Modulation of microsaccades by spatial frequency during object categorization

Matt Craddock [a,b,1], Frank Oppermann [a,c,1], Matthias M. Müller [a,*], Jasna Martinovic [d]

[a] Institute of Psychology, University of Leipzig, Germany
[b] School of Psychology, University of Leeds, UK
[c] Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Netherlands
[d] School of Psychology, University of Aberdeen, UK

ABSTRACT

The organization of visual processing into a coarse-to-fine information processing based on the spatial frequency properties of the input forms an important facet of the object recognition process. During visual object categorization tasks, microsaccades occur frequently. One potential functional role of these eye movements is to resolve high spatial frequency information. To assess this hypothesis, we examined the rate, amplitude and speed of microsaccades in an object categorization task in which participants viewed object and non-object images and classified them as showing either natural objects, man-made objects or non-objects. Images were presented unfiltered (broadband; BB) or filtered to contain only low (LSF) or high spatial frequency (HSF) information. This allowed us to examine whether microsaccades were modulated independently by the presence of a high-level feature – the presence of an object – and by low-level stimulus characteristics – spatial frequency. We found a bimodal distribution of saccades based on their amplitude, with a split between smaller and larger microsaccades at 0.4° of visual angle. The rate of larger saccades ($\geqslant 0.4°$) was higher for objects than non-objects, and higher for objects with high spatial frequency content (HSF and BB objects) than for LSF objects. No effects were observed for smaller microsaccades (<0.4°). This is consistent with a role for larger microsaccades in resolving HSF information for object identification, and previous evidence that more microsaccades are directed towards informative image regions.

1. Introduction

Object recognition is based on a cascade of feedforward and feedback mechanisms through the visual processing hierarchy (e.g. Bar et al., 2006; Hochstein & Ahissar, 2002; VanRullen, 2007). This cascade may follow a coarse-to-fine sequence in which spatial frequency information may be particularly important for coding information at different spatial and temporal scales (e.g. Bullier, 2001; Goffaux et al., 2010; Hegdé, 2008; Kauffmann, Ramanoel, & Peyrin, 2014). Initial, feedforward processing may rely on low spatial frequencies (LSF), which provide information about many features of the visual input in parallel, activating compatible nodes in a recognition network (e.g. Levin, Takarae, Miner, & Keil, 2001). However, the conscious identification of objects likely requires re-entrant processing (feedback mechanisms) with focused attention onto the location of decisive features of potential objects (e.g. Di Lollo, Enns, & Rensink, 2000; Evans & Treisman, 2005; Hochstein & Ahissar, 2002). High spatial frequency (HSF) information may provide more fine-grained details and boundaries necessary for object identification (e.g. Oliva & Schyns, 1997; Oliva & Torralba, 2006). While a single glance may rapidly capture LSFs in a visual scene, resolving HSFs and fine spatial detail may require microsaccades (Ko, Poletti, & Rucci, 2010; McCamy, Otero-Millan, Stasi, Macknik, & Martinez-Conde, 2014; Otero-Millan, Troncoso, Macknik, Serrano-Pedraza, & Martinez-Conde, 2008; Rucci, 2008; Rucci, Iovin, Poletti, & Santini, 2007; Turatto, Valsecchi, Tamè, & Betta, 2007). Microsaccades are small eye movements – typically up to 1° of visual angle – that occur frequently even during fixation (for reviews, see Martinez-Conde, Macknik, Troncoso, & Hubel, 2009; Martinez-Conde, Otero-Millan, & Macknik, 2013; Melloni, Schwiedrzik, Rodriguez, & Singer, 2009; Rolfs, 2009). The present study investigates how the occurrence of microsaccades depends on the spatial frequency and object information of the visual input.

Spatial frequency information at different scales contributes to object categorization in different ways. LSFs may be processed and reach higher-order areas faster than HSFs (Bar et al., 2006). LSFs provide coarse global image features associated with the rough shape and layout of objects, helping to determine, for example, scene category. Scene category can be extracted at the first glance as reflected in differential cerebral activity after 150 ms, even with visual exposures starting from 20 ms (Fabre-Thorpe, Richard, & Thorpe, 1998; Thorpe, Fize, & Marlot, 1996; VanRullen & Thorpe, 2001). This processing occurs without directly attending the target image and might thus rely on the first feedforward sweep of activation (Li, VanRullen, Koch, & Perona, 2002; Rousselet, Fabre-Thorpe, & Thorpe, 2002; Thorpe, Gegenfurtner, Fabre-Thorpe, & Bülthoff, 2001). LSF processing may form a major part of this initial feedforward sweep (Bullier, 2001).

When comparing pictures of objects filtered for spatial frequency content, intact unfiltered pictures as well as pictures containing both LSF and HSF information showed better performance compared with pictures only containing either LSF or HSF information from around 100 ms of exposure duration (Kihara & Takeda, 2010, 2012). Importantly, the categorization of LSF–only objects outperformed the categorization of HSF objects for the exposure durations of up to 250 ms, suggesting a prior for LSF information in early processing in this kind of categorization task (Kihara & Takeda, 2010). The differences did not change when attentional demands were increased, suggesting that the effects are based on the first feedforward processing (Kihara & Takeda, 2012).

However, the information extracted during feedforward processing does not always allow full, accurate identification of scenes and objects within them. For example, Evans and Treisman (2005) asked their participants to identify animal targets embedded in RSVP streams of distractors, with each image presented for 75–100 ms. The participants failed to identify the targets in more than half of the trials, and also often failed to localize the target correctly, suggesting that further processing is necessary. After the feedforward sweep comes re-entrant, feedback processing, which is likely directed at processing of HSFs. For example, consistent with the expectation that processing of HSF information follows processing of LSF information, coarse-to-fine, LSF-to-HSF image sequences of scenes elicit greater earlier activation in early occipital areas and both frontal and temporal areas compared to fine-to-coarse HSF-to-LSF sequences (Peyrin et al., 2010).

Eye movements in this period may be particularly important. Microsaccades follow a stereotypical pattern of inhibition and subsequent release after the onset of a visual stimulus, dropping significantly before rebounding to a new peak after approximately 200–400 ms (e.g. Engbert & Kliegl, 2003; Turatto et al., 2007). They are affected by a range of cognitive factors such as task difficulty and attention (Engbert, 2006; Engbert & Kliegl, 2003; Siegenthaler et al., 2014), and change neural processing (Bosman, Womelsdorf, Desimone, & Fries, 2009; Dimigen, Valsecchi, Sommer, & Kliegl, 2009; Martinez-Conde, Macknik, & Hubel, 2000, 2002; Troncoso et al., 2015). The amplitudes of saccades in these studies range from less than 1° of visual angle, which are typically defined as microsaccades (Martinez-Conde et al., 2013; Melloni et al., 2009), up to 1.5° or 2.0° (e.g. Engbert & Kliegl, 2003; Martinez-Conde, Macknik, Troncoso, & Dyar, 2006; Turatto et al., 2007; Yuval-Greenberg, Tomer, Keren, Nelken, & Deouell, 2008). Stimulus and fixation target size may also influence microsaccade amplitude ((McCamy, Najafian Jazi, Otero-Millan, Macknik, & Martinez-Conde, 2013; Otero-Millan, Macknik, Langston, & Martinez-Conde, 2013).

With regard to spatial frequency, there is evidence to suggest that HSF may increase the rate of microsaccades. Microsaccades occur at a higher rate during tasks which require high visual acuity (Ko et al., 2010), show directional biases during tasks that involve discrimination of visual detail (Turatto et al., 2007), and occur more frequently during foveation of faces or other salient objects (Otero-Millan et al., 2008). They also occur more frequently in more informative regions of visual scenes, such as those with high contrast and low spatial correlation (McCamy et al., 2014). Bonneh, Adini, and Polat (2015) tested microsaccade rates in response to passive viewing of transient Gabor patches with varying spatial frequency. They found that microsaccade latency following release from inhibition increased as spatial frequencies went from middle-level (2 cycles per degree) to higher (8 cycles per degree), which may have produced a later, smaller peak in microsaccade rate. However, microsaccade rates in passive viewing tasks may not reflect performance in more directed, active viewing tasks (e.g. McCamy et al., 2014).

Consistent with a role of microsaccades in object recognition, it has been demonstrated that the rebound peak in the saccade rate after the onset of a visual stimulus is modulated by high-level stimulus properties; for example, it is relatively elevated for objects compared to non-object stimuli (Hassler, Barreto, & Gruber, 2011; Keren, Yuval-Greenberg, & Deouell, 2010; Yuval-Greenberg et al., 2008). However, this evidence comes primarily from investigations of the relationship between microsaccades and a broadband peak in induced gamma band oscillations (~30 to 100 Hz), observed using the scalp-recorded electroencephalogram (EEG). This signal was considered to be a signature of the activation of an object representation and the binding of the activity of disparate populations of neurons, each representing distinct object features, into a single coherent percept (Tallon-Baudry & Bertrand, 1999). Several authors have convincingly demonstrated that an electrical, muscle-generated signal associated with microsaccades – the saccade spike potential (SSP) – underlies this effect (Hassler et al., 2011; Keren et al., 2010; Yuval-Greenberg et al., 2008). Thus, many of the reported modulations of induced gamma-band activity – for example, by object orientation (Martinovic, Gruber, & Müller, 2007, 2008) – were likely attributable to modulations of the underlying saccade rate in the critical window around 200–400 ms. Directly examining the saccade rate in this time window may thus reveal information regarding object recognition processes and role of eye-movements to resolve spatial frequency information.

In the present study, we use a living/non-living categorization task to probe the role of spatial frequency in object processing by varying the spatial frequency content of objects. We presented objects either as unfiltered, broadband (BB) images, or filtered to contain only LSF or HSF content. We chose spatial frequency ranges that corresponded to previous studies examining the different roles of HSF and LSF in object recognition (e.g. Bar et al., 2006). These ranges also correspond to the spatial frequency tuning curves observed in orbitofrontal and visual cortices (Fintzi & Mahon, 2013). We expected that we would observe the typical peaks in the saccade rate approximately 200–400 ms after stimulus onset. Given that microsaccades may have a role in resolving fine spatial detail, we expected to see higher rates for HSF and BB images than for images with LSF only. Additionally, we presented non-object trials with spatial frequency content matched to that of the object images. We expected that saccade peak rates would be reduced relative to object trials, in line with previous findings from EEG (e.g. Hassler et al., 2011; Yuval-Greenberg et al., 2008), and in free-viewing of blank scenes (Otero-Millan et al., 2013, 2008). Nevertheless, object versus non-object differences should also reveal whether differences in saccade rate are driven by high-level factors in combination with low-level stimulus properties, or low-level stimulus properties alone: spatial frequency differences on non-object trials would imply the latter.