Neuron
Article

Neural Activities in V1 Create a Bottom-Up Saliency Map

Xilin Zhang,[1] Li Zhaoping,[4] Tiangang Zhou,[5] and Fang Fang[1,2,3,*]
[1]Department of Psychology and Key Laboratory of Machine Perception (Ministry of Education)
[2]Center for Life Sciences
[3]IDG/McGovern Institute for Brain Research
Peking University, Beijing 100871, China
[4]Department of Computer Science, University College London, London WC1E 6BT, UK
[5]State Key Laboratory of Brain and Cognitive Science, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China
*Correspondence: ffang@pku.edu.cn
DOI 10.1016/j.neuron.2011.10.035

SUMMARY

The bottom-up contribution to the allocation of exogenous attention is a saliency map, whose neural substrate is hard to identify because of possible contamination by top-down signals. We obviated this possibility using stimuli that observers could not perceive, but that nevertheless, through orientation contrast between foreground and background regions, attracted attention to improve a localized visual discrimination. When orientation contrast increased, so did the degree of attraction, and two physiological measures: the amplitude of the earliest (C1) component of the ERP, which is associated with primary visual cortex, and fMRI BOLD signals in areas V1–V4 (but not the intraparietal sulcus). Significantly, across observers, the degree of attraction correlated with the C1 amplitude and just the V1 BOLD signal. These findings strongly support the proposal that a bottom-up saliency map is created in V1, challenging the dominant view that the saliency map is generated in the parietal cortex.

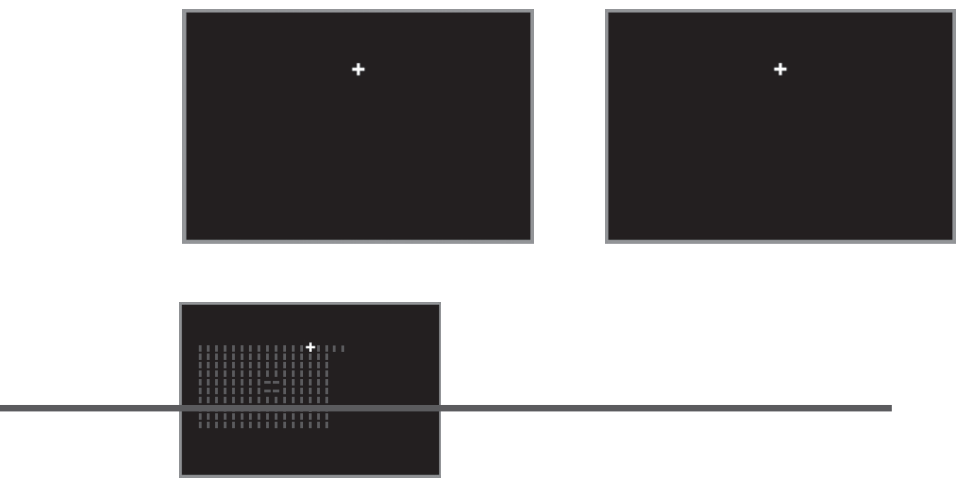
INTRODUCTION

Because neural resources are severely limited, only a very small fraction of visual inputs can reach all the way to perception. One of the main mechanisms of selection involves directing attention to a visual location, either overtly or covertly, without a shift in gaze. Attention may either be directed under voluntary control according to top-down goals, such as when directing gaze to an interesting book, or be attracted automatically by bottom-up stimuli, such as when the sudden appearance of a cat distracts one from reading. Throughout this study, we use the term salience to refer to this bottom-up attraction of exogenous attention. The regions of the brain responsible for top-down selection are well known, and include the frontal eye fields (FEF), dorsomedial prefrontal cortex, and posterior parietal cortex (PPC) (Corbetta and Shulman, 2002; Kastner and Ungerleider, 2000; Serences and Yantis, 2006). However, although bottom-up selection is typically faster and more potent (Jonides, 1981; Nakayama and Mackeben, 1989), there are controversies concerning the brain regions involved.

It is generally thought that the brain constructs a saliency map of visual space, with the activity at a location explicitly reporting the strength of its bottom-up attentional attraction (Koch and Ullman, 1985) so that it can be directly read out to guide attentional shifts before and after combining with top-down control factors. Based on neurophysiological and imaging studies, brain regions proposed to realize this saliency map have included the superior colliculus (Kustov and Robinson, 1996; Fecteau and Munoz, 2006), pulvinar (Shipp, 2004), parietal cortex (Bisley and Goldberg, 2010; Geng and Mangun, 2009; Gottlieb et al., 1998), V4 (Mazer and Gallant, 2003), and FEF (Serences and Yantis, 2007; Thompson and Bichot, 2005). However, neural activities in all these areas are also involved in top-down attentional direction. It is therefore unclear whether the observed neural correlates of saliency are relayed from brain regions upstream along the visual pathway, and whether they are the cause or the consequence of selection. In particular, because salient visual inputs typically enter awareness, it is difficult to determine whether the observed neural activities represent saliency as such, as opposed to being caused by the consequent perception of the selected stimuli.

A dominant view of the saliency map (Itti and Koch, 2001; Koch and Ullman, 1985; Wolfe, 1994) presumes that saliency results from pooling different visual features, being independent of whether the feature distinction making a location salient is in color, orientation, or other features. Hence, previous attempts to find the saliency map have typically concentrated in higher cortical areas, particularly the parietal cortex, whose neurons, unlike those in primary visual cortex (V1), are less selective to specific visual features.

By contrast, Li (1999, 2002) proposed that V1 (which, notably, projects directly and indirectly to all the previously proposed brain regions for the saliency map [Shipp, 2004]) creates a saliency map via intracortical interactions that are manifest in contextual influences (Allman et al., 1985). According to this theory, the saliency of a location is monotonically related to the highest neural response among all the V1 cells that cover that location with their spatial receptive fields (relative to the V1



responses to the other locations), regardless of the preferred feature of the most responsive neuron. Many psychophysical predictions arising from this proposal have been confirmed (Koene and Zhaoping, 2007; Zhaoping and May, 2007). One particularly interesting confirmation is that an eye of origin singleton, e.g., a bar presented to the left eye among many other bars presented to the right eye, can distract attention away from a very salient visual search target (e.g., a uniquely oriented bar presented to the right eye), even when observers cannot distinguish this eye of origin singleton from other items (Zhaoping, 2008). This supports the V1 theory, because the reason that observers cannot distinguish this singleton is that the eye of origin feature is not represented in any cortical area except V1. Indeed, Wolfe and Franzel (1988) reported that observers found it impossible to find a visual search target defined by its unique eye of origin. The apparent contradiction between the inaccessibility to search of the eye of origin feature, and yet its ability to attract attention can be resolved by realizing that attentional attraction by an input feature can be dissociable from the recognition of this feature needed for visual search.

To determine which cortical area realizes the saliency map, it is important to probe bottom-up attraction free from top-down influences (e.g., those arising from feature and object recognition). One way to do this is to use stimuli that are presented so briefly (and followed by a high contrast mask) that they are invisible. As

bars in a regular Manhattan grid in the lower visual field on a dark screen. All bars were identically oriented except for a foreground region of 2 × 2 bars of another orientation. The foreground region was at 7.2° eccentricity in either the lower left or the lower right quadrant. The orientation of the background bars was randomly chosen from 0° to 180°. There were five possible orientation contrasts between the foreground bars and the background bars: 0°, 7.5°, 15°, 30°, and 90°. A nonzero orientation contrast could possibly make the foreground region salient enough to attract attention. To isolate the bottom-up saliency signal, we minimized top-down influences by presenting the texture stimuli very briefly and subsequently masking them using a high luminance mask (Figure 1B). Subjects reported that they were unaware of the texture stimuli and could not detect even by forced choice which quadrant contained the foreground region. The percentages of correct detection (mean ± SEM) were 50.5 ± 0.8%, 50.0 ± 0.8%, 49.8 ± 0.8%, and 50.4 ± 0.7% for orientation contrasts of 7.5°, 15°, 30°, and 90°, respectively, statistically indistinguishable from the chance level (see Experimental Procedures).

Psychophysical Experiment

To assess the saliency (i.e., the degree of attentional attraction) of the invisible foreground region, we used a modified version

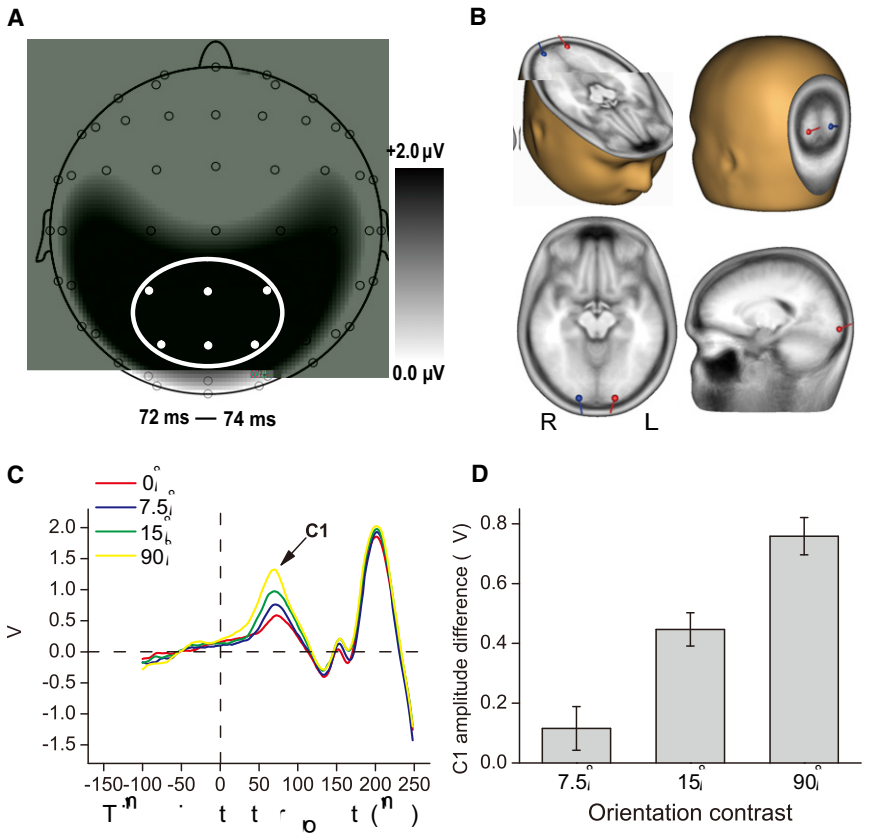


Figure 3. ERP Results

(A) C1 topography in response to the masked texture stimuli averaged over all orientation contrasts and subjects. Posterior electrodes, including CP1, CPz, CP2, P1, Pz, and P2 (within the white ellipse), had the largest C1 amplitudes. The latencies of the grand averaged C1 are between 72 and 74 ms.

(B) Dipole modeling of the intracranial sources of the C1 component. A symmetrical pair of dipoles located in V1 (Talairach coordinates: ± 18, −96, −10) could account for 89% of the variance in the C1 scalp voltage distribution over the interval 62–82 ms after texture stimulus onset (R: right hemisphere; L: left hemisphere).

(C) ERPs averaged over the six electrodes and all subjects in response to the masked texture stimuli.

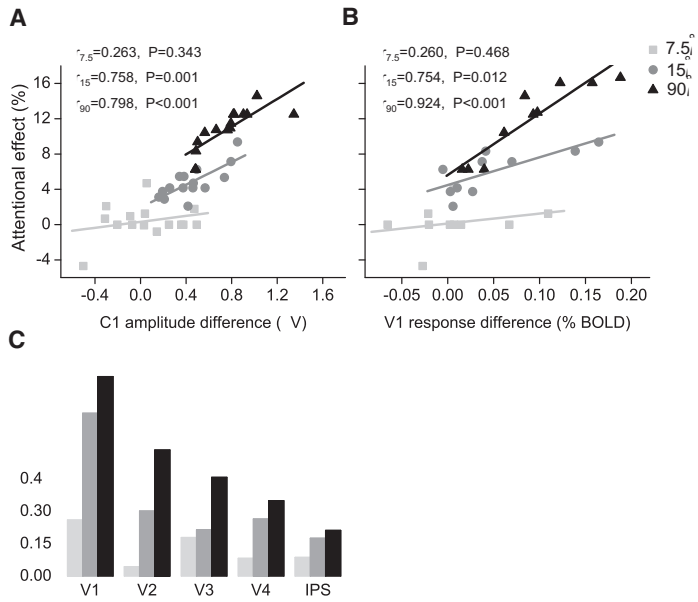
(D) C1 amplitude differences between orientation contrasts 7.5°, 15°, 90°, and 0°. Error bars denote 1 SEM calculated across subjects for each condition.

7.5°, 15°, and 90° (Figure 3D). C1 amplitude differences were submitted to one-way repeated-measures ANOVA, which showed that the main effect of orientation contrast was significant ($F_{2, 28} = 44.392$, $p < 0.001$). Post hoc paired t tests revealed that the C1 amplitude difference increased with the orientation contrast (7.5° versus 15°: $t_{14} = 4.793$, $p = 0.001$; 15° versus 90°: $t_{14} = 6.015$, $p < 0.001$), parallel to the attentional attraction in Figure 2. This suggests that the C1 amplitude and the attentional attraction might be closely related. An ERP experiment that was identical, except for relocating the stimuli from the lower to upper visual field, provided the same qualitative conclusion (Figure S2), while showing a reversal of the C1 polarity. This suggests that the C1 originates from V1 (Di Russo et al., 2002).

fMRI Experiment

The experimental protocol was similar to that of the ERP experiment, except that only trials with orientation contrasts 7.5°, 15°, and 90° were included. Behavioral data again confirmed that the texture stimuli were invisible to subjects (7.5°: 50.2 ± 1.1%; 15°: 50.4 ± 1.1%; 90°: 50.5 ± 0.9%). Contralateral and ipsilateral regions of interest (ROIs) in V1–V4 and IPS were defined as being the cortical areas that responded to the retinal inputs in the foreground region and its contralateral counterpart (that would always contain background bars). In V1–V4, texture stimuli with orientation contrasts of 15° and 90° generally evoked larger BOLD signals in the contralateral than the ipsilateral ROIs (Figure 4A). In other words, the foreground region evoked stronger neural activities than its contralateral counterpart. The differences between the peak BOLD signals at the contralateral ROIs and those at the ipsilateral ROIs are shown in Figure 4B and were submitted to a repeated-measures ANOVA with orientation contrast (7.5°, 15°, and 90°) and cortical area (V1–V4 and IPS) as within-subject factors. The main effect of orientation contrast was significant ($F_{2, 18} = 20.352$, $p < 0.001$), demonstrating that the peak amplitude difference increased with the orientation contrast. We also found a significant main effect of cortical area ($F_{4, 36} = 3.425$, $p = 0.041$) and a significant interaction between orientation contrast and cortical area ($F_{8, 72} = 3.221$, $p = 0.030$). Hence, the effect of orientation contrast decreased gradually from lower to higher cortical areas. This was confirmed in further analysis which showed that the main effect of orientation contrast was significant in V1–V4 (all $F_{2, 18} > 13.722$, $p < 0.010$), but not in IPS ($F_{2, 18} = 0.120$, $p = .840$). These findings revealed that neural activities in early visual areas were parallel to the attentional effect.

To examine several other areas of interest, including lateral geniculate nucleus (LGN) and FEF, we ran a supplementary fMRI experiment. This employed a similar design, but with an increased repetition time (TR) of 2 s to enable whole brain scanning. ROI analyses showed that the main effect of orientation contrast was significant in V1–V4, but not in IPS, LGN, and FEF. Furthermore, we performed a group analysis and did a whole-brain search with a general linear model (GLM) procedure (Friston et al., 1995) for cortical areas whose activities increased with the orientation contrast. Only early visual cortical areas were found (Figure S3).



correlation was found in other areas. These results indicate a close relationship between the attentional effect, V1 activities, and the C1 component.

DISCUSSION

We assume that the absence of awareness to an exogenous cue (and indeed the whole texture stimuli) maximally reduced various top-down influences, even if it did not completely abolish them. These influences include those arising from feature perception, object recognition, and subjects' intentions (Jiang et al., 2006). By contrast with most previous studies on visual saliency, this enabled us to observe a relatively pure saliency signal. This is particularly important because temporally sluggish fMRI signals typically reflect neural activities resulting from both bottom-up and top-down processes, even in the early visual cortical areas (Fang et al., 2008; Harrison and Tong, 2009; Ress and Heeger, 2003). We could then investigate whether the awareness-free saliency signal would be observed in IPS and/or in earlier visual areas. Human IPS (and its monkey analog) is associated with both top-down and bottom-up attention, and is a site at which correlates of saliency have been observed (Bisley and Goldberg, 2010; Geng and Mangun, 2009; Gottlieb et al., 1998). We found that the BOLD response to this invisible cue in V1–V4, but not in IPS, increased with the attentional cueing effect. Indeed, this resembled the saliency value of this cue that was the output of a V1 saliency model (Li, 1999, 2002). The cue-evoked C1 amplitude, believed to represent V1's sensory responses (Clark et al., 1995; Di Russo et al., 2002; Martínez et al., 1999), also increased with the saliency. More importantly, across observers, the cueing effect significantly correlated with the C1 amplitude, and with the BOLD signal in V1, but not elsewhere. This meant that the saliency map for individual subjects could be predicted from their V1 activities.

The most parsimonious account of our results is that V1 is more important than later cortical areas for realizing the saliency

One should note that according to the V1 saliency hypothesis, saliency of a visual location is determined by its highest evoked V1 response relative to those evoked by other locations. In other words, saliency is determined by the relative rather than absolute levels of V1 responses. This perspective is necessary to understand why V1 responses to a non-salient conjunctive search target in an inhomogeneous background (e.g., a red-vertical bar among many green-vertical and red-horizontal bars) is not necessarily lower than those to a salient pop-out target against a homogeneous background (e.g., a red-vertical bar among red-horizontal bars, Hegdé and Felleman, 2003). As explained in the analysis above, due to the intracortical iso-orientation suppression, and iso-feature (e.g., iso-color) suppression in general (Li, 1999), the V1 population responses to a homogeneous background are quite low, and lower than those to a less homogeneous background, such as the background for the conjunction target. Therefore, the unique feature target can be more salient than the unique conjunctive target even when the former evokes a lower V1 response, provided that the population responses to the homogeneous background of the unique feature target are sufficiently lower still. The dependence of saliency on the relative rather than the absolute levels of neural responses means that one has to look at the population responses, rather than a single neuron response, to assess saliency in a scene (Hegdé and Felleman, 2003). Alternatively, one may compare the relative saliency of two items from their evoked V1 responses only when they share the same or comparable background stimuli. The latter is the case in our cueing stimuli, in which different pop-out foregrounds share the same homogeneous background texture.

Our data suggest that the neural correlates of saliency observed in intermediate and higher cortical areas, such as V4 or the parietal cortex, may be relayed from V1 rather than created within these areas. Parietal regions are known to integrate bottom-up and top-down attentional guidance (Bisley and Goldberg, 2010). Meanwhile, consistent with the idea that saliency is computed outside V4, V4 lesions impair the selection of the non-salient but not the salient objects in the scene (Schiller and Lee, 1991), and modulations in V4 responses to salient locations are eliminated when monkey prepares a goal related saccade elsewhere (Burrows and Moore, 2009). Similarly, lesions of the frontal eye field disrupt visual pursuit (Lynch, 1987) but barely affect input-driven saccades to salient locations (Schiller et al., 1987). Because neural correlates of saliency in these areas are generally evoked by highly visible inputs, and because the saliency signal was absent in IPS in our data which generated saliency using invisible stimuli, it remains unclear whether saliency is only relayed to parietal regions when the visual input responsible is perceptually visible.

Note that we distinguish a cortical area (V1) creating the saliency map from those that read out or inherit the saliency values from earlier regions along the visual pathway. Hence, for example, superior colliculus and parietal cortex, both receiving inputs from V1 directly or indirectly, may be viewed as areas reading out the saliency values to execute attentional shifts or to combine with top-down factors (note that retinal drives to SC do not lead to visual evoked saccades) (Schiller, 1998). Meanwhile, retina and LGN cannot be viewed as saliency maps. Of course, saliency values can be computed from their population responses (as indeed in the proposal that this happens via V1 intracortical mechanisms). However, the responses in these regions lack the significant context dependence required for saliency (e.g., that a vertical bar is salient in a background of horizontal, but not vertical, bars).

Our findings can be viewed as identifying V1 as the neural substrate of the early component of attentional selection. There has been over half a century of debate about the extent to which exogenous attentional selection occurs early or late, i.e., before or after visual inputs is perceptually identified (see Yantis and Johnston [1990] for a review). In principle, both top-down and bottom-up selection could occur at early or late stages. Most evidence discriminating early versus late selection has come from behavioral studies, whereas physiological evidence from ERP and single unit recordings has mainly implicated the extra-striate cortices in early selection (Luck et al., 1994; Moran and Desimone, 1985). V1 neurons are tuned only to primitive features rather than complex objects, and they respond even to stimulus features that are invisible to awareness. Thus, identifying V1 as the neural substrate of saliency confirms that selection can occur before input identification and awareness. Locating bottom-up selection in V1 invites us to re-evaluate the brain network for attention control.

EXPERIMENTAL PROCEDURES

Subjects

A total of 22 human subjects (7 male, 20–35 years old) were involved in the study. All of them participated in the psychophysical experiment. Sixteen and ten of them participated in the ERP and fMRI experiments, respectively. One subject in the ERP experiment was excluded because of frequent eye blinks. All subjects were naive to the purpose of the study except for two subjects (two of the authors). They were right-handed, reported normal or corrected to normal vision, and had no known neurological or visual disorders. They gave written, informed consent in accordance, and our procedures and protocols were approved by the human subjects review committee of Peking University.

Stimuli

Each texture stimulus (Figure 1A) had a regular Manhattan grid of 15 × 29 low-luminance bars (3.4 cd/m^2), presented in the lower visual field on a dark screen (1.6 cd/m^2). Each bar was a rectangle of 0.075° × 0.75° in visual angle. The center-to-center distance between the bars was 1.13°. All bars were identically oriented except for a foreground region of 2 × 2 bars with another orientation in either the lower left or the lower right quadrant. The foreground region was centered at 7.2° eccentricity. The orientation of the background bars was randomly chosen from 0° to 180°. There were five possible orientation contrasts between the foreground bars and the background bars: 0°, 7.5°, 15°, 30°, and 90°. The mask stimulus (Figure 1B) had the same grid as the texture stimuli. Each mask element contained 12 intersecting high-luminance bars (120 cd/m^2) oriented from 0° to 165° at every 15° interval. The bars in the mask had the same size and shape as those in the texture stimuli.

Psychophysical Experiment

Visual stimuli were displayed on an IIYAMA color graphic monitor (model: HM204DT; refresh rate: 60 Hz; resolution: 1024 × 768; size: 22 inches) at a viewing distance of 57 cm. Subjects' head position was stabilized using a chin rest. A white fixation cross was always present at the center of the monitor.

Each trial began with the fixation. A texture stimulus was presented for

attention. Then a two-dot probe was presented for 50 ms at randomly either the foreground region (valid cue condition) or its contralateral counterpart (invalid cue condition) (Figure 1C) with equal probability. Subjects were asked to press one of two buttons to indicate whether the upper dot was to the left or right of the lower dot. The experiment consisted of ten blocks. Each block had 96 trials, from randomly interleaving 24 trials from each of the four orientation contrasts (7.5°, 15°, 30°, and 90°) between the foreground and background bars. The attentional effect for each orientation contrast was quantified as the difference between the accuracy of the probe task performance in the valid cue and the invalid cue conditions.

All subjects also underwent a two-alternative forced choice (2AFC) experiment to determine whether the masked foreground region was indeed invisible in a criterion-free way. The stimuli and procedure in this 2AFC experiment were the same as those in the attention experiment, except that no probe was presented. After the presentation of a masked texture stimulus, subjects were asked to make a forced choice response regarding which side (lower left or lower right) from the fixation they thought the foreground region appeared. They performed at chance level in this 2AFC experiment for all four orientation contrasts, providing an objective confirmation that the masked foreground region was indeed invisible.

ERP Experiment

The experimental setup and procedure were similar to those in the 2AFC experiment. There were four possible orientation contrasts (0°, 7.5°, 15°, and 90°) in the texture stimuli. The experiment consisted of 20 blocks of 80 trials, 400 trials for each orientation contrast.

Scalp EEG was recorded from 64 Ag/AgCl electrodes positioned according to the extended international 10–20 EEG system. Vertical electro-oculogram (VEOG) was recorded from electrodes placed above and below the right eye. Horizontal EOG (HEOG) was recorded from electrodes placed at the outer cantus of each eye. Electrode impedance was kept below 5 kΩ. EEG was amplified with a gain of 500 K, bandpass filtered at 0.05–100 Hz, and digitized at a sampling rate of 500 Hz. The signals on these electrodes were referenced online to the nose and were rereferenced offline to the average of two mastoids. Using Brain Vision Analyzer (Brain Products, Munich, Germany), eye-blink artifacts were semi-automatically corrected using the procedure described by Gratton et al. (1983). EEG epochs lasting 350 ms, starting at 100 ms before the texture stimulus onset, were made. They were selectively averaged according to the orientation contrast. Epochs with EEG or residual EOG exceeding ±50 μV at any electrode were excluded from the average. The average waveforms were low-pass filtered at 40 Hz and baseline corrected with respect to the average voltage during the 100-ms prestimulus interval.

The C1 response was apparent between 60 and 90 ms after stimulus onset. To select electrodes for the C1 amplitude and latency analysis, grand averaged ERPs were made by averaging across subjects and orientation contrasts. Posterior electrodes, including CP1, CPz, CP2, P1, Pz, and P2, had the largest C1 amplitudes. To quantify the C1 amplitude for each subject, the mean amplitude of the five sampling points around the C1 peak was first calculated for each of these six electrodes, and this mean was then averaged across the six electrodes. The C1 latency was the mean of the peak latencies across these six electrodes.

Estimation of the dipole sources was performed using the BESA algorithm as described by Clark and Hillyard (1996) and Frey et al. (2010). The C1 component was modeled based jointly on the grand-averaged waveforms elicited by texture stimuli with the four orientation contrasts. The waveform in the interval between 62 and 82 ms was simulated with two dipoles, one in each hemisphere, which were constrained to have mirror-symmetrical locations, but allowed to vary in orientation. The initial starting positions of dipoles were randomly chosen and using different starting locations yielded high similar dipole configurations.

fMRI Experiment

The event-related fMRI experiment consisted of four functional scans of 128 continuous trials. Each scan began with 6 s fixation and lasted 274 s. There were four types of trials—orientation contrast trials (7.5°, 15°, and 90°) and fixation trial. In an orientation contrast trial, a texture stimulus was presented for 50 ms, followed by a 100 ms mask and 1,850 ms fixation. Similar to the 2AFC experiment, subjects were asked to indicate the location of the foreground region, which was left to the fixation in one half of orientation contrast trials and right in the other half at random. In a fixation trial, only the fixation point was presented for 2 s. In a scan, there were 32 trials for each type of trial. The order of the trials was counterbalanced across four scans using M-sequences (Buracas and Boynton, 2002). These are pseudo random sequences that have the advantage of being perfectly counterbalanced n trials back, so that each type of trials was preceded and followed equally often by all types of trials, including itself.

Retinotopic visual areas (V1, V2, V3, and V4) were defined by a standard phase-encoded method developed by Sereno et al. (1995) and Engel et al. (1997), in which subjects viewed rotating wedge and expanding ring stimuli that created traveling waves of neural activity in visual cortex. A block-design scan was used to localize the ROIs in V1–V4 and IPS corresponding to the foreground region. The scan consisted of 12 12-s stimulus blocks, interleaved with 12 12-s blank intervals. In a stimulus block, subjects passively viewed images of colorful natural scenes, which had the same size as the foreground region in texture stimuli and were presented at the location of the foreground region (either left or right to fixation). Images appeared at a rate of 4 Hz.

MRI data were collected using a 3T Siemens Trio scanner with a 12-channel phase-array coil. In the scanner, the stimuli were back-projected via a video projector (refresh rate: 60 Hz; spatial resolution: 1,024 × 768) onto a translucent screen placed inside the scanner bore. Subjects viewed the stimuli through a mirror located above their eyes. The viewing distance was 83 cm. Blood oxygen level-dependent (BOLD) signals were measured with an echo-planar imaging sequence (TE: 30 ms; TR: 1000 ms; FOV: 186 × 192 mm^2; matrix: 62 × 64; flip angle: 90; slice thickness: 5 mm; gap: 0 mm; number of slices: 16, slice orientation: coronal). The fMRI slices covered the occipital lobe, most of the parietal lobe and part of the temporal lobe. A high-resolution 3D structural data set (3D MPRAGE; 1 × 1 × 1 mm^3 resolution) was collected in the same session before the functional scans. Subjects underwent two sessions, one for the retinotopic mapping and the other for the main experiment.

The anatomical volume for each subject in the retinotopic mapping session was transformed into a brain space that was common for all subjects (Talairach and Tournoux, 1988) and then inflated using BrainVoyager QX. Functional volumes in both sessions for each subject were preprocessed, including 3D motion correction, linear trend removal, and high-pass (0.015 Hz) (Smith et al., 1999) filtering using BrainVoyager QX. Head motion within any fMRI session was<2 mm for all subjects. The images were then aligned to the anatomical volume in the retinotopic mapping session and transformed into Talairach space. The first 6 s of BOLD signals were discarded to minimize transient magnetic saturation effects.

A general linear model (GLM) procedure was used for the ROI analysis. The ROIs in V1–V4 and IPS were defined as areas that responded more strongly to the natural scene images than blank screen ($p < 10$

BOL39729(c9(att)-394(39228m)-13(fo723288(in)-32t,20)tyl)9(pee)eLDto bya45(f158((moti(th)-277nd73glum61(514uelume4(f)-13(59(a45(f158((werxa(Hea

a foreground region (i.e., 2 × 2 bars) like in our experiments. The model mechanisms include (1), direct inputs to V1 neurons from each bar according to the classical receptive fields, and (2), interactions between V1 neurons by the intracortical connections implementing contextual influences (such as surround suppression) of the surround to the neural responses. At each grid location, the maximum response from all pyramidal model neurons was obtained. This maximum was averaged over all simulation time steps within 50 ms (simulated by five membrane time constant of the model neurons). The saliency of each grid location is the *Z*-score of this maximum obtained as follows: take the difference between this maximum and the average of the maximums over all grid locations and then divide it by the standard deviation of all the maximums (Li, 1999). Saliency in the foreground region is the maximum of the *Z*-scores over the 4 × 4 bar region centered on the foreground region. The result for each orientation contrast (7.5°, 15°, 30°, and 90°) as plotted in Figure 2 was obtained by averaging the foreground region saliency from 24 simulations for 24 different background bar orientations evenly distributed between 0° and 180°. The saliency of the foreground region should be directly related to the strength of its attentional attraction (i.e., its cueing effect).

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and can be found with this article online at doi:10.1016/j.neuron.2011.10.035.

ACKNOWLEDGMENTS

We are grateful to Peter Dayan for reading the manuscript with helpful comments and Yan Song for help with dipole source localization. This work was supported by the Ministry of Science and Technology of China (2011CBA00405 and 2010CB833903), the National Natural Science Foundation of China (Project 30925014, 30870762, 90920012, 30921064, and 90820307), the Fundamental Research Funds for the Central Universities, and the Gatsby Charitable Foundation.

Accepted: October 31, 2011
Published: January 11, 2012

REFERENCES

Ales, J.M., Yates, J.L., and Norcia, A.M. (2010). V1 is not uniquely identified by polarity reversals of responses to upper and lower visual field stimuli. Neuroimage *52*, 1401–1409.

Allman, J., Miezin, F., and McGuinness, E. (1985). Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons. Annu. Rev. Neurosci. *8*, 407–430.

Bao, M., Yang, L., Rios, C., He, B., and Engel, S.A. (2010). Perceptual learning increases the strength of the earliest signals in visual cortex. J. Neurosci. *30*, 15080–15084.

Bisley, J.W., and Goldberg, M.E. (2010). Attention, intention, and priority in the parietal lobe. Annu. Rev. Neurosci. *33*, 1–21.

Buracas, G.T., and Boynton, G.M. (2002). Efficient design of event-related fMRI experiments using M-sequences. Neuroimage *16*, 801–813.

Burrows, B.E., and Moore, T. (2009). Influence and limitations of popout in the selection of salient visual stimuli by area V4 neurons. J. Neurosci. *29*, 15169–15177.

Clark, V.P., and Hillyard, S.A. (1996). Spatial selective attention affects early extrastriate but not striate components of the visual evoked potential. J. Cogn. Neurosci. *8*, 387–402.

Clark, V.P., Fan, S., and Hillyard, S.A. (1995). Identification of early visually evoked potential generators by retinotopic and topographic analysis. Hum. Brain Mapp. *2*, 170–187.

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

Di Russo, F., Martínez, A., Sereno, M.I., Pitzalis, S., and Hillyard, S.A. (2002). Cortical sources of the early components of the visual evoked potential. Hum. Brain Mapp. *15*, 95–111.

Engel, S.A., Glover, G.H., and Wandell, B.A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. Cereb. Cortex *7*, 181–192.

Fang, F., Boyaci, H., Kersten, D., and Murray, S.O. (2008). Attention-dependent representation of a size illusion in human V1. Curr. Biol. *18*, 1707–1712.

Fecteau, J.H., and Munoz, D.P. (2006). Salience, relevance, and firing: a priority map for target selection. Trends Cogn. Sci. (Regul. Ed.) *10*, 382–390.

Frey, H.P., Kelly, S.P., Lalor, E.C., and Foxe, J.J. (2010). Early spatial attentional modulation of inputs to the fovea. J. Neurosci. *30*, 4547–4551.

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.B., Frith, C., and Frackowiak, R.S.J. (1995). Statistical parametric maps in functional imaging: a general linear approach. Hum. Brain Mapp. *2*, 211–224.

Geng, J.J., and Mangun, G.R. (2009). Anterior intraparietal sulcus is sensitive to bottom-up attention driven by stimulus salience. J. Cogn. Neurosci. *21*, 1584–1601.

Gilbert, C.D., and Wiesel, T.N. (1983). Clustered intrinsic connections in cat visual cortex. J. Neurosci. *3*, 1116–1133.

Gottlieb, J.P., Kusunoki, M., and Goldberg, M.E. (1998). The representation of visual salience in monkey parietal cortex. Nature *391*, 481–484.

Gratton, G., Coles, M.G.H., and Donchin, E. (1983). A new method for off-line removal of ocular artifact. Electroencephalogr. Clin. Neurophysiol. *55*, 468–484.

Harrison, S.A., and Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. Nature *458*, 632–635.

Hegdé, J., and Felleman, D.J. (2003). How selective are V1 cells for pop-out stimuli? J. Neurosci. *23*, 9968–9980.

Hupé, J.M., James, A.C., Girard, P., and Bullier, J. (2001). Response modulations by static texture surround in area V1 of the macaque monkey do not depend on feedback connections from V2. J. Neurophysiol. *85*, 146–163.

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

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

Jiang, Y., Costello, P., Fang, F., Huang, M., and He, S. (2006). A gender- and sexual orientation-dependent spatial attentional effect of invisible images. Proc. Natl. Acad. Sci. USA *103*, 17048–17052.

Jonides, J. (1981). Voluntary vs. automatic control over the mind's eye's movement. In Attention and Performance, *Volume XI*, M.I. Posner and O. Marin, eds. (Hillsdale, NJ: Lawrence Erlbaum Associates), pp. 187–205.

Kastner, S., and Ungerleider, L.G. (2000). Mechanisms of visual attention in the human cortex. Annu. Rev. Neurosci. *23*, 315–341.

Knierim, J.J., and Van Essen, D.C. (1992). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. J. Neurophysiol. *67*, 961–980.

Koch, C., and Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. Hum. Neurobiol. *4*, 219–227.

Koene, A.R., and Zhaoping, L. (2007). Feature-specific interactions in salience from combined feature contrasts: evidence for a bottom-up saliency map in V1. J. Vis. *7*, 1–14.

Kourtzi, Z., and Kanwisher, N. (2000). Cortical regions involved in perceiving object shape. J. Neurosci. *20*, 3310–3318.

Kustov, A.A., and Robinson, D.L. (1996). Shared neural control of attentional shifts and eye movements. Nature *384*, 74–77.

Li, Z. (1999). Contextual influences in V1 as a basis for pop out and asymmetry in visual search. Proc. Natl. Acad. Sci. USA *96*, 10530–10535.

Li, Z. (2002). A saliency map in primary visual cortex. Trends Cogn. Sci. (Regul. Ed.) *6*, 9–16.

Luck, S.J., Hillyard, S.A., Mouloua, M., Woldorff, M.G., Clark, V.P., and