Vision Research 60 (2012) 101–113

Contents lists available at SciVerse ScienceDirect

Vision Research

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

# On the effectiveness of noise masks: Naturalistic vs. un-naturalistic image statistics

Bruce C. Hansen [a,*], Robert F. Hess [b]

[a] *Department of Psychology & Neuroscience Program, Colgate University, Hamilton, NY 13346, USA*
[b] *McGill Vision Research Unit, McGill University, Montreal, QC H3A 1A1, Canada*

ARTICLE INFO

*Article history:*
Received 22 September 2011
Received in revised form 10 January 2012
Available online 31 March 2012

*Keywords:*
Natural scenes
Spatial masking
Power spectrum
Fractal noise
Contrast gain control

ABSTRACT

It has been argued that the human visual system is optimized for identification of broadband objects embedded in stimuli possessing orientation averaged power spectra fall-offs that obey the $1/f^{\beta}$ relationship typically observed in natural scene imagery (i.e., $\beta = 2.0$ on logarithmic axes). Here, we were interested in whether individual spatial channels leading to recognition are functionally optimized for narrowband targets when masked by noise possessing naturalistic image statistics ($\beta = 2.0$). The current study therefore explores the impact of variable $\beta$ noise masks on the identification of narrowband target stimuli ranging in spatial complexity, while simultaneously controlling for physical or perceived differences between the masks. The results show that $\beta = 2.0$ noise masks produce the largest identification thresholds regardless of target complexity, and thus do not seem to yield functionally optimized channel processing. The differential masking effects are discussed in the context of contrast gain control.

## 1. Introduction

Decades of research have lead to a popular view of the initial processes in human striate cortex that, in part, involves multiple sub-populations of striate neurons acting like non-linear "filters" (or "channels" in terms of psychophysical terminology). These filter-channels are argued to each extract a specific narrow band of spatial frequency and orientation content from our visual environment (e.g., Campbell & Robson, 1968; Carandini et al., 2005; De Valois, Albrecht, & Thorell, 1982; De Valois, Yund, & Hepler, 1982; Field & Tolhurst, 1986; Graham & Nachmias, 1971; Maffei & Fiorentini, 1973; Merigan & Maunsell, 1993; Pantle & Sekuler, 1968; Phillips & Wilson, 1984; Ringach, 2002; Shapley & Lennie, 1985; Wilson & Bergen, 1979). Further, numerous studies have reported large scale interactions between channels tuned to different spatial frequencies and orientations (e.g., Bauman & Bonds, 1991; Bonds, 1989; Bosking et al., 1997; DeAngelis et al., 1992; Fitzpatrick, 2000; Kersten, 1984; Legge & Foley, 1980; Meese & Holmes, 2010; Meier & Carandini, 2002; Morrone, Burr, & Maffei, 1982; Nelson et al., 1994; Olzak, 1985; Olzak & Thomas, 1991; Petrov, Carandini, & McKee, 2005; Ross & Speed, 1991). Based on such studies, we now know a great deal regarding how any given spatial channel interacts with others under conditions utilizing various spatial configurations of narrowband stimuli. However, we still know very little in terms of how such channels operate when interacting with a very broad range of spatial channels (broad in both spatial frequency and orientation). That is, previous simultaneous masking experiments (using narrowband overlay, lateral, or surround masking configurations) possess limited predictive power regarding how specific channels operate when processing the real-world environment (Olshausen & Field, 2005). Specifically, the natural environment is known to be broadband in both spatial frequency and orientation (reviewed in Hansen, Haun, and Essock (2008)), which means that at any given location within a scene, many visual channels are likely to be simultaneously active. Thus, the functional operation of a given channel will be weighted by the interdependent responses from a broad array of differently tuned channels (and not just a small sub-set of channels).

Given the above, if one wishes to understand how spatial channels may operate on a day-to-day basis, it is necessary to utilize simultaneous masking paradigms that employ masks that are broad in terms of both spatial frequency and orientation. Granted, numerous studies have utilized white noise masks in simultaneous masking paradigms designed to elucidate the response characteristics of spatial channels underlying the detection, discrimination, or identification of stimuli ranging from sinusoidal gratings to broadband stimuli such as letters and human faces (e.g., Alexander, Xie, & Derlacki, 1994; Burgess et al., 1981; Gold, Bennett, & Sekuler, 1999; Henning, Hertz, & Hinton, 1981; Legge et al., 1985; Majaj et al., 2002; Oruç & Barton, 2010; Oruç & Landy, 2009; Parish & Sperling, 1991; Pelli et al., 2006; Solomon & Pelli, 1994; Tjan et al., 1995). However, those studies typically employed white noise masks with the express aim of parsing an observer's performance from their 'intrinsic noise' as a 'pure' measure of observer ability (Pelli & Farell, 1999). Further, white noise masks possess constant contrast energy across all spatial frequencies and orientations, a property that is far from the typical distribution of contrast

---

* Corresponding author.
*E-mail address:* bchansen@colgate.edu (B.C. Hansen).

0042-6989/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.visres.2012.03.017

across spatial frequency in the natural environment. Particularly, the 2nd order luminance statistics of natural scene imagery have been extensively studied and shown to possess a property where the contrast (or power in the Fourier domain) at different spatial frequencies (averaged across orientation), $f$, falls with increasing $f$, following a $1/f^{\beta}$ relationship (e.g., Billock, 2000; Field, 1987; Field & Brady, 1997; Hansen & Essock, 2005; Kretzmer, 1952; Oliva & Torralba, 2001; Ruderman & Bialek, 1994; Simoncelli & Olshausen, 2001; Tolhurst, Tadmor, & Tang, 1992; Torralba & Oliva, 2003; van der Schaaf & van Hateren, 1996), with $\beta$ typically observed to be near 2.0 *on logarithmic axes*, or equivalently, an $\alpha$ exponent of 1.0 if assessing the amplitude spectrum – the square-root of the power spectrum (Billock, 2000; Burton & Moorhead, 1987; Field, 1987, 1993; Field & Brady, 1997; Hansen & Essock, 2005; Ruderman & Bialek, 1994; Thomson & Foster, 1997; Tolhurst, Tadmor, & Tang, 1992; van der Schaaf & van Hateren, 1996). What this means, relative to white noise (i.e., $\beta = 0.0$), is that stimuli with $\beta$ exponents near 2.0 possess more contrast at lower spatial frequencies and less at higher spatial frequencies. To better understand how spatial channels operate when processing our broadband environments, it therefore seems logical to not only incorporate broadband masks into simultaneous masking paradigms, but also to use broadband masks with power spectra $\beta$s near 2.0.

Interestingly, it has been suggested that there exists a correspondence between the prevalence of content at particular spatial scales in natural scenes (i.e., the $1/f^{\beta}$ relationship) and the shape and scale of human spatial filters, with those filters being well matched to optimally code the natural world (Brady & Field, 1995; Field, 1987; Simoncelli & Olshausen, 2001). Several lines of psychophysical research have explored the extent to which spatial channels are optimized to process natural and naturalistic stimuli across a broad array of tasks. Relevant to the current study are the tasks that involved the identification of image content within scenes where the $\beta$ exponents were varied. Such studies have consistently shown that humans are best at identification when the images possess $\beta$ exponents near 2.0, and worst at smaller or larger $\beta$s (Párraga, Troscianko, & Tolhurst, 2000, 2005; Tolhurst & Tadmor, 2000). That is, our visual systems seem to best process structural changes between objects when the luminance statistics of the scenes within which the objects are embedded closely match those typically observed on a day-to-day basis. However, since the target content in those studies was broadband, it is difficult to identify how any one spatial channel was influenced by the simultaneous activation of other differently tuned channels under naturalistic ($\beta = 2.0$) or non-naturalistic ($\beta$s much smaller or larger than 2.0) stimulation. Additionally, since the target content was suprathreshold, it remains unclear whether such $\beta = 2.0$ tuning for identification would be present when identification is limited by detection.

Motivated by the above, the current study sought to explore the effectiveness of different $\beta$ noise masks to interfere with the identification of variable contrast narrowband targets in a simultaneous noise masking paradigm. We chose noise masks for two primary reasons. First, the form of the power spectrum can be precisely controlled and second the detection thresholds for bandpass targets embedded in natural scenes have been shown (Bex, Solomon, & Dakin, 2009; Hansen & Essock, 2005) to largely depend on edge density (noise imagery lacks the presence of broadband edges). Given the vast array of possible targets and tasks, we chose those that have been typically employed in simultaneous noise masking paradigms. Specifically, we measured the identification of narrowband targets that varied in terms of spatial complexity, ranging from simple (i.e., Gabor patterns) to complex (i.e., bandpass filtered letters) as a function of target contrast embedded in fixed high contrast noise set to one of three different $\beta$s (namely, 0.0, 2.0, or 3.0). The current study therefore asks: are human spatial channels optimized to process various target stimuli to identification when presented against naturalistic backgrounds (i.e., when noise $\beta = 2.0$)?

It is important to note that in order to effectively test whether spatial channels are optimized in the presence of one set of 2nd-order luminance statistics compared to others, it is essential to control for any physical or perceived differences between the different noise masks. Accordingly, each experiment in the current study was designed to systematically control for possible low-level accounts related to differences in noise spectral density, limitations due to available stimulus information, and though unlikely, differences in perceived contrast. Given the need to eliminate multiple confounds, the current study employed narrowband stimuli fixed to one central spatial frequency. It is also important to note that none of the noise patterns employed in the current study possess the typical orientation biases known to occur in natural scenes, nor do they possess any of the higher-order statistical relationships carried by the phase spectra of natural scenes. Therefore, all implications for real-world perception drawn from the current study should be considered with those caveats in mind.

The design of the current study is as follows: Experiment 1 used Gabor targets and was designed to test whether the corresponding spatial channel showed evidence of optimization for those targets when embedded in variable $\beta$ noise masks. Experiment 1 also utilized notch filtering in order to control for physical differences in contrast at and near the central spatial frequency of the target stimuli. Additionally, we employed an ideal observer analysis to factor out task constraints due to the different noise masks. Experiment 2 was designed to demonstrate whether perceived contrast varies with the type of noise, and then to control for any differences in perceived contrast to factor it out as a possible explanation for the threshold elevation differences observed in Experiment 1. Experiment 3 sought to repeat Experiments 1 and 2, but with filtered letter stimuli to extend the findings of Experiment 1 and 2 to more spatially complex targets (i.e., letters).

The results from Experiments 1–3 do not support the notion of optimized channel processing for masks possessing $\beta$s set at 2.0. In fact they show the complete opposite – noise masks with a $\beta$ value of 2.0 interfere with target identification (Gabors and letters) much more than $\beta = 0.0$ and $\beta = 3.0$ noise. Lastly, the higher thresholds for $\beta = 2.0$ masks could not be explained by physical or perceived differences between the noise masks.

## 2. General method

### 2.1. Apparatus

All stimuli were presented on a 21″ Viewsonic (G225fB) monitor driven by a dual core Intel® Xeon® processor (1.60 GHz × 2) equipped with 4 GB RAM and a 256 MB PCIe × 16 ATI FireGL V7200 dual DVI/VGA graphics card with 8-bit grayscale resolution. The color management settings for the graphics card (i.e., 3D display settings) were adjusted such that the luminance "gain" of the green gun was twice that of the red gun, which was set to twice that of the blue gun. A bit-stealing algorithm (Bex, Mareschal, & Dakin, 2007; Tyler, 1997) was employed to yield 10.8 bits of luminance (i.e., grayscale) resolution (i.e., 1785 unique levels) distributed evenly across a 0–255 scale. Stimuli were displayed using a linearized look-up table, generated by calibrating with a ColorVision Spyder3 Pro sensor. Maximum luminance output of the display monitor was 100 $\text{cd/m}^{-2}$, the frame rate was set to 85 Hz, and the resolution was set to 1600 × 1200 pixels. Single pixels subtended .0134° of visual angle (i.e., 0.80 arc min.) as viewed from 1.0 m. Head position was maintained with an Applied Science Laboratories (ASL) chin and forehead rest.