Exp Brain Res (2016) 234:1769–1780
DOI 10.1007/s00221-016-4583-y

RESEARCH ARTICLE

Neural activities in V1 create the bottom-up saliency map of natural scenes

Cheng Chen[1,2] · Xilin Zhang[3] · Yizhou Wang[1,2] · Tiangang Zhou[4] · Fang Fang[2,5]

Received: 26 June 2015 / Accepted: 30 January 2016 / Published online: 15 February 2016

Abstract A saliency map is the bottom-up contribution to the deployment of exogenous attention. It, as well as its underlying neural mechanism, is hard to identify because of the influence of top-down signals. A recent study showed that neural activities in V1 could create a bottom-up saliency map (Zhang et al. in Neuron 73(1):183–192, 2012). In this paper, we tested whether their conclusion can generalize to complex natural scenes. In order to avoid top-down influences, each image was presented with a low contrast for only 50 ms and was followed by a high contrast mask, which rendered the whole image invisible to participants (confirmed by a forced-choice test). The Posner cueing paradigm was adopted to measure the spatial cueing effect (i.e., saliency) by an orientation discrimination task. A positive cueing effect was found, and the magnitude of the cueing effect was consistent with the saliency prediction of a computational saliency model. In a following fMRI experiment, we used the same masked natural scenes as stimuli and measured BOLD signals responding to the predicted salient region (relative to the background). We found that the BOLD signal in V1, but not in other cortical areas, could well predict the cueing effect. These results suggest that the bottom-up saliency map of natural scenes could be created in V1, providing further evidence for the V1 saliency theory (Li in Trends Cogn Sci 6(1):9–16, 2002).

Keywords Bottom-up saliency map · Visual attention · Natural scene · fMRI · Primary visual cortex

✉ Yizhou Wang
Yizhou.Wang@pku.edu.cn
http://www.idm.pku.edu.cn/staff/wangyizhou

✉ Fang Fang
ffang@pku.edu.cn
http://www.psy.pku.edu.cn/en/fangfang.html

1 National Engineering Laboratory for Video Technology, Cooperative Medianet Innovation Center, and School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, People's Republic of China

2 Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing 100871, People's Republic of China

3 Laboratory of Brain and Cognition, National Institute of Mental Health, US National Institutes of Health, Bethesda, MD 20892, USA

4 State Key Laboratory of Brain and Cognitive Science, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, People's Republic of China

5 Department of Psychology and Beijing Key Laboratory of Behavior and Mental Health, Peking-Tsinghua Center for Life Sciences, and IDG/McGovern Institute for Brain Research, Peking University, Beijing 100871, People's Republic of China

Introduction

Visual attention is essential for us to recognize complex natural scenes (Carrasco 2011; Yoshida et al. 2012). It enables us to select the most valuable information. Such a process of information selection could be executed by a top-down signal voluntarily, or triggered by a bottom-up salient stimulus automatically, or most likely, achieved through a combination of both top-down and bottom-up signals (Corbetta and Shulman 2002; Serences and Yantis 2007). The top-down process could be guided under a specific goal (Buschman and Miller 2007), while the bottom-up process is guided by a bottom-up saliency map (Zhang et al. 2012). The bottom-up saliency map is defined as a topographical map to describe and predict the distribution of attentional attraction based on a bottom-up visual input (Koch and

Ullman 1985). In contrast to the well-known fact that several higher brain regions, including frontal eye field (FEF) and posterior parietal cortex (PPC), are responsible for top-down signals (Hopfinger et al. 2000; Baluch and Itti 2011; Gilbert and Li 2013), there still exists controversy about the neural basis of the bottom-up saliency map.

Evidence from brain imaging and neurophysiology studies had shown that many regions could realize a saliency map. Subcortical structures such as superior colliculus (Fecteau and Munoz 2006) and pulvinar (Shipp 2004) were found to be able to construct a saliency map. Besides, Mazer and Gallant (2003) found that extrastriate ventral area V4 could realize a retinotopic saliency map to guide eye movement, providing evidence that brain activities in the ventral pathway could be correlated with the saliency map. Their conclusion was also supported by Asplund et al. (2010), who found that the ventral network can account for stimulus-driven attention. Geng and Mangun (2009) found that the anterior intraparietal sulcus (aIPS) was sensitive to the bottom-up influence driven by stimulus saliency. Moreover, FEFs were also found to play an important role in decoding the winner-take-all (WTA) stage of saliency processing (Bogler et al. 2011). Most of these findings were consistent with a dominant view, which argues that the final saliency map results from pooling different visual feature channels after each visual feature channel construct its own saliency map independently (Koch and Ullman 1985; Itti and Koch 2001). Accordingly, higher cortical areas, in which neurons are less selective to single features, are more likely to be possible candidates that realize the bottom-up saliency map.

On the other hand, Li (1999, 2002) proposed the V1 theory, which argued that neural activities in V1 could create the bottom-up saliency map via intracortical interactions that were manifested in contextual influences. By measuring the reaction time searching for a singleton that differs from its surrounding in more than one feature, Koene and Zhaoping (2007) found their results were consistent with the properties of some V1 neurons, which provided evidence for the V1 theory of the bottom-up saliency map. Evidence from a brain imaging study also supported the V1 theory (Zhang et al. 2012).

An important reason of the controversy is that most of these cortical areas receive both bottom-up and top-down signals, which makes it difficult to determine whether the saliency map they realized truly reflects the bottom-up attentional attraction or not. In order to investigate the neural basis of the bottom-up saliency map, it is important to probe bottom-up signals free from top-down signals. Recently, Zhang and his colleagues used invisible texture stimuli to investigate the neural basis of the bottom-up saliency map. In their study, stimuli were presented using backward masking, which enabled the absence of awareness to an exogenous cue. Therefore, top-down influences were maximally reduced in their experiments. They found that even when participants could not perceive the stimuli, the bottom-up saliency map could still attract participants' attention to improve their performance in a visual discrimination task. More importantly, they also found that the degree of attentional attraction correlated with the amplitude of the earliest component of the ERP as well as the V1 BOLD signal across participants (Zhang et al. 2012). Their findings strongly supported the V1 theory.

However, as far as we know, few studies have tested the V1 theory on natural scenes (Zhaoping and Zhe 2015). Compared with stimuli that consist of simple oriented bars, natural scenes contain richer image statistics which human visual system is highly tuned to. Therefore, natural scenes are optimal for automatic processing such as attention selection (Bogler et al. 2011). More importantly, some argued that the results of Zhang and his colleagues' study might attribute to the restricted stimulus set they used (Betz et al. 2013). Betz et al. (2013) claimed that neural activities in V1 were monotonically related to stimulus contrast (or luminance), other than saliency. Therefore, it is possible that the V1 saliency map found in Zhang and his colleagues' study was an intermediate result of saliency computational processing, instead of a final saliency map. The restricted stimulus used in their study, which was an array of oriented bars, made the intermediate result indistinguishable from the final saliency map. In order to clarify this issue, the V1 theory should also be tested on natural scenes.

In this study, we tested whether their conclusion can generalize to complex natural scenes. We measured the bottom-up saliency map from both psychophysical and physiological aspects. Top-down signals were maximally reduced in our experiment using backward masking. The attentional effect of the bottom-up saliency map and BOLD signals responding to invisible natural scenes were measured to understand the neural mechanism of the bottom-up saliency map. We found that when the degree of saliency increased, the attentional effect and the BOLD signal in V1 (but not other cortical areas) also increased, even if participants were unaware of stimuli. More importantly, the attentional effect correlated with the BOLD signal across participants only in V1.

Materials and methods

Participants

Sixteen human participants (nine females and seven males, 19–26 years old) participated in the experiment. All of

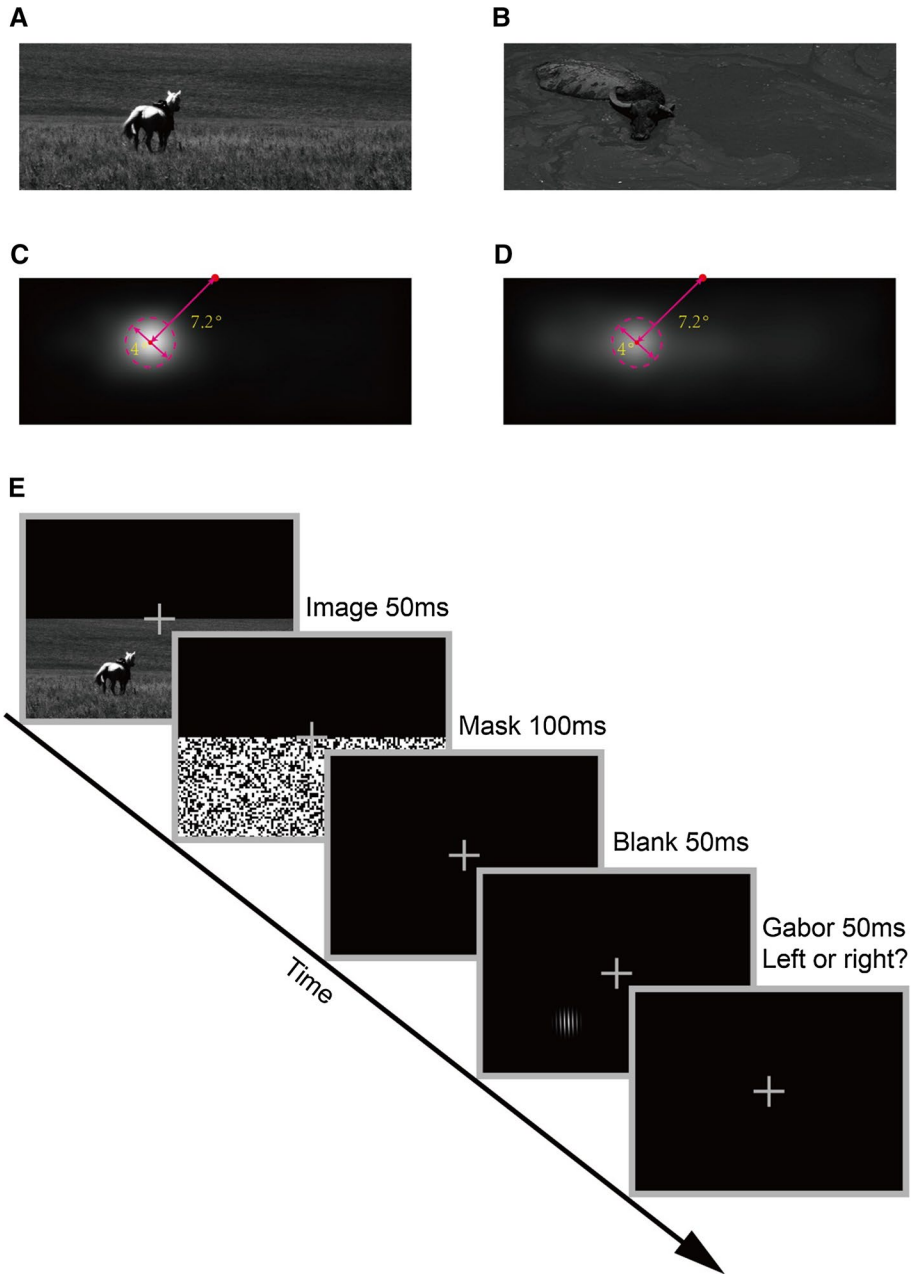


Fig. 1 Stimuli and psychophysical protocol. **a**, **b** Examples of high salient (**a**) and low salient (**b**) natural images. **c**, **d** The averaged saliency map of 25 high salient images (**c**) and 25 low salient images (**d**) with a salient region left to the fixation. Areas with a high luminance level had a high saliency. A round salient region could be seen in this map. The eccentricity of the center of the salient region was about 7.2°. The diameter of the salient region was about 4°. **e** Psychophysical protocol to measure the attentional effect of the bottom-up saliency maps of natural images. A low luminance natural image was presented for 50 ms which served as a cue, followed by a 100 ms mask and a 50 ms fixation screen. Then, a Gabor was presented for 50 ms at either the location of the salient region of the preceding image (valid cue condition) or its contralateral counterpart (invalid cue condition). Participants pressed one of the two buttons to indicate whether the orientation of Gabor is clockwise or counterclockwise to the *vertical*. The low luminance natural image was invisible to participants because of briefly presentation and backward masking. It was also confirmed by an additional 2-AFC experiment

them participated in the psychophysical experiment. Thirteen of them participated in the fMRI experiment. Two participants in the fMRI experiment were excluded from data analysis because of excessive head motion during fMRI scanning. All participants were naive to the purpose of the study except for one participant (one of the authors). All of them 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 with the procedures and protocols approved by the human participants review committee of Peking University.

Stimuli

A large number of natural images were collected from the Internet and several public datasets, including AIM (Bruce and Tsotsos 2005), Caltech101 (Fei-Fei et al. 2004), BSDS500 (Martin et al. 2001), and ImgSal (Li et al. 2013). All these images were scaled to the same size (11.63° × 31.03° of visual angle) with a mean low luminance (2.9 cd/m^2) (Figs. 1a, b, 2a). Then, we calculated the computational saliency map of each image by using a prominent bottom-up saliency model proposed by Itti et al. (1998). After that, we selected 50 images for the current

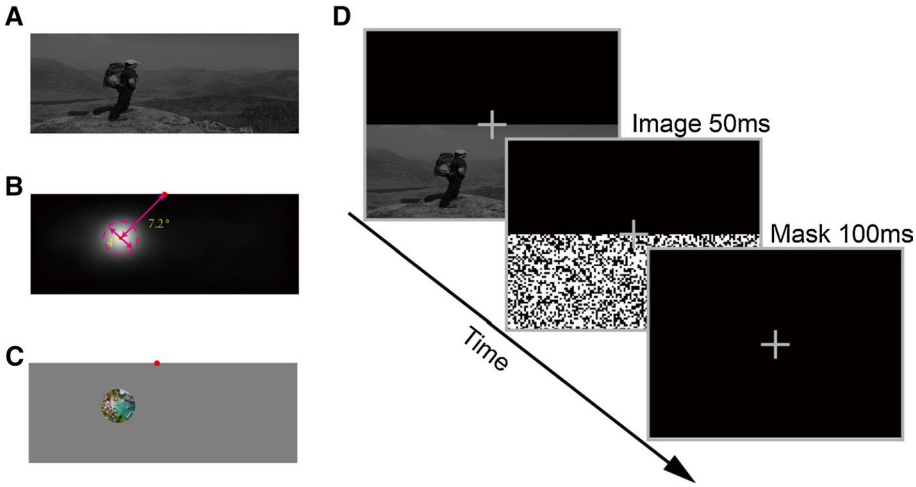


Fig. 2 Localizer and procedure of the fMRI experiment. **a**, **b** An example of the high salient natural images (**a**) and the averaged saliency map of the high salient images (**b**). **c** The localizer we used to define ROIs in the fMRI experiment. The colorful natural scene had the same size and was presented at the same location with the round salient region showed in (**b**). **d** The procedure of a salient image trial in the fMRI experiment. In a salient image trial, an image was presented on the lower half of the screen for 50 ms, followed by a 100 ms mask at the same position and 1850 ms fixation. Participants were asked to indicate the location of the salient region

experiment. Each of these images had only one round salient region centered at about 7.2° eccentricity in the lower left quadrant (i.e., left salient images). The diameter of the salient region was about 4° (Figs. 1c, d, 2b). Moreover, to measure the bottom-up saliency map quantitatively, we classified all images into two groups: the high salient group and the low salient group, based on the proposed saliency index in the following formulation.

$$\text{Index}(n) = \frac{S_{\text{I}}(n) - S_{\text{O}}(n)}{S_{\text{O}}(n)}.$$

In the above formulation, n denoted the index of an image. S_{I} denoted the averaged saliency value of the round salient region in Fig. 1c, d, and S_{O} denoted the averaged saliency value of the residual region (the rest of an image). A higher value of the Index indicated a higher saliency of the round region. We selected half of these left salient images with an upper 50 % Index as the high salient images and the rest as the low salient images. The averaged saliency map of the 25 high salient images is shown in Fig. 1c, and the low salient counterpart is shown in Fig. 1d.

Furthermore, in order to balance the location of the salient region, each of the left salient images was flipped horizontally along its vertical midline to generate a new image, which had a salient region in the lower right quadrant (i.e., right salient images). Notice that the content of a left salient image and its corresponding right salient image were exactly the same, the only difference between them was the location of the salient region. Thus, the stimuli used in the psychophysical and the fMRI experiment consisted of two groups: the high salient group and the low salient group. Each group contained 50 images, half of which were left salient images and the other half were right salient images. In both experiments, these stimuli were presented in the lower visual field on a dark screen (1.8 cd/m^2).

We tested the difference between the luminance of the low salient and high salient images. The results showed that there was no significant difference between the luminance of the low salient and high salient images. The averaged gray value within the round salient region was 85.19 ± 7.93 (Mean ± SE) for the low salient images and 96.62 ± 6.31 for the high salient images. Moreover, the averaged gray value of a whole image was 77.99 ± 8.35 for the low salient images and 81.66 ± 8.57 for the high salient images. Independent t tests showed that there was no significant difference between the low salient and high salient images, no matter in the averaged gray value within the round salient region ($t_{48} = 1.127$, $p = .265$, $\eta^2 = .026$) or the averaged gray value of a whole image ($t_{48} = .307$, $p = .760$, $\eta^2 = .002$).

Moreover, we also measured the distribution of the saliency index values. The means and standard deviations of the saliency index values of the low and high salient group were 4.03 ± 1.02 and 10.67 ± 4.20, respectively. Meanwhile, the range of the saliency index values of all the images was 2.03–19.07. Specifically, the minimum and maximum saliency index value of the low salient group was 2.03 and 5.77, respectively. The minimum and maximum saliency index value of the high salient group was 5.84 and 19.07, respectively.

Mask stimuli were high contrast checkerboards with randomly arranged checkers (Fig. 1e). The size of each

checker was about .25° × .25°. The luminance of a black checker was 1.8 cd/m^2, while the luminance of a white checker was 79 cd/m^2.

Psychophysical experiment

In the psychophysical experiment, visual stimuli were presented on a Gamma-corrected Iiyama HM204DT 22 inches monitor, with a spatial resolution of 1024 × 768 and a refresh rate of 60 Hz. The viewing distance was 83 cm. Participants' head position was stabilized using a chin rest and a head rest. A white cross was always presented at the center of the screen as a fixation, and participants were asked to fixate the cross throughout the experiment.

Each trial started with a fixation. A natural image was presented on the lower half of the screen for 50 ms, followed by a 100 ms mask at the same position, and another 50 ms fixation interval. Here, the bottom-up saliency map of the image could serve as a cue to attract spatial attention, while the mask could ensure that the image was invisible to participants. Then, a Gabor centered at about 7.2° eccentricity from the fixation was presented at either the lower left quadrant or the lower right quadrant with equal probability for 50 ms. The location of the Gabor was either at the salient region of the preceding image or its contralateral counterpart, thus indicating the valid cue condition or the invalid cue condition. The Gabor had a spatial frequency of 5.5 cpd (cycle per degree), and its diameter was 2.5° with full contrast. The Gabor orientated at ±1.5° away from the vertical. Participants were asked to press one of the two keys to indicate the orientation of the Gabor (left or right tilted). The duration of each trial was 2 s. Figure 1e shows the procedure of our experiment. The experiment consisted of ten runs. Each run contained 100 trials with two conditions: the high salient condition and the low salient condition. Images for the two conditions were randomly selected from the high and the low salient image groups correspondingly. The attentional effect of bottom-up saliency maps for each condition was quantified as the difference between the orientation discrimination task performance (discrimination accuracy) of the valid cue condition and the invalid cue condition.

Moreover, in order to check whether these natural images were indeed invisible to participants, participants were also asked to complete an additional two-alternative forced-choice (2-AFC) experiment before the main experiment. The additional 2-AFC experiment consisted of four runs. Each run contained 100 trials. We used the same images in the 2-AFC experiment as the main experiment. In the additional experiment, each trial began with either a low luminance image or a blank with equal probability, followed by a mask. Participants were asked to make a forced-choice response to judge whether there was an image presented before the mask or not. Performance at chance level in this experiment could provide an objective confirmation that the masked images were indeed invisible. Besides, participants were also asked about their subjective feelings of these images after the 2-AFC experiment and after each run of the main experiment to ensure that the stimuli were invisible to participants throughout the experiments.

FMRI experiment

The fMRI experiment was conducted to investigate the neural basis of the bottom-up saliency map. An event-related design was adopted. The experiment consisted of eight functional scans of 125 continuous trials. Each scan lasted 268 s, including a 6 s fixation at the beginning, a 12 s fixation at the end, as well as 125 trials. There were five types of trials, including four types of salient image trials: two degrees of saliency (high and low) × two locations of salient region (left and right), and fixation trials. In all types of trials, a white cross was always presented at the center of the screen. Participants were asked to fixate the cross throughout each scan. In a salient image trial, an image was presented on the lower half of the screen for 50 ms, followed by a 100 ms mask at the same position and 1850 ms fixation. Participants were asked to indicate the location of the salient region (Fig. 2d). In half of salient image trials, the location of the salient region was left to the fixation, while in the other half, the location was right to the fixation. The location of the salient region was counterbalanced in each run. In a fixation trial, only the fixation point was presented for 2000 ms. In each scan, there were 25 trials for each type. The order of these trials was counterbalanced across eight scans using M-sequences (Buracas and Boynton 2002). M-sequences were pseudo random sequences that had the advantage of being perfectly counterbalanced n trial 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 (Sereno et al. 1995; Engel et al. 1997), in which participants viewed rotating wedge and expanding ring stimuli that created traveling waves of neural activity in visual cortex. A block-design scan was used to define the regions of interest (ROIs) in LGN, V1–V4, LOC, IPS, and FEF corresponding to the salient region. The scan consisted of twelve 12 s stimulus blocks, interleaved with twelve 12 s blank intervals. In a stimulus block, participants passively viewed images with colorful natural scenes, which had the same size as the salient region in the natural images, and were presented at the location of the salient region (either left or right to the fixation; Fig. 2c). Images appeared at a rate of 8 Hz.

MRI data acquisition

In the scanner, the stimuli were back-projected via a video projector (refresh rate: 60 Hz; spatial resolution: 1024 × 768) onto a translucent screen placed inside the scanner bore. Participants viewed the stimuli through a mirror located above their eyes. The viewing distance (i.e., the distance from the mirror to eyes) was 83 cm. MRI data were collected using a 3T Siemens Trio scanner with a 12-channel phase-array coil. Blood oxygen level-dependent (BOLD) signals were measured with an echo-planar imaging sequence (TE: 30 ms; TR: 2000 ms; FOV: 186 × 192 mm^2; matrix: 62 × 64; flip angle: 90; slice thickness: 5 mm; gap: 0 mm; number of slices: 30; slice orientation: coronal). The fMRI slices covered the occipital lobe, most of the parietal lobes, 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 runs. All the participants underwent two sessions, one for the retinotopic mapping and the other for the main experiment.

MRI data processing and analysis

The anatomical volume for each participant in the retinotopic mapping session was transformed into the anterior commissure–posterior commissure (AC–PC) space and then inflated using Brain Voyager QX. Functional volumes in all the sessions for each participant were preprocessed, including 3D motion correction, linear trend removal, and high-pass (.015 Hz) filtering (Smith et al. 1999) using Brain Voyager QX. Head motion within any fMRI session was <2 mm for all participants except two participants excluded from further analysis because of excessive head motion. FMRI images were then aligned to the anatomical volume in the retinotopic mapping session and transformed into the AC–PC 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 ROI analysis. The ROIs in LGN, V1–V4, LOC, IPS, and FEF were defined by a localizer scan and retinotopic mapping scans ($p < 10^{-8}$, uncorrected). In order to compare the locations of the ROIs with those reported in other studies, the coordinates of each ROI were identified in the Talairach space. The coordinates of rLGN, lLGN, rFEF, and lFEF were (21 ± 3, −26 ± 3, −1 ± 3), (−23 ± 3, −27 ± 2, −1 ± 2), (37 ± 7, −10 ± 5, 47 ± 4), and (−37 ± 6, −12 ± 4, 46 ± 3), respectively, consistent with previous studies (Berman et al. 1999; Chen et al. 1999; Connolly et al. 2002; Kastner et al. 2004; Luna et al. 1998; O'Connor et al. 2002; Paus 1996). The Talairach coordinates of the ROIs and their sizes (i.e., number of voxels) can be found in Tables 1 and 2.

Table 1 ROI coordinates

	Right (mean ± SD)	Left (mean ± SD)
LGN	(21 ± 3, −26 ± 3, −1 ± 3)	(−23 ± 3, −27 ± 2, −1 ± 2)
V1	(7 ± 3, −85 ± 5, 7 ± 8)	(−7 ± 3, −86 ± 4, 1 ± 9)
V2	(12 ± 3, −88 ± 6, 13 ± 5)	(−10 ± 2, −91 ± 6, 12 ± 6)
V3	(19 ± 3, −84 ± 6, 13 ± 7)	(−16 ± 4, −89 ± 6, 14 ± 8)
V4	(27 ± 2, −63 ± 7, −11 ± 4)	(−30 ± 5, −65 ± 6, 12 ± 5)
LOC	(43 ± 7, −68 ± 5, 0 ± 4)	(−43 ± 5, −72 ± 7, 0 ± 5)
IPS	(27 ± 6, −58 ± 8, 46 ± 7)	(−26 ± 7, −61 ± 6, 42 ± 6)
FEF	(37 ± 7, −10 ± 5, 47 ± 4)	(−37 ± 6, −12 ± 4, 46 ± 3)

Table 2 ROI sizes

	Right	Left
LGN	94	89
V1	303	421
V2	503	525
V3	466	392
V4	253	239
LOC	908	1111
IPS	580	456
FEF	250	237

The event-related BOLD signals were calculated separately for each participant, following the method developed by Kourtzi and Kanwisher (2000). For each event-related scan, the time course of the MR signal intensity was first extracted by averaging the data from all the voxels within the predefined ROI. The average event-related time course was then calculated for each type of trial, by selectively averaging from stimulus onset and using the average signal intensity during the fixation trials as a baseline to calculate percent signal change. Specifically, in each scan we averaged the signal intensity across the trials for each type of trial at each of 12 corresponding time points starting from the stimulus onset. These event-related time courses of the signal intensities were then converted to time courses of percent signal change for each type of trials by subtracting the corresponding value for the fixation trials and then dividing by that value. Because M-sequences have the advantage that each type of trials was precede and followed equally often by all types of trials, the overlapping BOLD responses due to the short interstimulus interval were removed by this averaging procedure (Buracas and Boynton 2002). The resulting time course for each type of trials was then averaged across scans and participants.

Computational saliency model

We adopted a prominent computational saliency model proposed by Itti and his colleagues (Itti et al. 1998) to measure the bottom-up saliency map of each image. The model was based on the center–surround mechanism, and combined information from three channels: color, intensity, and orientation. By using this model, we could predicate the degree of saliency of each image based on the formulation we proposed.

Results

In order to reduce the influence of top-down signals maximally, we presented an image very briefly and followed by a high contrast mask (Fig. 1e). In the additional 2-AFC experiment, all the participants reported that they were unaware of the natural images. The percentages of correct detection (mean ± SEM) were 48.6 ± 1.5 and 50.9 ± 1.4 % for the high salient and the low salient groups, respectively. The results were statistically indistinguishable from chance level (one sample t test: $t_{15} = -.934$, $p = .365$, $\eta^2 = .055$; significant level $\alpha = .05$), which indicated that the natural images in both groups were invisible to participants.

Psychophysical experiment

In the main experiment, considering that the salient region of a natural image could serve as a cue to attract attention, the attentional effect of the bottom-up saliency map of invisible natural images was quantified as the difference between the accuracy of the Gabor orientation discrimination performance in the valid cue condition and that in the invalid cue condition. We found that the discrimination accuracy was higher in the valid cue condition than that in the invalid cue condition, for both high salient images (Valid: 81.31 ± .98 %; Invalid 72.88 ± .98 %; Fig. 3a) and low salient images (Valid: 77.86 ± .93 %; Invalid 76.54 ± .88 %; Fig. 3a). Thus, the attentional effect of the bottom-up saliency maps for high salient group and low salient group was 8.43 ± .33 and 1.32 ± .47 %, respectively (left panel in Fig. 3b).

We submitted the behavioral results to a two-way repeated-measure ANOVA with saliency and validity as within-participant factors. The main effect of validity was significant ($F_{1,15} = 281.068$, $p < .001$, $\eta_p^2 = .949$). However, the main effect of saliency was not significant ($F_{1,15} = .035$, $p = .853$, $\eta_p^2 = .002$). Moreover, the interaction between these two factors was also significant ($F_{1,15} = 93.419$, $p < .001$, $\eta_p^2 = .862$). The simple contrast analysis showed that for both high salient and low salient groups, participants' performance was significantly better in the valid cue condition than that in the invalid cue condition (high salient: $F_{1,15} = 328.554$, $p < .001$, $\eta_p^2 = .956$; low salient: $F_{1,15} = 7.742$, $p = .014$, $\eta_p^2 = .340$). The attentional effect of the high salient images was also significantly larger than that of the low salient images ($t_{15} = 9.665$, $p < .001$, $\eta^2 = .862$). The results indicated that the bottom-up saliency map exhibited a positive cueing effect even when an image was subjectively invisible, which suggested that participants' attention was attracted to the salient region of the invisible image, so that they performed better in the valid cue condition than in the invalid cue condition. We also calculated the saliency index of the high salient and the low salient images based on the proposed formulation. The saliency index predicted the degree of the attentional attraction

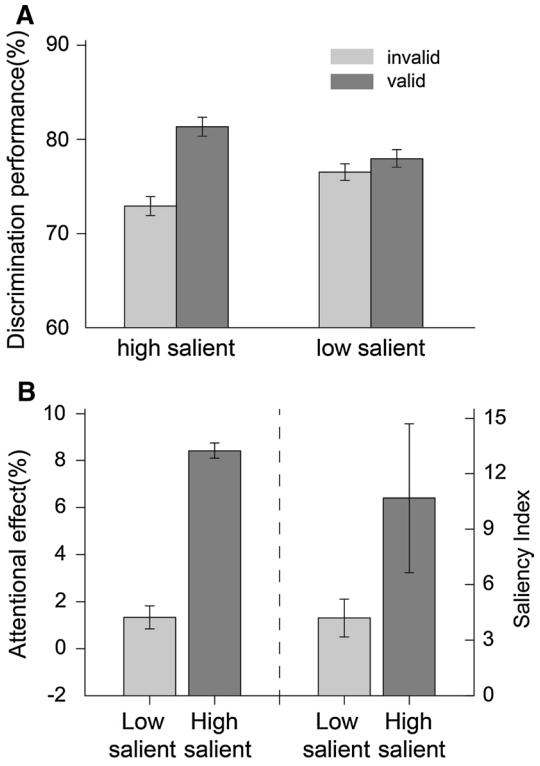


Fig. 3 Psychophysical results and computational saliency model prediction. **a** The *left two bars* and *right two bars* indicated the discrimination performance in the psychophysical experiment for the high salient and low salient images, respectively. *Light gray* indicated the performance of the invalid cue condition, while *dark gray* indicated the performance of the valid cue condition. *Error bars* denoted 1 SEM across participants for each condition. **b** *Left* The *left two bars* indicated the psychophysical attentional effects for the high salient and the low salient group, respectively. The attentional effect for each group was quantified as the difference between the orientation discrimination task performance of the valid cue condition and the invalid cue condition. *Error bars* denoted 1 SEM across participants for each condition. *Right* The *right two bars* indicated the saliency index calculated by a computational saliency model (Itti et al. 1998) for the high salient and the low salient group, respectively. The saliency index predicted the degree of attentional attraction. *Error bars* denoted 1 SD across all images in each group

of a bottom-up saliency map (right plane in Fig. 3b). Psychophysical data were consistent with the prediction from the computational model.

FMRI experiment

In the fMRI experiment, the percentages of correct detection were 50.1 ± .7 and 50.3 ± .7 % for the high salient and the low salient groups, respectively. The results were statistically indistinguishable from chance level (high salient: $t_{10} = .195$, $p = .850$, $\eta^2 = .004$; low salient: $t_{10} = .340$, $p = .741$, $\eta^2 = .011$). The behavioral data confirmed that the low luminance natural images were indeed subjectively invisible to participants. Contralateral and ipsilateral ROIs in LGN, V1–V4, and IPS were defined as the cortical areas that responded to retinal inputs in the salient region and its contralateral counterpart. LOC and FEF in two hemispheres could be activated equally well by stimuli presented in the left and the right visual fields in the localizer scan. Thus, instead of presenting data in ipsilateral and contralateral LOC and FEF, we directly analyzed event-related BOLD signals according to the degree of saliency (the high salient and the low salient groups) in these two cortical areas.

It was found that in V1–V4 and IPS, the natural images in both groups evoked larger BOLD signals in the contralateral ROIs compared with the ipsilateral ROIs (Fig. 4a). It meant that the salient region could evoke stronger neural activity than its contralateral counterpart. BOLD signal difference was quantified as the peak value difference of the BOLD signal in the contralateral ROI and that in the ipsilateral ROI (Fig. 4b). The BOLD signal differences of the high salient group and the low salient group were compared by paired *t* tests. We found that the BOLD signal difference of the high salient group was significantly higher than that of the low salient group ($t_{10} = 4.989$, $p = .001$, $\eta^2 = .713$; uncorrected) in V1. However, in LGN, V2–V4, and IPS, we did not observe any significant difference between the BOLD signal difference of the high salient group and that of the low salient group (LGN: $t_{10} = -.690$, $p = .506$, $\eta^2 = .045$; V2: $t_{10} = .194$, $p = .850$, $\eta^2 = .004$; V3: $t_{10} = -.159$, $p = .877$, $\eta^2 = .003$; V4: $t_{10} = -.125$, $p = .903$, $\eta^2 = .002$; IPS: $t_{10} = .540$, $p = .601$, $\eta^2 = .003$; uncorrected). Moreover, we also measured BOLD signal peak value difference between the high salient and the low salient groups in LOC and FEF. There was no significant difference between the high salient and the low salient groups in these two areas (LOC: $t_{10} = -.141$, $p = .891$, $\eta^2 = .002$; FEF: $t_{10} = -.690$, $p = .506$, $\eta^2 = .005$). As the attentional effect of the bottom-up saliency map of the high salient group was also significantly higher than that of the low salient group, these findings revealed that neural activity in V1 was parallel to the attentional effect.

Fig. 4 fMRI results. **a** Event-related BOLD signals averaged across participants in the contralateral and ipsilateral ROIS in LGN, V1–V4, and IPS. They were evoked by the bottom-up saliency map of natural images in the high salient and the low salient group. *Error bars* denoted 1 SEM calculated across participants at each time point. **b** Peak amplitude differences between the event-related BOLD signal in the contralateral ROI and that in the ipsilateral ROI in LGN, V1–V4, and IPS for the high salient and the low salient group. *Error bars* denoted 1 SEM calculated across participants. **c** *Left column* event-related BOLD signals averaged across participants in LOC and FEF. They were evoked by the bottom-up saliency map of natural images in the high salient and the low salient group. *Error bars* denoted 1 SEM calculated across participants at each time point. *Right column* Peak amplitude of the event-related BOLD signals in LOC and FEF for the high salient and the low salient group. *Error bars* denoted 1 SEM calculated across participants

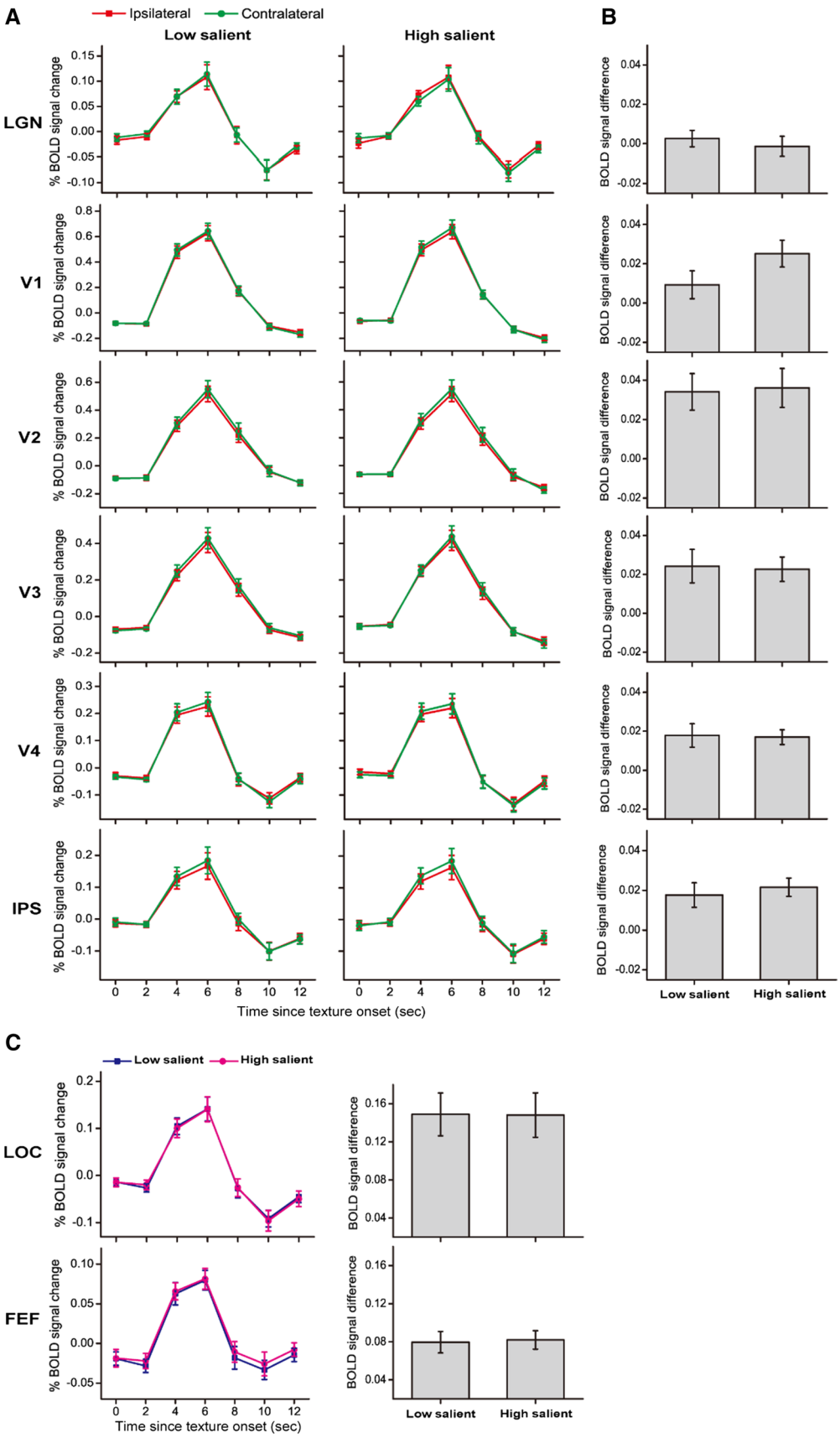
Correlation analysis

In order to further evaluate the role of neural activities of early visual cortical areas in realizing the bottom-up saliency map, we calculated correlation coefficients between our psychophysical and fMRI measures across individual participants. The attentional effect was significantly correlated with the BOLD signal difference in V1 for the high salient group ($r = .633$, $p < .05$), but not for the low salient group ($r = .372$, $p = .260$) (Fig. 5a). However, no significant correlation was found between the attentional effect and the BOLD signal difference in the other cortical areas (Fig. 5b). The results indicated a close relationship between the attentional effect and V1 neural activities.

Discussion

Using a modified cueing effect paradigm (Posner et al. 1980), we investigated the neural basis of the bottom-up saliency map of natural scenes. We found that even if participants were unaware of natural images, the attentional effect and the BOLD signal difference in V1 still increased with the degree of saliency. In addition, the attentional effect significantly correlated with the BOLD signal difference only in V1, but not other cortical areas. These findings suggest that the bottom-up saliency map of natural scenes is constructed in V1, thus providing a strong evidence to support the V1 theory (Li 1999, 2002).

The most interesting observation in our study is that we found V1 played an important role in creating the bottom-up saliency map of natural scenes. Our claim was based on the finding that neural activities (i.e., the BOLD signal difference) in V1 were closely correlated with the attentional attraction (i.e., bottom-up saliency) measured in the psychophysical experiment. Such a significant correlation was not found in other cortical areas. Specifically, neural activities induced by the bottom-up saliency map were not observed in LOC, IPS, or FEF, which indicated that


A
Ipsilateral
Contralateral
Low salient
High salient
LGN
V1
V2
V3
V4
IPS
% BOLD signal change
Time since texture onset (sec)
B
BOLD signal difference
Low salient
High salient
C
Low salient
High salient
LOC
FEF
% BOLD signal change
Time since texture onset (sec)
BOLD signal difference
Low salient
High salient

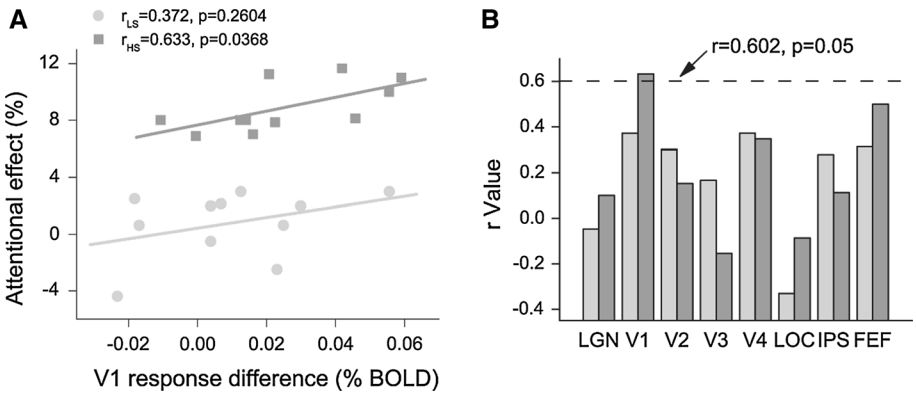


Fig. 5 Correlations between the psychophysical and the fMRI measures across individual participants. **a** Correlations between the attentional effect and the BOLD signal difference in V1 for the high salient (*dark gray*) and the low salient (*light gray*) group. **b** Correlation coefficients (the *r* values) between the attentional effect and the BOLD signal difference in LGN, V1–V4, LOC, IPS, and FEF for the high salient (*dark gray*) and the low salient (*light gray*) group

the observed neural activities in V1 were not attributed to signals feedback from these areas. It might be argued that the significant correlation between the attentional effect and the BOLD signal difference could be attributed to the larger variance of the computational saliency index values in the high salient group. Although the variance of the saliency index values was larger in the high salient group compared to the low salient group, the standard errors of the attentional effect were quite similar between these two groups. Moreover, the standard errors of the BOLD signal difference were also similar between these two groups (see Fig. 4b). Therefore, it is more likely that the significant correlation reflects a close link between the behavioral performance and the neural activities in V1, rather than an artifact caused by the large variance of the saliency index values in the high salient group.

We suggest that the underlying neural mechanism of our observation may be attributed to the lateral connections (Gilbert and Wiesel 1983) between V1 neurons. Our results are consistent with previous findings that the neural responses of V1 neurons were higher when their preferred stimuli popped out from background (Marcus and Van Essen 2002). More importantly, our observation supports the V1 saliency theory (Li 1999, 2002), which states that V1 creates the bottom-up saliency map. Our findings challenge the dominant view that the bottom-up saliency map is created in higher brain regions such as IPS and FEF (Koch and Ullman 1985; Geng and Mangun 2009; Bogler et al. 2011). It should be noted that we are not claiming that other cortical areas do not play a role in generating the bottom-up saliency map. However, their contributions are minor.

One important assumption of our study is that top-down signals were maximally reduced in our experiments. In the area of cognitive neuroscience, it has been proved and widely accepted that subjective awareness is determined by top-down signaling (Del Cul et al. 2007). Thus, rendering a stimulus invisible could maximally reduce top-down signals that evoked by stimuli. No matter in psychophysical or fMRI experiments, we found that the natural images were invisible to participants, which confirmed the assumption that top signals were maximally reduced in our experiments. It is quite important because several studies indicated that temporarily sluggish fMRI signals reflected neural activities resulting from both bottom-up and top-down processes, even in early visual cortex (Ress and Heeger 2003; Harrison and Tong 2009). Thus, blocking top-down signals make sure that we could observe a relative pure signal of the bottom-up saliency map of natural scenes in different brain regions. Some had argued that top-down information could modulate visual processing even when the stimuli were invisible (Dehaene et al. 1998; Jiang et al. 2006; Yang and Yeh 2011; Lin and Yeh 2015). Consistent with these studies, our study also found that invisible stimuli could generate an attentional effect even if top-down signals had been maximally reduced. However, a major distinction between these studies and our findings is that we aim to identify the brain area that creates the bottom-up saliency map, other than brain areas in which the representation of a saliency map could be observed. According to our results, we believe that neural activities in V1 could determine an early attentional selection even if participants are unaware of natural scenes. Meanwhile, we are not claiming that other areas could not represent a saliency map. These intermediate or higher brain areas are more likely to inherit or read out saliency signals from V1, rather than create a saliency map within themselves.

Our results extend evidence that supports the V1 theory from both psychophysical and physiological aspects. Instead of using textures consisted of simple oriented bars,

we used images contained natural scenes as our stimuli. Natural scenes contain richer naturalistic low-level features, including luminance, contrast, spatial frequency, curve, etc. These features are basic units that our visual system needs to deal with. An important concern of our study is that the observed differences in behavioral performance and neural activities between the high salient and the low salient group might be attributed to the difference in luminance, other than saliency, between these two groups (Betz et al. 2013). However, we found that there was no significant luminance difference between these two groups (see supplemental information). More importantly, our fMRI results showed that there was no significant difference between the BOLD signal difference of the high salient group and that of the low salient group in LGN, which is an area that highly sensitive to luminance. Therefore, it is more likely that the results found in V1 revealed a response to saliency rather than luminance. Our study is not only a critical complement to the previous study (Zhang et al. 2012), but also provides an important evidence for the V1 saliency map argument.

It is notable that our results seem to be in conflict with a recent study (Yoshida et al. 2012). The study found that attentional guidance over complex natural scenes was still preserved in the monkeys with blindsight from unilateral ablation of V1. However, in our experiments, we only investigated the role of cortical areas in creating the bottom-up saliency map. We did not exclude the possibility that some subcortical regions, such as pulvinar or superior colliculus, might be important in reading out the bottom-up saliency map. Moreover, there also exist direct connections between subcortical regions and FEF (Gilbert and Li 2013). Therefore, the observed attentional attraction with the absence of V1 might be induced by the bottom-up saliency signals constructed in subcortical regions, which is compatible with our results.

In conclusion, our findings suggest that V1 plays an important role in creating the bottom-up saliency map of natural scenes, providing further compelling evidence for the V1 theory, and challenging the dominant view that saliency map is constructed in higher cortical areas.

Acknowledgments This work was supported by MOST 2015CB351800, NSFC 61272027, NSFC 31230029, NSFC 31421003, and NSFC 61527804.

References

Asplund CL, Todd JJ, Snyder AP, Marois R (2010) A central role for the lateral prefrontal cortex in goal-directed and stimulus-driven attention. Nat Neurosci 13(4):507–514

Baluch F, Itti L (2011) Mechanisms of top-down attention. Trends Neurosci 34(4):210–224

Berman RA, Colby CL, Genovese CR, Voyvodic JT, Luna B, Thulborn KR, Sweeney JA (1999) Cortical networks subserving pursuit and saccadic eye movements in humans: an FMRI study. Hum Brain Mapp 8(4):209–225

Betz T, Wilming N, Bogler C, Haynes J, Konig P (2013) Dissociation between saliency signals and activity in early visual cortex. J Vis 13(14):1–12

Bogler C, Bode S, Haynes J (2011) Decoding successive computational stages of saliency processing. Curr Biol 21(19):1667–1671

Bruce N, Tsotsos J (2005) Saliency based on information maximization. NIPS 18:155–162

Buracas GT, Boynton GM (2002) Efficient design of event-related fMRI experiments using M-sequences. Neuroimage 16(3):801–813

Buschman TJ, Miller EK (2007) Top-down versus bottom-up control of attention in prefrontal and posterior parietal cortices. Science 315(5820):1860–1862

Carrasco M (2011) Visual attention: the past 25 years. Vis Res 51(13):1484–1525

Chen W, Zhu XH, Thulborn KR, Ugurbil K (1999) Retinotopic mapping of lateral geniculate nucleus in humans using functional magnetic resonance imaging. Proc Natl Acad Sci USA 96(5):2430–2434

Connolly JD, Goodale MA, Menon RS, Munoz DP (2002) Human fMRI evidence for the neural correlates of preparatory set. Nat Neurosci 5(12):1345–1352

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

Dehaene S, Naccache L, Le Clec'H G, Koechlin E, Mueller M, Dehaene-Lambertz G, Van de Moortele PF, Le Bihan D (1998) Imaging unconscious semantic priming. Nature 395(6702):597–600

Del Cul A, Baillet S, Dehaene S (2007) Brain dynamics underlying the nonlinear threshold for access to consciousness. PLoS Biol 5(10):e260

Engel SA, Glover GH, Wandell BA (1997) Retinotopic organization in human visual cortex and the spatial precision of functional MRI. Cereb Cortex 7(2):181–192

Fecteau JH, Munoz DP (2006) Salience, relevance, and firing: a priority map for target selection. Trends Cogn Sci 10(8):382–390

Fei-Fei L, Fergus R, Perona P (2004) Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories. In: IEEE CVPR, workshop on generative-model based vision

Geng JJ, Mangun GR (2009) Anterior intraparietal sulcus is sensitive to bottom-up attention driven by stimulus salience. J Cogn Neurosci 21(8):1584–1601

Gilbert CD, Li W (2013) Top-down influences on visual processing. Nat Rev Neurosci 14(5):350–363

Gilbert CD, Wiesel TN (1983) Clustered intrinsic connections in cat visual cortex. J Neurosci 3(5):1116–1133

Harrison SA, Tong F (2009) Decoding reveals the contents of visual working memory in early visual areas. Nature 458(7238):632–635

Hopfinger JB, Buonocore MH, Mangun GR (2000) The neural mechanisms of top-down attentional control. Nat Neurosci 3(3):284–291

Itti L, Koch C (2001) Computational modelling of visual attention. Nat Rev Neurosci 2(3):194–203

Itti L, Koch C, Niebur E (1998) A model of saliency based visual attention for rapid scene analysis. IEEE Trans Pattern Anal Mach Intell 20:1254–1259

Jiang Y, Costello P, Fang F, Huang M, He S (2006) A gender and sexual orientation-dependent spatial attentional effect of invisible images. Proc Natl Acad Sci USA 103(45):17048–17052

Kastner S, O'Connor DH, Fukui MM, Fehd HM, Herwig U, Pinsk MA (2004) Functional imaging of the human lateral geniculate nucleus and pulvinar. J Neurophysiol 91(1):438–448

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

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

Kourtzi Z, Kanwisher N (2000) Cortical regions involved in perceiving object shape. J Neurosci 20(9):3310–3318

Li Z (1999) Contextual influences in V1 as a basis for pop out and asymmetry in visual search. Proc Natl Acad Sci USA 96(18):10530–10535

Li Z (2002) A saliency map in primary visual cortex. Trends Cogn Sci 6(1):9–16

Li J, Levine MD, An X, Xu X, He H (2013) Visual saliency based on scale-space analysis in the frequency domain. IEEE Trans Pattern Anal Mach Intell 35(4):996–1010

Lin SY, Yeh SL (2015) Unconscious grouping of Chinese characters: evidence from object-based attention. Lang Linguist 16(4):517–533

Luna B, Thulborn KR, Strojwas MH, McCurtain BJ, Berman RA, Genovese CR, Sweeney JA (1998) Dorsal cortical regions subserving visually guided saccades in humans: an fMRI study. Cereb Cortex 8(1):40–47

Marcus DS, Van Essen DC (2002) Scene segmentation and attention in primate cortical areas V1 and V2. J Neurophysiol 88(5):2648–2658

Martin D, Fowlkes C, Tal D, Malik J (2001) A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Computer vision. IEEE ICCV, vol 2, pp 416–423

Mazer JA, Gallant JL (2003) Goal-related activity in V4 during free viewing visual search: evidence for a ventral stream visual salience map. Neuron 40(6):1241–1250

O'Connor DH, Fukui MM, Pinsk MA, Kastner S (2002) Attention modulates responses in the human lateral geniculate nucleus. Nat Neurosci 5(11):1203–1209

Paus T (1996) Location and function of the human frontal eye field: a selective review. Neuropsychologia 34(6):475–483

Posner MI, Snyder CRR, Davidson BJ (1980) Attention and the detection of signals. J Exp Psychol Gen 109(2):160–174

Ress D, Heeger DJ (2003) Neuronal correlates of perception in early visual cortex. Nat Neurosci 6(4):414–420

Serences JT, Yantis S (2007) Spatially selective representations of voluntary and stimulus-driven attentional priority in human occipital, parietal, and frontal cortex. Cereb Cortex 17(2):284–293

Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell RBH (1995) Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science 268(512):889–893

Shipp S (2004) The brain circuitry of attention. Trends Cogn Sci 8(5):223–230

Smith AM, Lewis BK, Ruttimann UE, Ye FQ, Sinnwell TM, Yang Y, Duyn JH, Frank JA (1999) Investigation of low frequency drift in fMRI signal. Neuroimage 9(5):526–533

Yang YH, Yeh SL (2011) Accessing the meaning of invisible words. Conscious Cogn 20(2):223–233

Yoshida M, Itti L, Berg D, Ikeda T, Kato R, Takaura K, White B, Munoz D, Isa T (2012) Residual attention guidance in blindsight monkeys watching complex natural scenes. Curr Biol 22(15):1429–1434

Zhang X, Zhaoping L, Zhou T, Fang F (2012) Neural activities in V1 create a bottom-up saliency map. Neuron 73(1):183–192

Zhaoping L, Zhe L (2015) Primary visual cortex as a saliency map: a parameter-free prediction and its test by behavioral data. PLoS Comput Biol 11(10):e1004375