Vision Research 107 (2015) 49–57

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Rapid scene categorization: Role of spatial frequency order, accumulation mode and luminance contrast

Louise Kauffmann [a,b,*], Alan Chauvin [a,b], Nathalie Guyader [c], Carole Peyrin [a,b,*]

[a] Univ. Grenoble Alpes, LPNC, F-38040 Grenoble, France
[b] CNRS, LPNC, F-38040 Grenoble, France
[c] Univ. Grenoble Alpes, GIPSA-lab, F-38402 Grenoble, France

ARTICLE INFO

Article history:
Received 25 July 2014
Received in revised form 8 November 2014
Available online 10 December 2014

Keywords:
Coarse-to-fine
Fine-to-coarse
Low spatial frequency
High spatial frequency
Root mean square contrast normalization

ABSTRACT

Visual analysis follows a default, predominantly coarse-to-fine processing sequence. Low spatial frequencies (LSF) are processed more rapidly than high spatial frequencies (HSF), allowing an initial coarse parsing of visual input, prior to analysis of finer information. Our study investigated the influence of spatial frequency processing order, accumulation mode (i.e. how spatial frequency information is received as an input by the visual system, throughout processing), and differences in luminance contrast between spatial frequencies on rapid scene categorization. In Experiment 1, we used sequences composed of six filtered scenes, assembled from LSF to HSF (coarse-to-fine) or from HSF to LSF (fine-to-coarse) to test the effects of spatial frequency order. Spatial frequencies were either successive or additive within sequences to test the effects of spatial frequency accumulation mode. Results showed that participants categorized coarse-to-fine sequences more rapidly than fine-to-coarse sequences, irrespective of spatial frequency accumulation in the sequences. In Experiment 2, we investigated the extent to which differences in luminance contrast rather than in spatial frequency account for the advantage of coarse-to-fine over fine-to-coarse processing. Results showed that both spatial frequencies and luminance contrast account for a predominant coarse-to-fine processing, but that the coarse-to-fine advantage stems mainly from differences in spatial frequencies. Our study cautions against the use of contrast normalization in studies investigating spatial frequency processing. We argue that this type of experimental manipulation can impair the intrinsic properties of a visual stimulus. As the visual system relies on these to enable recognition, bias may be induced in strategies of visual analysis.

1. Introduction

The human visual system is constantly involved in the perception and categorization of complex stimuli such as natural scenes. In the spatial domain, scenes are classically described in terms of pixel luminance. In the Fourier domain, a dual representation of a scene is created using the amplitude and phase spectra (Field, 1987; Ginsburg, 1986; Hughes, Nozawa, & Kitterle, 1996; Tolhurst, Tadmor, & Chao, 1992). The amplitude spectrum corresponds to the distribution of luminance contrast across spatial frequencies and orientations, and the phase spectrum corresponds to the spatial relation between spatial frequencies. Luminance contrast refers to the magnitude of luminance variation in a stimulus relative to its mean luminance (Shapley & Enroth-Cugell, 1984). The visual system uses low-level features, such as spatial frequencies and luminance contrast to enable recognition, and from a neurobiological point of view, we now know that cells in the primary visual cortex respond to luminance contrast, spatial frequencies and orientations (Boynton, 2005; De Valois, Albrecht, & Thorell, 1982; De Valois, Yund, & Hepler, 1982; Poggio, 1972; Shams & Von der Malsburg, 2002; Shapley & Lam, 1993). Many studies have also highlighted the importance of the amplitude spectrum in scene categorization (Guyader, Chauvin, Peyrin, Hérault, & Marendaz, 2004; Oliva & Torralba, 2001; Torralba & Oliva, 2003). Overall, these studies support current influential models of scene perception (Bar, 2003; Bar et al., 2006; Bullier, 2001; Hegdé, 2008; Kauffmann, Ramanoël, & Peyrin, 2014; Peyrin et al., 2010; Schyns & Oliva, 1994). According to these models, visual analysis is based on the parallel extraction of different attributes at different spatial frequencies in scenes, and follows a

* Corresponding authors at: Laboratoire de Psychologie et Neurocognition, Université Pierre Mendès France, BP 47, 38040 Grenoble Cedex 09, France (L. Kauffmann). CNRS UMR 5105, Université Pierre Mendès France, BP 47, 38040 Grenoble Cedex 09, France (C. Peyrin).

E-mail addresses: louise.kauffmann@upmf-grenoble.fr (L. Kauffmann), carole.peyrin@upmf-grenoble.fr (C. Peyrin).

http://dx.doi.org/10.1016/j.visres.2014.11.013
0042-6989/

predominant and default coarse-to-fine processing sequence. Low spatial frequencies (LSF), containing the coarse information on a visual stimulus, are rapidly conveyed by magnocellular pathways to the occipital cortex and then access high-order cortical areas, in order to activate plausible interpretations of the visual input. This initial coarse analysis is then used to guide the later processing of high spatial frequencies (HSF) which are conveyed more slowly by parvocellular pathways and provide finer information.

Several behavioral studies, and a few neuroimaging studies, have investigated coarse-to-fine processing during scene perception (De Cesarei & Loftus, 2011; Musel, Chauvin, Guyader, Chokron, & Peyrin, 2012; Musel et al., 2014; Parker, Lishman, & Hughes, 1992; Peyrin et al., 2010; Schyns & Oliva, 1994). Evidence of a predominant coarse-to-fine processing was originally provided by psychophysical studies using hybrid scenes (Schyns & Oliva, 1994). Hybrid stimuli are made up by superimposing two images of scenes that contain different spatial frequencies and different semantic information (e.g. a highway scene in LSF superimposed on a city scene in HSF). When presentation times were very short (30 ms), perception of these hybrid scenes was dominated by LSF information. When presentation times were longer (150 ms), perception of hybrids was dominated by HSF information. This suggests that LSF take precedence over HSF in the visual time-course. More recent studies (Musel et al., 2012, 2014; Peyrin et al., 2005, 2010) explicitly simulated different time courses of spatial frequency processing during scene categorization using sequences of scene images in which the spatial frequency content differed from one image to the other, going either from low-to-high spatial frequencies (coarse-to-fine processing) or from high-to-low spatial frequencies (fine-to-coarse processing). These studies showed that sequences depicting coarse-to-fine processing were categorized more rapidly than those depicting fine-to-coarse processing. This suggests that the presentation order of spatial frequencies strongly influences the speed of scene categorization, and when LSF are presented first in the sequence, this may particularly facilitate the process. Recent event-related brain potential (ERP) studies have suggested that the accumulation of spatial frequency information could also influence the perception of scenes, irrespective of the presentation order of spatial frequencies. De Cesarei and Codispoti (2011) and De Cesarei, Mastria, and Codispoti (2013) investigated how spatial frequencies influence the identification of neutral and emotional scenes. They also used sequences containing images of scenes in which the amount of spatial frequency information increased progressively from one image to the next. To be precise, they presented sequences in which the first scene was either LSF or HSF, and the entire scene was gradually revealed by progressively adding either HSF or LSF information. This procedure allowed the authors to investigate the effects of the addition of spatial frequency information, according to the type of spatial frequency content (either LSF or HSF) which had been initially processed. Behavioral results showed that the identification rate of scenes increased as spatial frequency information was added, irrespective of the spatial frequency content initially presented in the sequence (LSF or HSF), and no differences in behavioral performances were observed between sequences starting with LSF and HSF information. These results suggested that scene identification did not critically depend on the initial processing of LSF – it appeared to rely more on the addition of spatial frequency information. However, divergences between the above mentioned studies may result from methodological differences in the accumulation mode of spatial frequency in the sequence of scene images. The accumulation mode refers here to how spatial frequency information is received as an input by the visual system, throughout the sequences (e.g. in the previously cited examples, either successively or additively). They also raised the question of whether the visual system would benefit from the reinjection of spatial frequency information relating to previous inputs during the processing of spatial frequency sequences. To our knowledge, no study has as yet directly investigated how the accumulation mode of spatial frequency in the sequence influences the well-established advantage of LSF over HSF during rapid scene categorization.

The first experiment in the present study aimed to investigate rapid scene categorization depending on both the presentation order of spatial frequencies (LSF before HSF or HSF before LSF) and the accumulation mode of spatial frequencies in the sequence (successive or additive presentation of different spatial frequency bands). In order to do this, we used dynamic scenes composed of six images of a scene filtered in different spatial frequencies, going from LSF to HSF or from HSF to LSF. This allowed us to test the effects of spatial frequency order. Dynamic scenes also depicted either a successive or additive processing. This allowed us to test the effects of spatial frequency accumulation mode. Successive sequences therefore started with either an LSF or an HSF filtered scene. They subsequently shifted to a higher or a lower spatial frequency band, and ended with an HSF or LSF filtered scene, respectively. Additive sequences started with an LSF or an HSF filtered scene, but this time, either HSF or LSF information was added. The spatial frequency content therefore increased from one image to the next, and sequences ended with an almost intact (or non-filtered) scene. Participants were asked to perform a categorization task on these stimuli (indoor vs. outdoor). Based on recent models of visual perception, we expected to observe a coarse-to-fine advantage, i.e. more rapid categorization when LSF (rather than HSF) were presented first. However, if the addition of spatial frequencies throughout the sequence constitutes an advantage irrespective of the order of presentation of spatial frequency, additive sequences should be categorized more rapidly than successive sequences in sequences which begin with either LSF or HSF information. Interaction between the accumulation mode of spatial frequencies and the order of spatial frequencies during sequences ought to reduce the coarse-to-fine advantage in additive sequences (compared to successive sequences).

In the same context, differences in luminance contrast have been shown to exert a strong influence on speed of visual processing. For example, reaction times decrease as luminance contrast increases (Harwerth & Levi, 1978). The luminance contrast in scenes decreases as spatial frequency increases, following a $1/f^{\alpha}$ function (Field, 1987). LSF are characterized by a high luminance contrast, and HSF are characterized by a lower luminance contrast. The temporal precedence of LSF over HSF (i.e. coarse-to-fine processing) during scene categorization could therefore be explained by differences in contrast rather than in spatial frequency content. In order to avoid any confusion between the influence of spatial frequency content and that of luminance contrast in scene perception, recent studies equalize both the mean luminance and the luminance contrast of the filtered stimuli used (see, for example, Goffaux et al., 2011; Mu & Li, 2013; Vlamings, Goffaux, & Kemner, 2009) by attributing the same mean luminance and the same root mean square (RMS) contrast to all filtered stimuli. The RMS contrast corresponds to the standard deviation of luminance values and has been shown to be the most reliable indicator of the visibility of broadband filtered images (Bex & Makous, 2002). However, the specific role of luminance contrast in the spatial frequency processing of scenes has never been systematically investigated. In a second experiment, we investigated the specific role of luminance contrast, spatial frequencies, and their interaction during the coarse-to-fine processing of scenes. In order to do so, we used dynamic scenes adapted from Experiment 1, and manipulated the spatial frequency content and luminance contrast of the images composing the sequences.