center for Research, ISPA-Instituto Universitario, 1149-041, Lisbon, Portugal.

**Corresponding author: Dan Wang (Email: wangdanmails@gmail.com).*

† These authors contributed equally.

Abstract

Task-irrelevant yet salient stimuli can elicit automatic, bottom-up attentional capture and compete with top-down, goal-directed processes for neural representation. However, the temporal dynamics underlying this competition, and how they influence early visual processing, remain poorly understood. Here, we combine electroencephalography with Rapid Invisible Frequency Tagging (RIFT) to non-invasively and simultaneously track early visual responses to target and distractor. Both target and distractor evoke stronger initial RIFT responses than nontarget, reflecting top-down and bottom-up attentional effects on early visual processing. Importantly, the presence of a distractor attenuates the initial RIFT response to the target, reflecting competition during the initial stages of visual processing. RIFT responses to the distractor eventually decrease below responses to the target and nontarget, representing active suppression of task-irrelevant stimuli. We show that the dynamic interplay between top-down control and bottom-up saliency directly impacts early visual responses, thereby illuminating a complete timeline of attentional competition in visual cortex.

Introduction

Imagine driving down a busy road, focusing on the surrounding traffic, when a flashing billboard suddenly catches your eye and briefly distracts you from the roadway ahead. This illustrates how attentional control arises from the interaction between two competing processes: bottom-up control driven by saliency (e.g., the flashing billboard) whereby attention is automatically captured by elements that stand out from the environment[1-3], and top-down control (e.g., maintaining focus on the roadway) which directs attention based on goals and intentions[4]. It is widely accepted that both processes contribute to attentional selection[5]. According to the biased competition framework[6-8], objects in the visual field compete for neural representation in visual cortex. This competition is initially driven by bottom-up salience during the early feedforward sweep of sensory processing and is subsequently shaped by top-down signals, likely conveyed via feedback connections from higher-level cortical areas[9,10]. It remains unclear, however, how these processes unfold over time within early visual cortex. Here, we test whether initial bottom-up salience signals and subsequent top-down control mechanisms are both reflected in early visual cortex responses to competing stimuli.

To determine how top-down and bottom-up processes unfold over time in early visual cortex, we employed Rapid Invisible Frequency Tagging (RIFT) while participants performed the additional singleton task[1,2,10]. In this task, participants search for a shape singleton target among nontarget items (e.g., a green diamond among green circles). On some trials, one of the nontarget items is a salient but irrelevant color singleton distractor (e.g., a red circle). Typically, response times increase on "distractor present" trials compared to "distractor absent" trials, indicating that the distractor captured attention in a bottom-up way. By utilizing RIFT, we are able to track, in time, how biased competition unfolds between the bottom-up salience of the distractor and the top-down relevance of the target in early visual cortex. Specifically, we can test (1) whether the presence of a salient task-irrelevant distractor reduces early visual cortex

responses to a concurrent task-relevant target, and (2) whether subsequent top-down control mechanisms further reduce responses to a salient but task-irrelevant distractor.

RIFT works by modulating the luminance of one or more visual stimuli at distinct high frequencies (e.g., 60 Hz and 64 Hz), which elicits frequency-matching periodic activity in the EEG signal originating from early sensory areas (V1/V2)[11-18]. These periodic responses enable highly time-resolved and spatially specific tracking of attention in the early visual cortex[11,14,15,18]. Importantly, because stimulus luminance is modulated at frequencies far above the critical flicker fusion threshold (e.g., ~40 Hz)[19], the flicker is imperceptible to observers and does not perceptually interfere with the ongoing task[20]. Together, these properties make RIFT a powerful tool for disentangling visually evoked responses to concurrently presented visual events. Here, this technique allows us to test whether—and how—the competition between target and distractor stimuli unfolds in early visual cortex. As we will show, RIFT responses to the distractor initially increase and compete with target processing, attenuating RIFT responses to targets. Importantly, this initial competition predicts subsequent behavioral performance. The RIFT responses of the distractor eventually fall below those of targets and nontargets, indicating active suppression of task-irrelevant stimuli. These findings are consistent with the biased competition framework and provide a complete timeline of attentional competition between top-down control and bottom-up saliency in the visual cortex.

Results

Behavioral results

Our experimental paradigm included two search conditions (distractor present and distractor absent). Participants were instructed to search for a unique shape singleton target and respond as quickly and accurately as possible to the orientation of the line segment inside it (Figure 1A). To evaluate whether the presence of distractor affected behavioral performance,

we conducted a paired-sample t-test comparing mean response times (RTs) for correct trials between the distractor present and absent conditions. This revealed that RTs were significantly slower when the distractor was present (*Mean* = 924 ms, *SD* = 93 ms) than when it was absent (*Mean* = 873 ms, *SD* = 97 ms; $t(23) = 8.582$, $p < 0.001$, Cohen's $d = 1.752$; Figure 1B). Additionally, accuracy was significantly lower when the distractor was present (*Mean* = 69.42%, *SD* = 8.60%) than when it was absent (*Mean* = 79.26%, *SD* = 9.39%; $t(23) = 9.556$, $p < 0.001$, Cohen's $d = 1.951$; see supplementary Figure 1).

Validation of frequency-specific neural responses

We verified whether our frequency-tagging manipulation successfully elicited corresponding frequency-specific neural responses by calculating the coherence between the EEG signal and the corresponding tagging frequencies. Because we randomly shifted the phase of the 60 Hz tag relative to the 64 Hz tag, we computed coherence spectrograms separately for each frequency (see Methods section). The resulting spectrograms showed clear peaks at 60 Hz (Figure 1C) and 64 Hz (Figure 1D) following flicker onset, with the strongest responses over parietal and occipital electrodes and no propagation to frontal electrodes (see topography insets, individual traces in supplementary Figure 2; supplementary Figure 3), confirming successful retrieval of the tagging signals from the EEG.

RIFT responses to the distractor and nontarget

We tested whether attentional capture by the distractor was reflected in the RIFT responses. Compared to the nontarget, the distractor evoked significantly stronger coherence in an initial time window ($p = 0.036$; [cluster extent: ~150 to ~350 ms]; cluster-based permutation test) and significantly weaker coherence in a later time window ($p = 0.019$; [cluster extent: ~640 to ~900 ms]; cluster-based permutation test; Figure 2A, left). To further characterize the temporal dynamics of the visual processing at the early time window (0-600

ms), we conducted a one-sided t-test to examine whether the time-to-peak of the coherence traces differed between distractor and nontarget. However, the time-to-peak of the coherence trace for the distractor was not significantly different from that for the nontarget (p = 0.103, $t(23)$ = 1.302, Cohen's d = 0.266; Figure 2A, right).

Coherence measures thus indicated that more attention was allocated to the salient distractor than to the nontarget items initially, as indexed by a stronger stimulus-specific RIFT responses, signifying early attentional capture. Critically, attention to the salient distractor later fell below that directed to the nontarget items, reflecting attentional disengagement and suppression. This pattern compellingly illustrates the temporal unfolding of attentional allocation in early visual cortex, revealing how attention is first captured by the salient distractor and subsequently withdrawn and even suppressed over time.

RIFT responses to target and nontarget

To examine whether attentional selection of the target was reflected in the RIFT responses when no distractor was present, we compared target-evoked coherence with the coherence evoked by the nontarget within the same trial. Target-related coherence was significantly stronger than coherence for nontarget in an initial time window (p = 0.0001; [cluster extent: ~75 to ~540 ms]; cluster-based permutation test; Figure 2B, left). The time-to-peak of the coherence trace did not differ significantly between the two during the early time period (0-600 ms; p = 0.132, $t(23)$ = 1.145, Cohen's d = 0.234; Figure 2B, right). These results suggest that when no salient competitor is present, attention is initially allocated to the only salient item in the display, allowing for fast and accurate selection of the salient target.

RIFT responses to target with and without distractor

We tested whether the presence of a distractor affected early visual processing of the target. To this end, we compared coherence evoked by the target in the presence versus absence

of a distractor. Coherence was significantly higher when the distractor was absent in an initial time window (p = 0.015; [cluster extent: ~50 to ~300 ms]; cluster-based permutation test; Figure 2C, left). Furthermore, the coherence for the target reached its peak earlier in the early time window (0-600 ms) when the distractor was absent compared to when it was present (p = 0.021, $t(23)$ = 2.149, Cohen's d = 0.439; Figure 2C, right). These results demonstrate that the presence of a distractor results in less attention being allocated to the target, consistent with the notion of biased competition[6].

RIFT responses to targets and distractor

To investigate the dynamics of attentional competition between the target and the distractor, we statistically compared their coherence traces in the distractor present condition. No significant differences in coherence were observed between the target and the distractor at the beginning of the trial (Figure 2D, left). However, within the early time window (0–600 ms), the peak coherence occurred significantly later for the target than for the distractor (p = 0.013, $t(23)$ = 2.395, Cohen's d = 0.489; Figure 2D, right). Critically, target-evoked coherence significantly exceeded distractor-evoked coherence in a later time window (p = 0.002; [cluster extent: ~390 to ~840 ms]; cluster-based permutation; Figure 2D, left). Notably, this effect emerged prior to the behavioral response (see Supplementary Figure 4).

To test whether a distractor modulates attentional competition while accounting for potential overall differences in RIFT responses between target-present and target-absent trials (e.g., due to differences in overall arousal), we performed an interaction analysis across trial types (see Supplementary Analyses, Supplementary Figure 5). This analysis showed that the presence of a distractor increases early competition for attentional resources with the target, followed by a later shift of attention toward target processing in distractor present trials.

Taken together, these findings suggest that the target and distractor are in direct competition for attentional resources during the early stage of processing, with the distractor initially capturing attention and thereby delaying visual processing of the target. Over time, the RIFT responses to the distractor were attenuated and even suppressed relative to nontarget items, thereby resolving the competition between the distractor and target, enabling the selection of the target. These results highlight that RIFT is well-suited to measure changes in neuronal excitability in early visual cortex associated with attentional competition during visual search, as indicated by the synchronization of visual neuronal responses with the external periodic input.

Correlation between RIFT responses and behavioral RTs

To investigate whether the measured RIFT responses to target and distractor were related to participants' behavior, we computed trial-wise correlations between Trial-Ensemble Phase Similarity (TEPS, a phase-based, single-trial measure of the tagging signal; see Methods section) and RTs. We employed time-resolved correlations rather than a single fixed time window (e.g., TEP differences between the target and distractor) to avoid defining an arbitrary a priori analysis window and follow a more data-driven approach. This trial-by-trial approach captures transient fluctuations in the relative coherence of target- and distractor-related responses that may be obscured in across-trial averages. Because we were specifically interested in how the competition between the target and the distractor was resolved over time, we calculated the difference in RIFT responses between the target and distractor within the same trial and correlated this differential RIFT response with the RTs to the target across trials (Figure 3). We observed a significant negative correlation between TEPS and RTs ($p = 0.003$; [cluster extent: ~290 to ~520 ms]; cluster-based permutation; Figure 3), indicating that the greater the RIFT responses to the target compared to the distractor, the faster the participant

responded to the target. This finding demonstrates that the RIFT responses capture behaviorally relevant processes.

Discussion

The present study employed Rapid Invisible Frequency Tagging (RIFT) with EEG to examine how competition between top-down attention (to the target) and bottom-up attention (to the distractor) unfolds over time in visual cortex. In line with the biased competition framework[6, 21], the results indicate that early in processing the distractor competed with the target, whereas at later stages the target prevails. Crucially, the outcome of this competition is a clear neural representation in visual cortex of the selected (winning) object, accompanied by a diminished representation of the non-selected (losing) objects.

In the current task, participants correctly responded to the target well above chance level, indicating that ultimately the target is selected and wins the competition. However, the strength of this study lies in how the RIFT responses to both the target and distractor jointly provide insight into how this competition is resolved, up to the point of the eventual selection of the target. It is evident that in distractor-absent trials, there is basically no competition (the target is the only salient element in the display) and the RIFT responses show that in this condition target processing dominates from the earliest moment onwards, giving rise to fast and accurate responses.

However, due to the limited processing capacity of the visual system, competition arises when both the target and distractor are simultaneously present, both competing for neural representation. The results show that, early on—during the initial feedforward sweep of sensory processing—there is a competition between target and distractor: RIFT responses to the target were decreased, while responses to the distractor exceeded those elicited by nontarget items. A potential concern is that across-trial comparisons (e.g., targets in distractor-present vs.

distractor-absent trials, or salient distractors vs. nontargets) might reflect factors other than attentional competition, such as trial-wise differences in overall attentional state (e.g., arousal). However, such an account cannot explain the reduced early target-related RIFT responses in distractor-present trials. Arousal would predict enhanced, not decreased, target responses. Importantly, our interaction analysis (see supplementary Analyses; supplementary Figure 5) showed that the distractor shifts the balance between target- and distractor-related responses, providing further evidence for early attentional competition at initial stage.

Later in time, RIFT responses to the salient distractor dropped not only below responses to the target, but also below responses to nontarget items. This suggests that salient but task-irrelevant objects in the environment not only cease to attract attention over time but may also become suppressed to facilitate the neural representation of the task-relevant objects. Critically, within-subject, trial-wise correlations reveal that larger RIFT responses to target compared with distractor (a direct measure of competition in the early visual cortex) are associated with faster reaction times to find the target, highlighting the functional role of these early sensory modulations for the successful completion of goal-directed behavior.

When directly comparing distractor- and target-related RIFT responses within the same trial, we did not observe a reliable early enhancement of distractor responses relative to the target. This lack of a clear early difference may be attributable to the stimulus configuration used in the present study. The experimental paradigm contains both a shape-singleton (the target) and a color-singleton (the salient distractor). To optimize the signal-to-noise ratio of the RIFT signal, we used relatively large shapes for which the inside areas were tagged, combined with relatively thin colored outlines that were not tagged. Their comparable salience likely minimized bottom-up differences. Moreover, in the present design the distractor was consistently defined by color whereas the target was defined by shape. Although this configuration should not affect our theoretical conclusions, it may also modulate the time

course or the strength of the observed competition effects. Nevertheless, the color singleton clearly competed with the target for attentional resources, as evidenced by (1) slower reaction times, (2) reduced target-related RIFT responses when a salient distractor was present, and (3) a delayed time-to-peak of the target RIFT response in distractor-present trials. Together, these findings indicate early attentional competition between target and distractor.

Overall, the pattern of RIFT responses is consistent with stimulus-driven accounts of perceptual competition, which propose that during the initial feedforward sweep of visual processing attention is automatically captured by the most salient element in the display[2,10,22]. Only later, through feedback signals, top-down processing allows attention to be disengaged from the distractor[23]. The below-baseline RIFT response of the distractor indicates that disengagement even involves suppression, in line with the notion of reactive suppression[24].

A recent study by Klink et al.[25] provides similar evidence supporting initial capture followed by rapid disengagement. In their research, Klink and colleagues examined neural responses in area V4 of macaque monkeys performing an eye movement-based version of the same paradigm that was used here (i.e., the additional singleton paradigm). Eye-tracking data suggested that the salient distractor was effectively ignored, as the monkeys' eyes moved directly to the target. Yet, the neuronal activity of V4 neurons showed a different picture: Early on, during the initial stage, there was attentional enhancement at the location occupied by the salient distractor. This initial enhancement was followed by suppression, occurring about 150 ms later. These data show that even though behaviorally there appears to be successful inhibition of the salient distractor, this inhibition was preceded by attentional capture, providing evidence for the fast disengagement hypothesis[23]. Thus, the pattern of neuronal activity in V4 neurons reported by Klink et al.[25] is consistent with the current pattern of findings in early visual cortex, obtained non-invasively in human subjects. However, the suppression of the distractor seems to occur relatively later in time in our study, which might be due to differences

in species, task differences and/or measurement techniques. In particular, EEG-based RIFT captures population-level synchronization in early visual cortex and is likely less sensitive than single-unit or multi-unit recordings which are better at revealing smaller modulations that occur early. In addition, differences in training history and response requirements, especially the extensive training of non-human primates, may further contribute to the more extended temporal dynamics observed in the present study.

Also using the additional singleton paradigm, Lin et al.[26] recently recorded human intracranial signals covering multiple brain regions (but with very limited coverage in early visual cortex). They were able to dissociate distractor-specific representations from target signals in the high-frequency range (60–100 Hz). Consistent with the current findings, Lin et al. found that initially salient distractors were processed around 220 ms after stimulus onset, while at the same time, target-related processing was attenuated. Their findings highlight the competition for neural representation between target and distractor, consistent with the biased competition framework[6,21]. The present work extends the findings of Lin and colleagues (non-invasively), by specifically revealing that biased competition between target and distractor stimuli even influences early visual cortex responses.

Using the same paradigm, previous studies have recorded scalp EEG to examine shifts of attention toward targets and distractors[27-30]. Specifically, the N2pc component of the event-related potential was used to track the allocation of attention to lateralized positions in the search array. For example, in Experiment 2 of Hickey et al.[27], both the distractor and the target (on different trials) elicited an N2pc when they appeared on opposite sides of the array. Critically, however, the pattern of N2pc responses indicated that attention was initially captured by the salient distractor before shifting to the target. Based on these EEG findings, the present study a priori defined the early time window (around 200 ms) to capture the initial stage of attentional capture reflected in RIFT responses. This is consistent with a stimulus-driven

account of attentional capture[2,10] and with our current results. Importantly, however, our approach extends prior work by simultaneously measuring neural responses to both the target and the salient distractor within the same trial. This allowed us to directly assess their competitive interaction in visual cortex and link these neural dynamics to behavioral performance.

The temporal profile of RIFT responses suggests that suppression of the distractor relative to the nontarget items might be necessary for successful processing of the target. In other words, consistent with the biased competition framework, the competition between the distractor and the target must be resolved in favor of the target to enable its processing. Similar findings have been reported previously[14,25,26,31,32]. As previously argued distractor suppression may be necessary for attentional disengagement and is likely driven by top-down control signals originating from higher-order cortical regions such as the inferior frontal gyrus, prefrontal cortex, and area V4[10,33]. These regions have been implicated in facilitating attentional shifts away from distractors and toward goal-relevant stimuli[25,31,34].

One potential concern is that differences in eye movements toward the tagged stimuli might influence RIFT responses, since neural responses tend to be stronger for stimuli presented near the fovea[35]. To address this concern, we removed trials in which fixation was not adequately maintained. In addition, trial-wise correlation analyses between gaze bias and RIFT responses for both targets and salient distractors revealed no significant correlations (see Supplementary Figure 6), suggesting that gaze position did not predict RIFT responses. Moreover, previous studies have shown that RIFT responses are not affected by small eye movements around fixation[11,14], which further supports the conclusion that our findings reflect attentional process rather than fixation instability or gaze shifts.

We interpret the changes in RIFT-evoked neuronal excitability as reflecting changes in the responsiveness of early visual cortex. Consistent with this view, previous studies have taken RIFT response modulations to indicate changes in neural excitability within early visual areas, including both primary and secondary regions[13,14]. Supporting this interpretation, studies combining RIFT with magnetoencephalography (MEG) have consistently localized RIFT responses to early visual areas V1 and V2[13,14,16, 36,37]. Taken together, this body of evidence provides a strong basis for linking our observed RIFT responses to early visual cortex activity. The present findings demonstrate that both top-down and bottom-up factors shape even the earliest stages of visual processing when stimuli compete for representation.

In summary, the present study demonstrates that, during visual search, a salient distractor initially competes with target processing in early visual cortex, reflecting a strong bottom-up drive. Subsequently, reactive suppression of the distractor is accompanied by a relative enhancement of the neural response to the target, which may jointly underlie successful target selection. Using RIFT with EEG, the current study reveals the dynamic interplay between bottom-up salience and top-down control in resolving attentional competition within early visual cortex.

Methods

Participants

To determine the appropriate sample size, we conducted a priori power analysis using G*Power 3.1[38]. Assuming a paired-samples t-test, an alpha level of 0.05, an effect size (dz) of 0.6, and a desired power of 0.8, the analysis indicated that 24 participants would be required. This sample size is comparable to those used in prior studies on attentional capture with EEG (18 participants)[27-28] and rhythmic sensory stimulation (24 participants)[11,39]. Four participants were replaced: three due to an excessive proportion of saccades (93.4%, 63.9%, and 52.8%,

respectively), one due to below-chance search task performance (48% accuracy). The final sample thus consisted of 24 participants (*mean* age = 22.83 years, *SD* = 2.87; 22 females). All participants had normal or corrected-to-normal vision and reported no history of epilepsy or cognitive impairments. Written informed consent was obtained prior to participation, and participants received either monetary compensation or course credit. The study was approved by the Ethics Committee of Utrecht University.

Apparatus

Stimuli were presented using a ProPixx projector (VPixx Technologies Inc., QC, Canada; 960 × 540 pixels, 480 Hz refresh rate) in a rear-projection format (screen size: 48 × 27.2 cm). All stimuli were created using MATLAB 2021 (The MathWorks, Inc.) with the PsychToolbox extension[40]. The viewing distance was maintained at 72 cm using a chin and forehead rest. Gaze was tracked using an EyeLink SR (SR Research, Ontario, Canada) eye tracker, which recorded data from both eyes at a sampling rate of 500 Hz.

EEG data were recorded using a 64-channel ActiveTwo BioSemi system (BioSemi B.V., Amsterdam, The Netherlands) at a sampling rate of 2048 Hz. To monitor eye movements and detect ocular artifacts, two additional electrodes were placed: one above the left eye to record vertical eye movements and one on the outer canthus of the left eye to capture horizontal eye movements. Before the experiment, signal quality across all channels was assessed and optimized using BioSemi ActiView software, ensuring stable and high-quality recordings.

Procedure

In the main experiment (depicted in Figure 1A), participants were instructed to maintain their gaze fixation on the central cross throughout the entire experiment. Each trial began with the presentation of a placeholder display for a randomly varying duration between 1200 and 1450 ms. Following the placeholder display, a search display appeared for a fixed duration of

1300 ms. Participants were instructed to identify whether the line segment inside the unique (target) shape (circle or diamond) was vertical (press "P") or horizontal (press "Q") as quickly and accurately as possible with left and right index finger respectively. Upon participant response, the color of the central cross changed from white to black. Participants completed 40 practice trials to familiarize the experimental procedure. The main experiment consisted of 1152 trials, divided into 8 blocks, and lasted approximately one hour. Prior to the experiment, a 9-point calibration was conducted to ensure accurate gaze measurements. This calibration was repeated after every two experimental blocks to maintain gaze tracking precision throughout the session.

Stimuli

All stimuli were displayed on a uniform dark background with an RGB value of (20, 20, 20). The placeholder display comprised six shapes, each formed by superimposing a diamond (4.3° × 4.3° square rotated 45°) onto a circle (radius = 2.1°). The sizes of the circle and diamond matched those of the stimuli used in the subsequent search task. The outline of each shape was white (RGB: 255, 255, 255), while the inner area was filled with a mid-gray color (RGB: 127.5, 127.5, 127.5). These six shapes were evenly spaced along an imaginary circle (radius = 6°) centered around the fixation cross (1.2° in length; RGB: 255, 255, 255).

In the search task, six items were displayed in the same spatial arrangement as in the placeholder display, ensuring spatial consistency throughout the experiment. In the distractor present condition, the array consisted of one shape singleton target, one color singleton (salient) distractor, and four nontargets. In the distractor absent condition, the array contained one shape singleton target and five nontargets. The nontargets always shared the same color as the target and the same shape as the distractor. The target was either a circle or a diamond. When the target was a circle, all distractors were diamonds and vice versa. The outline color of the target was either green (RGB: 0, 131, 0) and that of the color-salient distractor was red (RGB: 255,

0, 0), or vice versa. By varying both the shape and the color assignment across trials, participants could not proactively prepare for target or distractor features before the onset of the search array. Each item in the search array contained a central white line segment (3.1° in length), oriented either horizontally or vertically. The inner area of the target and one distractor (salient or nontarget, depending on condition) was luminance-modulated (excluding the white line segment) at either 64 Hz or 60 Hz, making them perceptually indistinguishable from mid-gray[11]. The inner areas of the remaining nontargets were filled with mid-gray. The target appeared randomly at one of the six positions with equal probability, with the constraint that the distractor was never placed directly adjacent to the target.

Tagging manipulation

We implemented RIFT stimulation from specific spatial locations in the visual field (corresponding to the target and distractor stimuli) by sinusoidally modulating the luminance of the inner area of stimuli at high temporal frequency[11,41]. Tagging was applied throughout the entire duration of the search display. The two tagging frequencies (60Hz and 64Hz) and the tagged locations were counterbalanced across target and distractor stimuli to ensure that any difference in RIFT responses between stimuli could not be attributed to differences between tagging frequencies and locations. In the distractor present condition, the target was tagged with one frequency (either 60 Hz or 64 Hz) and the distractor with the other. In the distractor absent condition, the target was tagged with one frequency and one of the nontargets with the other, with the nontarget tagging matched to the location of the distractor in the distractor present condition.

To improve the temporal resolution of the RIFT responses given the inherent trade-off between time and frequency resolution (i.e., the Heisenberg uncertainty principle for signals), we implemented a phase randomization procedure as described below. The 64 Hz tagging signal was phase-locked to the onset of the search display in all trials. In contrast, the phase of

the 60 Hz tagging signal was randomized at one of eight equally spaced phase offsets within a cycle (excluding 0° to avoid overlap with the 64 Hz component), randomly assigned on each trial. These phase offsets were recorded and later used to reconstruct a phase-aligned EEG signal for subsequent analysis (see RIFT responses section). Decoupling the 60 Hz and 64 Hz signals based on phase facilitates the separation of tags during EEG analysis, specifically when using methods that quantify the degree of phase alignment across trials (i.e. coherence, see RIFT responses section). Practically, it enables the use of broader bandpass filters when isolating the tagged responses, thereby enhancing temporal resolution without compromising frequency specificity.

EEG pre-processing

All data analysis was conducted in MATLAB using the Fieldtrip toolbox[42]. The EEG data was first re-referenced to the average of all channels (excluding poor channels determined by five default bad channels [T7, T8, Fp1, Fp2, Tp7] and additional visual inspection: median = 1.5 channels). Data was high-pass filtered (0.01Hz), then line noise and its harmonics were removed using a DFT filter (50, 100, 150Hz). Data was segmented into trials ranging from 0.8 s before to 1.2 s after search display onset. An ICA was performed to remove oculomotor artifacts, and trials with other motor artifacts were removed from further EEG analysis as per visual inspection (*mean* = 6.49 %). Baseline correction was performed by averaging (and then subtracting from the signal) a window ranging from 0.8 s to 0.2 s before the onset of the search display.

Eye-tracking analysis

The time window of interest spanned 1.2 seconds following the onset of the search display. Due to head movements after eye calibration, one participant had one block excluded, and another participant had two blocks excluded before eye tracking data analysis. Blink

correction was performed using custom code adapted from Hershman et al.[43]. To ensure that EEG responses to the frequency-tagged stimuli (target and distractor) were not confounded by large eye movements, we implemented eye movement exclusion criteria. A circular region of interest (ROI) with a radius of 3.5 dva (no stimuli was presented at this area) was defined around the central fixation point. Trials in which participants' gaze deviated outside this ROI for more than 50 ms were classified as saccade trials and excluded from further EEG analysis. Participants with more than 50% of their trials marked as saccade trials were excluded from group-level analyses. As a result, three participants were replaced. On average, 13.65% of trials per participant were removed based on the saccade criterion.

RIFT responses

To quantify the degree to which the EEG signal reflects the tagging signals, we computed magnitude-squared coherence[11,44], a dimensionless measure ranging from 0 to 1 that reflects the consistency of two signals in both magnitude and phase. Coherence was computed between a reference sinusoid (sampled at 2048 Hz) and the neural responses to the tagged stimuli, separately for each frequency, EEG channel and participant. To calculate coherence for a specific frequency of interest, segmented trials (*N*) were bandpass filtered (±1.9 Hz) around the respective tagging frequency using a two-pass, fourth-order Butterworth filter with a Hamming taper. The filtered time-series data were then subjected to a Hilbert transform to extract the instantaneous magnitude (*M(t)*) and phase ($\phi(t)$) of the signal. The set of all instantaneous magnitudes of the filtered responses ($\overrightarrow{Mx}(t)$ and the reference sinusoid ($\overrightarrow{My}(t)$ across all n trials, as well as the differences between their instantaneous phases across all n trials ($\Delta\overrightarrow{\phi xy}(t)$ were used to compute time-varying coherence (see Equation 1). Notably, when calculating coherence at 60 Hz, individual EEG trials were first phase (re)-aligned by temporally shifting them by a maximum of 16.67 ms (34 samples in EEG) based on the phase

at which they were presented. This ensured accurate estimation of coherence and causes only to minimal temporal smearing (see Tagging manipulation section, above).

$$coh\,(t) = \frac{\left|\sum_{tr=1}^{n} \overrightarrow{Mx}\,(t)\,\overrightarrow{My}(t)\,e^{i\Delta\overrightarrow{\phi xy}(t)}\right|^2}{n\sum_{tr=1}^{n}\overrightarrow{Mx}(t)^2\;\;\overrightarrow{My}(t)^2} \qquad \text{(Equation 1)}$$

For each participant, six channels were selected based on the highest coherence averaged across the two tagging frequencies during the 1.2 s following search display onset[11,45]. The same sensor set was used across all conditions and trial types, ensuring that within-trial comparisons were not biased toward either tagging frequency. This approach provides a stable and sensitive estimate of the RIFT signal while avoiding arbitrary or frequency-dependent sensor selection. Notably, previous studies have shown that the exact number of top channels selected does not substantially affect the results[11]. To account for the fact that frequency-tagged stimuli evoke spatially specific neural responses that vary depending on their location on the screen, channel selection was performed separately for each of the 6 tagging locations (see scalp topographies for the six locations in the supplementary Figure 7)[16]. Coherence traces were then averaged across the top six selected channels, six tagging locations, and two tagging frequencies to produce a single coherence trace per condition for each participant, which was used for all subsequent EEG analyses. Coherence spectrograms (Figure 1C & 1D) were computed across frequencies from 56.8 Hz to 67.2 Hz in 0.8 Hz steps.

To examine the trial-wise correlation between RIFT responses and behavioral performance (i.e., within participants) we calculated Trial-Ensemble Phase Similarity (TEPS) as a single trial measure of RIFT. TEPS quantifies the phase similarity between each individual trial and the average phase of all other trials, using a leave-one-trial-out approach. This measure ranges from 1 (perfect alignment) to –1 (perfect opposition), capturing how closely a trial's phase follows the group-level phase dynamics over time (see Equation 2). Specifically, for each trial *n* and time point *t*, we extracted the instantaneous phase $\phi_K(t)$ from the bandpass-filtered

EEG signal (±1.9 Hz) via the Hilbert transform. We then calculated the circular mean phase across all other trials $\overline{\phi}_{-k}(t)$ and defined TEPS as the cosine of the phase difference between $\phi_K(t)$ and $\overline{\phi}_{-k}(t)$.

$$TEPS_k = \cos(\,\phi_K(t) - \overline{\phi}_{-k}(t)\,) \qquad \text{(Equation 2)}$$

Statistical Analysis

For the behavioral analysis, we excluded trials with reaction times shorter than 200 ms and used a paired-sample t-test to compare the mean response times (RTs) across conditions. For the time-to-peak analysis, the early time window (0–600 ms) was defined based on the end of the first peak of the averaged target, distractor, and nontarget time courses, which occurred at approximately 600 ms. Although post hoc, condition-blind ROI definitions may not be entirely bias-free[46], individual-level analyses showed that all participants reached their first peak before 600 ms (Figure 2, right panels). Note that this region of interest was defined based on the combined responses from all experimental conditions, and therefore did not favor the time-to-peak responses of one condition over another. Within this window, paired-sample *t*-tests were conducted to compare conditions.

For the coherence analysis, all comparisons were conducted within a fixed-length time window of 1.2 s following search display onset. Stimuli remained on the screen for the full duration of this window regardless of response time, ensuring equal trial lengths across conditions. To statistically compare coherence traces between conditions, individual traces were subtracted between conditions and resulting individual coherence-difference traces were subjected to a cluster-based permutation test[47]. First, one-sample t-tests at the group level were performed for each time point to identify clusters where coherence traces differed significantly from zero ($p < 0.05$). Individual clusters were defined as one or more consecutive significant time points. For each cluster, the sum of t-values across all included timepoints (i.e., the t-mass)

was computed and used as the cluster-level statistic. A null distribution of t-mass values was created by flipping the sign of a random selection of difference traces across 10,000 permutations and repeating the cluster identification procedure described above, except for only including the largest cluster in the null-distribution. Observed clusters were considered statistically significant if their t-mass exceeded the 95th percentile of t-masses within the null distribution.

To statistically assess correlations between neural responses in early visual cortex and behavioral performance, we computed trial-wise Pearson correlations between single trial RIFT responses (TEPS) and response times (including only trials with correct responses). We normalized the TEPS value for each location by subtracting the mean TEPS value at that location from the raw TEPS value for each trial, in order to eliminate differences in RIFT responses across locations (see supplementary Figure 7). For each participant, we calculated the correlation between TEPS at each time point and RTs between trials, resulting in a time-resolved series of correlations. One-sample t-tests at the group level were performed at each time point to identify clusters where the correlation coefficients differed significantly from zero ($p < 0.05$). The statistical significance (t-mass) of the time-resolved correlations was assessed using the same cluster-based permutation procedure described above for the coherence analysis.

Statistics and Reproducibility

The sample consisted of 24 participants. Written informed consent was obtained prior to participation. The study was approved by the Ethics Committee of Utrecht University. All ethical regulations relevant to human research participants were followed. The study employed a within-subject design in which all participants completed all experimental conditions. Biological replicates were defined as individual participants. All data analyses were conducted using MATLAB, and EEG analyses were performed using the FieldTrip toolbox. Behavioural

data and coherence time-to-peak measures were compared between conditions using paired-sample t-tests. Data are reported as mean ± SD. Statistical significance was defined as '*' for $p < 0.05$, '**' for $p < 0.01$, and '***' for $p < 0.001$. Differences between coherence time courses across conditions were assessed using cluster-based permutation tests based on t-values. In addition, correlations between coherence measures and behavioral performance were evaluated using permutation tests based on t-values. Significant time windows and exact p-values are reported.

Data Availability

The source data to create all the figures presented in the manuscript and raw data have been deposited in Open Science Framework (OSF): https://osf.io/szuxa/[48]. All other data are available from the corresponding author upon reasonable request.

Code Availability

The analysis script for this experiment is publicly available at OSF: https://osf.io/szuxa/[48].

References

1. Theeuwes, J. Cross-dimensional perceptual selectivity. *Percept. Psychophys.* **50**, 184–193 (1991).
2. Theeuwes, J. Perceptual selectivity for color and form. *Percept. Psychophys.* **51**, 599–606 (1992).
3. Theeuwes, J. Response to commentaries to Luck et al. (2021): Progress toward resolving the attentional capture debate. *Vis. Cogn.* **29**, 637–643 (2021).
4. Folk, C. L. & Remington, R. Selectivity in distraction by irrelevant featural singletons: Evidence for two forms of attentional capture. *J. Exp. Psychol. Hum. Percept. Perform.* **24**, 847–858 (1998).
5. Luck, S. J., Gaspelin, N., Folk, C. L., Remington, R. W. & Theeuwes, J. Progress toward resolving the attentional capture debate. *Vis. Cogn.* **29**, 1–21 (2021).
6. Desimone, R. & Duncan, J. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* **18**, 193–222 (1995).
7. Luck, S. J., Chelazzi, L., Hillyard, S. A. & Desimone, R. Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J. Neurophysiol.* **77**, 24–42 (1997).

8. Tsotsos, J. K. et al. Modeling visual attention via selective tuning. *Artif. Intell.* **78**, 507–545 (1995).

9. Beck, D. M. & Kastner, S. Top-down and bottom-up mechanisms in biasing competition in the human brain. *Vis. Res.* **49**, 1154–1165 (2009).

10. Theeuwes, J. Top–down and bottom–up control of visual selection. *Acta Psychol.* **135**, 77–99 (2010).

11. Arora, K. et al. Dissociating external and internal attentional selection. *iScience* **28**, 112282 (2025).

12. Dietz, L. et al. Anticipated relevance prepares visual processing for efficient memory-guided selection. *bioRxiv* (2025).

13. Duecker, K., Gutteling, T. P., Herrmann, C. S. & Jensen, O. No evidence for entrainment: Endogenous gamma oscillations and rhythmic flicker responses coexist in visual cortex. *J. Neurosci.* **41**, 6684–6698 (2021).

14. Duecker, K. et al. Guided visual search is associated with target boosting and distractor suppression in early visual cortex. *Commun. Biol.* **8**, 912 (2025).

15. Ferrante, O., Zhigalov, A., Hickey, C. & Jensen, O. Statistical learning of distractor suppression downregulates prestimulus neural excitability in early visual cortex. *J. Neurosci.* **43**, 2190–2198 (2023).

16. Minarik, T., Berger, B. & Jensen, O. Optimal parameters for rapid (invisible) frequency tagging using MEG. *NeuroImage* **281**, 120389 (2023).

17. Seijdel, N., Marshall, T. R. & Drijvers, L. Rapid invisible frequency tagging (RIFT): A promising technique to study neural and cognitive processing using naturalistic paradigms. *Cereb. Cortex* **33**, 1626–1629 (2023).

18. Zhigalov, A., Herring, J. D., Herpers, J., Bergmann, T. O. & Jensen, O. Probing cortical excitability using rapid frequency tagging. *NeuroImage* **195**, 59–66 (2019).

19. Landis, C. Determinants of the critical flicker-fusion threshold. *Physiol. Rev.* **34**, 259–286 (1954).

20. Spaak, E., Bouwkamp, F. G. & de Lange, F. P. Perceptual foundation and extension to phase tagging for rapid invisible frequency tagging (RIFT). *Imaging Neurosci.* **2**, 1–14 (2024).

21. Reynolds, J. H. & Desimone, R. Interacting roles of attention and visual salience in V4. *Neuron* **37**, 853–863 (2003).

22. Theeuwes, J. Attentional capture and control. *Annu. Rev. Psychol.* **76**, 1–24 (2025).

23. Theeuwes, J., Atchley, P. & Kramer, A. F. On the time course of top-down and bottom-up control of visual attention. *Atten. Perform.* **18**, 104–124 (2000).

24. Geng, J. J. Attentional mechanisms of distractor suppression. *Curr. Dir. Psychol. Sci.* **23**, 147–153 (2014).

25. Klink, P. C., Teeuwen, R. R. M., Lorteije, J. A. M. & Roelfsema, P. R. Inversion of pop-out for a distracting feature dimension in monkey visual cortex. *Proc. Natl Acad. Sci.* **120**, e2210839120 (2023).

26. Lin, R. et al. Neural evidence for attentional capture by salient distractors. *Nat. Hum. Behav.* **8**, 932–944 (2024).

27. Hickey, C., McDonald, J. J. & Theeuwes, J. Electrophysiological evidence of the capture of visual attention. *J. Cogn. Neurosci.* **18**, 604–613 (2006).

28. Hickey, C., Van Zoest, W. & Theeuwes, J. The time course of exogenous and endogenous control of covert attention. *Exp. Brain Res.* **201**, 789–796 (2010).

29. Schubö, A. Salience detection and attentional capture. *Psychol. Res.* **73**, 233–243 (2009).

30. Wang, B., van Driel, J., Ort, E. & Theeuwes, J. Anticipatory distractor suppression elicited by statistical regularities in visual search. *J. Cogn. Neurosci.* **31**, 1535–1548 (2019).

31. Cosman, J. D., Lowe, K. A., Zinke, W., Woodman, G. F. & Schall, J. D. Prefrontal control of visual distraction. *Curr. Biol.* **28**, 1330–1330 (2018).

32. Forschack, N., Gundlach, C., Hillyard, S. & Müller, M. M. Dynamics of attentional allocation to targets and distractors during visual search. *NeuroImage* **264**, 119759 (2022).

33. Born, S., Kerzel, D. & Theeuwes, J. Evidence for a dissociation between the control of oculomotor capture and disengagement. *Exp. Brain Res.* **208**, 621–631 (2011).

34. de Fockert, J. W. & Theeuwes, J. Role of frontal cortex in attentional capture by singleton distractors. *Brain Cogn.* **80**, 367–373 (2012).

35. Wandell, B. A., Dumoulin, S. O. & Brewer, A. A. Visual field maps in human cortex. *Neuron* **56**, 366–383 (2007).

36. Schneider, M., Tzanou, A., Uran, C. & Vinck, M. Cell-type-specific propagation of visual flicker. *Cell Rep.* **42**, 112492 (2023).

37. Zhigalov, A. & Jensen, O. Alpha oscillations do not implement gain control in early visual cortex but rather gating in parieto-occipital regions. *Hum. Brain Mapp.* **41**, 5176–5186 (2020).

38. Faul, F., Erdfelder, E., Lang, A. G. & Buchner, A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* **39**, 175–191 (2007).

39. Duncan, D. H., Forschack, N., van Moorselaar, D., Müller, M. M. & Theeuwes, J. Learning modulates early encephalographic responses to distracting stimuli: a combined SSVEP and ERP study. *J. Neurosci.* **45**, 1234–1245 (2025).

40. Kleiner, M. et al. What's new in psychtoolbox-3. *Perception* **36**, 1–16 (2007).

41. Drijvers, L., Jensen, O. & Spaak, E. Rapid invisible frequency tagging reveals nonlinear integration of auditory and visual information. *Hum. Brain Mapp.* **42**, 1138–1152 (2021).

42. Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J. M. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* **2011**, 156869 (2011).

43. Hershman, R., Henik, A. & Cohen, N. A novel blink detection method based on pupillometry noise. *Behav. Res. Methods* **50**, 107–114 (2018).

44. Pan, Y., Frisson, S. & Jensen, O. Neural evidence for lexical parafoveal processing. *Nat. Commun.* **12**, 5234 (2021).

45. Hustá, C., Meyer, A. & Drijvers, L. Using rapid invisible frequency tagging (RIFT) to probe the neural interaction between representations of speech planning and comprehension. *Neurobiol. Lang.* **1**, 1–16 (2025).

633 46. Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. & Baker, C. I. Circular analysis in systems
634 neuroscience: the dangers of double dipping. *Nat. Neurosci.* **12**, 535–540 (2009).

635 47. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG and MEG data. *J. Neurosci.*
636 *Methods* **164**, 177–190 (2007).

637 48. Wang, D. et al. Dynamic competition between bottom-up saliency and top-down goals in early
638 visual cortex [Data set]. *OSF* https://doi.org/10.17605/OSF.IO/SZUXA (2026).

639 ## Acknowledgments

640 J.T. is supported by a European Research Council (ERC) advanced grant 833029
641 [LEARNATTEND] and by a NWO Open competition grant 25 406.21.GO.034. D.W. is
642 supported by China Scholarship Council scholarship 202308510050.

643 ## Author contributions

644 D.W., S.G., S.V.d.S., and S.C. conceptualized the study. D.W., S.G., K.A., S.V.d.S., and S.C.
645 developed the methodology and experimental design. S.V.d.S., S.G., and S.C. supervised the
646 project. D.W., K.A., and S.C. performed the visualization. D.W., S.G., and S.C. wrote the
647 original draft, and all authors (D.W., S.G., S.V.d.S., S.C., K.A., and J.T.) reviewed and edited
648 the manuscript.

649 ## Competing interests

650 The authors declare no competing interests.

651 ## Figure captions

652 **Figure 1**. **Experimental paradigm, behavioral results, and validation of frequency-specific neural responses.**
653 **(A)** Experimental paradigm. After an initial display consisting of placeholders, the search display was presented.
654 Participants were instructed to search for a unique shape singleton target (here a diamond among circles) and
655 respond as quickly and accurately as possible to orientation of the line segment inside it. On half of the trials, a
656 color distractor was present (here: red among green). Target and distractor (or one of the nontargets) were
657 frequency tagged (luminance-modulated) at 60 Hz and 64 Hz throughout the search display (see Tagging
658 manipulation for details). Note: the figure is not to scale; no outlines were visible around the flickering regions in

659 the actual experiment. **(B)** Behavioral results. Participants were slower to find the target when a distractor was
660 present. The blue bar represents reaction times in the distractor present condition, while the orange bar represents
661 the distractor absent condition. Each dot indicates the mean response time of an individual participant (n=24
662 biologically independent samples). ***$p < 0.001$. **(C)** Time-frequency plot of coherence (phase realigned 60 Hz),
663 averaged across participants and individuals top 6 channels. *Top right inset:* Scalp topography of average 60 Hz
664 coherence across the 1.2-second after search display onset. **(D)** Time-frequency plot of coherence (64 Hz),
665 averaged across participants and individuals top 6 channels. *Top right inset:* Scalp topography of average 64 Hz
666 coherence across the 1.2-second after search display onset.

667

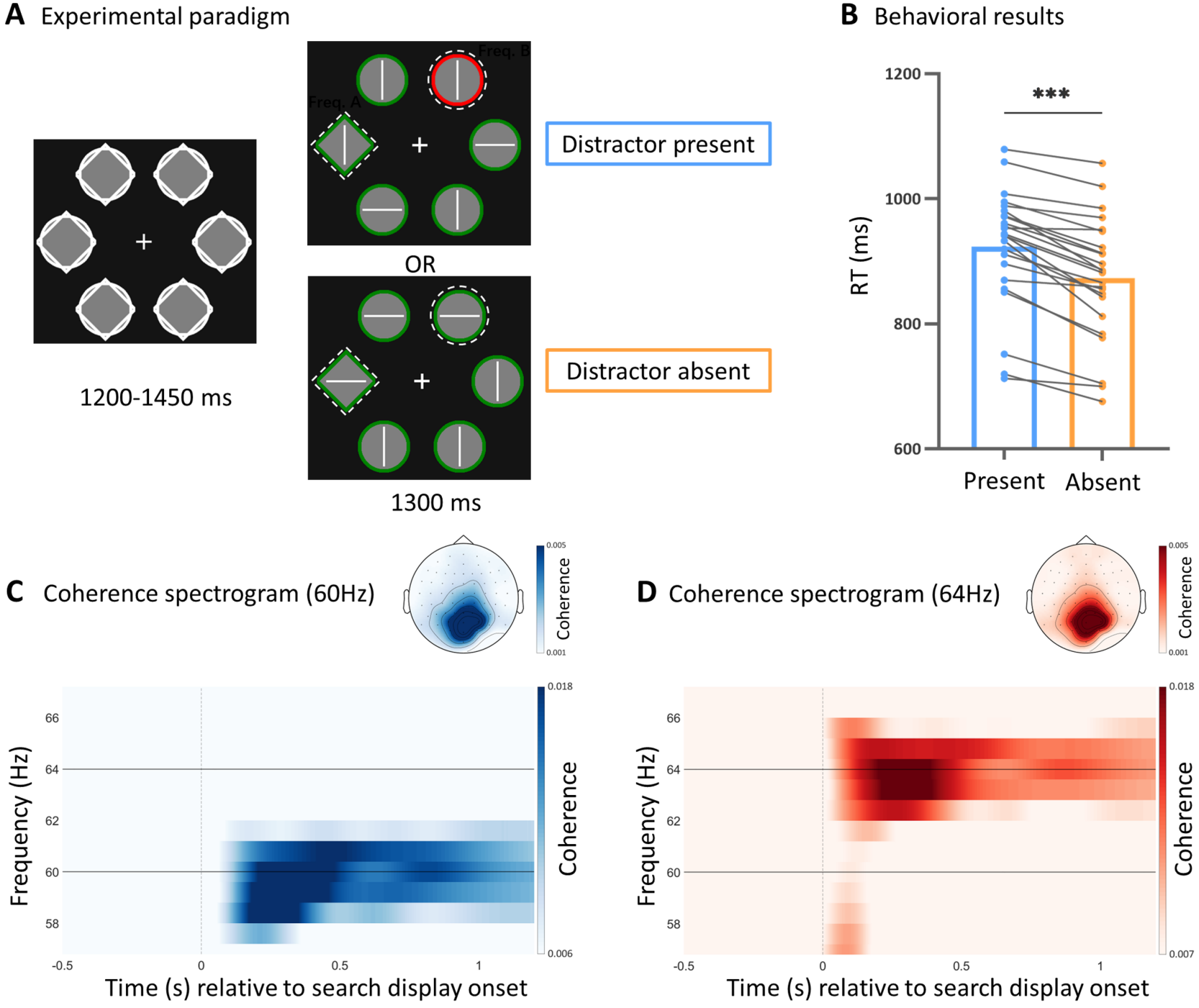
668 **Figure 2. RIFT responses across experimental conditions. (A).** Coherence time-series of distractor (blue
669 dashed), nontarget (orange dashed) and the difference (purple solid). Shaded areas represent 95% confidence
670 intervals of the mean. Significant clusters (from cluster based-permutation tests) are indicated by horizontal solid
671 black lines. Right bar graphs show the time-to-peak analysis of the coherence trace (0-600 ms) for each condition.
672 Each dot represents the peak time point of an individual participant (n=24 biologically independent samples). *p
673 < 0.05. **(B).** Coherence time-series of target without distractor (orange solid), nontarget (orange dashed) and the
674 difference (purple solid). **(C).** Coherence time-series of target with distractor (blue solid), target without distractor
675 (orange solid) and the difference (purple solid). **(D).** Coherence time-series of target with distractor (blue solid),
676 distractor (blue dashed) and the difference (purple solid).

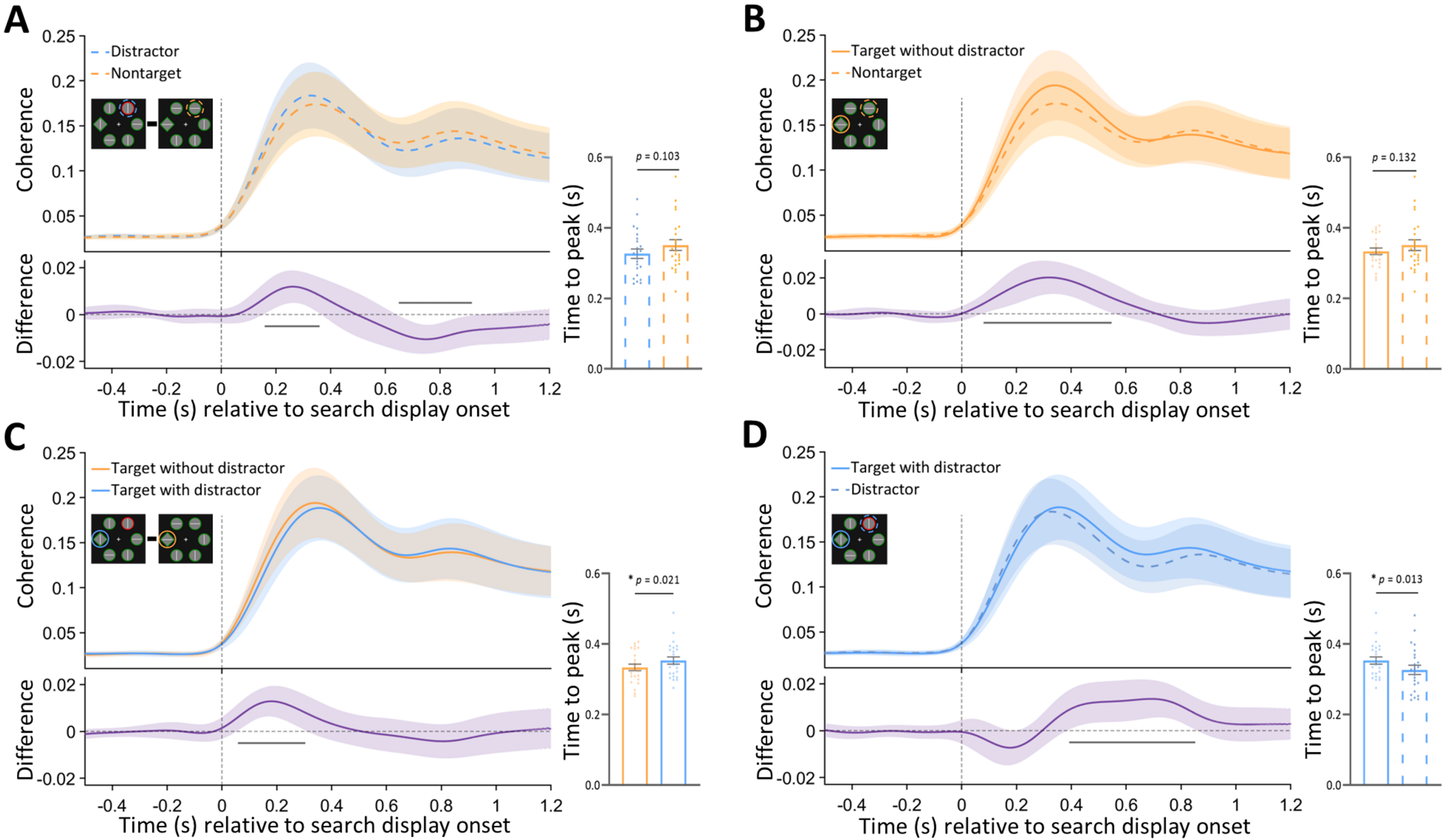
677

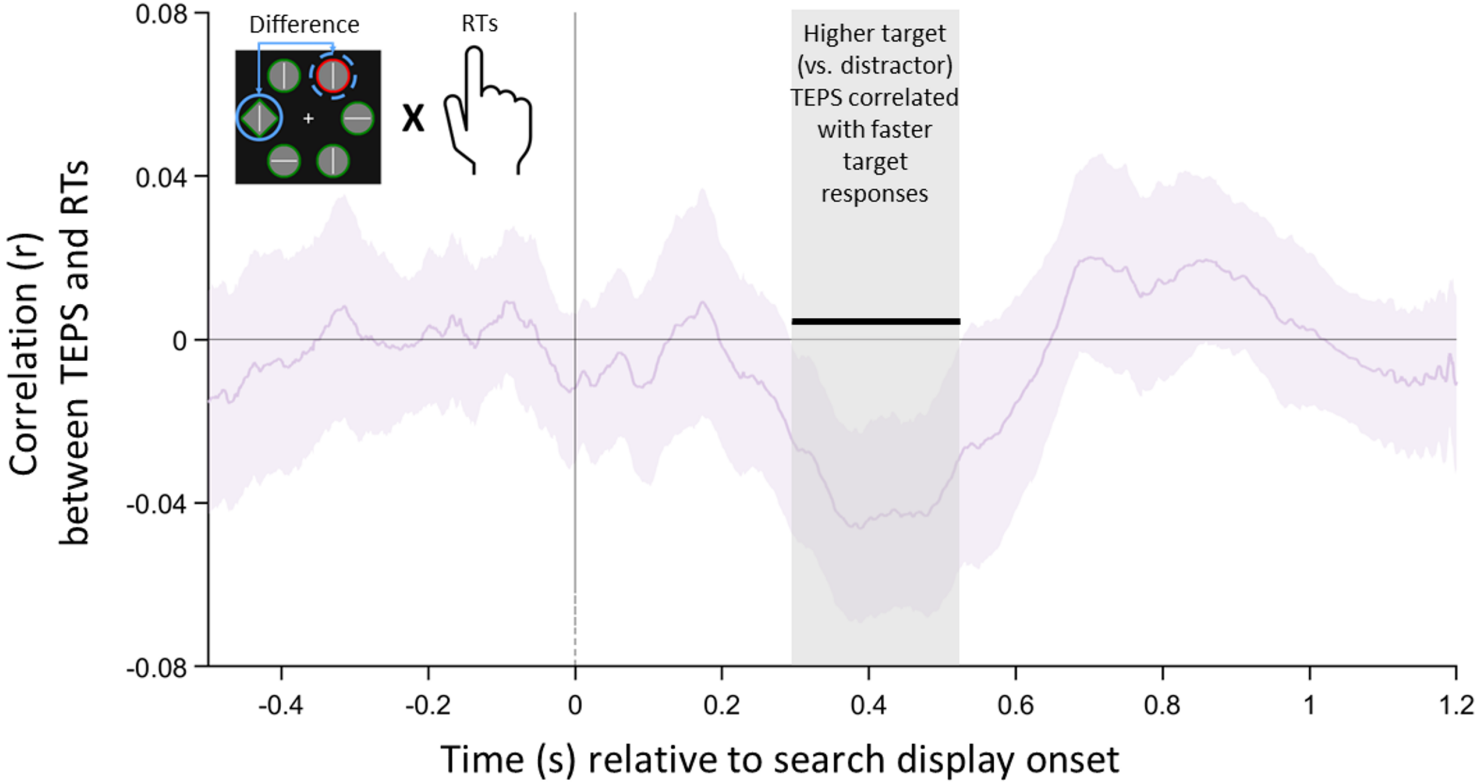
678 **Figure 3. Time-resolved, trial-wise correlation between RTs and the TEPS difference between target and**
679 **distractor within the same trial.** Only correct trials in the distractor present condition used. Shaded areas
680 represent 95% confidence intervals of the mean (n=24 biologically independent samples). Significant clusters
681 (from cluster based-permutation tests) are indicated by horizontal solid black lines.

682

683


A Experimental paradigm
Freq. B
Freq. A
Distractor present
OR
Distractor absent
1200-1450 ms
1300 ms
B Behavioral results
1200
1000
800
600
RT (ms)
***
Present
Absent
C Coherence spectrogram (60Hz)
0.005
Coherence
0.001
0.018
66
64
62
60
58
Frequency (Hz)
Coherence
0.006
-0.5
0
0.5
1
Time (s) relative to search display onset
D Coherence spectrogram (64Hz)
0.005
Coherence
0.001
0.018
66
64
62
60
58
Frequency (Hz)
Coherence
0.007
-0.5
0
0.5
1
Time (s) relative to search display onset


A
Distractor
Nontarget
Coherence
Difference
Time (s) relative to search display onset
Time to peak (s)
p = 0.103
B
Target without distractor
Nontarget
p = 0.132
C
Target without distractor
Target with distractor
* p = 0.021
D
Target with distractor
Distractor
* p = 0.013


Difference
RTs
X
Higher target (vs. distractor) TEPS correlated with faster target responses
0.08
0.04
0
-0.04
-0.08
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
1.2
Correlation (r) between TEPS and RTs
Time (s) relative to search display onset