Vision Research 48 (2008) 1374–1382

Contents lists available at ScienceDirect

Vision Research

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

Centre-surround effects on perceived orientation in complex images

Erin Goddard [a,b,*], Colin W.G. Clifford [a], Samuel G. Solomon [b]

[a] *School of Psychology, The University of Sydney, Sydney, Australia*
[b] *Disciplines of Physiology, Anatomy and Histology, School of Medical Sciences and Bosch Institute, The University of Sydney, Sydney, Australia*

ARTICLE INFO

Article history:
Received 21 October 2007
Received in revised form 22 February 2008

Keywords:
Tilt illusion
Contextual modulation
Natural scenes
Visual cortex
Spatial vision

ABSTRACT

Using the simultaneous tilt illusion [Gibson, J., & Radner, M. (1937). Adaptation, after-effect and contrast in the perception of tilted lines. *Journal of Experimental Psychology, 12*, 453–467], we investigate the perception of orientation in natural images and textures with similar statistical properties. We show that the illusion increases if observers judge the average orientation of textures rather than sinusoidal gratings. Furthermore, the illusion can be induced by surrounding textures with a broad range of orientations, even those without a clearly perceivable orientation. A robust illusion is induced by natural images, and is increased by randomising the phase spectra of those images. We present a simple model of orientation processing that can accommodate most of our observations.

1. Introduction

Our current understanding of the visual system is based to a large extent on the measurement of physiological and perceptual responses to reduced, simplified stimuli, such as sinusoidal gratings. In a recent review of the current understanding of V1, Olshausen and Field (2005) suggested that for the highly non-linear visual system, sinusoidal gratings have no particular significance, and that the body of experimental work on V1 is biased by the number of studies which use sinusoidal gratings and other reduced stimuli to infer the response properties of V1 cells. While the processing of bars, spots and gratings by V1 cells may be relatively well described, we cannot necessarily use these neural responses to predict responses to more complex visual stimuli, including natural scenes (Burr, Morrone, & Maffei, 1981; David, Vinje, & Gallant, 2004).

There is psychophysical evidence to suggest that the visual system is tuned to scene statistics that are characteristic of natural scenes. One such statistical commonality among natural scenes is their amplitude spectra, which typically vary with spatial frequency (f) such that

$$\textit{amplitude}\ (f) \propto f^{-\alpha} \tag{1}$$

where, across natural images, α tends to fall within a range of about 0.7–1.7, with an average value of approximately 1. For example, average α values of 0.94 (van der Schaaf & van Hateren, 1996), 1.16 (Tadmor & Tolhurst, 1994) and 0.91 (Ruderman & Bialek, 1994) have been reported.

Stimuli with this spatial frequency structure are more easily discriminated on the basis of spatial frequency content (Tadmor & Tolhurst, 1994). Adaptation to a series of natural images selectively reduces sensitivity to low and medium spatial frequencies (Webster & Miyahara, 1997), and perceived contrast is maximally suppressed by a surrounding stimulus when the inducing stimulus has an α value of 1, regardless of the α value of the central test patch (McDonald & Tadmor, 2006). Thus not only are simple, reduced stimuli unlike those encountered in the natural world, they are also unlikely to optimally engage mechanisms of the visual system which are under scrutiny.

These findings highlight the need for models of the visual system that can account for physiological and perceptual responses to natural stimuli. To relate the processing of simple visual stimuli to that of complex stimuli, we should test whether models that can account for responses to simpler stimuli can predict perceptual and physiological responses to them. Here we ask whether an underconstrained model of a visual illusion, the simultaneous tilt illusion, which is well documented with bars and gratings, can also be used to account for illusions induced by more complex textures.

The simultaneous tilt illusion refers to the impact of nearby lines or gratings on the perceived orientation of contours. The impact of the nearby stimulus depends on the orientation difference between it and the test: small differences repulse the perceived orientation of the test away from that of the surrounding stimulus, while larger differences can attract (Wenderoth & Johnstone, 1988; Westheimer, 1990). The neural basis for this illusion has been modelled as contextual modulation of those cells whose respon-

---

[*] Corresponding author.
E-mail address: ering@psych.usyd.edu.au (E. Goddard).

0042-6989/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2008.02.023

siveness contributes to the perceived orientation of the test, where the presence of the surrounding stimulus either reduces their responsiveness, shifts their orientation preference, or broadens their bandwidth. In each case this contextual modulation is dependent upon the orientation of the surrounding lines or gratings, and each manipulation is capable of shifting the population response and predicting the tilt illusion with simple stimuli (Clifford, Wenderoth, & Spehar, 2000; Coltheart, 1971; Gilbert & Wiesel, 1990; Jin, Dragoi, Sur, & Seung, 2005). Here we ask whether a simple model of orientation processing, similar to these, can account for the tilt illusion with textures, which share some properties that are typical of natural images.

We produced textures whose two-dimensional Fourier spectra were defined by a $1/f$ distribution of spatial frequencies at each orientation, similar to that of natural images, and a broad range of orientations. We measured the tilt illusion using these broadband textures as the inducing and test stimuli, and asked whether our model, which predicts the illusion for gratings, could account for our results. We generated predictions of this model for both gratings and textures. We evaluated the extent to which the model of the illusion with gratings can predict the illusion with textures, by tracking the performance on textures of the parameter sets that provide the best 10% of model fits for gratings. The model, which considers the amplitude spectra of these textures, predicts most of our observations. We then measured the tilt illusion induced when using segments of natural images as surrounding stimuli. These image segments, unlike broadband textures, contain clear contours: the tilt illusion induced by them was mostly, but not entirely, predicted by their amplitude spectra.

2. Methods

2.1. Visual stimuli

Images were generated and displayed with Matlab (v7.0, MathWorks, Natick, MA) using Open GL and routines from PsychToolbox (Brainard, 1997; Pelli, 1997) on a G5 Macintosh computer driving a nVidia GeForce 6600LE graphics card. The images were displayed on a gamma-corrected ViewSonic G810-6 cathode ray tube monitor, refreshed at a rate of 75 Hz and viewed from a distance of 0.57 m. To remove visual cues to vertical the subject viewed the screen in a darkened room, and matt black cardboard with a circular window covered all but a circular central portion (radius 10.75 deg) of the screen. The perceived tilt of a central circular surface (radius 1.5 deg) was measured in the presence of an abutting annular surface (outer radius 7.25 deg). The remainder of the screen was held at the mean luminance of $\sim$50 cd/m^2.

In the first experiment the centre and annular surfaces were sinusoidal gratings (spatial frequency 3 cycles/deg). The spatial phase of both the central and annular gratings was randomly selected on each stimulus presentation. In the subsequent experiments with broadband textures, central and annular surfaces were the real component of the inverse Fourier transform of a square matrix of complex numbers; the amplitude of these complex numbers defined the amplitude spectrum of the surface; the angle of these numbers in the complex plane defining the phase spectrum. Each point in the phase spectrum was drawn from a uniform random distribution ranging from 0 to 2π, and was regenerated for every stimulus presentation. The two-dimensional amplitude spectrum was specified as the product of a $1/f$ distribution over spatial frequency for each orientation, and a Gaussian distribution over orientation; frequency being the polar distance, and orientation the polar angle relative to the central element of the matrix.

The power at each orientation was defined by a wrapped normal distribution with a standard deviation ranging from 3.125 to 50 deg and a peak at vertical. At large standard deviations, these distributions are non-zero at the orientation orthogonal to the peak, so they must be wrapped; this is not possible analytically, and was approximated as described by Dakin, Mareschal, and Bex (2005). We also used annular surfaces with 'notched' orientation structure, which were specified by a flat distribution, from which a wrapped normal distribution was subtracted. The amplitude spectrum was then restructured (using *fftshift2* in Matlab) before combination with the phase spectrum. We used OpenGL commands to appropriately rotate the central and annular surfaces on the video card.

Natural image segments used in the final experiment were selected from 100,000 randomly chosen 256×256 pixel segments taken from the first 1000 images in van Hateren's database (van Hateren & van der Schaaf, 1998). Each image segment was decomposed using Fourier analysis into its amplitude and phase spectra. We compared the amplitude spectrum of each image segment with each of the amplitude spectra used to generate the broadband textures. For each pair of amplitude spectra, we assessed their similarity by calculating the correlation coefficient between the pixel values in the two spectra. Four of the sixteen image segments with the highest correlation coefficients were selected for each orientation bandwidth, such that each image segment was of a separate scene (r values ranged from 0.82 to 0.94). These images can be downloaded from van Hateren's database (van Hateren & van der Schaaf, 1998) and located using the image numbers and segment locations in Appendix 1. Annular surfaces always had the amplitude spectrum of the original natural image segment, and either the original phase spectrum (natural image condition), or a randomly generated phase spectrum (phase randomised condition). The central (test) surface was a broadband texture with a wrapped normal distribution of orientations of 12.5 deg standard deviation, as described above.

Since we randomly generated the phase matrix of each broadband texture, there was some variation in the distribution of pixel intensities of each broadband texture, even between textures with the same amplitude spectrum. To remove this variation, the surfaces were normalised such that the root mean square (RMS) contrast of the pixel intensities was 0.25. RMS contrast has been shown to correlate well with the detectability of visual stimuli, and is used here to approximately normalise the surfaces for perceived contrast (Bex & Makous, 2002).

2.2. Subjects and procedure

A total of 11 subjects (18–39 years old, 3 male) participated, including two of the authors (EG, CC). At least 4 subjects participated in each experiment, and each experiment included at least 3 subjects who were naïve to the purpose of the investigation. All subjects had normal or corrected to normal vision. In accordance with the guidelines of the Human Research Ethics Committee of The University of Sydney, human subjects gave informed written consent before participating. All subjects were free to withdraw from the study at any time.

On each trial, a central fixation spot, displayed for 0.3 s, preceded the stimulus, which was also presented for 0.3 s. Subjects were required, by responding with a key-press, to report the average orientation of the central patch as tilted clockwise or counterclockwise of vertical. In each experiment, the orientation of the surround denotes how far the texture in the surrounding annulus was rotated from vertical. Each session contained equal numbers of trials for annuli that were tilted clockwise or counterclockwise of vertical, to avoid a build up of adaptation to one average orientation. The orientation at which subjects were equally likely to report the central test patch as tilted clockwise or counterclockwise of vertical (the point of subjective vertical) was estimated in sessions of 120 trials. Each session included four randomly interleaved Bayesian adaptive staircases (Kontsevich & Tyler, 1999), two with clockwise- and two with counterclockwise-annuli, each consisting of 30 trials. The staircases provided two estimates of the tilt illusion in each session (each estimate is half the difference between perceived vertical obtained from one staircase with clockwise- and another with counterclockwise-annuli). For each observer, each data point is the average of at least two sessions (ie. at least eight staircases).

2.3. Model

Perceived average orientation is computed from a population of orientation selective detectors whose receptive fields lie over the central surface, and whose tuning is sensitive to the pattern in the annular surface. We call the mechanism that provides this sensitivity to the annular surface the 'surround'. The detectors' response to the central stimulus (C) is divided by its surrounds response to the annular stimulus (S), so the net response (R) is

$$R = \frac{C}{1 + k \cdot S} \tag{2}$$

where k is the overall strength of the surround. C is the correlation of the detectors' tuning curve and the orientation distribution in the central stimulus (θ_c); S is the correlation of the tuning curve of the surround and the distribution of orientations in the annular surface (θ_s). The tuning curve of each detector, $vm(\theta_{\max}, \alpha)$, is a von Mises distribution (Mardia & Jupp, 2000) with a peak at the preferred orientation of that detector ($\theta_{\max}$), and a full width at half height of α. The model does not allow for variation in bandwidth across detectors (Clifford, Wenderoth, & Spehar, 2000), so

$$C_{(\theta_{\max},\alpha,\theta_c)} = vm_{(\theta_{\max},\alpha)} * \theta_c \tag{3}$$

where * denotes correlation. The tuning curve of the surround is the difference of two von Mises distributions with different bandwidths (β and γ), which generally produces a surround whose sensitivity across orientations looks like a 'Mexican hat'. The tuning curve of the surround is symmetric around the preferred orientation of the detector. Each detector is therefore maximally suppressed by surrounds of its preferred orientation; other orientations suppress the response to a lesser extent, and can instead facilitate the response.

$$S_{(\theta_{\max},\beta,\gamma,q,\theta_s)} = (vm_{(\theta_{\max},\beta)} - q \cdot vm_{(\theta_{\max},\gamma)}) * \theta_s \tag{4}$$

The model is analogous to the surrounds defined in physiological investigations of V1 of cat and primate, but we do not mean it to be seen as a direct implementation, and there are some potential discrepancies. In V1 the orientation of the surround that is most effective at suppressing responses can depend on the orientation of the stimulus used to evoke responses from the classical receptive field, at least for complex cells; this may not be the case for simple cells (Cavanaugh, Bair, & Movshon,