



Electrophysiological correlates of top-down effects facilitating natural image categorization are disrupted by the attenuation of low spatial frequency information



Adrienn Aranka Rokszin^a, Dóra Győri-Dani^a, László G. Nyúl^b, Gábor Csifcsák^{c,*}

^a Doctoral School of Education, Faculty of Arts, University of Szeged, Petőfi Sándor sgt. 30-34, 6722 Szeged, Hungary

^b Department of Image Processing and Computer Graphics, Faculty of Science and Informatics, University of Szeged, Árpád tér 2, 6720 Szeged, Hungary

^c Department of Cognitive and Neuropsychology, Institute of Psychology, Faculty of Arts, University of Szeged, Egyetem u. 2, 6722 Szeged, Hungary

ARTICLE INFO

Article history:

Received 16 June 2015

Received in revised form 15 December 2015

Accepted 18 December 2015

Available online 19 December 2015

Keywords:

Categorization

Spatial frequency

Top-down facilitation

ERP

ABSTRACT

The modulatory effects of low and high spatial frequencies on the posterior C1, P1 and N1 event-related potential (ERP) amplitudes have long been known from previous electrophysiological studies. There is also evidence that categorization of complex natural images relies on top-down processes, probably by facilitating contextual associations during the recognition process. However, to our knowledge, no study has investigated so far how such top-down effects are manifested in scalp ERPs, when presenting natural images with attenuated low or high spatial frequency information. Twenty-one healthy subjects participated in an animal vs. vehicle categorization task with intact grayscale stimuli and images predominantly containing high (HSF) or low spatial frequencies (LSF). ERP scalp maps and amplitudes/latencies measured above occipital, parietal and frontocentral sites were compared among the three stimulus conditions. Although early occipital components (C1 and P1) were modulated by spatial frequencies, the time range of the N1 was the earliest to show top-down effects for images with unmodified low spatial frequency spectrum (intact and LSF stimuli). This manifested in ERP amplitude changes spreading to anterior scalp sites and shorter posterior N1 latencies. Finally, the frontocentral N350 and the centroparietal LPC were differently influenced by spatial frequency filtering, with the LPC being the only component to show an amplitude and latency modulation congruent with the behavioral responses (sensitivity index and reaction times). Our results strengthen the coarse-to-fine model of object recognition and provide electrophysiological evidence for low spatial frequency-based top-down effects within the first 200 ms of visual processing.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Traditionally, theories of visual object perception assume a hierarchical architecture behind human object recognition, in which feedforward activation spreads from lower-level (posterior) to higher-level (anterior) cortical areas (Grill-Spector and Malach, 2004; Logothetis and Sheinberg, 1996). This specific type of information processing has been shown to predict the remarkably good categorization performance of rapidly presented complex natural images (Serre et al., 2007). However, evidence also suggests that object recognition is influenced by top-down modulation of lower-level visual representations (Gilbert and Li, 2013), enabling the disambiguation of the visual input (Epshtein et al., 2008). Such effects involve expectation based on prior knowledge (Summerfield and Egner, 2009), or the contextual facilitation of object processing (Bar, 2004; Biederman et al., 1982).

Several studies of top-down influences focus on the role of the magnocellular (M) and parvocellular (P) subcortical channels that are respectively sensitive to low and high spatial frequencies (LSF and HSF), and propose a spatial frequency-based coarse-to-fine processing model of visual object recognition (Bar, 2003; Bullier, 2001). According to a very influential model, coarse 'gist-like' low spatial frequency representation of the object is conveyed rapidly to the orbitofrontal cortex (OFC), where representations of the most possible objects are activated, triggering predictions about object identity (Bar, 2003; Kveraga et al., 2007). This information is subsequently projected back in a top-down manner to occipital and occipitotemporal regions, enabling the disambiguation of the visual input (Bar et al., 2006). In addition, low spatial frequencies are also involved in the processing of contextual information (Peyrin et al., 2004; Torralba and Oliva, 2003). By activating strong contextual associations, low spatial frequency-based predictions about the visual environment were also hypothesized to facilitate object identification in a top-down manner (Bar, 2003, 2004; Kveraga et al., 2011).

Electroencephalography (EEG) was used in order to assess the underlying dynamics of neural activity behind visual object processing.

* Corresponding author.

E-mail address: gaborcsifcsak@yahoo.co.uk (G. Csifcsák).

The two-state interactive model of visual object recognition (Schendan and Kutas, 2007; Schendan and Lucia, 2010) suggests that in the first state (0–200 ms post-stimulus), object-sensitive visual areas are activated in a bottom-up manner, primarily reflecting sensory/perceptual processes (feature detection, figure-ground segregation or structural encoding). Accordingly, event-related potentials (ERPs) appearing before 200 ms (the posterior components C1, P1 and N1) are sensitive to physical properties such as spatial frequency, contrast and luminance. Studies using artificial stimuli (e.g., sinusoidal gratings, isoluminant color contrast stimuli) found a close correspondence between the magnitude of the occipital C1 (a negative component, typically peaking between 70 and 100 ms) and P1 (with a maximum between 80 and 140 ms) components and information conveyed by the parvocellular and magnocellular pathways, respectively (e.g., Baseler and Sutter, 1997; Elleberg et al., 2001; Foxe et al., 2008; Hansen et al., 2011; Rudvin et al., 2000).

In the case of more complex stimuli, depicting living/non-living isolated objects, Craddock et al. (2013) observed greater P1 amplitudes for HSF than LSF images. In that study, the posterior N1 (negative component peaking between 140 and 200 ms at occipital-parietal sites) was also found to be larger for images containing high spatial frequencies only. Although the N1 component emerges in State 1 of the two-state interactive model (Schendan and Kutas, 2007; Schendan and Lucia, 2010), it is not only associated with stimulus-dependent feedforward processing, but also with top-down influences such as those associated with task demands (Hopf et al., 2002), prior expectations (Melloni et al., 2011) and proactive control mechanisms (Roberts et al., 2014). In addition, shifts in N1 latencies were also shown to reflect top-down attention factors (Pollux et al., 2011).

Evolving after 200 ms post-stimulus, State 2 is predominantly characterized by interactions of feedforward and feedback processing between object-sensitive regions and higher-level cortical areas. In this time window, the frontocentral N350 reflects 'object model selection', i.e. the matching of perceptual information to long-term memory representations (Schendan and Kutas, 2007). In studies investigating visual object recognition, a positive waveform called the 'late positive complex' (LPC) peaking after 400 ms at centroparietal sites was associated with secondary classification processes such as the selection of response category, evaluation of category decisions or matching items with memory representations (Craddock et al., 2013; Schendan and Kutas, 2007; Schendan and Maher, 2009; Schendan and Lucia, 2010). This waveform is most probably composed of multiple components, including the modality-independent, response-locked 'positive-choice response' (P-CR) and the 'late slow wave' (SW), and in many respects shows similarities to the classic P300 (Falkenstein et al., 1994; Dien et al., 2004; Rushby et al., 2005). Since the above cited visual studies did not differentiate between these subcomponents with principal component analysis (PCA), herein we also refer to it as the LPC in order to enable comparing our results to those of Craddock et al. (2013).

Despite the fact that Bar's (2003, 2004) model emphasizes the role of low spatial frequency-induced top-down effects in visual stimulus processing, and that there are several studies dealing with the sensitivity of posterior event-related potentials to spatial frequencies, we have not found an electrophysiological study that linked the two, or was able to confirm or disprove Bar's theory by analyzing event-related potentials. Neither have we found ERP studies dealing with the effects of spatial frequencies on the processing of complex natural images containing contextual information. This is rather surprising, given the vast amount of literature pointing to the facilitatory role of context on object recognition and classification (Biederman et al., 1982; Bar, 2004; Joubert et al., 2008). Therefore, the aim of the current study was to investigate how the attenuation of low and high spatial frequencies in natural images modifies neural activity during categorization. In order to focus on top-down effects associated with frontal areas, ERP amplitude analysis was not restricted to occipitotemporal

regions, but extended to the entire scalp. Furthermore, ERP latency shifts associated with top-down processes were also examined (Pollux et al., 2011).

Based on previous electrophysiological studies, we predicted that the C1 component (Elleberg et al., 2001; Foxe et al., 2008; Hansen et al., 2011) will be enhanced in amplitude for images containing high spatial frequencies. As for the P1 component, our hypotheses were less straightforward, since both higher (Elleberg et al., 2001; Hansen et al., 2011) and lower (Craddock et al., 2013, 2015) P1 amplitudes were reported for LSF vs. HSF stimuli. This contradiction might be explained by differences in the image set (artificial stimuli vs. pictures of isolated objects), task (passive viewing vs. object categorization) or the analyzed scalp region (occipital midline vs. occipital/occipitotemporal). Given that the stimuli and the task in the current study corresponded better to those reported by Craddock et al. (2013, 2015), we anticipated the P1 and the subsequent N1 to be enlarged for stimuli containing high spatial frequencies (intact and HSF images). Furthermore, we hypothesized that top-down effects, manifested by ERP amplitude changes spreading to anterior scalp regions and shortening of the posterior N1 latency, would also be observed in the 140–200 ms time window. Given the N350 and the LPC index post-sensory processes in the two-state interactive model (Schendan and Kutas, 2007; Schendan and Lucia, 2010), we expected the amplitudes/latencies of these components primarily to correspond to our participants' classification performance (accuracy reflected by the sensitivity index (d'), reaction times) (Craddock et al., 2013).

2. Methods

2.1. Participants

Twenty-one healthy adults (mean age = 23.905, SD = 3.477, 13 females) participated in an animal vs. vehicle categorization task. All participants had normal or corrected-to-normal vision. None of the subjects suffered from any developmental, psychiatric, or neurological disorders. Ethical approval for this study has been granted by the Review Board of the Institute of Psychology, University of Szeged. All individuals provided signed, informed consent, and received no financial compensation for their participation.

2.2. Stimuli and procedure

Images were selected from a commercially available collection (Corel Photo Library). Stimuli in both categories were chosen to be as varied as possible. The category of animals included mammals, birds, insects, fish and reptiles, while the category of vehicles consisted of cars, trucks, trains, civil or military airplanes and boats. It is important to emphasize that while previous studies mostly presented isolated objects on a homogeneous background, we used complex images with objects embedded in natural scenes and man-made environments, since we rarely encounter isolated objects outside the laboratory. This way, top-down facilitation of object classification would not only depend on the rough, gist-like processing of shapes, but also on the presence of contextual cues, which is an important element of the above described model of Bar (2004). All pictures had a resolution of 75 pixels/in., and size of 256 × 256 pixels.

Three stimulus types were used: intact stimuli and their modified versions with attenuated high or low spatial frequencies (see Fig. 1 for examples). Color extraction and spatial frequency filtering were applied using Adobe Photoshop CS5 (Adobe Systems Inc., San Jose, USA) to match images used in a previous study (Bar et al., 2006). Low-pass filtering was done with a Gaussian blur filter (6.1 pixel kernel), whereas high-pass filtering was performed with a radius of 0.3 pixels. All images were luminance-adjusted with the SHINE toolbox (<http://www.mapageweb.umontreal.ca/gosseliff/shine/>).

Download English Version:

<https://daneshyari.com/en/article/929787>

Download Persian Version:

<https://daneshyari.com/article/929787>

[Daneshyari.com](https://daneshyari.com)