Exp Brain Res (2013) 227:367–377
DOI 10.1007/s00221-013-3517-1

RESEARCH ARTICLE

The neural signature of spatial frequency-based information integration in scene perception

Tonglin Mu · Sheng Li

Received: 2 December 2012 / Accepted: 7 April 2013 / Published online: 19 April 2013

Abstract Spatial frequency-based information plays an important role in visual perception. By combining behavioral and electroencephalogram (EEG) measurements, we investigated the mechanisms of the interaction and information integration between different spatial frequency bands. The observers performed a scene categorization task on hybrid images that were generated by combining the low spatial frequency (LSF) component of one image with the high spatial frequency (HSF) component of another image. The results showed that the recognition of the HSF component was interfered by the non-attended LSF component at semantic level. The strength of the semantic interference was modulated by the physical similarity between the LSF and HSF components. Analyses of the EEG data revealed an early anterior N1 component (122 ms from stimulus onset) that was related to the observed interaction of the semantic and physical information between the LSF and HSF components. These findings demonstrate that the semantic information from different spatial frequency bands can be integrated at early stage of the perceptual processing. This early integration is likely to occur at frontal areas in order to initiate top-down facilitation.

T. Mu · S. Li (✉)
Department of Psychology, Peking University, 5 Yiheyuan Road, Haidian, Beijing 100871, China
e-mail: sli@pku.edu.cn

S. Li
Key Laboratory of Machine Perception, Ministry of Education, Peking University, Beijing 100871, China

S. Li
PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing 100871, China

Keywords Spatial frequency · Scene perception · Context effect · Congruency · ERP · Anterior N1

Introduction

Spatial frequency is one of the basic visual features that the visual system is specialized to analyze. A series of behavioral and physiological studies have demonstrated the existence of spatial frequency channels in early visual system (Blakemore and Campbell 1969; Campbell and Robson 1968; De Valois et al. 1982; Wilson and Wilkinson 1997). Anatomically, the segregation of two major visual pathways, the magnocellular pathway (M-pathway) and the parvocellular pathway (P-pathway), has been identified throughout the hierarchy of visual processing from retina to visual cortex. One of the functional differences between the two pathways is related to the processing of spatial frequency-based information. That is, the M-pathway is more sensitive to low spatial frequency (LSF) information, whereas the P-pathway is more specialized on the processing of high spatial frequency (HSF) information (Merigan and Maunsell 1993). More importantly, there were cumulating evidences from behavioral experiments that link such mechanism to the functions of higher level visual perception.

For example, previous literatures have suggested a coarse-to-fine principle that the processing of the fine scale information carried by HSF is proceeded by that of the coarse scale information carried by LSF (Bar 2003; Bullier 2001; Hegde 2008; Hughes et al. 1996; Parker et al. 1992, 1997; Schyns and Oliva 1994; Goffaux et al. 2011; Peyrin et al. 2010). Meanwhile, other theories beyond the coarse-to-fine principle have also been proposed. Particularly, there were increasing evidences showing that the

information from different spatial frequency bands can be used flexibly during the perception of scenes and objects. For example, by using hybrid images that combined the LSF component of one image with the HSF component of another image, researchers have demonstrated that sensitization to different spatial frequencies can influence the usage of the spatial frequency bands for visual categorization (Oliva and Schyns 1997; Schyns and Oliva 1999; Ozgen et al. 2006; Morrison and Schyns 2001). The visual system also takes the advantage of the diagnostic information from specific spatial frequency bands in a given task context (Rotshtein et al. 2010), such as the usefulness of the HSF information in subordinate categorization tasks (Collin 2006; Collin and McMullen 2005). Despite the inconsistency in the role played by the spatial frequency-based information in higher level visual perception, these proposals have suggested a parallel mechanism that agrees with the anatomical segregation of the M- and P-pathways in the visual system. However, the scenes and objects in natural environment never appear in a single frequency band. The information from different spatial frequency bands must be integrated at a certain level of visual processing before the moment of full recognition.

A recent behavioral study by Kihara and Takeda (2010) investigated the integration of multiple spatial frequency bands in scene perception. The study adopted a scene categorization task and tested observers' performance for different image types (intact, LSF, HSF, and LSF/HSF combined) and exposure durations (30–250 ms). Critically, the study compared the accuracy for the LSF/HSF combined image condition with the estimated accuracy based on the combination of the LSF image and HSF image conditions. The latter one reflected the behavioral performance when different spatial frequency bands (i.e., LSF and HSF) contributed independently without functional integration. The results showed higher accuracy for the LSF/HSF combined image condition and this advantage appeared only after about 100 ms from the stimulus onset, demonstrating the benefit of the information integration of the different spatial frequency bands. Therefore, the results of the study suggested that the integration begins as early as 100 ms after the stimulus onset in scene perception. Nevertheless, evidences from neurophysiology and brain imaging that support this proposal are still lacking.

The aim of the present study was to investigate the neural signature of the information integration between different spatial frequency bands in scene perception. We adopted a behavioral paradigm of scene categorization task on hybrid images (Oliva and Schyns 1997; Rotshtein et al. 2007, 2010; Schyns and Oliva 1994, 1999) and tested the semantic interference from the LSF component on the HSF component of the hybrid image. We chose the semantic interference as the evidence of the information integration between the two spatial frequencies due to two reasons. First, semantic interference can be directly measured in the scene categorization tasks, and it thus could serve as a reliable indicator of the information integration. Second, the behavioral and neural mechanism of semantic interference has been widely studies in the literature. The interpretation of the present findings can be benefited by comparing with previous results. The present study consisted of two experiments. In "Experiment 1", we tested the behavioral effect on the semantic interference. In "Experiment 2", we combined behavioral and electroencephalogram (EEG) measurements aiming to identify the neural signature for such information integration that has been revealed behaviorally in Kihara and Takeda's study.

Experiment 1

Methods

Observers

Ten naive observers (6 males, mean age: 23.18 ± 1.60) were participated in the experiment. All observers had normal or corrected to normal vision and gave written informed consent. The study was approved by the local ethics committee.

Stimuli

Image datasets from LabelMe project (Russell et al. 2008) were used in the experiments. The images belong to six semantic categories that can be grouped either as natural scenes or man-made scenes. The three natural scene categories were forests, coasts, and open countries. The three man-made scene categories were high ways, streets, and buildings. There are totally 1,096 natural images (326 forests, 360 coasts, and 411 open countries) and 1,216 man-made images (260 high ways, 292 streets, and 664 buildings) in the database. All images were resized to the resolution of 200 × 200 pixels and transformed to gray scale. The fixation point was at the center of the hybrid image. The stimuli were presented on the center of the screen with a gray background (intensity of 128). The size of the stimuli was 7.2° × 7.2°.

We defined physical similarity and semantic similarity between different images. The physical similarity between two images was defined by two criteria: the correlation of pixel-wise intensity and the correlation of energy distribution. The correlation of pixel-wise intensity of two images was the Pearson correlation coefficient between their corresponding pixel intensity values. For the correlation of

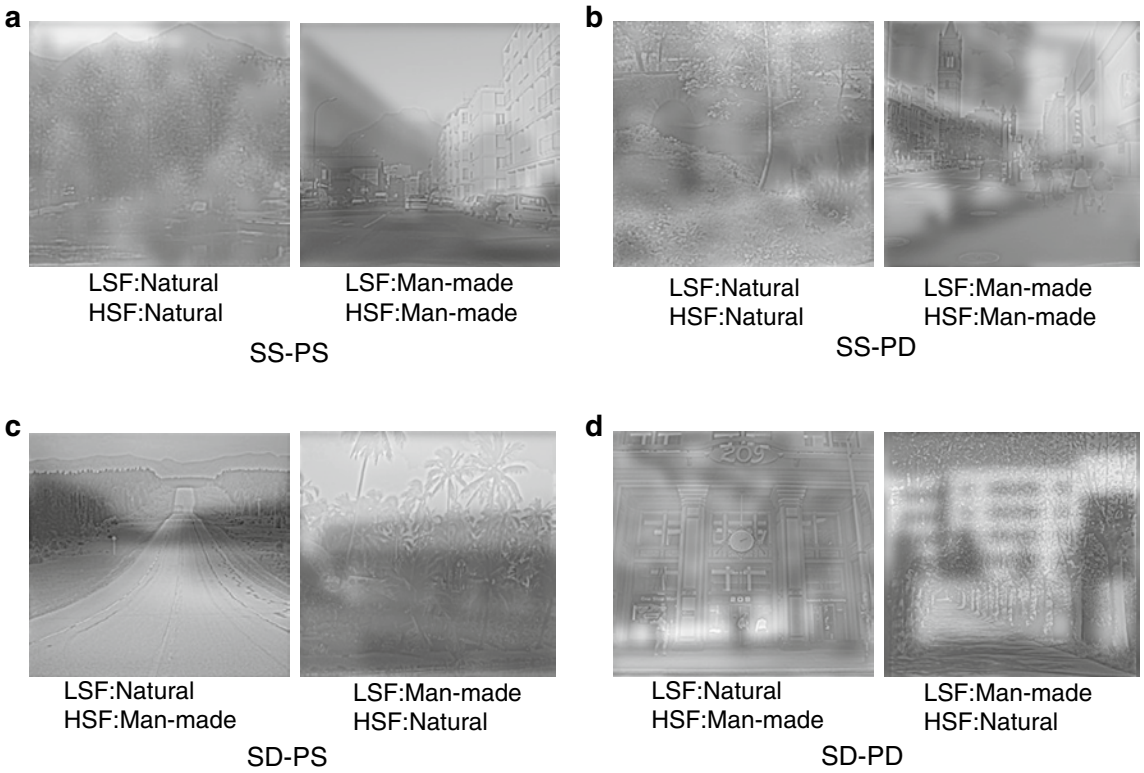

Fig. 1 Examples of hybrid images for the four stimulus conditions. Two hybrid images (one natural scene in HSF and one man-made scene in HSF) are displayed for each condition. **a** SS–PS condition. **b** SS–PD condition. **c** SD–PS condition. **d** SD–PD condition. *SS–PS* semantically and physically similar, *SS–PD* semantically similar and physically dissimilar, *SD–PS* semantically dissimilar and physically similar, *SD–PD* semantically and physically dissimilar

energy distribution, for each pixel, its orientation was defined as the angle between the line connecting the pixel to the fixation and the line connecting the fixation to the mid-point of the upper edge of the image. The intensity of each pixel reflected the energy of the pixel. For each orientation, the energy value was obtained by averaging the intensity values for the pixels along this orientation. The coefficient value for two images was calculated as the correlation between the energy values along all orientations (0°–359°) of the two images. The Pearson correlation coefficient between the corresponding mean energy values across orientations of two images was defined as the correlation of energy distribution. An image pair was defined as physically similar (PS) if both correlation coefficient values were higher than 0.6 (top 5 % in the dataset) and physically dissimilar (PD) image if both correlation coefficient values were <0.05 (bottom 5 % in the dataset). We defined two images as semantically similar (SS) if they were from the same category (e.g., both images were from forest category) and semantically dissimilar (SD) if they were from different groups (e.g., one image was from the natural scene group, and the other one was from the man-made scene group).

All images were transformed into spatial frequency domain using a Fast Fourier Transform algorithm and filtered using 2D Gaussian filters. The cutoff frequency of the high pass filter was 3.3 cycle/°, and the cutoff frequency of the low pass filter was 1.1 cycle/° (Rotshtein et al. 2007, 2010; Skottun 2000). The signal decay is 6 dB for both the high pass and low pass filters. Root mean square (RMS) contrast normalization was applied to both the LSF and HSF images (mean intensity: 128; SD: 40).

We generated four types of hybrid images (Oliva and Schyns 1997; Schyns and Oliva 1994) by combining the LSF component of one image with the HSF component of another image (Fig. 1). The two images can be (1) SS and PS (SS–PS), (2) SS but PD (SS–PD), (3) SD but PS (SD–PS), and (4) SD and PD (SD–PD). Finally, we generated 929 hybrid images for the SS–PS condition, 955 hybrid images for the SS–PD condition, 410 hybrid images for the SD–PS condition, and 590 hybrid images for the SD–PD condition. There were possible repetitions of the scene images. Our image selection procedure ensured that a scene image appeared only once under the SS condition or under the SD condition, but it is possible that an image appeared in both SS and SD conditions.

Design

Observers were instructed to attend to the HSF component of the hybrid images (what was drawn with lines) and make a judgment on whether each image was a natural scene or a man-made scene by pressing one of two buttons.[1] The purpose of the study was to investigate the spatial frequency-based information integration by testing the semantic interference of the LSF component over the HSF component. The coarse-to-fine principle was taken into account when we chose this order rather than the opposite one (i.e., the interference of the HSF over LSF). According to the coarse-to-fine principle, the LSF information was processed precede that of the HSF information. Therefore, we would expect that when the observer was paying attention to the HSF component of an image, the LSF information on the same image would be automatically processed. In this way, we should observe semantic interference effect between the two spatial frequency bands.

Each trial lasted for 2,500 ms. A trial started with a stimulus presentation for 120 ms (long exposure condition) or 30 ms (short exposure condition) followed by a 150-ms mask. Previous studies have suggested that the visual system may process the spatial frequency components differentially under different exposure times. Schyns and Oliva (1994) have demonstrated that subjects were more likely to recognize the image in LSF under short exposure time and the image in HSF under long exposure time. We used two exposure times in "Experiment 1" aiming to investigate whether the exposure time had effect on the spatial frequency-based information integration. The experiment consisted of four blocks: two short exposure blocks and two long exposure blocks. Four stimulus conditions were presented in a random order within each block. Each condition had 50 trials in each block, and there were 800 trials in total.

Results

Observers' behavioral performance is shown in Fig. 2. The values of accuracy and response time for each condition are shown in Table 1. A repeated measures ANOVA (exposure time × semantic similarity × physical similarity) on accuracy showed that the accuracy for the long exposure condition was significantly higher than the accuracy for the short exposure condition [$F(1,9) = 63.06$, $p < 0.001$]. There was also main effects for the semantic similarity [SS > SD, $F(1,9) = 20.21$, $p = 0.001$] and the physical similarity [PS < PD, $F(1,9) = 60.00$, $p < 0.001$]. We also observed significant three-way interaction [$F(1,9) = 7.04$, $p < 0.05$] and therefore analyzed the data separately for the long and short exposure conditions.

For the long exposure condition (Fig. 2a), the ANOVA analysis revealed significant effects for the semantic similarity [SS > SD, $F(1,9) = 41.71$, $p < 0.001$], the physical similarity [PS < PD, $F(1,9) = 27.23$, $p = 0.001$], and their interaction [$F(1,9) = 10.04$, $p < 0.05$]. Planned comparison analysis revealed that the accuracy of the SS condition was significantly higher than that of the SD condition under the PS condition [$F(1,9) = 71.8$, $p < 0.001$] and significantly higher than that of the SD condition under the PD condition [$F(1,9) = 9.61$, $p < 0.05$]. To further determine the interference from the non-attended LSF information under different physical similarity conditions, we defined a semantic congruency index as the difference between accuracies in the SD and SS conditions under each physical similarity condition. The analysis revealed that the semantic congruency index under PS condition is significantly higher than that of the PD condition [$t(9) = 8.52$, $p < 0.001$]. This result indicates that the non-attended LSF component exhibits stronger semantic interference if its physical property is similar to that of the HSF component.

For the short exposure condition (Fig. 2c), the ANOVA revealed significant effect of physical similarity [PS < PD, $F(1,9) = 53.39$, $p < 0.001$] and its interaction with semantic similarity [$F(1,9) = 47.96$, $p < 0.001$], but not for the semantic similarity [$F(1,9) = 2.41$, $p = 0.16$]. Planned comparison analysis revealed that the accuracy of the SS condition was significantly higher than that of the SD condition under the PS condition [$F(1,9) = 26.19$, $p < 0.001$] and marginally lower than that of the SD condition under the PD condition [$F(1,9) = 4.97$, $p = 0.053$]. Semantic interference analysis revealed that the semantic congruency index under the PS condition is significantly higher than that of the PD condition [$t(9) = 5.83$, $p < 0.001$].

The absence of the semantic similarity effect for the short exposure condition was likely due to the higher accuracy for the SD compared with the SS condition under the PD condition (marginally significant, $p = 0.053$). However, as both the long exposure and short exposure conditions exhibited significant interaction effects and semantic congruency effects, we would suggest that these two exposure conditions exhibited similar effects of semantic interference.

[1] We have conducted a control experiment on a categorization task in which 14 observers were instructed to attend the LSF component of the hybrid images. The results showed that the accuracies of the two semantically different conditions (i.e., SD–PS [$t(13) = -0.36$, $p = 0.07$] and SD–PD [$t(13) = -1.57$, $p < 0.05$]) were not significantly higher than chance level (the accuracy of SD–PD condition was actually significantly <50 %). The results suggested that the observers were not able to correctly categorize the scenes based solely upon the LSF information in a hybrid image given the interference from the HSF component.

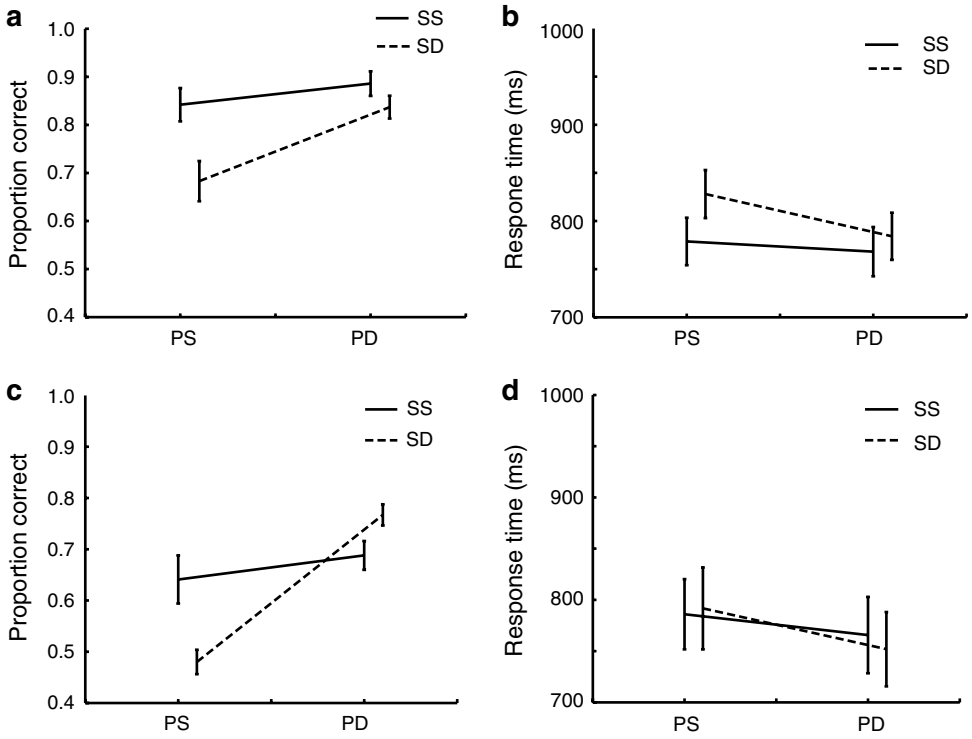

Fig. 2 Behavioral results of the scene categorization task in "Experiment 1". **a** The mean accuracy for the long exposure condition. **b** The mean response time of the correct trials for the long exposure condition. **c** The mean accuracy for the short exposure condition. **d** The mean response time of the correct trials for the short exposure condition. *Error bars* represent the standard error of means across observers

Table 1 Behavioral accuracy and response time in "Experiment 1"

	SS–PS	SS–PD	SD–PS	SD–PD
Long exposure				
Accuracy (±SEM)	84.2 % (±3.5 %)	88.6 % (±2.6 %)	68.3 % (±4.2 %)	83.7 % (±2.4 %)
Response time (±SEM)	779 (±25) ms	768 (±26) ms	828 (±25) ms	784 (±25) ms
Short exposure				
Accuracy (±SEM)	64.1 % (±4.7 %)	68.8 % (±2.8 %)	48 % (±2.4 %)	76.7 % (±2.1 %)
Response time (±SEM)	786 (±34) ms	766 (±37) ms	792 (±40) ms	752 (±36) ms

We also recorded observers' response time during the categorization task (Fig. 2b, d). A repeated measures ANOVA (exposure time × semantic similarity × physical similarity) on response time of the correct trials revealed no significant three-way interaction [$F(1,9) = 0.325$, $p = 0.58$]. There were significant main effects on semantic similarity [SS < SD, $F(1,9) = 9.36$, $p < 0.05$], physical similarity [PS > PD, $F(1,9) = 24.39$, $p < 0.001$], and their interaction [$F(1,9) = 13.26$, $p < 0.01$]. For the long exposure condition (Fig. 2b), the ANOVA analysis revealed significant effects for the semantic similarity [SS < SD, $F(1,9) = 18.82$, $p < 0.01$], physical similarity [PS > PD, $F(1,9) = 21.27$, $p < 0.01$], but not their interaction [$F(1,9) = 3.80$, $p = 0.19$]. For the short exposure condition (Fig. 2d), the ANOVA analysis revealed significant effects for the physical similarity [$F(1,9) = 11.00$, $p < 0.05$], but not for the semantic similarity [$F(1,9) = 0.28$, $p = 0.61$] and their interaction [$F(1,9) = 1.30$, $p = 0.30$].

These results mirrored the behavioral accuracy results by demonstrating that the recognition of the HSF component was interfered by the non-attended LSF component as the semantic and physical information interacted between the two components. Short stimulus presentation time and low behavioral accuracy were also known to reduce the EEG signal to noise ratio. To maintain the EEG signal level, we used only the long exposure time in "Experiment 2".

Experiment 2

Methods

Observers

Eighteen naive observers (9 males, mean age: 21.65 ± 2.42) were participated in the experiment. All

observers had normal or corrected to normal vision and gave written informed consent. The study was approved by the local ethics committee.

Stimuli

In "Experiment 2", a new stimulus condition was added. Under this condition, a stimulus only contains the HSF component of a scene image (HSF-only). The HSF-only condition served as a baseline condition in which no semantic interference occurred during the categorization task.

Design

Observers' task in "Experiment 2" was identical to "Experiment 1". The length of each trial ranged from 2,000 to 3,000 ms. A trial started with stimulus presentation for 120 ms followed by an 80-ms mask. Observers were instructed to attend to the HSF component of the hybrid images and make a judgment on whether each image was a natural scene or a man-made scene by pressing one of two buttons.

EEG recording and analysis

EEG data were acquired from a 64-channel EEG cap (Brain Products, Munich, Germany). Electrode impedance was kept below 5 kΩ for scalp channels. Electrooculograms (EOGs) were recorded with electrodes placed lateral to the external canthi of the left eye and above the right eye to capture ocular activity. An external electrode placed on the tip of nose was set to be the reference. The electrode AFz was chosen to be the ground. EEG data were recorded at a sampling rate of 1,000 Hz and were re-referenced offline with the mean of the left and right mastoids. EEG data were epoched and baseline corrected by subtracting the average of the prestimulus (200 ms) data. The baseline corrected data were filtered with a finite impulse response (FIR) filter (pass band cutoff frequency: 0.016 Hz; stop band cutoff frequency: 100 Hz; decay of stop band: 24 dB/octave). Event-related potentials (ERPs) were obtained by averaging the filtered data based on conditions and electrodes. The group peak point of an ERP component was identified from the averaged curve across observers. For each observer, the peak point of the ERP component was identified as the point with the maximum value within a 60 ms time window centered at the group peak point. The amplitude of the ERP component for each observer was then calculated by averaging eleven time points centered at the observer's peak point.

Behavioral results

The values of accuracy and response time in "Experiment 2" are shown in Table 2. As shown in Fig. 3a, a repeated measures ANOVA (semantic similarity × physical similarity) on accuracy showed significant effects for the semantic similarity [SS > SD, $F(1,17) = 37.41$, $p < 0.001$], the physical similarity [PS < PD, $F(1,17) = 42.10$, $p < 0.001$], and their interaction [$F(1,17) = 10.39$, $p < 0.01$]. Planned comparison analysis revealed that the accuracy of the SS condition was significantly higher than that of the SD condition under the PS condition [$F(1,17) = 11.96$, $p < 0.01$]. Under the PD condition, the accuracies of the SS and SD conditions were not significantly different [$F(1,17) = 0.06$, $p = 0.82$]. The analysis also showed that the semantic congruency index under the PS condition is significantly higher

Table 2 Behavioral accuracy and response time in "Experiment 2"

	SS–PS	SS–PD	SD–PS	SD–PD
Accuracy (±SEM)	79.8 % (±2 %)	86.1 % (±2.2 %)	61.5 % (±4.3 %)	79.2 % (±2.6 %)
Response time (±SEM)	849 (±24) ms	840 (±25) ms	888 (±31) ms	851 (±23) ms

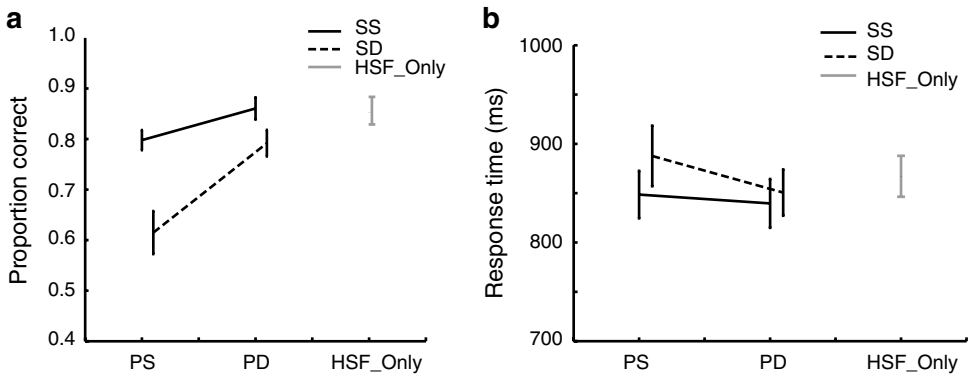

Fig. 3 Behavioral results of the scene categorization task in "Experiment 2". **a** The mean accuracy. **b** The mean response time of the correct trials. *Error bars* represent the standard error of means across observers

than that of the PD condition [$t(17) = 3.22$, $p < 0.01$]. These results replicated the findings from "Experiment 1", showing that the recognition of the HSF scenes was interfered by the non-attended LSF information. Similar physical property between the HSF and LSF information led to stronger semantic interference.

For all correct trials, a repeated measures ANOVA (semantic similarity × physical similarity) on response time showed significant main effects on semantic similarity [SS < SD, $F(1,17) = 22.83$, $p < 0.001$] and physical similarity (PS > PD, $F(1,17) = 8.42$, $p < 0.01$], but not their interaction [$F(1,17) = 2.94$, $p = 0.11$].

EEG results

For the EEG data, we concentrated our analyses on three groups of electrodes: the frontal electrodes (Fz, F1, F2, FCz, FC1, FC2), the parieto-occipital electrodes (P1, P2, P3, P4, PO3, PO4), and the occipital electrodes (O1, O2, Oz). For each observer, the averaged ERP data from all electrodes in each group were used in statistical analyses.

For the frontal electrodes, there was a N1 component centered at 122 ms after the stimulus onset (Fig. 4a). A repeated measures ANOVA (semantic similarity × physical similarity) showed significant interaction effect [$F(1,17) = 5.05$, $p < 0.05$] for the anterior N1 component (Fig. 4b). There was a marginally significant main effect for the semantic similarity [$F(1,17) = 4.07$, $p = 0.06$] but no significant main effect for the physical similarity [$F(1,17) = 0.00$, $p = 0.97$] on this anterior N1 component. Analysis on the semantic congruency index revealed similar effect as from the behavioral results. The index under the PS condition is significantly higher than that of the PD condition [$t(17) = 2.25$, $p < 0.05$]. Furthermore, the amplitude of the N1 component under the HSF-only condition was significantly lower than that of the hybrid conditions [averaged across four conditions, $t(17) = 3.65$, $p < 0.01$], suggesting that the observed interaction effect of N1 component for the hybrid images was due to the interference from the non-attended LSF component.

For the parieto-occipital electrodes, we found significant main effect for the semantic similarity [$F(1,17) = 5.85$, $p < 0.05$] for P2 component (latency: 247 ms, Fig. 4c, d). There were neither significant effect for the physical similarity [$F(1,17) = 0.71$, $p = 0.41$] nor for the interaction [$F(1,17) = 0.43$, $p = 0.52$]. We did not find any significant effect for the occipital electrodes before 300 ms after the stimulus onset. Previous studies have suggested that ERP component around 200 ms after the stimulus onset in parieto-occipital areas is modulated by higher visual processing such as object recognition and scene perception (Codispoti et al. 2006; VanRullen and Thorpe 2001). The observed semantic similarity effect in this area was consistent with previous findings and suggested that the processing of the categorical information in higher visual areas was preceded by the frontal analyses of the semantic information. This finding agrees with the proposal that top-down influence plays an important role in scene perception.

For the occipital electrodes (Fig. 4e, f), we found significant main effect for the semantic similarity [$F(1,17) = 9.21$, $p < 0.01$] for the P3 component (latency: 344 ms), but not for the physical similarity [$F(1,17) = 1.77$, $p = 0.20$] or their interaction [$F(1,17) = 0.52$, $p = 0.48$]. It was possible that this late significant semantic effect was due to feedback modulation. However, further investigations are required to confirm this interpretation.

There was possibility that the observed early semantic effect in frontal electrodes was related to the final decision made by the observers. We therefore averaged the ERP data based on the behavioral choices (natural vs. man-made) for each observer and identified the component that is correlated with observer's categorization decision (Fig. 5). The analysis revealed no significant difference between two choice conditions before 282 ms from the stimulus onset. Analysis on the amplitude of P3 component of the frontal electrodes showed significant difference between natural and man-made choices [$t(1,17) = 2.71$, $p < 0.05$]. This result is consistent with previous literatures showing that the P3 component reflects the behavioral choices in perceptual decision tasks (Johnson and Olshausen 2003; Philiastides and Sajda 2006).

Discussion

Our study investigated the behavioral and neural signatures of the information integration between the LSF and HSF components in scene perception. The behavioral results showed significant interaction between semantic and physical similarities of the LSF and HSF components. Meanwhile, the ERP results revealed an early anterior N1 component in frontal area that corresponded to the observed behavioral effect. The findings advance our understanding of spatial frequency-based information integration in scene perception.

Our results demonstrate that, although being irrelevant to the task, both the semantic and physical information of the non-attended LSF component of a scene image are processed when observers make perceptual judgment. The analyses of the semantic similarity revealed significant interference from semantically incongruent and non-attended LSF information (Rotshtein et al. 2010). Particularly, the interference was stronger if the physical properties (i.e., the image statistics) of the LSF component are similar to that of the attended HSF component. These results demonstrate that, despite the nature of task irrelevance and the

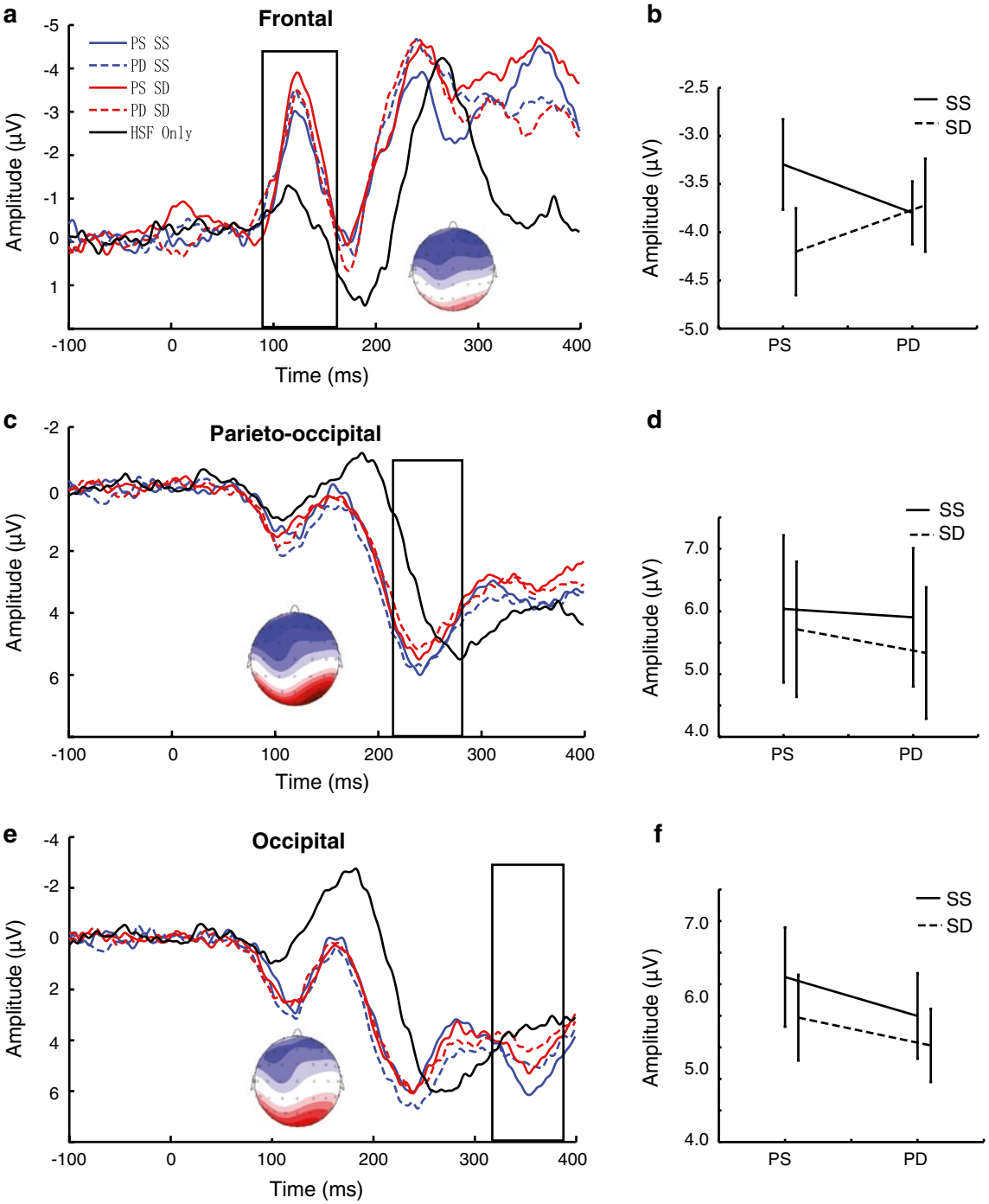

Fig. 4 ERP results in "Experiment 2". ERP *curved* were averaged across observers based on stimulus conditions. **a** The ERP of the frontal electrodes (Fz, F1, F2, FCz, FC1, FC2). The N1 component (centered at 122 ms after the stimulus onset) was marked with the *black rectangle*. **b** The amplitude of the N1 component in frontal electrodes for different conditions. **c** The ERP of the parieto-occipital electrodes (P1, P2, P3, P4, PO3, PO4). The P2 component (centered at 247 ms after the stimulus onset) was marked with the *black rectangle*. **d** The amplitude of the P2 component in parieto-occipital electrodes for different conditions. **e** The ERP of the occipital electrodes (O1, O2, Oz). The P3 component (centered at 344 ms after the stimulus onset) was marked with the *black rectangle*. **f** The amplitude of the P3 component in occipital electrodes for different conditions. The geographical *map* next to the ERP *curves* is the scalp distribution of the selected component in its defined time window. *Error bars* represent the standard error of means

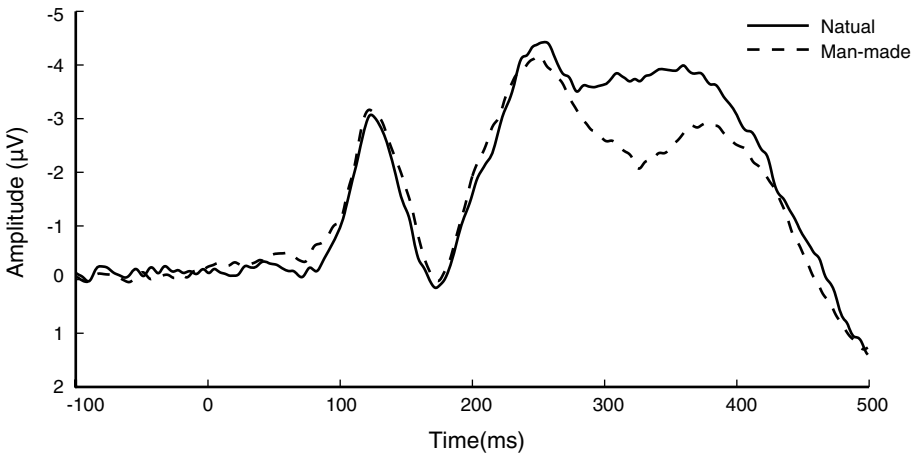

Fig. 5 ERP results based on the behavioral choices from the frontal electrodes. We averaged the ERP data based on observers' behavioral choices (*solid curve* for the natural group vs. *dashed curve* for the man-made group). The two *curves* showed significant difference in amplitude from 282 ms after the stimulus onset

lack of focused attention, both the semantic and physical information of the LSF component are processed and cannot be voluntarily ignored. The information from different spatial frequency bands can be integrated even if one of the spatial frequency bands is manipulated to be task irrelevant. Our results also indicate that both the physical information at sensory level and the semantic information at decision level are required for the spatial frequency-based information integration in scene perception.

Furthermore, our ERP results provide supporting evidence that the neural signature of the spatial frequency-based information integration appears around 120 ms from the stimulus onset. Previous behavioral study has shown that the integration of spatial frequency-based information occurs around 100 ms after the stimulus onset (Kihara and Takeda 2010). However, there was no evidence in terms of neural correlates supporting this proposal. In our study, the earliest ERP component that shows significant interaction between the semantic and physical similarities is the N1 peak of the frontal electrodes. The information comparison and integration between different spatial frequency bands are the necessary processes underlying this interaction effect. The results provide direct evidence that the integration of information from different spatial frequency bands can occur at the early stages of perceptual processing at frontal areas in scene perception.

Contextual congruency effect is a well-observed phenomenon in scene perception (Davenport and Potter 2004). The corresponding ERP component of such effect was suggested to relate to N400 sentence congruency effect (Ganis and Kutas 2003) or earlier semantic effect around 300 ms after the stimulus onset (Mudrik et al. 2010). These studies used incongruent objects in a picture (e.g., a player holds a watermelon in a basketball game) and examined the latency of the congruency effect with EEG recording. The results suggested that the earliest congruency effect was around 300 ms after the stimulus onset and related to semantic integration that was activated in later processing stages. Our results, however, suggest that the semantic congruency effect in scene perception can occur as early as 122 ms from the stimulus onset. The discrepancy between our results and previous findings could be due to the different experimental paradigms. But more importantly, the non-attended component of a scene image was in its LSF band in our study, whereas both the objects and backgrounds in previous studies were not band filtered. The lower ERP amplitude of the HSF-only condition compared with the hybrid conditions also indicates the interference from the LSF component during the recognition. The earlier congruency ERP component in our study suggests that the integration of spatial frequency-based information helps to facilitate the perceptual processing in order to support fast decisions.

Bar and colleagues (Bar et al. 2006; Bar 2003) have suggested that visual perception is facilitated by top-down influence. Particularly, Bar's model proposed that the LSF component of a natural image is processed firstly by frontal cortex, in order to provide an initial guess that guides the subsequent processing of detailed HSF information in sensory areas. This proposal was supported by a recent study (Peyrin et al. 2010) that revealed early (140–160 ms) frontal ERP activation when compared LSF–HSF with HSF–LSF image sequences. This ERP activation was accompanied with increase fMRI activity in frontal cortex under the same comparison, suggesting an early influence of frontal areas during coarse-to-fine analysis of scenes. Our results support the idea that top-down signal plays an important role in scene perception. The earlier semantic effect at the frontal area and the later semantic effect at the parieto-temporal area are consistent with the proposal of top-down facilitation in visual perception. However, our results demonstrate that both the LSF and HSF semantic information have been processed at the frontal area around 120 ms from the stimulus onset. A possible explanation for this

seemingly contradiction is that the observers in our experiment were to pay attention to the HSF component of the hybrid images. As attention is known to enhance and accelerate visual information processing (Carrasco and McElree 2001), it is possible that the semantic information of the HSF component can be processed in details as early as the LSF information given specific task requirement. Further investigations are required to clarify this issue.

In conclusion, our study provides direct evidence for the neural signature of the integration of spatial frequency-based information. The semantic and physical information from the LSF and HSF bands can be integrated as early as around 120 ms after the stimulus onset at frontal areas. This early integration can benefit the recognition performance by facilitating the subsequent perceptual analyses through top-down influences.

Acknowledgments We thank Man Song for help with the EEG data analysis. This work was supported by the National Natural Science Foundation of China (No. 31271081, 31230029, 31070896), the National High Technology Research and Development Program of China (863 Program) (No. 2012AA011602), and the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry.

References

Bar M (2003) A cortical mechanism for triggering top-down facilitation in visual object recognition. J Cogn Neurosci 15(4):600–609. doi:10.1162/089892903321662976

Bar M, Kassam KS, Ghuman AS, Boshyan J, Schmid AM, Dale AM, Hamalainen MS, Marinkovic K, Schacter DL, Rosen BR, Halgren E (2006) Top-down facilitation of visual recognition. Proc Natl Acad Sci USA 103(2):449–454. doi:10.1073/pnas.0507062103

Blakemore C, Campbell FW (1969) On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. J Physiol 203(1):237–260

Bullier J (2001) Integrated model of visual processing. Brain Res Brain Res Rev 36(2–3):96–107

Campbell FW, Robson JG (1968) Application of Fourier analysis to the visibility of gratings. J Physiol 197(3):551–566

Carrasco M, McElree B (2001) Covert attention accelerates the rate of visual information processing. Proc Natl Acad Sci 98(9):5363–5367. doi:10.1073/pnas.081074098

Codispoti M, Ferrari V, Junghofer M, Schupp HT (2006) The categorization of natural scenes: brain attention networks revealed by dense sensor ERPs. Neuroimage 32(2):583–591. doi:10.1016/j.neuroimage.2006.04.180

Collin CA (2006) Spatial-frequency thresholds for object categorisation at basic and subordinate levels. Perception 35(1):41–52

Collin CA, McMullen PA (2005) Subordinate-level categorization relies on high spatial frequencies to a greater degree than basic-level categorization. Percept Psychophys 67(2):354–364

Davenport JL, Potter MC (2004) Scene consistency in object and background perception. Psychol Sci 15(8):559–564. doi:10.1111/j.0956-7976.2004.00719.xPSCI719

De Valois RL, Albrecht DG, Thorell LG (1982) Spatial frequency selectivity of cells in macaque visual cortex. Vis Res 22(5):545–559

Ganis G, Kutas M (2003) An electrophysiological study of scene effects on object identification. Brain Res Cogn Brain Res 16(2):123–144

Goffaux V, Peters J, Haubrechts J, Schiltz C, Jansma B, Goebel R (2011) From coarse to fine? Spatial and temporal dynamics of cortical face processing. Cereb Cortex 21(2):467–476. doi:10.1093/cercor/bhq112

Hegde J (2008) Time course of visual perception: coarse-to-fine processing and beyond. Prog Neurobiol 84(4):405–439. doi:10.1016/j.pneurobio.2007.09.001

Hughes HC, Nozawa G, Kitterle F (1996) Global precedence, spatial frequency channels, and the statistics of natural images. J Cogn Neurosci 8(3):197–230. doi:10.1162/jocn.1996.8.3.197

Johnson JS, Olshausen BA (2003) Timecourse of neural signatures of object recognition. J Vis 3(7):499–512. doi:10:1167/3.7.43/7/4

Kihara K, Takeda Y (2010) Time course of the integration of spatial frequency-based information in natural scenes. Vis Res 50(21):2158–2162. doi:10.1016/j.visres.2010.08.012

Merigan WH, Maunsell JH (1993) How parallel are the primate visual pathways? Annu Rev Neurosci 16:369–402. doi:10.1146/annurev.ne.16.030193.002101

Morrison DJ, Schyns PG (2001) Usage of spatial scales for the categorization of faces, objects, and scenes. Psychon Bull Rev 8(3):454–469

Mudrik L, Lamy D, Deouell LY (2010) ERP evidence for context congruity effects during simultaneous object-scene processing. Neuropsychologia 48(2):507–517. doi:10.1016/j.neuropsychologia.2009.10.011

Oliva A, Schyns PG (1997) Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. Cogn Psychol 34(1):72–107. doi:10.1006/cogp.1997.0667

Ozgen E, Payne HE, Sowden PT, Schyns PG (2006) Retinotopic sensitisation to spatial scale: evidence for flexible spatial frequency processing in scene perception. Vis Res 46(6–7):1108–1119. doi:10.1016/j.visres.2005.07.015

Parker DM, Lishman JR, Hughes J (1992) Temporal integration of spatially filtered visual images. Perception 21(2):147–160

Parker DM, Lishman JR, Hughes J (1997) Evidence for the view that temporospatial integration in vision is temporally anisotropic. Perception 26(9):1169–1180

Peyrin C, Michel CM, Schwartz S, Thut G, Seghier M, Landis T, Marendaz C, Vuilleumier P (2010) The neural substrates and timing of top-down processes during coarse-to-fine categorization of visual scenes: a combined fMRI and ERP study. J Cogn Neurosci 22(12):2768–2780. doi:10.1162/jocn.2010.21424

Philiastides MG, Sajda P (2006) Temporal characterization of the neural correlates of perceptual decision making in the human brain. Cereb Cortex 16(4):509–518. doi:10.1093/cercor/bhi130

Rotshtein P, Vuilleumier P, Winston J, Driver J, Dolan R (2007) Distinct and convergent visual processing of high and low spatial frequency information in faces. Cereb Cortex 17(11):2713–2724. doi:10.1093/cercor/bhl180

Rotshtein P, Schofield A, Funes MJ, Humphreys GW (2010) Effects of spatial frequency bands on perceptual decision: it is not the stimuli but the comparison. J Vis 10(10):25. doi:10.1167/10.10.25

Russell BC, Torralba A, Murphy KP, Freeman WT (2008) LabelMe: a database and web-based tool for image annotation. Int J Comput Vis 77(1–3):157–173. doi:10.1007/s11263-007-0090-8

Schyns PG, Oliva A (1994) From blobs to boundary edges—evidence for time-scale-dependent and spatial-scale-dependent scene recognition. Psychol Sci 5(4):195–200

Schyns PG, Oliva A (1999) Dr. Angry and Mr. Smile: when categorization flexibly modifies the perception of faces in rapid visual presentations. Cognition 69(3):243–265

Skottun BC (2000) The magnocellular deficit theory of dyslexia: the evidence from contrast sensitivity. Vis Res 40(1):111–127

VanRullen R, Thorpe SJ (2001) The time course of visual processing: from early perception to decision-making. J Cogn Neurosci 13(4):454–461

Wilson HR, Wilkinson F (1997) Evolving concepts of spatial channels in vision: from independence to nonlinear interactions. Perception 26(8):939–960