CHAPTER 12

Processing of second-order stimuli in the visual cortex

Curtis L. Baker, Jr.\textsuperscript{1,*} and Isabelle Mareschal\textsuperscript{2}

\textsuperscript{1} Department of Ophthalmology, McGill University, 687 Pine Ave. W. H4-14, Montreal, PQ H3A 1A1, Canada
\textsuperscript{2} Center for Neural Science, Room 809, New York University, New York, NY 10003, USA

Abstract: Naturally occurring visual stimuli are rich in examples of objects delineated from their backgrounds simply by differences in luminance, so-called first-order stimuli, as well as those defined by differences of contrast or texture, referred to as second-order stimuli. Here we provide a brief overview of visual cortical processing of second-order stimuli, as well as some comparative background on first-order processing, concentrating on single-unit neurophysiology, but also discussing relationships to human psychophysics and to neuroimaging. The selectivity of visual cortical neurons to orientation, spatial frequency, and direction of movement of first-order, luminance-defined stimuli is conventionally understood in terms of simple linear filter models, albeit with some minor nonlinearities such as thresholding and gain control. However, these kinds of models fail entirely to account for responses of neurons to second-order stimuli such as contrast envelopes, illusory contours, or texture borders. Second-order stimuli constructed from sinusoidal components have been used to analyze the neurophysiological mechanisms of such responses; these experiments demonstrate that the same neuron can exhibit three distinct kinds of tuning to spatial frequency, and also to orientation. These results can be understood in terms of a type of nonlinear ‘filter $\rightarrow$ rectify $\rightarrow$ filter’ model, which has been widely used in human psychophysics. Finally, several general issues will be discussed, including potential artifacts in experiments with second-order stimuli, and strategies for avoiding or controlling for them; caveats about definitions of first- vs. second-order mechanisms and stimuli; the concept of form–cue invariance; and the functional significance of second-order processing.

Introduction

Our visual world is most simply thought of in terms of local variations in luminance, and thus it is not surprising that until recently most theoretical conceptions of visual processing, as well as stimuli used to characterize it, were cast in terms of responses to local luminance differences. Nevertheless, we readily detect objects which differ from their background in other aspects than luminance. In the natural image in Fig. 1A, the up-turned fishing boats are clearly delineated from their background by luminance differences, but the fine texture of the foreground vegetation is distinguished from the coarse stripes of the boat ramp much less by luminance differences than by changes in texture and contrast.

Simple visual stimuli whose principal attributes can be characterized as variations in local luminance are referred to as first-order stimuli. A first-order stimulus which has proven particularly useful for visual science experiments is a sine wave grating (Fig. 1B): note it is spatially one-dimensional, and in a static image is entirely specified by parameters of orientation, spatial frequency, spatial phase, and contrast. A smoothly moving grating is additionally characterized by its temporal frequency, in Hertz (cycles/s); thus the speed or velocity of its motion is the ratio of its temporal frequency to its spatial frequency, with dimensions of degrees per second. Notice that attributes of this stimulus, such as its orientation and spatial frequency, are entirely specified by changes in luminance across space.

Visual stimuli which portray variations in stimulus properties other than luminance, such as contrast

\textsuperscript{*} Corresponding author: C.L. Baker Jr., Department of Ophthalmology, McGill University, 687 Pine Ave. W. H4-14, Montreal, PQ H3A 1A1, Canada. Tel.: +1-514-842-1231, ext. 4819; Fax: +1-514-843-1691; E-mail curtis@astra.vision.mcgill.ca
or texture, are referred to as second-order stimuli — also sometimes called ‘non-Fourier’ stimuli (Chubb and Sperling, 1988; Cavanagh and Mather, 1989). An example of such a stimulus which has proven experimentally useful is the contrast envelope stimulus (Fig. 1C), consisting of a finely detailed pattern (carrier — here a sinewave grating at a high spatial frequency), whose contrast varies on a coarse spatial scale (envelope — here a sinewave grating of low spatial frequency). In this example, the carrier and envelope waveforms have the same orientation, though this does not necessarily have to be the case. Note that the envelope of this stimulus is not defined by a variation in luminance across space, but rather
by a variation in the change of luminance across space, i.e. a second-order statistic — the rate of change of the rate of change of luminance.

In this review, we introduce some theoretical background regarding filtering models of cortical receptive fields, building a foundation with conventional first-order processing concepts before delving into experimental data and models for second-order responses. Second-order processing will be introduced with a variety of examples from the literature, followed by a summary of our analysis of its nature at the single neuron level using contrast envelope stimuli. A type of model will be discussed, which not only provides a framework for understanding these results, but also suggests a linkage to a wide variety of other kinds of visual processing. While our principal emphasis will be on single unit neurophysiology, relationships to human psychophysics and brain imaging experiments will also be pointed out.

First-order processing

Spatial selectivity

Early visual cortex neurons are selective for a remarkable variety of stimulus attributes. Like neurons earlier in the visual pathway, they respond only to stimuli presented within a small region of visual space, the neuron’s receptive field. The seminal experiments of Hubel and Wiesel (1962) additionally demonstrated a narrow selectivity for orientation of simple luminance-defined stimuli such as bars or edges. Later studies demonstrated a similarly impressive selectivity for the orientation and also the spatial frequency of sinewave gratings (Campbell et al., 1969). These and other kinds of stimulus selectivity are easily understood for at least one kind of cortical neuron, termed ‘simple cells’ by Hubel and Wiesel (1962): the receptive fields of these neurons have spatially segregated sub-regions, alternating in excitatory and inhibitory effect (Fig. 2A). A light bar presented in an excitatory region will elicit a transient discharge at its onset, an ‘On-response’; it will also elicit a transient discharge when removed from an inhibitory region (presumably due to a rebound from inhibition), termed an ‘Off-response’. Dark bars have the opposite effects, eliciting On-responses from inhibitory zones and Off-responses from excitatory zones (Ferster, 1988; Hirsch et al., 1998). In visual cortex receptive fields, these excitatory and inhibitory regions are elongated and occur in alternation. The stimulus orientation selectivity of such a neuron can be understood in terms of the competing influences of these regions, if it is assumed that they are linearly additive (Hubel and Wiesel, 1962), prior to a simple threshold. Thus if a light bar or a grating (Fig. 2B) is presented at an orientation different from that of the elongated sub-regions, their excitatory and inhibitory influences will cancel one another out. But an oriented bar or grating at the same orientation (Fig. 2C) will cause consistent stimulation of these sub-regions, which will synergistically sum to a strong response. Hubel and Wiesel (1962) first demonstrated this correspondence between the orientation of the simple cell sub-regions and the neurons’ preferred stimulus orientation.

Selectivity for spatial frequency of sinewave gratings can be understood similarly. A maximal response is obtained only if the spacing of light and dark regions of a grating correspond well with the spacing of excitatory and inhibitory regions (Fig. 2C), otherwise there will be a failure of effective summation, even if the orientation is optimal (Fig. 2D). But the sinewave grating which maximally activates a simple type cell will be one whose light and dark bars are in register with the alternating excitatory and inhibitory sub-regions (Fig. 2C).

Such a correspondence between spacing of sub-regions and optimal grating spatial frequency has been demonstrated quantitatively (Movshon et al., 1978; Andrews and Pollen, 1979; DeAngelis et al., 1993; McLean et al., 1994).

Velocity tuning

These neurons are also selective for the speed, and often the direction of motion, of moving bars or gratings: it turns out that this spatiotemporal selectivity can also be understood in an analogous manner. Since the stimuli are spatially elongated, and presented at the neuron’s preferred orientation, they have only one spatial dimension which is relevant: the luminance profile along the orthogonal direction (i.e. the direction of motion). Therefore these spatiotemporal stimuli can be represented as two-dimen-
Fig. 2. Spatial selectivity of cortical neurons for sinewave gratings understood in terms of linear summation. (A) Cartoon model of oriented, excitatory and inhibitory sub-regions of simple-type receptive field in visual cortex. (B) Sinewave grating at differing orientation elicits poor summation of sub-regions, and no response. (C) An optimal grating is one whose dark and light bars have orientation and spacing corresponding to the neuron’s sub-regions, eliciting maximally effective summation. (D) Grating of correct orientation, but nonoptimal spatial frequency gives poor response because bars are not aligned with neuron’s sub-regions.

volutional images, as functions of time and of one spatial dimension. Such ‘space–time diagrams’ are shown in Fig. 3, for drifting bars (top row) and gratings (bottom row), which are drifting rightwards (left column) or leftwards (right column). Thus considered in this way, motion becomes a kind of orientation, but in space–time rather than space–space; higher speeds correspond to orientations closer to horizontal (note that a stationary stimulus would have a vertical orientation). Then it is easy to see by analogy that selectivity for velocity (speed and direction) can be obtained with a spatiotemporal filter which is oriented, as indicated with overlaid cartoons of excitation and inhibition in Fig. 3.

What does it mean to have a receptive field which is ‘space–time oriented’? It means that the temporal response of the neuron varies systematically with position in the receptive field. Start by considering only the excitation (+) in Fig. 3A (with time running downwards, the latency of the neuron’s response is the distance from the top of the diagram) and in this case as we proceed from left to right, the peak response is at progressively greater latencies. The inhibitory responses are constructed similarly, but in an alternating spatial relationship. As illustrated in Fig. 3, such a space–time oriented filter would have a net cancellation of excitation and inhibition in response to stimuli moving in the opposite direction to its space–time orientation (right panels), and a good summation of response for those moving at its preferred velocity (left panels). Such a linear spatiotemporal

CICERO/GALAYA B.V./CASANOVA 12: pp. 1-21
filtering account of visual motion processing has been proposed as an integral part of computational models of biological motion detection (Adelson and Bergen, 1985). The temporally delayed responses may arise from a population of lateral geniculate neurons, ‘lagged cells’ (Saul and Humphrey, 1990, 1992).

Neuronal properties understood in terms of the linear filter model

The linear filter model has provided a basis on which to understand important relationships between a neuron’s responses to different stimuli, including bars, gratings, and random noise. The space–time filter function can be experimentally measured with ‘reverse-correlation’, which entails taking the average space–time stimulus history preceding each neuronal spike, using a white noise stimulus — an example (Baker and Boulton, 1990) is shown in Fig. 4A. McLean et al. (1994) demonstrated that the space–time orientation of such plots could quantitatively predict a neuron’s preferred velocity to moving bars. The ratio of a neuron’s optimal temporal frequency to its optimal spatial frequency, in response to drifting sinewave gratings, can also be a good predictor of its optimal velocity for a moving bar (Baker, 1990) (Fig. 4B).

Of course cortical neurons do not behave linearly in every respect (Bonds, 1992), but some of the best-
known nonlinearities can be accommodated with simple modifications to the linear filter model. The most obvious nonlinearity is a threshold — cortical neurons usually have relatively little spontaneous activity, and give responses only when sufficiently excited — this can be modeled as a simple threshold operation following the linear spatiotemporal filter.

Many cortical neurons, termed ‘complex cells’, differ from simple-type cells in having On- and Off-responses which are not spatially segregated; nevertheless they also show a strong selectivity for orientation, spatial frequency, and velocity. However their selectivity can be understood in similar terms, if they add up the responses of a small pool of nearby simple-type cells, having the same selectivity but slightly differing receptive field positions (Hubel and Wiesel, 1962).

A number of other nonlinear behaviors of early cortical neurons could be accounted for with simple modifications to the linear filter model. For example, a cortical neuron maintains its selectivity to a given stimulus parameter across a range of contrasts (Albrecht et al., 1984; Skottun et al., 1987), and its response to an optimal stimulus is inhibited by superimposing a nonoptimal stimulus (Morrone et al., 1982; Bonds, 1989; DeAngelis et al., 1993). Such findings have been interpreted to indicate the operation of a nonlinear gain control mechanism which serves to ‘normalize’ the response to a stimulus, relative to other stimuli nearby in space or time. An example of a simple model which might account for such results is one (Heeger, 1992) in which the cell’s response gain is scaled by the pooled responses of nearby neurons, having different stimulus selectivities but the same receptive field location (Fig. 5A).

**Physiological basis of linear filter model**

It is easy to imagine that the excitatory and inhibitory sub-regions of a simple cell arise from linear summation of aligned sets of lateral geniculate nucleus (LGN) afferents (Hubel and Wiesel, 1962). This idea has been extended with evidence that excitatory sub-regions may receive both excitation for LGN On-cells and inhibition from LGN Off-cells (and conversely for inhibitory sub-regions), a ‘push–pull’ model which gives more truly linear behavior (Ferster, 1988; Carandini et al., 1999). This idea has been extended to direction selectivity, with lagged-type LGN cells providing space–
threshold other neurons

space-time oriented filter

threshold

Fig. 5. Quasi-linear models of first-order direction selectivity. (A) Conventional linear space–time oriented filter model, with subsequent half-squaring, and gain control from pooled responses of similar neurons selective for other orientations and spatial frequencies. (B) Alternative model, suggested by experiments on two-flash motion at large SOAs and white noise analysis, in which signals from quadrature-phase linear filters are combined nonlinearly.

However, a simple linear filter model has been challenged by evidence for a crucial role of lateral inhibitory interactions in shaping orientation selectivity (Ringach et al., 1997). Another kind of challenge has been evidence for nonlinear summation (e.g. shunting inhibition, facilitation), based on intracellular recording (Borg-Graham et al., 1998), integrative properties of NMDA receptors (Rivadulla et al., 1999), direction-selectivity for apparent motion at very large onset asynchronies (Baker and Cynader, 1994), and nonlinear white noise analysis (Baker and Boulton, 1990) (see Fig. 5B).

In spite of these and other caveats, the linear filter model has proven to be a useful metaphor for understanding a variety of spatial and temporal integrative behaviors of many early visual cortex neurons. The best characterized nonlinearities can be modeled as part of an adaptive gain control which affects the overall response gain of the neuron, but with its fundamental stimulus selectivity being controlled by the linear filter (Carandini et al., 1999).

**Human psychophysics**

A very large body of human psychophysics has supported the idea that the visual system can be considered an ensemble of such linear spatial filters, selective for orientation and spatial frequency. In such a ‘bank of filters’ scheme, threshold for stimulus detection is determined by the most sensitive filter (e.g. Campbell and Robson, 1968; Blakemore and Campbell, 1969); for review see Graham (1980). The contrast sensitivity function of the visual system is interpreted as representing the envelope of sensitivities of these filters.

Motion detection thresholds can be understood similarly (Levinson and Sekuler, 1975). For suprathreshold stimuli, the theory successfully predicts that the spatial limits of apparent motion (‘Dmax’) scale inversely with spatial frequency content of the displaced stimulus (Bischof and DiLollo, 1990; Cleary and Braddick, 1990; Boulton and Baker, 1993b).

Contrast gain control has also proven a useful concept in models of first-order motion psychophysics, for example in how apparent motion is affected by unequal durations (Bischof and DiLollo, 1996) or differing contrasts (Georgeson and Scott-Samuel, 1999) on successive exposures.

**Second-order processing: detection of borders defined by contrast or texture**

**Form–cue invariance**

The earliest report of neuronal responses to second-order motion was by Albright (1992), who showed that neurons in primate area MT/V5 responded to...
moving regions of dynamic noise, on a static noise background (Fig. 6B). The same neurons also responded to luminance-defined stimuli, with the same preferred direction of motion, leading to the notion of ‘form–cue invariance’ — the property of a neuron being selective to a stimulus attribute (e.g. direction of motion) regardless of the ‘cue’ defining it (here, either temporal contrast, or luminance contrast). Presumably neurons having such form–cue invariance would be functionally useful in mediating responses to overall objects regardless of the cue (or, in our terms, carrier) by which they are segregated from the background.

Similar findings of form cue invariance for bar-shaped stimuli have been reported in other brain areas, for different stimuli and stimulus attributes. Geesaman and Andersen (1996) demonstrated a similar invariance for complex motion attributes in MT/V5 neurons, using stimuli defined by a variety of carrier types. Chaudhuri and Albright (1997) described responses to dynamic noise bars in primate V1 — rarely direction selective (for either kind of stimulus), but having the same selectivity to orientation of the bar, whether defined by luminance or by dynamic noise. In area 18 neurons of the cat, Leventhal et al. (1998) found single neurons with consistent direction selectivity to bars defined either by luminance, or by gratings or texture on a gray background of the same luminance (Fig. 6C,D).

Using drifting periodic contrast modulation stimuli (Fig. 1C), Zhou and Baker (1993, 1996) also found single neurons in cat areas 17 and 18, which
responded both to luminance and contrast-defined stimuli with a consistent preferred direction of motion. For a given neuron, the preferred contrast envelope spatial frequency was similar to, but often significantly lower than, the optimal spatial frequency for luminance gratings (Zhou and Baker, 1996; Mareschal and Baker, 1998b). Area 18 neurons were shown to have consistent optimal orientations for envelopes and luminance gratings (Mareschal and Baker, 1998a, 1999). Unlike the above studies, however, the use of narrowband carrier waveforms led to finding surprisingly narrow tuning to both carrier spatial frequency and orientation (Zhou and Baker, 1996; Mareschal and Baker, 1999); these findings will be described much more extensively below. This specificity of a neuron’s response to carrier parameters may provide an important qualification to the notion of form–cue invariance, though it does not preclude the possibility that other neurons might exist whose envelope responses are invariant to the nature of the carrier.

Illusory contours, texture borders

A rather differently motivated line of research involved searching for neural correlates of human perception of ‘illusory contours’, formed for example by abutting gratings which have an abrupt discontinuity at their juncture (Fig. 6E); these stimuli have no luminance-defined edge along the illusory contour, and so they might be considered as ‘second-order’ (see below). Neurons have been reported in cat area 18 (Peterhans and von der Heydt, 1991) and in primate area V2 (Grosof et al., 1993), to respond to motion of such illusory contours with a similar selectivity to orientation and direction of motion as for luminance edges.

Yet another line of research has investigated possible mechanisms of texture segregation, using stimuli having a central region differing from its background in local texture properties (Fig. 6F) — again, with the same net luminances in the figure and ground, these could be considered second-order stimuli. For example, Olavarria et al. (1992) demonstrated responses in primate MT/V5 neurons using textures of small lines, differing in orientation, and responses to texture borders were also found in primate area V1 (Nothdurft et al., 1999).

Human psychophysics

Henning et al. (1975) introduced contrast envelope stimuli, like that shown in Fig. 1C, arguing that their detection could not be accounted for in terms of linear filter schemes. Several subsequent studies (e.g. Derrington and Badcock, 1985; Ledgeway and Smith, 1994b) indicated that contrast envelopes were detected by a different mechanism.

Chubb and Sperling (1988) introduced ‘drift-balanced’ motion stimuli which were constructed such that their motion would be invisible to any motion detector based on linear filtering: this was achieved by using a dynamic noise carrier which has no coherent motion signals. These authors also introduced early versions of ‘filter → rectify → filter’ models (see below), and proposed an analogous kind of processing for spatial vision. This work inspired many subsequent studies of such ‘non-Fourier’ processing, in spatial vision tasks (Sutter et al., 1989; Graham and Sutter, 1998) as well as motion perception (see Smith, 1994).

A continuing question has been whether second-order stimulus attributes are processed by a distinct mechanism from that which handles first-order ones. Numerous studies have provided evidence for such separate processing, but the question has remained controversial (Taub et al., 1997) (see below). Particularly compelling psychophysical evidence has emerged from recent work showing that adaptation to first- or second-order stimuli selectively affects thresholds only for target stimuli of the same type, but that cross-adaptation is absent (Nishida et al., 1997), and that subthreshold facilitation is similarly specific to the type of stimulus (Schofield and Georgeson, 1999).

Analysis of the mechanism of second-order processing

Use of contrast envelope stimuli to characterize second-order responses in single neurons

To demonstrate and characterize second-order responses from early visual cortex neurons, we employed contrast envelope stimuli (Fig. 1C) consisting of a high spatial frequency sinewave carrier whose contrast is modulated by a lower spatial frequency,
Fig. 7. Responses of linear model and cortical neuron to first- and second-order stimuli. (A) Sincwave grating stimulus which is optimal for a linear model (superimposed cartoon receptive field). (B) Post-stimulus time histograms of cortical simple cell’s response to drifting grating, at a series of temporal frequencies in both preferred and nonpreferred directions. Note modulated discharge, indicative of a simple type cell, preference for temporal frequencies of 3–12 Hz, and very little direction selectivity. (C) Contrast envelope stimulus, which would fail to activate a linear receptive field. (D) Response of same neuron as in A, to contrast envelope drifting at a series of temporal frequencies. Note that response is again modulated, but with a greater degree of direction selectivity and tuning to lower temporal frequencies of about 1.5–6 Hz. Post-stimulus time histograms (B and D) from Mareschal and Baker (1998b).
drifting sinewave envelope (e.g. Zhou and Baker, 1993; Mareschal and Baker, 1998a). In most of these experiments the carrier was stationary, and the carrier and envelope were of the same orientation (i.e. the stimuli were spatially one-dimensional). Fig. 7 shows images of a luminance grating (A) and contrast envelope (C), each with a cartoon spatially linear receptive field overlaid on the stimuli, to illustrate how such a neuron should respond well to the luminance grating, but not the contrast envelope: in the latter case, the effects of the fine-grain light and dark regions of the carrier would cancel out within the much larger summation sub-regions of the linear receptive field. To put this another (equivalent) way, the carrier spatial frequency is so high as to be outside the passband of the luminance-responsive linear receptive field.

Each neuron’s receptive field was first characterized conventionally using sinewave gratings, establishing its preferred orientation, spatial frequency, temporal frequency and direction selectivity. Then contrast envelope stimuli were tested, setting the envelope parameters in accordance with those found optimal for the luminance gratings, and testing a series of high carrier spatial frequencies. At least half the neurons in area 18 of the cat, and a minority of those in area 17, were found to respond significantly to contrast envelopes (Zhou and Baker, 1993). In most cases, a neuron’s response to the envelope stimulus was weaker than to luminance gratings, but both kinds of stimuli elicited consistent preferred directions of motion, though with differing degrees of direction selectivity (Mareschal and Baker, 1998b; Zhou and Baker, 1996). Both simple and complex type cells were envelope responsive, though more frequently for complex cells. Simple cells showed modulated discharges to drifting stimuli of either type (Fig. 7B,D), demonstrating that the nonlinearity underlying second-order responses is not related to that which distinguishes complex from simple type cells.

Three kinds of spatial frequency tuning in the same neuron

A theoretically important early result was that envelope responses were highly dependent on carrier spatial frequency, typically being tuned to frequencies 5–30 times greater than for luminance gratings (Zhou and Baker, 1996; Mareschal and Baker, 1999) (see Fig. 8A,B). Bandpass tuning was also found for envelope spatial frequency (Fig. 8C), for frequencies similar to (though often somewhat lower than) those for luminance gratings (Mareschal and Baker, 1998b). Bandwidths of all three types of spatial tuning were similarly narrow (Mareschal and Baker, 1999), such that the carrier bandwidth was unambiguously well outside the luminance passband of the neuron (e.g. in Fig. 8, the tuning curves in A and B are nonoverlapping).

Optimal carrier spatial frequencies for these neurons were typically 1–3 cycles per degree (cpd), sometimes as high as 4 cpd, which is very high relative to the cat’s spatial acuity of 5–10 cpd (Cleland et al., 1979; Ulrich et al., 1981). In this respect, the cat is a highly advantageous choice of species for studies of second-order responses, since practical use of spatial frequencies which are comparably high for the monkey would be awkward.

Across the population of sampled neurons, the optimal carrier spatial frequency was not in any characteristic fixed ratio to either the optimal luminance or envelope spatial frequencies (Zhou and Baker, 1996; Mareschal and Baker, 1999). A given individual neuron’s optimal carrier spatial frequency was not dependent on the envelope spatial frequency at which it was measured. The independence of these tunings to luminance, carrier, and envelope spatial frequencies suggest that they are mediated by separate mechanisms.

Three kinds of orientation tuning in the same neuron

While our earlier studies employed spatially one-dimensional stimuli produced by lookup table animation, our later use of digital movies (Brainard, 1997) allowed us to vary independently the orientations of carrier and envelope in our second-order stimuli (Mareschal and Baker, 1998a, 1999). Fig. 9 shows orientation tuning curves as polar plots, with distance from the origin indicating strength of response. Luminance gratings always elicited narrow orientation tuning (Fig. 9A) as expected; second-order stimuli also showed similarly narrow tuning to the envelope waveform (the carrier orientation was held constant). Most interesting was that many neurons also showed a bandpass tuning to the carrier (with envelope orientation held constant (Fig. 9C)).
A) Luminance S.F. Tuning

B) Carrier S.F. Tuning

C) Envelope S.F. Tuning

Fig. 8. Three kinds of spatial frequency tuning in a single A18 neuron. In each case, average response of the neuron is plotted against spatial frequency, with examples of stimuli shown below the graphs. (A) Response to luminance grating is sharply tuned, peaking about 0.15 cpd. (B) Response to contrast envelope, with fixed envelope spatial frequency and varying carrier frequency; response is again very sharp, tuned to much higher frequency of about 1.5 cpd. (C) Response to contrast envelope, with fixed carrier frequency and varying envelope frequency. Optimum is about 0.09 cpd similar to, but slightly lower than for luminance grating in A. Reprinted with the permission of Cambridge University Press, from Mareschal and Baker (1999).

though often with a much broader bandwidth. This result suggests that the mechanism of carrier selectivity is almost certainly cortical (Mareschal and Baker, 1998a, 1999).

Fig 10 shows scatterplots of the measured optimal orientations, across the population of sampled neurons. Neurons’ optimal orientation for the envelope was very similar to that for luminance gratings (Fig. 10A), consistent with a ‘form–cue invariance’ (Albright, 1992); but a given neuron’s preferred carrier orientation was totally unpredictable from its optimal envelope or luminance orientation (Fig. 10B). The independence of carrier orientation tuning again suggests that the carrier is processed by a different set of neurons than those which determine selectivity for the envelope component, or for luminance gratings.

Temporal properties

A large proportion of neurons in cat area 17 and 18 are at least partially direction-selective to luminance gratings, and this was seen also in the same neurons’ responses to contrast envelopes. A given neuron always showed a consistent preferred direction for luminance and envelope motion, though the quantitative degree of directionality could differ considerably (Zhou and Baker, 1996; Mareschal and Baker, 1998b; Fig. 7B,D). Envelope responses were temporally bandpass, with optimal temporal frequencies of typically 2–6 Hz; the bandwidths of tuning for envelope and luminance motion were quantitatively similar. A given neuron’s optimal temporal frequency for envelope motion was similar to, but often quantitatively somewhat less, than that for luminance grating motion; in addition, simple type cells also exhibited much greater temporal phase lags to contrast envelope responses than to luminance gratings (Mareschal and Baker, 1998b).

These findings of slower temporal properties for second-order than for first-order motion are consistent with the majority of human psychophysical evidence — for example, a lower temporal acuity for contrast envelopes than for luminance grat-
Fig. 9. Three kinds of orientation and directional tuning of an A18 neuron. Data are shown as polar plots: distance from origin represents magnitude of neuron’s response, angle represents stimulus orientation. In each case, three examples of the stimulus are shown alongside the polar plot. (A) Response to luminance-defined sinewave grating. (B) Response to contrast envelope stimulus, varying envelope orientation while holding carrier orientation fixed. (C) Same as B, but varying carrier orientation, with envelope orientation fixed. Reprinted with the permission of Cambridge University Press, from Mareschal and Baker (1999).

A model of envelope-responsive neurons

That a neuron can show different preferred spatial and temporal frequency preferences, and differing degrees of direction selectivity, to luminance and envelope stimuli suggests that its firing reflects a combination of inputs from two distinct pathways or ‘streams’, one mediating first-order and the other second-order responses. An adequate model for the one handling luminance grating responses could be a conventional linear filter model like that discussed above, but such a model would be entirely incapable of producing envelope responses for reasons outlined earlier.

Contrast envelopes can be thought of as analogous to AM (‘amplitude modulation’) radio signals, in which the amplitude of a radio frequency carrier has been modulated by a much lower frequency audio signal; in an AM radio, the audio signal is ‘demodulated’ by rectification followed by low-pass filtering to smooth out the irregularities introduced by the high frequency carrier. While a signal-processing scheme of this kind might mediate neuronal responses to contrast envelopes, some important modifications are needed to account for the physiological data. Firstly, the spatial and temporal tuning characteristics for envelope responses require that the filter following the rectifier be band-pass rather than low-pass. Secondly, to achieve the tuning to carrier spatial frequency and orientation, we need a front-end filter prior to the rectification (analogous to the radio frequency tuner which selects which station we receive). The intervening rectification ensures that the fine-grain positive and negative portions of the carrier-frequency signal do not cancel one another when smoothed by the later filter. Thus we are led to a ‘filter → rectify → filter’ (F→R→F) model, for the second-order stream mediating a cortical neuron’s response to contrast envelopes (Fig. 11).

This is a very attractive idea because the results of our neurophysiological data map on to it so straightforwardly. The very different tunings for orientation and spatial frequency of the carrier and envelope are
Fig. 10. Relationships of different kinds of orientation tuning. (A) Optimal orientation to luminance grating, compared to optimal orientation of envelope of a contrast modulation stimulus. Each point represents measurements from one neuron; solid line shows unity ratio, dashed lines represent orthogonal relative orientations. Note that most neurons showed very similar orientation preference for luminance gratings and envelopes. (B) Similar plot, for orientation of carrier of contrast modulation. Note the lack of correlation between the two kinds of tuning, with many neurons preferring carrier orientations close to orthogonal to the envelope (darkened symbols). From Mareschal and Baker (1998a).

Fig. 11. Model of cortical neuron response, having separate parallel signal-processing pathways for first- and second-order stimuli. In the top path, luminance-defined stimuli are processed conventionally with a linear spatiotemporal filter. The bottom path processes second-order stimuli, and has a ‘filter → rectify → filter’ cascade consisting of early linear filtering subunits, a nonlinearity (e.g. rectification), and a late linear filter. The early filters are shown superimposed on a contrast envelope stimulus, to illustrate how they could mediate selectivity for carrier properties, while the late filter is shown on the full-wave rectified stimulus to illustrate its selectivity for properties of the envelope. The neuron’s response reflects a combination of both pathways.
easily understood: the selectivity for stimulus parameters of the carrier reflect the properties of the first filter, while the selectivity for envelope attributes reflects the tuning of the second filter.

Fig. 11 illustrates the complete model with separate streams or paths for processing first- and second-order stimuli. A simple luminance grating can elicit a response only via the first-order pathway: regardless of its spatial frequency, it can never satisfy the conflicting demands of the highly discrepant spatial frequency tunings of the early and late filters of the F→R→F stream. Similarly, a contrast envelope stimulus will exert an effect on the final output only via the F→R→F stream, for the reasons discussed already.

Many findings from human psychophysics can be similarly understood in terms of a F→R→F scheme. For example, second-order operations which are selective for carrier spatial frequency and orientation (Green, 1986; Ledgeway and Smith, 1994b; Dakin and Mareschal, 2000) may indicate characteristics of the early filter, while specificity to attributes of the envelope could reflect properties of the late filter (Nishida et al., 1997). But in addition to second-order processing, F→R→F models have also been used in various forms to understand spatial vision tasks such as detection of texture borders (Landy, 1991; Graham and Sutter, 1998), illusory contours (Wilson, 1999), terminators (Loffler and Orbach, 1999), and non-Cartesian stimuli (Gallant et al., 1993; Gallant et al., 1996; Wilson, 1999). By making either the early or late filters selective for other attributes, this kind of model might conceivably handle an even wider variety of stimuli, such as ‘theta motion’ (Zanker, 1993) if the early filters are selective to direction or speed of motion, or second-order stereopsis (Hess and Wilcox, 1994) if the late filters are disparity-selective. Appropriately arranged early and late filters could be designed for specific responses to optic flow invariants (expansion–contraction, rotation, etc.), or for contour integration (Field et al., 1993).

Possible physiological substrates of second-order processing

In the model of Fig. 11, the upper branch which handles first-order processing could be any of the quasi-linear neurons of early visual cortical areas, though the low preferred spatial frequencies in our data suggest a greater contribution from second-tier areas (A18, V2). Our findings of selectivity for carrier spatial frequency and orientation suggest that the early filters of the F→R→F branch must be cortical, and the quite high optimal carrier frequencies suggest quasi-linear neurons in first-tier areas (A17, V1) as very likely candidates.

The more interesting question is the neural substrate of the later filters of the F→R→F branch: neurons corresponding to this stage, if they exist, would in this scheme be exclusively responsive to second-order stimuli. They would not respond to full-field sinewave gratings, of any spatial frequency or orientation, but would respond to contrast envelope stimuli, with the same selectivities already described for spatial frequency and orientation of both carrier and envelope, and in many cases also to direction of envelope motion. Neurons having all these properties have never been reported, though this might well be a consequence of not having used appropriate ‘search stimuli’, and use of protocols (such as ours) in which neurons are always first characterized with sinewave gratings.

The possible existence of neurons exclusively responsive to second-order stimuli in humans is supported by psychophysics showing separable adaptation and masking subthreshold facilitation of first- and second-order stimuli (Nishida et al., 1997; Schofield and Georgeson, 1999). A possible extrastriate locus is suggested by brain imaging experiments (Smith et al., 1998) and by descriptions of brain-damaged patients having selective deficits for either first- or second-order motion (Vaina et al., 1999). A higher order site would also be consistent with the slower temporal processing characteristic of second-order processing.

Utility of stimuli composed of sinusoidal components

While it might seem counterintuitive to use stimuli comprised of sinusoidal components to analyze nonlinear mechanisms of second-order processing, this strategy has actually proven invaluable. The use of a band-limited carrier, at a spatial frequency well above the neuron’s measured spatial frequency response to simple gratings, provides a way to ensure...
definitively that a neuron’s response is genuinely second-order, and not simply mediated by the same linear filtering assumed to mediate conventional grat- ing responses. Using narrowband carriers also led to the finding of surprisingly narrow carrier spatial frequency tuning (Zhou and Baker, 1993; Zhou and Baker, 1996), and later to demonstration of independent orientation tuning for carrier and envelope (Mareschal and Baker, 1998a, 1999). Similarly, sinusoidal envelope waveforms allow measurement of narrow tunings to envelope spatial and temporal frequency. For both carrier and envelope, sinusoidal signals permit useful quantitative comparisons of bandwidths as well as optimal values to luminance responses in the same and in other neurons. Finally, the close relationship of sinusoidal signals to linear filtering theory facilitates interpretation of the data in terms of a nonlinear model composed of linear elements.

Is it really second order?

Perceived change-of-position

The ability to perceive a visual stimulus, or for a neuron to respond to it, might not necessarily require a separate specialized, low-level mechanism. For example, human perception of second-order motion might be mediated by a high-level appreciation of the change-of-position of salient stimulus attributes, sometimes called ‘feature-tracking’ (e.g. Lu and Sperling, 1995; Derrington and Ukkonen, 1999; Seiffert and Cavanagh, 1999). However, much evidence argues otherwise: motion after-effects to second-order motion (Ledgeway, 1994a; Ledgeway and Smith, 1994a; Nishida et al., 1994; Nishida and Sato, 1995), a very different vulnerability to added noise (Bex and Baker, 1999), and direction-selective single neuron responses as discussed above.

Early nonlinearity

A more serious and frequently touted possibility is that a subtle nonlinearity prior to the linear filtering of first-order processing might accomplish the necessary demodulation (i.e. by giving a slightly different gain to light and dark regions of the carrier, preventing cancellation) (e.g. Burton, 1973; Taub et al., 1997). To plausibly accommodate the body of psychophysical and neurophysiological findings which led to the notion of linear summation, such an ‘early nonlinearity’ would have to be slight. Using a luminance-nulling technique to psychophysically measure the contribution of such early nonlinear demodulation in human vision, (Scott-Samuel and Georges, 1999) demonstrated that second-order motion could be mediated this way only at quite high stimulus contrasts, and at temporal frequencies above about 8 Hz. Their findings are consistent with those of (Holliday and Anderson, 1994) using a different method, who reported genuine second-order motion at low temporal frequencies but operation of a different mechanism at higher rates. (Note that our single neuron responses to contrast envelopes showed optimal temporal frequencies of 2–6 Hz.)

Such a simple early nonlinearity could not account for most neurophysiological reports of second-order responses. The necessity of only a slight nonlinearity implies that responses would be extremely weak to second-order stimuli, whereas at least some neurons can give responses roughly equal in magnitude to those from comparable first-order stimuli (e.g. (Leventhal et al., 1998); see Fig. 7B,D). Such an early nonlinearity scheme would predict the same tuning to envelope spatial frequency as to luminance spatial frequency, and would fail to predict the selectivity to carrier spatial frequency and orientation which we observe.

First-order intrusion

Another potential pitfall is that responses to second-order stimuli might be mediated by their first-order attributes, i.e. responses mediated by the carrier rather than by the envelope. A very simple and instructive example is that of a conventional complex type cell (or, an ‘energy model’ (Adelson and Bergen, 1985) of first-order motion), which would give a strong response to an envelope stimulus whose carrier is its optimal grating: literally a response to a second-order stimulus, but critically lacking in selectivity for attributes of the envelope such as its direction of motion, spatial frequency or orientation. In general, one must ask whether a second-order stimulus is actually being detected by a first-order mechanism.
For A17/18 neurons, we could cleanly avoid this problem by constraining ourselves to carrier spatial frequencies well outside a given neuron’s measured spatial frequency pass-band (Zhou and Baker, 1994; Mareschal and Baker, 1998b). Effective strategies for human psychophysics include use of stochastic carriers (Chubb and Sperling, 1988; Smith and Ledgeway, 1997), tasks requiring judgments of envelope-specific attributes, and inference from the pattern of the data that only envelope attributes could have mediated performance (Boulton and Baker, 1993b; Baker and Hess, 1998). This remains a particularly challenging problem for the design of brain imaging experiments on second-order processing.

The serious potential for confusion to be generated by use of the same terminology (first- and second-order), for both stimuli and mechanisms, is quite comparable to the analogous unfortunate situation for luminance vs. color. Just as we must beware that a chromatic stimulus might inadvertently activate a luminance mechanism (e.g. via optical chromatic aberrations), so also the response to a second-order stimulus can sometimes be due to ‘cross-activation’ of a first-order mechanism (e.g. Smith and Ledgeway, 1997).

When used with reference to stimuli, the terms first- and second-order best refer to stimulus attributes (e.g. orientation or direction of motion), rather than to stimuli themselves. Indeed, many stimuli may have both first- and second-order attributes, for example in the carrier and envelope of a contrast modulation.

_Caveats about the definition of ‘second-order’_

Contemporary vision researchers would typically use the term ‘second-order’ to denote stimuli whose principal attributes, such as direction of motion, are not defined by first-order statistics of luminance or color (Cavanagh and Mather, 1989) — this image statistical definition has it origins in work on texture segregation (Julesz, 1981). In view of recent human psychophysics demonstrating that quasi-linear apparent motion appears to use only luminance and not chromatic mechanisms, whereas nonlinear (second-order) motion does use chromatic information (Baker et al., 1998; Yoshizawa et al., 2000), it seems unwise to exclude color as one of the defining properties of second-order stimuli.

A related approach might be to define second-order stimuli in terms of nonlinear processing, with first-order corresponding to linear filtering. However, it then becomes a ‘trash-can category’ which is overly inclusive. For example, cross-orientation inhibition (Morrone et al., 1982; Bonds, 1989) is certainly highly nonlinear (and, literally, a ‘second-order’ interaction), but it is thought to reflect operation of a contrast gain control, of linear (first-order) filtering mechanisms. One way around this is to tie the definition to F→R→F processing (Wilson, 1999); however, while the F→R→F model has proven an extremely useful metaphor for understanding second-order processing, it seems undesirable to have a rigidly model-dependent definition.

Many of these semantic problems are complicated by the now prevailing usage of the terms ‘first- and second-order’ to refer both to visual stimuli and to mechanisms which are specific to their processing. This usage is in retrospect undesirable, but now seems too entrenched to attempt prescriptive reform. As noted already, this situation and its problems are reminiscent of those with using ‘luminance’ and ‘chromatic’ to refer both to stimuli and to mechanisms.

Our interim (working) definition of second-order stimuli would be those whose principal attributes are defined by variations in contrast or texture. This definition leaves ill-defined a number of stimuli whose status seems presently uncertain, for example spatial differentials of local motion (e.g. rotation, expansion–contraction), ‘theta motion’ (Zanker, 1993), cyclopean motion (Smith and Scott-Samuel, 1998), plaid coherence, contour integration (Field et al., 1993), non-Cartesian stimuli (Gallant et al., 1996; Wilson, 1999), and ‘salience-defined’ (also termed ‘third-order’) stimuli (Lu and Sperling, 1995). In future, when we have learned much more about the neuronal processing of these stimuli, it may be feasible to construct more mature definitions which are biologically rather than mathematically motivated.

_Function of second-order processing_

One could argue that second-order processing is of very limited functional significance, since ‘pure’ second-order stimuli are ecologically rare — most
natural contours contain luminance as well as textural differences. In this view, second-order processing might be a marginally useful evolutionary accident which merely supplements abundant luminance-defined information, and might be critically useful for only a few special situations. Such a limited role seems unlikely, however, in view of the lack of difficulty in finding neurons which respond to such stimuli, and the impressive specificity of those neurons to characteristics of second-order stimuli (particularly, to carrier spatial frequency). But there are computationally motivated reasons why second-order information might be important.

While first-order luminance information is ecologically abundant, it is often highly unreliable in defining objects due to the vagaries of illumination — indeed the ambiguities introduced by shadows and by gradients of shading are an important part of what makes vision a difficult problem (Marr, 1982). False edges provided by shadows can be disambiguated by differences of texture, color, motion, or disparity — all second-order cues, or their close relatives. Indeed, we might turn the above argument on its head, and propose that a good computational strategy for image understanding might be to ignore first-order information entirely, and only use second-order cues! Perhaps it should not be surprising that many neurons in extrastriate cortical areas respond relatively poorly to sinewave gratings (e.g. Gallant et al., 1996).

Natural images have first-order statistics like those of fractals, with an amplitude that falls with the reciprocal of spatial frequency (Field, 1987). But random fractal images which satisfy this relationship look very unnatural, because they lack the rich spatial ‘structure’ of natural images (Field, 1994), from edges, terminations, occlusions, etc. Natural images also differ from random fractals in being ‘sparse’ (i.e. high kurtosis in the luminance distribution), having a differentiation of local regions varying in contrast (Field, 1994). Thus it seems a possibility that second-order processing mechanisms might play an important role in representing the higher-order statistical structure in natural images. For example, a $F \rightarrow R \rightarrow F$ model responds best to images that are sparse within a (carrier) spatial frequency/orientation passband.

Thus asking, “What is second-order analysis good for?”, might be akin to earlier, similar questions about orientation-selective and spatial frequency-selective neurons or psychophysical channels: the answer might not simply be to encode one particular class of stimulus or solve one specific problem, but rather to provide a general-purpose, low-level representation, which serves as a ‘basis set’ for representing higher-order structural features occurring in natural images, for example $F \rightarrow R \rightarrow F$ operators could play an important role in detection of the pattern motion of plaids (Wilson et al., 1992), and contribute to higher order form recognition (Wilkinson et al., 1998; Wilson, 1999).

Natural images contain visual information at a wide range of spatial scales, from small striations of the bark on a tree, to a forest of trees. While the visual system can represent this information with neurons spanning a large range of receptive field sizes or spatial frequencies, it must also be able to encode attributes of large-scale structures defined by finely scaled, textured surfaces, as well as sparse images containing objects which are largely devoid of texture. Second-order processing may provide a common basis for this kind of representation.

Acknowledgements

We wish to thank David Field, Fred Kingdom, Stacey Maher, and Nick Scott-Samuel for useful comments on the manuscript. This research was supported by Canadian MRC (MA 9685) and NSERC (OPG0001978) grants to C.B., an FCAR fellowship to I.M., and supplementary funds from FRSQ.

References


1 While color has often been considered first-order, as noted previously, our findings suggest its processing is more closely related to second-order operations (Baker et al., 1998; Yoshizawa et al., 2000).


Graham, N. (1980) Spatial-frequency channels in the human vi-


