Neural evidence for attentional capture by salient distractors

Rongqi Lin[1,2,3,4,5,6], Xianghong Meng[7], Fuyong Chen[8], Xinyu Li[5], Ole Jensen[9], Jan Theeuwes[6], and Benchi Wang[1,2,3,4]

[1]Key Laboratory of Brain, Cognition and Education Sciences (South China Normal University), Ministry of Education, China

[2]Institute for Brain Research and Rehabilitation, South China Normal University, China

[3]Center for Studies of Psychological Application, South China Normal University, China

[4]Guangdong Key Laboratory of Mental Health and Cognitive Science, South China Normal University, China

[5]Department of Psychology, Zhejiang Normal University, China

[6]Department of Experimental and Applied Psychology, Vrije Universiteit Amsterdam, the Netherlands

[7]Department of Neurology, Shenzhen University General Hospital, China

[8]Department of Neurosurgery, The university of Hongkong Shenzhen Hospital, China

[9]Centre for Human Brain Health, School of Psychology, University of Birmingham, United Kingdom

Author notes

RL, XM, and FC equally contributed to the current research. RL, XM, FC, XL, and BW designed the experiment, RL, XM and FC collected the data, RL and BW analyzed the data. RL, OJ, JT, and BW wrote the paper. All authors approved the final version of the manuscript for submission and declared no competing financial interests. This research was supported by the National Science and Technology Innovation 2030 Major Program (2022ZD0204802), and the National Natural Science Foundation of China grant (32000738) to BW, the Sanming Project of Medicine in Shenzhen (SZSM202003006) to XM. Correspondence should be addressed to Benchi Wang, Institute for Brain Research and Rehabilitation, South China Normal University, Zhongshan Road West 55, Guangzhou, China, 510000. Email: wangbenchi.swift@gmail.com. The processing codes and derived iEEG data will be made available upon reasonable request.

Abstract

Salient objects often capture our attention, serving as distractors and hindering our current goals. It remains unclear when and how salient distractors interact with our goals and our knowledge on the neural mechanisms responsible for attentional capture is limited to a few brain regions recorded from non-human primates. Here we conducted a multivariate analysis on human intracranial signals covering most brain regions, and successfully dissociated distractor-specific representations from target-arousal signals in the high-frequency (60-100 Hz) activity. We found that salient distractors were processed rapidly around 220 ms, while target-tuning attention was attenuated simultaneously, supporting initial capture by distractors. Notably, neuronal activity specific to the distractor representation was strongest in superior and middle temporal gyrus, amygdala, and anterior cingulate cortex, while there were smaller contributions from parietal and frontal cortices. These results provide neural evidence for attentional capture by salient distractors engaging a much larger network than previously appreciated.

Keywords: Attentional capture; salient distractor; intracranial EEG; IEM decoding

Introduction

Rapidly orienting towards salient objects in a crowded environment has played a crucial role in human evolution as it permits us to quickly identify potential prey, mates or predators. This inherent human trait also makes us vulnerable as our attention can be captured by salient objects around us. For instance, we are often distracted by our phone's buzzer when focusing on a task at hand. Similarly, when performing visual search tasks in a laboratory setting, certain types of stimuli that stand out from the environment (e.g., a red circle surrounding by green stimuli), known as salient objects, automatically capture our attention[1,2], regardless of their relevance to our immediate goals[3–6], giving rise to the phenomenon of attentional capture. Research over the past few decades has revealed behavioral evidence on how salient objects (distractors) capture our attention and the properties of salient distractor processing[3] (for review, see refs.[1,7]). Despite this, there has been a fierce debate regarding whether or not salient objects capture attention automatically. Although this debate has not yet been resolved by behavioral studies, the present study with intracranial electroencephalography (iEEG) recordings might help, by providing direct evidence regarding the neural dynamics and brain regions in the human brain responsible for attentional capture and the ways in which salient distractors interact with our current goals.

Circumstantial evidence supporting salient distractor processing in the brain is mainly provided by neurophysiological studies conducted in non-human primates. These studies have shown that the neuronal activity increases in the visual, parietal, and frontal cortex when salience signals are present within the receptive field[8–13]. This activity decreases after monkeys have been trained to ignore the salience signals that

act as distractors[14–16]. Although these experiments provide important insight into the neural mechanisms underlying salient distractor processing, they typically involve recording neuronal activity from small and often restricted retinotopic locations (i.e., receptive fields)[17,18], making it difficult to track the target (current goals) and distractors simultaneously and thus limiting our understanding of how they interact. Similarly, noninvasive electrophysiological studies in humans that measure event-related potential (ERP) components (the N2pc and Pd) to characterize neural responses to salient distractors, also evaluate target or distractor processing individually[19]. That is, it is important to note that the assessment of target and distractor processing in these studies[17-19] never occurs within the same group of trials, i.e., they are never evaluated within the same timeframe. Therefore, these findings deserve to be complemented by human intracranial studies uncovering the full network involved in the simultaneous processing of the distractors and targets in relation to attentional capture.

Developing a comprehensive theory of attentional capture requires a deep understanding of how the human brain handles salient distractors. To explore this, we examined iEEG recordings from eighteen neurosurgical patients performing a visual search task[4,20], which offered a unique opportunity to investigate the neuronal responses related to salient distractor processing in the human brain with optimal anatomical precision and high temporal resolution. Moreover, a well-established inverted encoding model[21–23] (IEM, which can well track spatial attention in neuroscientific studies; for review, see ref.[24]) was applied to track the spatial tunings originating from the target and the distractor simultaneously in high-frequency activity (60-100 Hz). This representational specific multivariate approach enabled us

to dissociate the processing of salient distractors from that of targets, and provided the time course for the target and distractor processing, in order to examine when and how distractors interacted with the current goals. Additionally, by tracking neural responses towards salient distractors across various brain regions, we identified whether distinct brain regions exhibited activity specific to the distractor representation, and how they interact with each other.

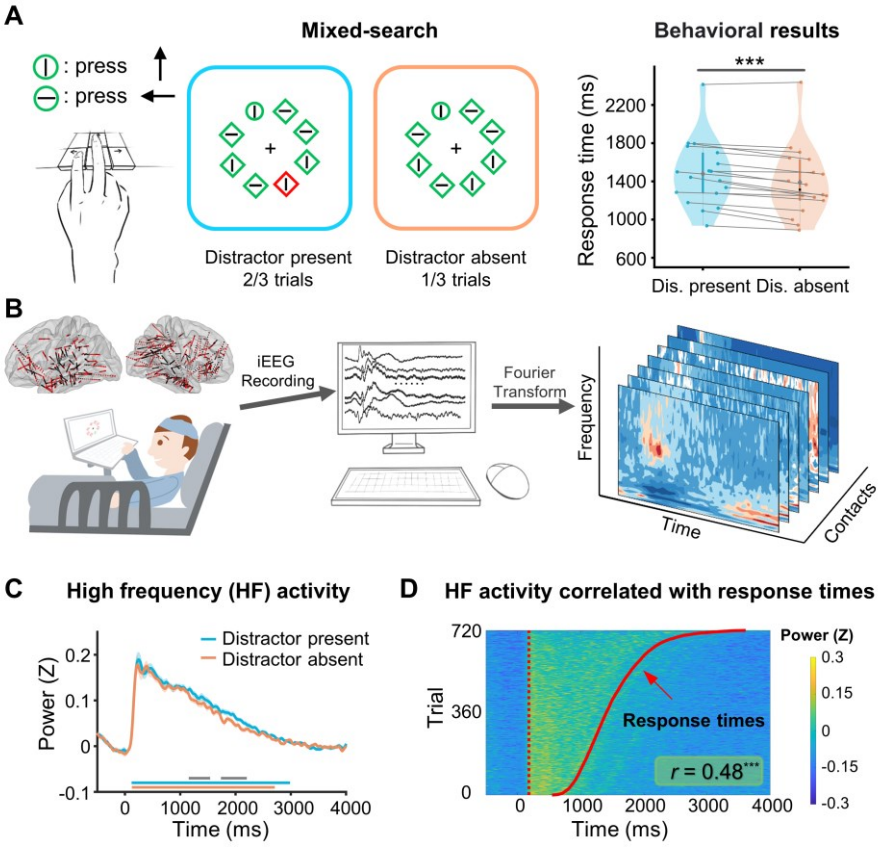


Figure 1. (A) Display setup and possible locations for search elements. Participants searched for one unique shape (target) among search elements (e.g., circle among diamonds), and indicated whether the line segment inside the target was vertical or horizontal by pressing the 'up' or 'left' key respectively, and as fast as possible. The rightmost panel shows the behavioral results, solid dots indicate individual response times (RTs) respectively. (B) Intracranial EEG data was recorded while participants performed the visual search task, from which high-frequency (HF; 60-100 Hz) power was extracted. The left side indicates possible locations for the contacts located in various brain areas. Red dots indicate locations of responsive contacts. (C) The observed HF power between the distractor present and absent conditions, with data variance represented by ±1 SEM. Significant areas are highlighted by solid lines, and the gray line marks the significant difference between conditions (cluster-based

permutation test, $p < .01$). (D) Trial-by-trial (Pearson) correlations between the time for archiving the peak of HF activity and RTs. The dashed red lines indicate the start of HF activity by visual inspection (180ms), and solid red lines indicate the trial-specific RTs. *$p < .05$, **$p < .01$, ***$p < .001$.

Results

In our visual search task, participants searched for a unique shape (target) among search elements while keeping fixation at the center and responded to the orientation of the line segment inside the target (see Fig. 1A). The target was present on each trial; while in the distractor present condition, a uniquely colored distractor was randomly present in two-thirds of the trials (see Materials and Methods for details) and participants were asked to ignore it. We also tested another session before the present experiment, in which only the target was present throughout the whole session for practicing; its results can be found in Supplementary Information (SI; Fig. S1). As shown in Fig. 1A, a paired t-test showed that the mean response times (RTs) were slower when a salient distractor was present (1480 ms) compared to when it was absent (1390 ms), $p < .001$ (FDR corrected). Consistent with other studies[4,20] (for review, see refs.[1,25]), the slower RTs in the distractor present condition provide evidence for attentional capture.

High-frequency activity

We first identified 1275 responsive contacts (after removing 7.5% of responsive contacts that were identified as potentially correlated with epileptic seizure, see Materials and Methods for details) that showed high-frequency (HF; 60-100 Hz) activity during the experiment (see Table S1 for individual numbers). The HF activity occurred in the 120 – 2986 ms and 124 – 2710 ms intervals for the distractor present and absent conditions respectively. Note that, in the presence of a salient distractor,

the initial burst of HF activity reduced slowly, resulting in higher power in the 1154 – 1540 ms and from 1736 – 2204 ms relative to the distractor absent condition (see Fig. 1C), reflecting a prolonged visual search process due to the involvement of the salient distractor. This conclusion was based on the observation that HF activity is typically stronger during challenging searches that produce slower responses[26], as well as on the association of HF activity with neuronal processing including sustained attention (for review, see ref.[27]). Results for other low frequency bands can be found in Fig. S2A and S2B.

We further examined the relationship between HF activity and participants' RTs. To do this, we initially sorted individual trial-specific RTs and then calculated their average across participants. Subsequently, we conducted a Pearson correlation analysis. The results revealed a significant trial-by-trial correlation between the peak time of HF activity and RTs, $r = 0.48$, $p < .001$ (see Fig. 1D). In sum, HF activity is correlated with our behavioral responses, which might reflect the processing of targets and salient distractors. This would be examined by further isolating the target and distractor neuronal responses using a multivariate analysis.

IEM reconstruction in High-frequency activity

To track the processing of targets and salient distractors, an inverted encoding model (IEM) was applied[23] (see Fig. 2A and Materials and Methods for details). Overall, spatial attention towards the target and salient distractor were reconstructed from HF activity, beginning approximately 200 ms after the onset of the search array, and persisted over time (see Fig. S3 for reconstructed HF CTFs across time for each of eight locations). As shown in Fig. 2B, the positive target-tuning CTF slopes (reflecting the spatial tuning of population-level HF activity towards the target) were

significant in the 230 – 338 ms and the 348 – 590 ms intervals in the distractor absent condition; while in the distractor present condition, target-tuning CTF slopes were significant above zero in the 308 – 538 ms and the 560 – 772 ms intervals (cluster-based permutation test, $p < .01$). Noted that when a salient distractor was introduced in the distractor present condition, compared to the distractor absent condition, target-tuning CTF slopes were notably reduced in the early time window (236 – 302 ms; cluster-based permutation test, $p < .01$), and the latency was significantly later, $t(17) = 6.43$, $p < .001$, Cohen's d = 1.52. This implies that spatial attention was captured by the salient distractor, leading to less effective processing of the target.

The current results, depicted in Fig. 2C left panel, indicate the simultaneous tracking of spatial attention towards the salient distractor and the target. The distractor-tuning CTF slopes (reflecting the spatial tuning of population-level HF activity towards the distractor) were significant above zero in the 218 – 546 ms interval. It is worth noting that, the magnitude of distractor-tuning CTF slopes was stronger than that of target-tuning CTF slopes in the early time window (228 –302 ms; cluster-based permutation test, $p < .01$), and its latency was significantly earlier, $t(17) = 4.32$, $p < .001$, Cohen's d = 1.02, indicating an initial and rapid processing of salient distractor, suggesting a cortical mechanism that swiftly reacts to a pop-out signal[14]. Additionally, distractor-related attention declined quickly, as its CTF slopes were lower compared to the target ones in the late time window from 568 ms to 664 ms (cluster-based permutation test, $p < .01$). No activity was detected for the non-responsive contacts (Fig. 2C right panel), and we were not able to reconstruct early spatial attention (before 400 ms) towards the target from other low-frequency bands (see Fig. S2C and S2D in SI for details). In sum, this indicates that the initial

processing of salient distractor is short-lived, but still capturing our attention.

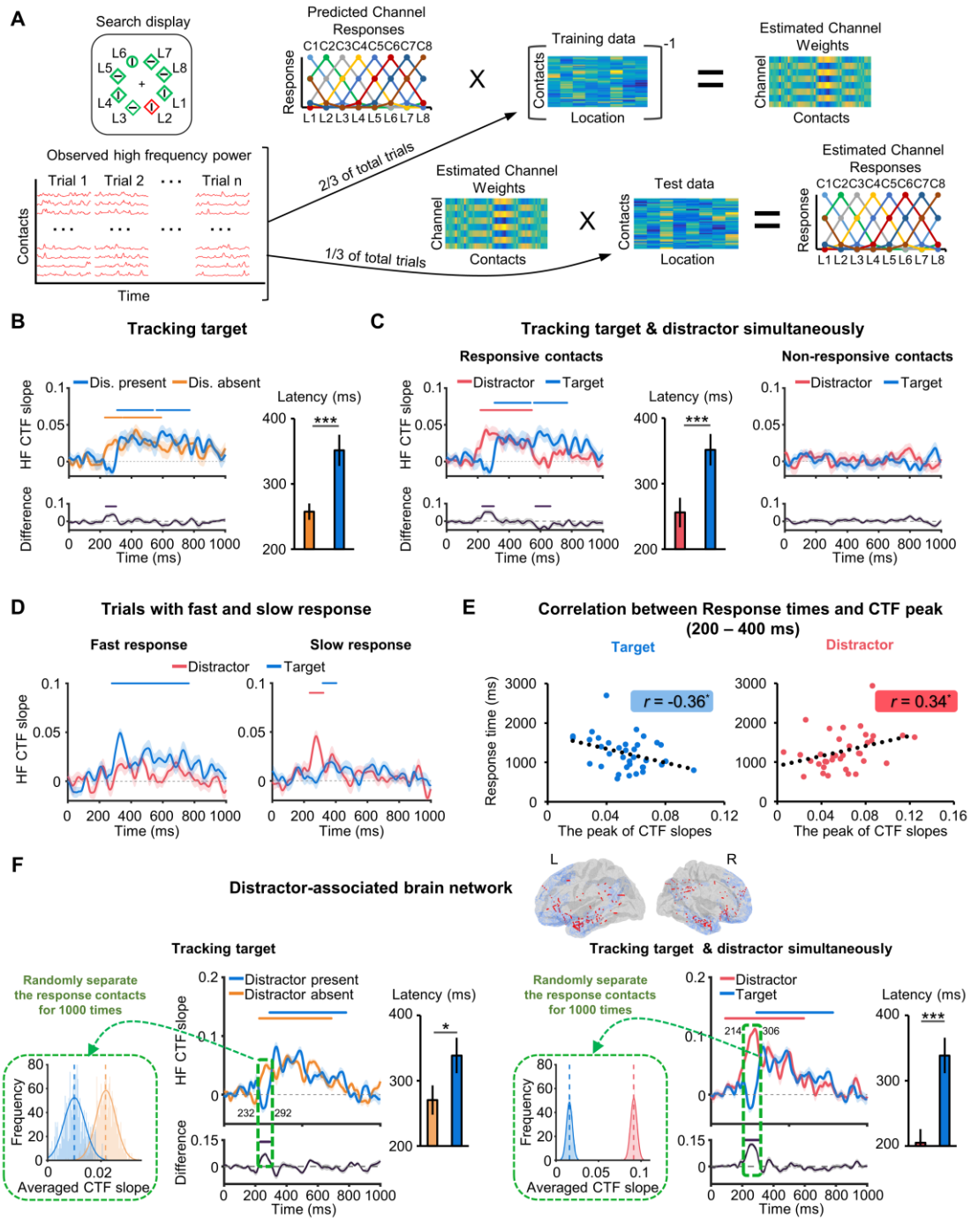


Figure 2. (A) Illustration of the inverted encoding model (IEM; see Materials and Methods for details). The search elements are situated at eight potential locations (L1 to L8) around a fixation cross, forming a virtual circle. We modelled High-frequency (HF) power at each responsive contact by calculating the weighted sum of eight spatially selective channels (C1 to C8), each channel tuned for one of the eight possible locations. We then split the data into two parts to obtain the estimated weights and calculated the channel response for each location. (B) It shows the spatial selectivity (measured as CTF slopes) for

the target in both the distractor present and absent conditions, including absent-minus-present difference in the magnitude of CTF slopes and the latency of the CTF slopes. (C) The left and right panels show the spatial selectivity for the target and salient distractors in the distractor present condition for responsive and non-responsive contacts, respectively. They also display the distractor-minus-target difference in the magnitude of CTF slopes and the latency of the CTF slopes. (D) The left and right panels show the spatial selectivity for the target and salient distractors in the distractor present condition for fast and slow trials. We divided all trials according to fast and trials per participant (median split). Following this, we applied the IEM model to the fast and slow trials per participant and conducted further analysis. (E) The correlation between the peak of CTF slopes and behavioral RTs for the target and distractor separately. For fast and slow trials per participant, we separately calculated the mean RTs and selected the peak of CTF slopes within the early time window (200 – 400 ms). (F) The results obtained using the random-group-strategy (see Fig. S5 and the corresponding text in the SI for details) for responsive contacts located in the distractor-associated brain network. The left panel show the spatial selectivity (measured as CTF slopes) for the target in the distractor present and absent conditions, including absent-minus-present difference in the magnitude of CTF slops and the latency of the CTF slopes, and the distribution of the target-tuning averaged CTF slopes for the distractor present and absent conditions from 1000 times IEM reconstructions. The right panel shows the spatial selectivity for the target and salient distractors in the distractor present condition, including distractor-minus-target difference in the magnitude of CTF slopes, the latency of the CTF slopes, and the distribution for the target- and distractor-tuning averaged CTF slopes from 1000 times IEM reconstructions. The green dashed outlines indicate the time windows used to estimate averaged CTF slopes for calculating the distribution. The data variance is represented by ±1 SEM. Significant CTF selectivity (or in difference) areas are highlighted by solid lines (cluster-based permutation test, $p < .01$). *$p < .05$, **$p < .01$, ***$p < .001$.

Further analyses were conducted to investigate whether early processing of distractors and targets were related to participants' behavioral responses. We first divided all trials according to fast and slow responses for each participant (median split). Following this, we applied the IEM model to the fast and slow trials per participant and observed that for fast trials, the target-tuning CTF slopes were significant above zero in the 278 – 764 ms interval (cluster-based permutation test, $p < .01$), whereas no significant effects were observed for the distractor. It is noteworthy

that for slow trials, the target-tuning CTF slopes were significant above zero in the 318 – 404 ms interval, while the positive distractor-tuning CTF slopes emerged earlier in the 236 – 322 ms interval (see Fig. 2D). This indicates that early distractor-related processing delays the behavioral responses.

Additionally, we examined the correlation between the peak of CTF slopes and behavioral RTs for targets and distractors separately. For fast and slow trials per participant, we calculated the mean RTs and selected the peak of CTF slopes within the early time window (200 – 400 ms). Results showed a negative correlation between mean RTs and the peak of target CTF slopes, $r = -0.36$, $p = .03$, but a positive correlation between mean RTs and the peak of distractor CTF slopes, $r = 0.34$, $p = .042$ (see Fig. 2E). These findings further support the notion that early processing of distractors drives subsequent behavioral responses.

In the present study, salient distractors were never positioned at the same location as the target. That is, they occupied different locations in each distractor present trial. Given that we were simultaneously tracking both the target and salient distractor, one might question whether the CTF slopes for salient distractors exhibit a co-variation with those for the target. In such a scenario, one would anticipate the correlation in the latency and peak of the CTF slopes between tracking the target and tracking the salient distractor. However, our findings do not support this hypothesis, as no such correlation was observed, both ps > 0.24 (see Fig. S4A).

Moreover, by creating 8 classifiers in Support Vector Machine (SVM) according to HF power for each target and salient distractor location in the distractor present condition, we were able to decode the processing of target and salient distractor locations. The SVM decoding results showed that the classification accuracy for the

248 salient distractor exceeded chance-level (1/8) in the 216 – 470 ms interval; while that
249 for the target exceeded chance-level during the intervals of 294 – 446 ms, 542 – 832
250 ms and 844 – 926 ms (see Fig. S4B; additional information regarding methods can be
251 found in SI). This reanalysis based SVMs rather IEMs confirms the early and rapid
252 detection of salient distractors.

253 **Using IEM reconstruction to identify distractor-associated brain network**

254 Using the IEM model, we were able to simultaneously track the processing of the
255 target and salient distractor across all responsive contacts. However, it's possible that
256 certain brain areas contribute more significantly to the processing and interaction
257 between the target and salient distractor. To explore this possibility, we grouped
258 responsive contacts into six smaller brain networks based on large-scale cerebral
259 networks[28] to further localize certain brain areas specific to salient distractor
260 processing. The sole purpose of utilizing this brain network template was to group
261 responsive contacts into functionally defined brain areas.

262 It should be noted that the number of responsive contacts in each brain network
263 varied greatly across participants (see Table S2 and S3). This variability made it
264 impractical to apply the IEM model separately to each brain network for every
265 participant. To mitigate this issue, we employed a random-group-strategy, where we
266 randomly grouped all responsive contacts across participants into groups of at least
267 eight contacts each per brain network. We assumed that the responses in contacts
268 from different participants within the same brain area should be similar. If spatial
269 attention towards salient distractors is governed by a single brain area, it should be
270 reconstructed in IEM model regardless of how the contacts within this brain area are
271 distributed, whether within or across participants (see Fig. S5). We validated the

random-group-strategy by dividing all responsive contacts into 20 groups, mimicking a group of 20 participants; and replicated the critical results from the raw group of 18 participants with high statistical power (refer to Fig. S5 and the corresponding text in SI for more detailed information).

To determine which brain areas are mainly involved in the processing and interaction between the target and salient distractor, the random-group-strategy (validated on all responsive contacts) was applied for six different brain networks separately based on large-scale cerebral networks[28]. Overall, we identified two brain networks (linked to the Default and Limbic system) that were primarily engaged in handling both the target and salient distractor (see Fig. S6 and S7, more detailed results regarding each brain network can be found in SI). Thus, in further analysis, we combined these areas and referred to them as the *distractor-associated brain network*. The results showed in this distractor-associated brain network replicated those found for all responsive contacts (see Fig. 2F, details can be found in SI). Critically, in the distractor present condition, compared to the distractor absent condition, target-tuning CTF slopes were significantly lower in the early time window (232 – 292 ms; cluster-based permutation test, $p < .01$). Additionally, in the distractor present condition, the magnitude of distractor-tuning CTF slopes was stronger than that of target-tuning CTF slopes in the early time window (214 – 306 ms; cluster-based permutation test, $p < .01$). It would also be interesting to incorporate the target-associated brain network as a focal point in future studies.

Noted that, importantly, this critical result observed in the distractor-associated network was not attributed to arbitrary placement of contacts. To address this concern, we iterated the grouping of contacts 1000 times to generate distributions of the

296 averaged CTF slopes (from 232 ms to 292 ms) for the target in the distractor present
297 and absent conditions, and that of the averaged CTF slopes (from 214 ms to 306 ms)
298 for the distractor and target in the distractor present condition (see Materials and
299 Methods for further details). The results revealed that the target-related distribution in
300 the distractor present condition differed significantly from that in the distractor absent
301 condition (chi-test, $p < .01$), with averaged CTF slopes smaller for target-tuning in the
302 distractor present condition (see Fig. 2F left panel); and the distractor-related
303 distribution differed significantly from the target-related distribution (chi-test, p
304 $< .01$), with averaged CTF slopes smaller for target-tuning than distractor-tuning (see
305 Fig. 2F right panel). This confirms that the results were not due to arbitrary placement
306 of contacts.

307 **An extended network in human brain supporting salient distractor processing**

308 Although we successfully identified the distractor-associated brain network,
309 within this brain network it was unclear which brain regions dominated salient
310 distractor processing the most and how salient distractor was exchanged between
311 these regions. Noted that, the number of contacts for each brain region varied
312 significantly and some of these regions did not have a sufficient number of contacts to
313 apply the IEM model. Thus, to investigate this further, we used a leave-one-out
314 procedure to determine the contribution to distractor-related CTF slopes from each
315 brain region. This approach was adopted based on a previous study that used a similar
316 method to investigate the contribution to the neural representations from different
317 brain regions[29]. We assumed that if a particular brain region made the greatest
318 contribution to the CTF slopes, excluding that region would cause the slopes to
319 decline most dramatically, and *vice versa* (see Fig. 3A). This procedure was repeated

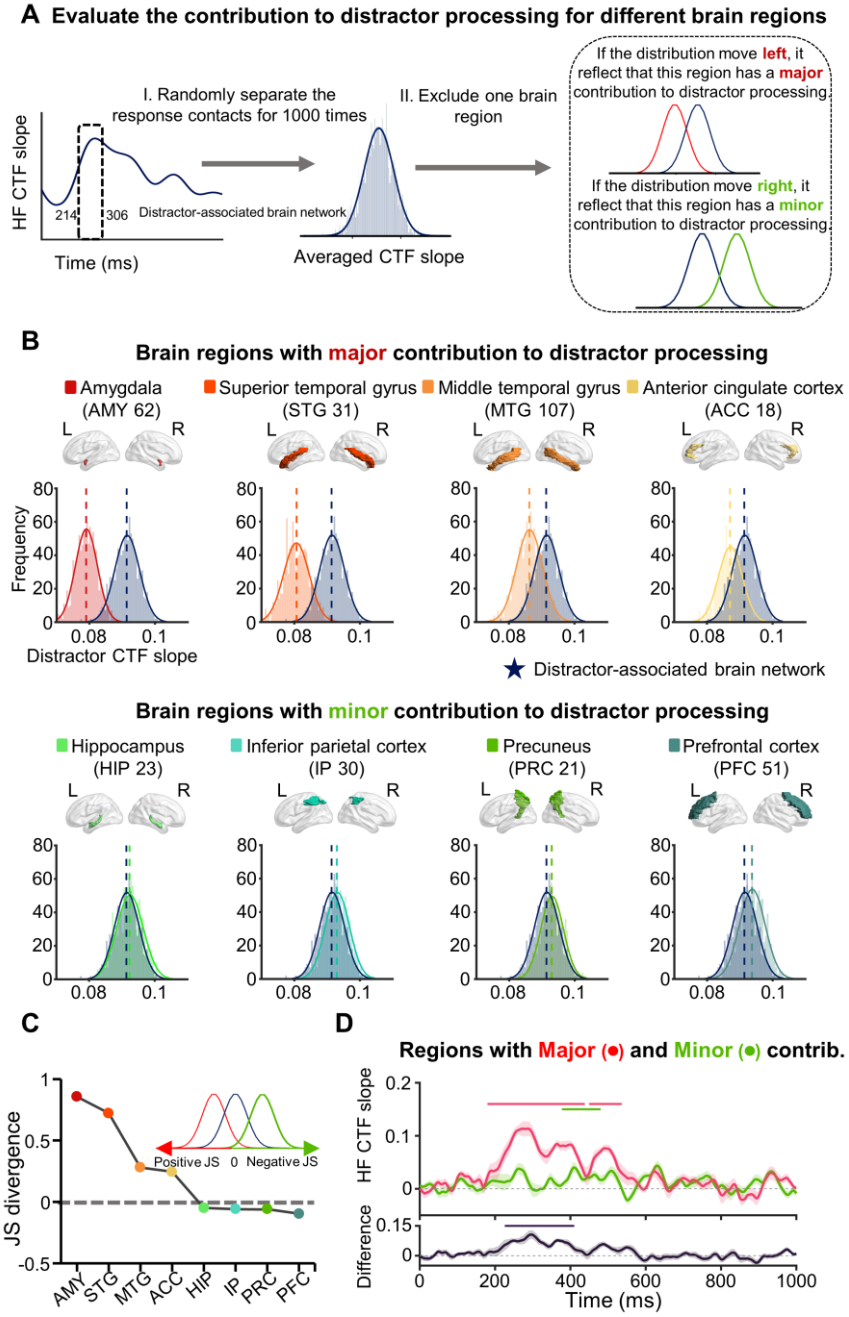
for 14 brain regions (which at least included 10 responsive contacts, defined by AAL template) located in the distractor-associated brain area. We first calculated the distribution of the averaged CTF slopes (from 214 ms to 306 ms, a time window showing significant CTF decoding for distractors in the distractor-associated brain network, and then created the same distribution after excluding each region. As shown in Fig. 3B, chi-square tests revealed significant differences between the distribution from the distractor-associated area and the distribution created ($p < .05$, FDR corrected) for STG (superior temporal gyrus), AMY (amygdala), MTG (middle temporal gyrus), ACC (anterior cingulate cortex), PRC (precuneus), IP (inferior parietal), HIP (hippocampus), and PFC (prefrontal cortex). The contribution from different brain regions were strongest from the AMY, and decreased gradually from STG, MTG, to ACC (see Fig. 3B and 3C). They all played a major role in salient distractor processing as their exclusion led to a decline in CTF slopes. In contrast, the PRC, IP, HIP, and PFC offered only minor contributions to this process as their exclusion led to an increase in CTF slopes[1].

We further computed Jensen-Shannon (JS) divergence (where a positive JS value indicated a major contribution to salient distractor processing, see Materials and Methods section for details) for these brain regions, measuring the distance between created distributions, showing that JS divergence decreased as the contribution from each region reduced (from AMY to PFC, see Fig. 3C). In brief, the current analysis

[1]Note that, we did not discuss the other six brain regions (medial orbitofrontal cortex, lateral orbitofrontal cortex, inferior temporal gyrus, posterior cingulate cortex, para hippocampus, and entorhinal) as they did not show significant change in salient distractor processing when removing them, but this does not discount their potential involvement in salient distractor processing, and warrants further investigation.

indicates that salient distractor processing is contributed to by various brain regions at different levels.

It is essential to appreciate that all these variations in CTF slopes, whether originating from major or minor regions, are linked to each other. We defined major regions as those after we excluded their contacts resulting in a relatively smaller CTF slope in response to salient distractors. Conversely, the minor regions were defined as those after we excluded their contacts resulting in a relatively larger CTF slope. In regard to the minor contribution regions, three interpretations are possible. Signals related to distractors from these minor regions may exhibit positive, close-to-zero, or negative CTF slopes. To further investigate which scenario is operative, we employed the IEM model to track the attention towards salient distractors for both types of regions. The results revealed that distractor-related attention was successfully reconstructed from 182 ms to 436 ms and from 452 ms to 534 ms for the brain regions with major contributions, and from 380 ms to 478 ms for the brain regions with minor contributions (cluster-based permutation test, $p < .01$; see Fig. 3D top panel). Crucially, when directly comparing with the minor-contribution regions, major-contribution regions generated bigger distractor-tuning CTF slopes (from 228 ms to 408 ms; cluster-based permutation test, $p < .01$; Fig 3D lower panel). This confirms that salient distractors were represented more strongly and earlier in the major-contribution regions before engaging the minor-contribution regions.



360

361 ***Figure 3***. (A) Illustration for evaluating the contribution to salient distractor processing of different
362 brain regions located in the distractor-associated brain network. Black dashed outlines indicate the time
363 window used to estimate averaged distractor-related CTF slopes for calculating their distribution,
364 which was obtained from 1000 times IEM reconstructions for 1000 created random groups. By
365 excluding one brain region to obtain the new distribution, we can determine whether this region has a
366 relatively major or minor contribution to salient distractor processing. If the distribution moves left, it
367 reflects a relatively major contribution and if it moves right, it reflects a relatively minor contribution.
368 The major and minor contribution were defined based on their internal relationship. (B) Brain regions
369 (including the number of contacts) with significant contributions (examined in the chi square test, see

Materials and Methods for details) to salient distractor processing and the created distributions after excluding one of them. (C) The Jensen-Shannon (JS) divergence after excluding one of major- and minor-contribution regions. If the JS divergence is positive, it indicates that the distribution for the distractor-associated area is shifted to the left after excluding this brain region, reflecting a major contribution. Conversely, a negative value suggests a movement towards the right, reflecting a minor contribution. (D) Spatial selectivity for salient distractors in the distractor present condition for the major- and minor-contribution regions, including the major-minus-minor difference in the magnitude of CTF slopes. The data variance is represented by ±1 SEM. Significant CTF selectivity (or in difference) areas are highlighted by solid lines (cluster-based permutation test, $p < .01$).

We further used the leave-one-out procedure to examine whether any specific brain region within the major-contribution regions processed salient distractor firstly by excluding one brain region and conducting the IEM decoding again. As shown in Fig. 4A, the results showed that distractor-related attention was successfully reconstructed around the similar time intervals, i.e., 182 – 418 ms and 460 – 534 ms, 182 – 432 ms and 440 – 528 ms, 164 – 436 ms and 458 – 556 ms, and 156 – 436 ms and 450 – 544 ms for leaving the AMY, STG, MTG, and ACC out respectively. No significant difference was observed for their latencies, $p = .875$, suggesting that salient distractors were processed in these brain regions simultaneously.

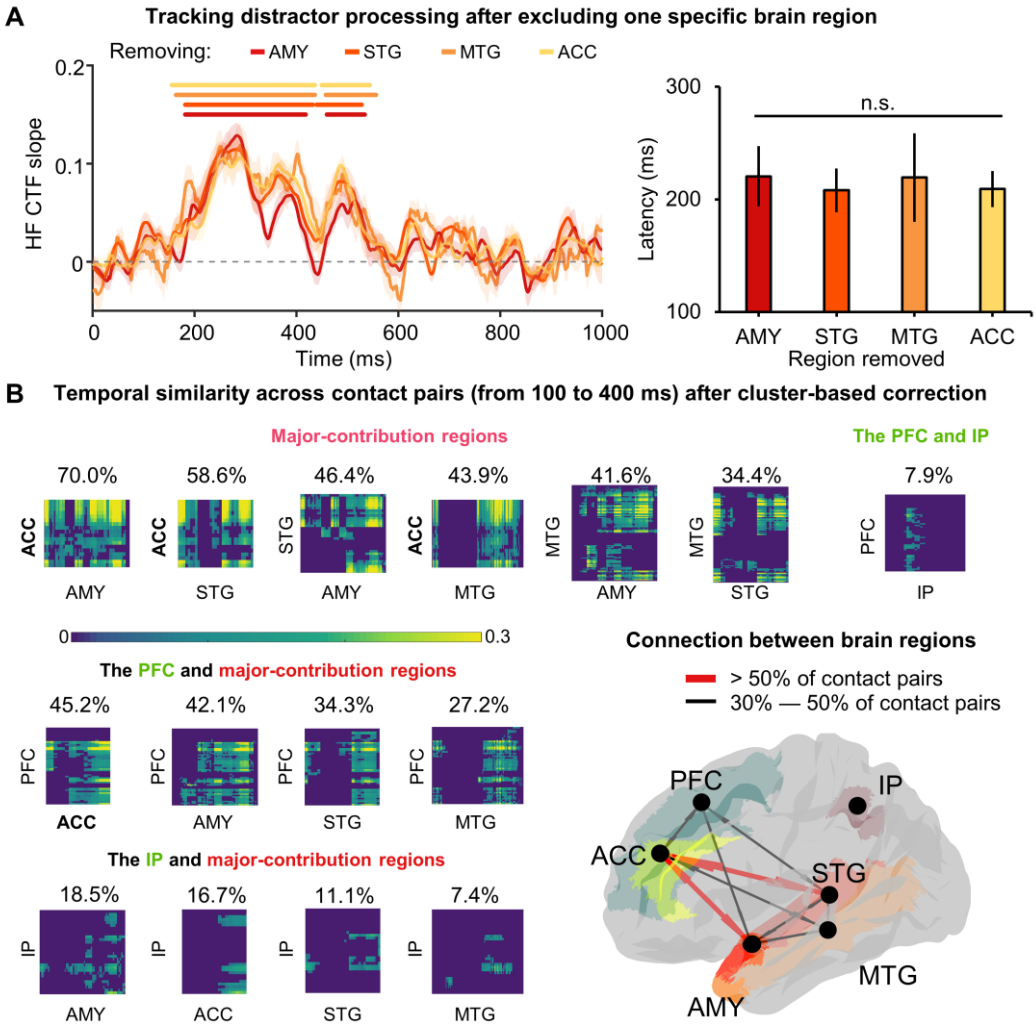


Figure 4. (A) Spatial selectivity for the salient distractors in the distractor present condition for the major-contribution regions after excluding each brain region, including their difference in the latency of the CTF slopes. The data variance is represented by ±1 SEM. Significant CTF selectivity areas are highlighted by solid lines (cluster-based permutation test, $p < .01$). (B) Temporal similarity (from 100 to 400 ms) across contact pairs between brain regions. The percentage presented above each sub-figure indicate how many contact pairs between brain regions have significant connections (after cluster-based correction, $p < .01$). The right-bottom panel illustrate the connections between brain regions, the red thick lines indicate the percentage of connections between regions is above 50%, and the black thin lines indicate that the percentage of connections between regions is between 30% and 50%.

Functional connections across contacts between brain regions when reacting to salient distractors

During salient distractor processing, the AMY, STG, MTG, and ACC might act as an integrated network to support salient distractor processing. We examined this by

analyzing the temporal correlation (from 100 to 400 ms) of HF power across contact pairs between brain regions (i.e., between-regions similarity, refer to the Materials and Methods section for details). Fig. 4B illustrate significant similarities across contact pairs after cluster-based correction ($p < .01$). These results provide a tentative estimate for the dominating connectivity supporting salient distractor processing, in particular the connection between the ACC and AMY (70% of contact pairs) was strongest, followed by the connection between the ACC and STG (58.6% of contact pairs), and then the AMY and STG (46.4% of contact pairs). This suggest that the ACC form a hub to communicate with temporal lobe structures (see Fig 4B right panel for the illustration) in support of salient distractor processing. We also observed the connections between the PFC and IP (7.9% of contact pairs), but this coupling was weaker than what we observed for the major-contribution regions (see SI for the connections within other minor-contribution regions). Further analysis revealed that the connections between the PFC and major-contribution regions was stronger (about 37.2% of contact pairs on average) than that between the IP and major-contribution regions (about 13.4% of contact pairs on average), suggesting that the PFC may play a role in linking major- and minor-contribution regions when reacting to salient distractors.

Discussion

The present study advances the understanding of attentional capture by salient distractors and its neural architecture in complex environments by analyzing intracranial electroencephalography (iEEG) signals in humans. High-frequency (HF) activity was linked to behavioral responses in a visual search task. From this HF activity, target- and distractor-tuning attention were successfully reconstructed and

dissociated using a time-resolved multivariate approach (IEM decoding)[21–23]. Importantly, when a salient distractor was introduced, attention was initially captured around 220 ms, with distractor-tuning attention boosted and target-tuning attention attenuated simultaneously. The contribution to this distractor-associated processing was dominated by the amygdala (AMY), superior temporal gyrus (STG), middle temporal gyrus (MTG), and anterior cingulate cortex (ACC). The connectivity analysis suggest that the ACC might play an important role as a hub coordinating the processing of salient distractors in temporal lobe structures. This provides us with neural evidence for attentional capture by salient distractors engaging a much larger network than previously appreciated. In particular frontal and parietal cortex might play a secondary role compared to the ACC and temporal lobe structures.

In our study, we were unable to reconstruct spatial attention towards the target in the early time window (before 400 ms) from power in the alpha and beta band (see Fig. S2B). Interestingly, we were able to reconstruct distractor-tuning attention from power in the HF band (60 – 100 Hz), a finding that has not previously been reported in human studies. Neurophysiology recordings in monkeys have demonstrated that the coherence in gamma activities between parietal and frontal cortex mediates bottom-up attention in a single-feature search task[30], and inactivation of parietal cortex reduces the power in (high-) gamma activity in prefrontal cortex and simultaneously causes the reduction of salience selection[31]. Moreover, HF activity has previously been shown to be correlated with a wide range of cognitive functions, such as memory[32,33], attention[34], learning[35], and language comprehension[36,37], underscoring its importance."

Gaining insight into the neural mechanisms for handling salient distractors is

essential for any theory of attentional capture[7,38–41]. Over the last 30 years there has been a fierce debate whether physically salient stimuli capture attention automatically or whether it is possible to prevent capture by salient stimuli[1,7]. The stimulus-driven attention posits that salient stimuli can automatically capture attention regardless of whether the salient stimulus is a target or a distractor[1,20,42]. Our findings are consistent with such an interpretation as the processing of salient distractors occurred early in time (around 220 ms) and simultaneously reduced target selection. Recently, it has been proposed that under specific conditions (which were not tested here), capture by salient stimuli may be prevented via inhibitory top-down control processes[43,44]. Further studies may be able to identify the neural suppression mechanisms that may prevent the occurrence of attentional capture by salient stimuli, in advance to address the longstanding debate.

Previous studies involving scalp EEG revealed an ERP component (N2pc) when a salient distractor was presented laterally, signifying neural responses towards salient distractors[19,45]. These findings provide insight into how the human brain processes salient distractors, while they do not necessarily reflect any simultaneous interaction between the processing of the target and salient distractors, as only one of them was investigated at a time. That is, they typically investigated distinct sets of trials to track these processes separately, but they never used the same set of trials to concurrently assess both target and distractor processing. As a result, these studies[19,45] isolated target and distractor processing in different timeframes. Notably, the ERP component N2pc (emerging around 250 ms) occurs later compared to the distractor-tuning attention (around 220 ms) reconstructed from HF activity in the present study. As such, the attentional capture phenomenon identified in this study might reflect a

different mechanism than the N2pc. The comparison between scalp and intracranial EEG signals warrants further investigations.

Importantly, the superior and middle temporal gyrus, amygdala, and anterior cingulate cortex (i.e., major-contribution regions) demonstrated activity specific to the distractor representation. The ACC, associated with novelty processing, has been observed to transiently activate in response to novel stimuli (i.e., salient stimuli that stand out from the visual field) in both intracranial recordings[46] and fMRI measurements[47,48]. This is consistent with the ACC being assumed to be one of the core regions in salience processing[49,50]. Apart from this region, the findings from the STG and MTG relating to salient distractor processing are in line with previous research showing the involvement of the temporal cortex in early attentional selection, adjusting the attentional priority map[51–53] (for review, see ref.[54]). These studies have focused on goal-driven selection instead of salience-based selection, yet they implied that temporal area could be involved in early salience processing, which we confirmed in the current study. In our study, the stronger connection between the ACC and temporal lobe structures might indicate that the ACC acts as a hub to coordinate the processing of salient distractors, warranting further investigations on the interaction between the salient-distractor-associated key regions.

Surprisingly, the amygdala was also observed to be involved in salient distractor processing. Previous studies indicate that the amygdala plays an important role in biasing attention to alerting signals[55]. This may indicate that the amygdala determines the prioritization of salient distractor in general, as salience signals also alerts us the potential threats in complex environments. Overall, we argue that the processing of salient distractor emerges from the interaction between these major-contribution

regions, and they may act as an integrated network to support salient distractor processing.

Given by the strong focus on frontal and parietal cortex in previous studies conducted in non-human primates[15,16], it is surprising that these regions only contribute in a small and late albeit robust way to the salient distractor processing. One possible reason could be that neurons in frontal and parietal regions merely respond selectively to salient stimuli's locations and not to their features[9]. Similarly, the hippocampus might be linked to storing information about novel stimuli[56], such as the color of the salient distractors, while the precuneus might be associated with spatial attention towards the location of the salient distractors[57]. In any circumstance, our data suggest that these minor-contribution regions do not directly contribute to processing salient distractors, although they are part of the integrated network supporting the processing of salient distractors.

Temporoparietal junctions (TPJ), as a core brain region in the ventral attention network that is typically recruited by unexpected events (e.g., exogenous cue in the Posner task[58], 1980), are generally thought of as the brain region responsible for salience processing[59,60], driving our initial exogenous attention. However, in the present study, we did not identify a brain region located at temporoparietal junctions, though we identified parietal and temporal brain regions responsible for salient distractor processing. To better understand the role of TPJ in salient distractor processing, experiments should be done with sufficient electrode contact recordings in this region.

In sum, our findings suggest that the theories on the network contributing to salient distractor processing need to be revised to embrace the role of ACC

communicating with temporal lobe structures. In future, this is best done by multi-sites recordings that allow us to investigate the network dynamics associated with salient distractor processing.

Materials and Methods

Participants

Eighteen patients (9 males and 9 females, ages 21– 58) at Shenzhen University General Hospital undergoing treatment for medication-resistant epilepsy participated in the present study. All patients had clinical depth electrodes implanted solely for diagnostic purposes as part of their evaluation for neurosurgical epilepsy treatment. Each depth electrode (0.8 mm in diameter) had 8, 10, 12, 14, or 16 contacts that were 1.5 mm apart and 2 mm in contact length. The present study was conducted according to the latest version of the Declaration of Helsinki and approved by the Medical Ethics Committee of Shenzhen University General Hospital. All participants gave verbal or written informed consent to participate in research.

Paradigm and design

We adopted a visual search task[20,41], in which participants had to maintain fixation throughout the trial. When the search array was briefly presented (4s or until response), they were asked to search for one unique shape (target) among the search elements (a circle among diamonds, or a diamond among circles), and to indicate whether the line segment (0.3° × 1.4°) inside the target was vertical or horizontal by pressing the ‘up’ or ‘left’ key respectively, and as fast as possible (see Fig. 1A). These unfilled shapes (2° × 2°) with either a red or green outline were presented on an imaginary circle with a radius of 4°, centered at the fixation cross against a black

546 background. If participants did not respond or pressed the wrong key, warning
547 messages were given afterwards.

548 We introduced two conditions: the *distractor present* and *absent* conditions; in
549 both conditions, the target was present on each trial, and was equally likely to be a
550 circle or a diamond (see Fig. 1). One-third of the trials were distractor absent trials;
551 and two-thirds of the trials were distractor present trials, in which a uniquely colored
552 distractor (i.e., a distractor singleton) was present, which had the same shape as other
553 neutral elements, but a different color (red or green with an equal probability). These
554 two types of trials were mixed within each block (i.e., a mixed-search design). The
555 target and distractor (if present) appeared at each possible location with equal chance.
556 After practicing, six blocks of 720 trials were tested. There was a practice session in
557 which only a target was present (see Fig. S1), 3 to 6 days before the actual start of the
558 experiment; the results for this session can be found in SI as well.

559 **Recordings and preprocessing**

560 The intracranial EEG (iEEG) data were acquired using a Nihon Kohden system at
561 a sampling rate of 1000 or 2000 Hz. A common contact, used as the reference for the
562 online recording data, was placed subcutaneously and recorded simultaneously with
563 the depth electrodes. The offline preprocessing was performed in Matlab 2016
564 (MathWorks Inc), using EEGLAB[61], ERPLAB[62], along with in-house Matlab code.
565 Contacts within each electrode that were deemed noisy/corrupted contacts upon visual
566 inspection, were excluded from further analysis. Signals were re-referenced to the
567 average activity across all clean contacts, and then resampled at 500 Hz. Common
568 referencing stands as one of the most widely adopted referencing schemes for EEG
569 recordings, as well in iEEG studies[63–65]. While some degree of dependency might be

introduced between the contacts by the common reference scheme, this would be minimal given the high number of contacts (approximately 143 contacts per subject). We removed 50 Hz power line noise prior to subsequent analyses using Hamming-windowed FIR filters (order of 180), and then band-pass filtered the data from 0.3 to 200 Hz using a second-order Butterworth filter. For behavior analysis, incorrect trials and trials on which the response times (RTs) were larger or smaller than 2.5 standard deviations from the average RTs per participant were excluded.

Eye movements were not monitored in the present study given by the complex clinical environment. While this may initially raise concerns, it is important to note that the present paradigm is intentionally designed to enable participants to perform the task without making eye movements. Importantly, previous studies employing the same paradigm and using an eye-tracker to control fixation (e.g., see refs.[41,66]). Moreover, other studies (e.g., see refs.[67,68]) employing a modified paradigm in which participants were forced to make eye movements toward the target, showed that saccade latency towards the salient distractor was around 230 ms, which was slower than the onset of HF activity (around 120 ms, see Fig. 1C) and the latency for distractor-related processing (around 220 ms, see Fig. 2C and 2F) observed in our study. These findings indicate that eye movements are not likely to modulate HF activity as the early, rapid detection of salient distractors reflected by the brain activity occur earlier than potential eye movements.

Electrode reconstruction and localization

We first registered post-operative computed tomography (CT) images to pre-operative T1-weighted MR images using FreeSurfer v6.0.0[69], following the iElvis pipeline[70]. We inspected the quality of the registration and manually labeled each

contact location on the T1-registered CT images, which were then mapped onto a standard MNI space (Fig. 1B). To further determine the exact contact locations belonging to different brain areas and brain regions, we assigned contact MNI coordinates to each brain area according to large-scale cerebral networks (Yeo-7 template)[28], and to each brain region according to FreeSurfer's automatic parcellation. We went through each contact and assigned the closest cortical/subcortical label.

High-frequency activity

High-frequency activity (HF; 60-100 Hz) was considered to be the key electrophysiological marker of local neural population activity[71–73], and to be correlated with attention[74], which is the focus of our interest. To extract HF activity, preprocessed iEEG data were broken into event-related epochs (-3 s to 6 s relative to the search onset; avoiding edge artifacts from wavelet convolution), and then were convolved with a set of Morlet wavelets with frequencies ranging from 1 to 150 Hz in 80 logarithmically spaced steps. The number of cycles of each wavelet was logarithmically spaced between 4 and 20 to strike a good balance between temporal and frequency precision. The HF power values within our focused epochs (-0.5 s to 4 s relative to the search onset) were z-scored by subtracting the average value and dividing by the standard deviation across all trials.

To accurately compare HF power between conditions, we subtracted the mean HF power from -400 ms to -100 ms prior to the search onset for each condition separately and ran a cluster-based permutation test against a null-distribution shuffled from 1000 iterations (see the below section of cluster-based permutation test for details) to make the comparison between conditions. We further correlated the time to reach the peak of HF activity with the response times (RTs) obtained from the search

task via Pearson correlation.

Responsive contacts

Responsive contacts were identified by comparing the post stimulus HF response (averaged over a time window of from 100 ms to 400 ms) to the pre-stimulus baseline (averaged over a time window of from -400 ms to -100 ms) using two-tailed, paired t-tests. The *p*-values from all recording contacts (across all participants) were pooled together to control the false discovery rate (FDR)[75]. Contacts that showed significant HF responses ($p_{HF} < .05$) were regarded as the candidates for responsive contacts. Then, we compared the HF response for each time point after stimulus onset with the pre-stimulus baseline using paired t-tests ($p_{HF} < .05$). For each contact, if we could identify a time window of successive significant activities which a temporal length of over 200 ms, the contact was treated as a responsive contact[32,76]. This procedure was performed across participants, resulting in 1379 contacts (see Table S1 for individual numbers). Noted that, no significant differences were observed in HF activity and the target-and distractor-related CTF slopes between all responsive contacts and the responsive contacts after excluding those that exhibited seizure-onset patterns (SOPs) as identified by neurosurgeons; and the responsive contacts showing SOPs demonstrated similar HF activity as the unaffected ones (see Fig. S10). Nevertheless, we chose a more conservative approach by excluding the responsive contacts showing SOPs (7.5% of all responsive contacts), resulting in 1275 responsive contacts remaining for further analysis.

Inverted encoding model

To reconstruct the attention towards the target and/or salient distractor, we

applied inverted encoding model (IEM)[23] to estimate spatial channel-tuning functions (CTFs) from the HF response across all responsive contacts over time (see Fig. 2A). In accordance with previous studies[22–24], we specified an explicit model to represent how neural populations encode spatial information. This model focused exclusively on the spatial selectivity of the multivariate neural responses. To accomplish this, we constructed a basis set of 8 spatial channels corresponding to the 8 locations employed in the present study. The response profile of each spatial channel across element locations in the visual display was modelled as a half sinusoid raised to the seventh power:

$$R = \sin(0.5\Theta)^7,$$

Where Θ is the element location (0-359°), and R is the response of the spatial channel in arbitrary units. Note that, to ensure the peak response of each spatial channel centered over one of the eight element locations (22.5°, 67.5°, etc.), the response profile was circularly shifted. The tuning curve for spatial channels was selected as a half sine wave function because it maintains an approximately Gaussian shape (see Fig. 2A), which mimics attentional response in a circular space. We opted for the 7th power, which aligns with previous studies[22–24].

We assumed that the HF power at each contact reflected the weighted sum of eight spatially selective channels (i.e., neuronal populations), each tuned to a different angular location (corresponding to the possible locations of the search elements). We then partitioned our data randomly into independent sets of training data (2/3 trials) and test data (1/3 trials) following a cross-validation routine. The training data (B; m contacts × n locations) were used to estimate weights that approximated the possible contributions of the eight spatial channels to the observed HF responses measured at

each contact and location. We defined C (k channels × n locations) as a matrix of the predicted response of each spatial channel (estimated by the basic function for that channel) for each location; and W (m contacts × k channels) as a weight matrix characterizing a linear mapping from channel space to contact space. We further described the relationships between B, C, and W in a linear model as follows:

$$B = WC$$

Using the weight matrix (W) obtained via least-squares estimation, we inverted the model to transform the observed test data into estimated channel responses (i.e., reconstructed CTFs). This IEM routine was iterated 10 times to minimize the influence of idiosyncrasies specific to any trial assignment, and obtain an averaged channel-response profiles. We then calculated the slope of the CTFs via linear regression to quantify selective attentional processing for different locations over time. Higher slope values indicate greater attentional selectivity while lower values indicate less attention selectivity. We reconstructed target-tuning and distractor-tuning CTFs according to the target location and salient distractor location respectively across participants. Noted that, the IEM model assumed that each contact contained a large number of spatial-selective neurons mainly tuned to a specific spatial location. Thus, the spatial tuning of a given contact was reflected by the combined response from the neurons detected by the contact. We analyzed this by binning according to 8 locations and calculated the corresponding weights for each contact. Therefore, the calculated CTF slopes through these weights represented the spatial tuning at the population level, originating from a group of spatial-selective neurons.

To examine the CTF onset latency, we used a jackknife-based routine[77]. CTF onset latency was measured as the earliest time at which the CTF slope reached 50%

of its maximum amplitude within the first significant cluster for CTF selectivity. The latency difference (D) between conditions was measured as the difference in the latency of CTF slopes averaged across all participants. To estimate the standard error of the latency difference, SE_D, we created subsamples that included all participants except one, and calculated the latency difference, D_{-i} (for $i = 1, \ldots, N$, where N is the sample size) for each subsample. The jackknife estimate of the standard error (SE_D) was calculated as

$$SE_D = \sqrt{\frac{N-1}{N}\sum_{i=1}^{N}(D_{-i} - \bar{J})^2},$$

where $\bar{J}$ is the mean of the latency differences obtained for all subsamples (i.e., $\bar{J} = \sum D_{-i} / N$).

A jackknifed t statistic, t_j, was then calculated as

$$t_j = \frac{D}{SE_D},$$

which follows an approximate t distribution with N-1 degrees of freedom under the null hypothesis. Our jackknife test for latency differences was two-tailed.

Cluster-based permutation test

To correct multiple comparisons, we used a cluster-based permutation test against a null-distribution shuffled from 1000 iterations (following Monte Carlo randomization procedure). Specifically, a one-sample t-test was performed against zeros to identify above-chance CTF selectivity, and time windows with t-values larger than a threshold ($p = .05$) were combined into contiguous clusters based on adjacency. The cluster statistics was defined as the sum of the t values within each cluster. Moreover, the IEM procedure was repeated 1000 times, however the location label

was randomized within each group of trials so that the labels were independent of the observed responses in each contact. We identified the largest clusters for random location labels per iteration as described above, forming a null distribution of clusters. Clusters were determined to be significant if the cluster statistics was larger than the 99th percentile of the null distribution.

To make a between-conditions comparison on CTF selectivity or HF activity, paired t-tests were performed between conditions (i.e., distractor present and absent conditions). We followed above routine to determine cluster statistics, but only the null distribution was formed by randomly permuting condition labels for 1000 times in order to get the largest clusters per iteration. Again, clusters were determined to be significant if the cluster statistics was larger than the 99th percentile of the null distribution.

The distribution of the averaged CTF slopes and chi-square test

The critical comparison in the present study was made between the distractor-tuning and target-tuning CTF slopes in the distractor present condition. To avoid any potential doubt that our results from random groupings were selected arbitrarily, we determined the first significant cluster when comparing the distractor-tuning CTF slopes with target-tuning CTF slopes as the critical time window (see Fig. 3), and then averaged the CTF slopes for distractor- and target-tuning separately in this time window. Later, we iterated the grouping of contacts for the distractor- and target-related IEM reconstruction 1000 times, obtaining distributions of the averaged CTF slopes for the distractor and target. To examine whether the distractor-related distribution was different from the target-related distribution, we applied the chi-square test after creating 50 groups of CTF slopes for each distribution and applied

FDR correction if necessary.

The Jensen-Shannon divergence

The Jensen–Shannon divergence is a measure of the distance between two probability distributions. It can be thought of as a measure of degree of similarity between the two distributions. The JS divergence is a symmetrized version of the Kullback-Leibler (KL) divergence, which allows it to be used to compare any two probability distributions, regardless of which distribution is considered the "reference" and which is considered the "comparison".

The JS divergence between two probability distributions, P and Q, is defined as:

$$JSD(P|Q) = (KL(P|M) + KL(Q|M))/2,$$

where M is the "average" distribution, defined as:

$$M = (P + Q)/2,$$

and KL(P|Q) is the KL divergence between P and Q, defined as:

$$KL(P|Q) = \sum p(i) * \log(p(i) / q(i)),$$

where the sum is taken over all possible values of i, p(i) and q(i) are the probabilities of i in the distributions P and Q, respectively. If the distribution moves left, we gave the JS value a positive sign.

Temporal similarity analysis

The similarity across contact pairs was estimated using HF power in the distractor present condition. In order to increase the signal to noise ratio[29], an averaged HF power was obtained from five randomly selected trials, and we iterated this routine 30 times. For each iteration, we correlated the time series of the averaged HF power

757 within a 100 – 400 ms time window for each contact pair between different brain
758 regions[78], obtaining Spearman's r values. Noted that, the distractor processing was
759 found to be significant in the distractor-associated brain network in the 214 – 306 ms
760 interval, so a larger time window was chosen to include this critical period. The
761 temporal similarity was calculated by averaging the temporal correlations
762 (Spearman's r values) across 30 iterations for each distractor location which then was
763 averaged across all locations. We Fisher Z-transformed all the correlation values for
764 further analyses and applied the same cluster-based permutation test as mentioned
765 above for CTF selectivity.

766 Reference

767 1. Theeuwes, J. Top–down and bottom–up control of visual selection. *Acta*
768 *Psychologica* **135**, 77–99 (2010).

769 2. Itti, L. & Koch, C. Computational modelling of visual attention. *Nat Rev Neurosci*
770 **2**, 194–203 (2001).

771 3. Theeuwes, J. Perceptual selectivity for color and form. *Perception & Psychophysics*
772 **51**, 599–606 (1992).

773 4. Theeuwes, J. Exogenous and endogenous control of attention: The effect of visual
774 onsets and offsets. *Perception & Psychophysics* **49**, 83–90 (1991).

775 5. Anderson, B. A., Laurent, P. A. & Yantis, S. Value-driven attentional capture. *Proc.*
776 *Natl. Acad. Sci. U.S.A.* **108**, 10367–10371 (2011).

777 6. Schmidt, L. J., Belopolsky, A. V. & Theeuwes, J. Attentional capture by signals of
778 threat. *Cognition and Emotion* **29**, 687–694 (2015).

779 7. Luck, S. J., Gaspelin, N., Folk, C. L., Remington, R. W. & Theeuwes, J. Progress
780 toward resolving the attentional capture debate. *Visual Cognition* **29**, 1–21 (2021).

781 8. Constantinidis, C. & Steinmetz, M. A. Neuronal Responses in Area 7a to Multiple-
782 stimulus Displays: I. Neurons Encode the Location of the Salient Stimulus.
783 *Cerebral Cortex* **11**, 581–591 (2001).

9. Constantinidis, C. Posterior Parietal Cortex Automatically Encodes the Location of Salient Stimuli. *Journal of Neuroscience* **25**, 233–238 (2005).

10. Katsuki, F. & Constantinidis, C. Early involvement of prefrontal cortex in visual bottom-up attention. *Nat Neurosci* **15**, 1160–1166 (2012).

11. Bichot, N. P., Schall, J. D. & Thompson, K. G. Visual feature selectivity in frontal eye fields induced by experience in mature macaques. *Nature* **381**, 697–699 (1996).

12. Schall, J. D. & Hanes, D. P. Neural basis of saccade target selection in frontal eye field during visual search. *Nature* **366**, 467–469 (1993).

13. Bichot, N. P. & Schall, J. D. Priming in Macaque Frontal Cortex during Popout Visual Search: Feature-Based Facilitation and Location-Based Inhibition of Return. *J. Neurosci.* **22**, 4675–4685 (2002).

14. 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. U.S.A.* **120**, e2210839120 (2023).

15. Ipata, A. E., Gee, A. L., Gottlieb, J., Bisley, J. W. & Goldberg, M. E. LIP responses to a popout stimulus are reduced if it is overtly ignored. *Nat Neurosci* **9**, 1071–1076 (2006).

16. Cosman, J. D., Lowe, K. A., Zinke, W., Woodman, G. F. & Schall, J. D. Prefrontal

802 Control of Visual Distraction. *Current Biology* **28**, 414-420.e3 (2018).

803 17. Moran, J. & Desimone, R. Selective attention gates visual processing in the
804 extrastriate cortex. *Science* **229**, 782–784 (1985).

805 18. Treue, S. & Trujillo, J. C. M. Feature-based attention influences motion processing
806 gain in macaque visual cortex. *Nature* **399**, 575–579 (1999).

807 19. Hickey, C., McDonald, J. J. & Theeuwes, J. Electrophysiological evidence of the
808 capture of visual attention. *Journal of cognitive neuroscience* **18**, 604–613 (2006).

809 20. Theeuwes, J. Perceptual selectivity for color and form. *Perception & Psychophysics*
810 **51**, 599–606 (1992).

811 21. Brouwer, G. J. & Heeger, D. J. Decoding and Reconstructing Color from Responses
812 in Human Visual Cortex. *Journal of Neuroscience* **29**, 13992–14003 (2009).

813 22. Sprague, T. C. & Serences, J. T. Attention modulates spatial priority maps in the
814 human occipital, parietal and frontal cortices. *Nat Neurosci* **16**, 1879–1887 (2013).

815 23. Foster, J. J., Sutterer, D. W., Serences, J. T., Vogel, E. K. & Awh, E. Alpha-Band
816 Oscillations Enable Spatially and Temporally Resolved Tracking of Covert Spatial
817 Attention. *Psychol Sci* **28**, 929–941 (2017).

818 24. Sprague, T. C., Saproo, S. & Serences, J. T. Visual attention mitigates information
819 loss in small- and large-scale neural codes. *Trends in Cognitive Sciences* **19**, 215–

820 226 (2015).

821 25. Luck, S. J., Gaspelin, N., Folk, C. L., Remington, R. W. & Theeuwes, J. Progress
822 toward resolving the attentional capture debate. *Visual Cognition* **29**, 1–21 (2021).

823 26. Ossandon, T. *et al.* Efficient 'Pop-Out' Visual Search Elicits Sustained Broadband
824 Gamma Activity in the Dorsal Attention Network. *Journal of Neuroscience* **32**,
825 3414–3421 (2012).

826 27. Clayton, M. S., Yeung, N. & Cohen Kadosh, R. The roles of cortical oscillations in
827 sustained attention. *Trends in Cognitive Sciences* **19**, 188–195 (2015).

828 28. Yeo, B. T. *et al.* The organization of the human cerebral cortex estimated by
829 intrinsic functional connectivity. *Journal of neurophysiology* (2011).

830 29. Kunz, L. *et al.* Hippocampal theta phases organize the reactivation of large-scale
831 electrophysiological representations during goal-directed navigation. *Sci. Adv.* **5**,
832 eaav8192 (2019).

833 30. Buschman, T. J. & Miller, E. K. Top-Down Versus Bottom-Up Control of Attention
834 in the Prefrontal and Posterior Parietal Cortices. *Science* **315**, 1860–1862 (2007).

835 31. Chen, X. *et al.* Parietal Cortex Regulates Visual Salience and Salience-Driven
836 Behavior. *Neuron* **106**, 177-187.e4 (2020).

837 32. Norman, Y. *et al.* Hippocampal sharp-wave ripples linked to visual episodic

recollection in humans. *Science* **365**, eaax1030 (2019).

33. Solomon, E. A. *et al.* Widespread theta synchrony and high-frequency desynchronization underlies enhanced cognition. *Nature communications* **8**, 1704 (2017).

34. Banaie Boroujeni, K., Oemisch, M., Hassani, S. A. & Womelsdorf, T. Fast spiking interneuron activity in primate striatum tracks learning of attention cues. *Proceedings of the National Academy of Sciences* **117**, 18049–18058 (2020).

35. Gueguen, M. C. *et al.* Anatomical dissociation of intracerebral signals for reward and punishment prediction errors in humans. *Nature communications* **12**, 3344 (2021).

36. Goldstein, A. *et al.* Shared computational principles for language processing in humans and deep language models. *Nature neuroscience* **25**, 369–380 (2022).

37. Khalighinejad, B., Herrero, J. L., Mehta, A. D. & Mesgarani, N. Adaptation of the human auditory cortex to changing background noise. *Nature communications* **10**, 2509 (2019).

38. Gaspelin, N. & Luck, S. J. Inhibition as a potential resolution to the attentional capture debate. *Current Opinion in Psychology* **29**, 12–18 (2019).

39. Geng, J. J. Attentional Mechanisms of Distractor Suppression. *Curr Dir Psychol*

856 *Sci* **23**, 147–153 (2014).

857 40. Theeuwes, J., Bogaerts, L. & van Moorselaar, D. What to expect where and when:
858 how statistical learning drives visual selection. *Trends in cognitive sciences* (2022).

859 41. Wang, B. & Theeuwes, J. Statistical regularities modulate attentional capture.
860 *Journal of Experimental Psychology: Human Perception and Performance* **44**, 13–
861 17 (2018).

862 42. Jonides, J. & Yantis, S. Uniqueness of abrupt visual onset in capturing attention.
863 *Perception & psychophysics* **43**, 346–354 (1988).

864 43. Folk, C. L., Remington, R. W. & Johnston, J. C. Involuntary covert orienting is
865 contingent on attentional control settings. *Journal of Experimental Psychology:*
866 *Human perception and performance* **18**, 1030 (1992).

867 44. Gaspelin, N., Leonard, C. J. & Luck, S. J. Direct evidence for active suppression of
868 salient-but-irrelevant sensory inputs. *Psychological science* **26**, 1740–1750 (2015).

869 45. Wang, B., Van Driel, J., Ort, E. & Theeuwes, J. Anticipatory Distractor
870 Suppression Elicited by Statistical Regularities in Visual Search. *Journal of*
871 *Cognitive Neuroscience* **31**, 1535–1548 (2019).

872 46. Baudena, P., Halgren, E., Heit, G. & Clarke, J. M. Intracerebral potentials to rare
873 target and distractor auditory and visual stimuli. III. Frontal cortex.

874 *Electroencephalography and Clinical Neurophysiology* **94**, 251–264 (1995).

875 47. Clark, V. P., Fannon, S., Lai, S., Benson, R. & Bauer, L. Responses to Rare Visual
876 Target and Distractor Stimuli Using Event-Related fMRI. *Journal of*
877 *Neurophysiology* **83**, 3133–3139 (2000).

878 48. Kiehl, K. A., Laurens, K. R., Duty, T. L., Forster, B. B. & Liddle, P. F. Neural
879 sources involved in auditory target detection and novelty processing: An event-
880 related fMRI study. *Psychophysiology* **38**, 133–142 (2001).

881 49. Uddin, L. Q. Salience processing and insular cortical function and dysfunction. *Nat*
882 *Rev Neurosci* **16**, 55–61 (2015).

883 50. Menon, V. & Uddin, L. Q. Saliency, switching, attention and control: a network
884 model of insula function. *Brain Struct Funct* **214**, 655–667 (2010).

885 51. Sani, I. *et al.* The human endogenous attentional control network includes a ventro-
886 temporal cortical node. *Nat Commun* **12**, 360 (2021).

887 52. Stemmann, H. & Freiwald, W. A. Evidence for an attentional priority map in
888 inferotemporal cortex. *Proceedings of the National Academy of Sciences* **116**,
889 23797–23805 (2019).

890 53. Bogadhi, A. R., Bollimunta, A., Leopold, D. A. & Krauzlis, R. J. Spatial Attention
891 Deficits Are Causally Linked to an Area in Macaque Temporal Cortex. *Current*

892 *Biology* **29**, 726-736.e4 (2019).

893 54. Ramezanpour, H. & Fallah, M. The role of temporal cortex in the control of
894 attention. *Current Research in Neurobiology* **3**, 100038 (2022).

895 55. Peck, C. J., Lau, B. & Salzman, C. D. The primate amygdala combines information
896 about space and value. *Nat Neurosci* **16**, 340–348 (2013).

897 56. Knight, R. T. Contribution of human hippocampal region to novelty detection.
898 *Nature* **383**, 256–259 (1996).

899 57. Rao, H., Zhou, T., Zhuo, Y., Fan, S. & Chen, L. Spatiotemporal activation of the
900 two visual pathways in form discrimination and spatial location: A brain mapping
901 study. *Hum. Brain Mapp.* **18**, 79–89 (2003).

902 58. Posner, M. I. Orienting of Attention. *Quarterly Journal of Experimental*
903 *Psychology* **32**, 3–25 (1980).

904 59. Corbetta, M. & Shulman, G. L. Control of goal-directed and stimulus-driven
905 attention in the brain. *Nat Rev Neurosci* **3**, 201–215 (2002).

906 60. Vossel, S., Geng, J. J. & Fink, G. R. Dorsal and Ventral Attention Systems: Distinct
907 Neural Circuits but Collaborative Roles. *Neuroscientist* **20**, 150–159 (2014).

908 61. Delorme, A. & Makeig, S. EEGLAB: an open source toolbox for analysis of single-
909 trial EEG dynamics including independent component analysis. *Journal of*

910 *Neuroscience Methods* **134**, 9–21 (2004).

911 62. Lopez-Calderon, J. & Luck, S. J. ERPLAB: an open-source toolbox for the analysis
912 of event-related potentials. *Front. Hum. Neurosci.* **8**, (2014).

913 63. Liu, J. *et al.* Stable maintenance of multiple representational formats in human
914 visual short-term memory. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 32329–32339 (2020).

915 64. Bijanzadeh, M. *et al.* Decoding naturalistic affective behaviour from spectro-spatial
916 features in multiday human iEEG. *Nature Human Behaviour* **6**, 823–836 (2022).

917 65. Grossman, S. *et al.* Convergent evolution of face spaces across human face-
918 selective neuronal groups and deep convolutional networks. *Nature*
919 *communications* **10**, 4934 (2019).

920 66. Feldmann-Wüstefeld, T., Weinberger, M. & Awh, E. Spatially Guided Distractor
921 Suppression during Visual Search. *J. Neurosci.* **41**, 3180–3191 (2021).

922 67. Gaspelin, N., Leonard, C. J. & Luck, S. J. Suppression of overt attentional capture
923 by salient-but-irrelevant color singletons. *Atten Percept Psychophys* **79**, 45–62
924 (2017).

925 68. Wang, B., Samara, I. & Theeuwes, J. Statistical regularities bias overt attention.
926 *Atten Percept Psychophys* **81**, 1813–1821 (2019).

927 69. Fischl, B., Sereno, M. I. & Dale, A. M. Cortical surface-based analysis: II: inflation,

928 flattening, and a surface-based coordinate system. *Neuroimage* **9**, 195–207 (1999).

929 70. Groppe, D. M. *et al.* iELVis: An open source MATLAB toolbox for localizing and
930 visualizing human intracranial electrode data. *Journal of Neuroscience Methods*
931 **281**, 40–48 (2017).

932 71. Mukamel, R. *et al.* Coupling Between Neuronal Firing, Field Potentials, and fMRI
933 in Human Auditory Cortex. *Science* **309**, 951–954 (2005).

934 72. Parvizi, J. & Kastner, S. Promises and limitations of human intracranial
935 electroencephalography. *Nat Neurosci* **21**, 474–483 (2018).

936 73. Watson, B. O., Ding, M. & Buzsáki, G. Temporal coupling of field potentials and
937 action potentials in the neocortex. *Eur J Neurosci* **48**, 2482–2497 (2018).

938 74. Jensen, O., Kaiser, J. & Lachaux, J.-P. Human gamma-frequency oscillations
939 associated with attention and memory. *Trends in Neurosciences* **30**, 317–324
940 (2007).

941 75. Benjamini, Y. & Yekutieli, D. The Control of the False Discovery Rate in Multiple
942 Testing under Dependency. *The Annals of Statistics* **29**, 1165–1188 (2001).

943 76. Norman, Y., Yeagle, E. M., Harel, M., Mehta, A. D. & Malach, R. Neuronal
944 baseline shifts underlying boundary setting during free recall. *Nat Commun* **8**, 1301
945 (2017).

946 77. Miller, J., Patterson, T. & Ulrich, R. Jackknife-based method for measuring LRP
947 onset latency differences. *Psychophysiology* **35**, 99–115 (1998).
948 78. Wang, L., Kuperberg, G. & Jensen, O. Specific lexico-semantic predictions are
949 associated with unique spatial and temporal patterns of neural activity. *eLife* **7**,
950 e39061 (2018).
951