Vision Research 51 (2011) 1397–1430

Contents lists available at ScienceDirect

Vision Research

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

Review

Beyond multiple pattern analyzers modeled as linear filters (as classical V1 simple cells): Useful additions of the last 25 years

Norma V. Graham *

Dept. of Psychology, Mail Code 5501, Columbia University, NY, NY 10027, USA

ARTICLE INFO

Article history:
Received 3 May 2010
Received in revised form 7 February 2011
Available online 15 February 2011

Keywords:
FRF
Second order
Contrast gain
Contrast normalization
Contrast comparison
Surround suppression
Lateral connections
Contour integration

ABSTRACT

This review briefly discusses processes that have been suggested in the last 25 years as important to the intermediate stages of visual processing of patterns. Five categories of processes are presented: (1) Higher-order processes including *FRF* structures; (2) Divisive contrast nonlinearities including contrast normalization; (3) Subtractive contrast nonlinearities including contrast comparison; (4) Non-classical receptive fields (surround suppression, cross-orientation inhibition); (5) Contour integration.

0. Introduction

We were asked for the 50th anniversary issue of *Vision Research* to highlight new knowledge on important questions open 25 years ago and on which progress had (or had not) been made. In a happy coincidence for me, 25 years ago I had just completed the draft of a book (published as Graham, 1989, summarized in a short paper Graham, 1992). I am reasonably certain, therefore, of what was known 25 years ago about a set of questions in pattern vision, or at least of what I thought was known.

The *simple multiple-analyzers model* shown in Fig. 1 top panel seemed at that time to be a very good model of pattern vision, particularly when you limited your attention to experiments using visual patterns of *near-threshold contrast*. In this model there were multiple analyzers, each of which was selectively sensitive on at least one of the multiple dimensions of pattern vision. These dimensions included spatial frequency, spatial position, orientation, direction of motion, and a number of others. To get from these multiple analyzers to the observer's response the model used a *decision rule* that was just a very simple combination of the multiple analyzers' outputs, e.g.: the observer says the pattern is vertical if and only if the analyzer producing the biggest output is the analyzer having peak sensitivity at the vertical orientation.

An aside about terms and the glossary: Many terms used in the main text without much definition are described more fully in the glossary. These terms appear in italics at least when they are first introduced. (Some italicized terms are not in the glossary but are italicized for momentary emphasis, or because they are titles of other sections in this review, or for other conventional reasons.)

The physiological substrate for an analyzer might be considered to be either a single neuron, or a set of neurons that are homogeneous in some sense (e.g. all sensitive to vertical orientation but in different spatial positions). To minimize blatantly neurophysiological terms when talking about concepts used to explain behavior, the word *unit* will be used here to mean a more abstract entity analogous to a single neuron, and the word *channel* will be used here to mean a more abstract entity analogous to a set of neurons that are homogeneous in some sense. The word *receptive field*, although it has its origin in the neurophysiological literature, is less blatantly neural, and both units/channels and neuron/neurons will be said to have *receptive fields*.

Twenty-five years ago the analyzers were generally based on the classical model of one of the types of neurons Hubel and Wiesel had discovered in cortical area V1 (striate cortex), the type called *simple cells*. According to the classical model, a simple cell adds and subtracts the weighted amount of stimulation of the excitatory and inhibitory areas in its *receptive field*. Since a neuron's output is spikes, and since spike rates lower than zero do not exist, a *half-wave rectification* or similar nonlinearity was assumed to change any below-zero result of the addition and subtraction into zero.

* Fax: +1 212 854 3609.
E-mail address: nvg1@columbia.edu

0042-6989
doi:10.1016/j.visres.2011.02.007

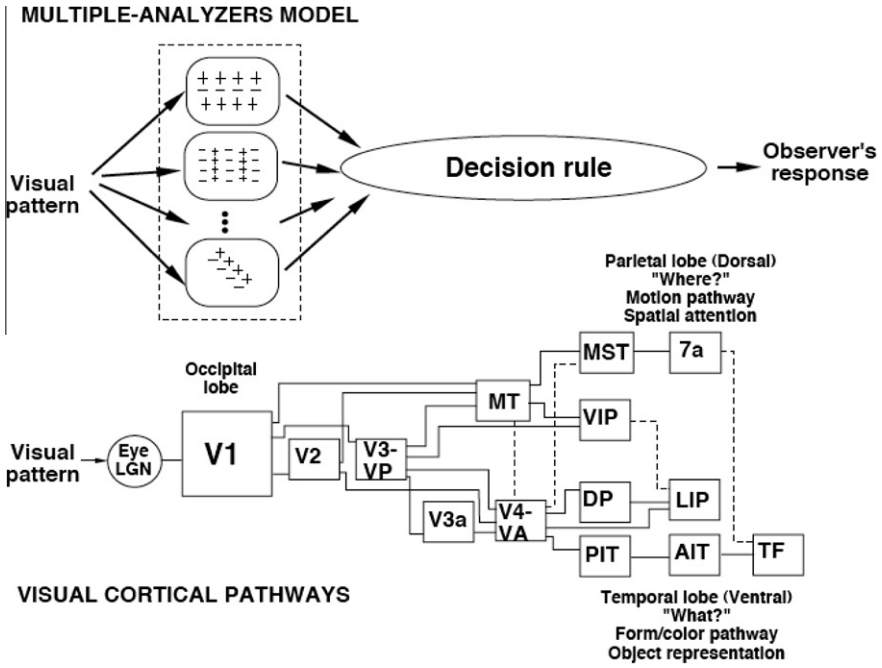

Fig. 1. Simple multiple-analyzers model (top). Simplified sketch of visual pathways (bottom, based on Movshon (1990)).

(This rectification was left implicit frequently, while referring to the model as a linear system. This common practice has led to some confusion.) Thus we will define a *classical V1 simple cell* as a *linear system* (an adding and subtracting device) followed by a *half-wave rectification*.

This classical V1 simple cell model was in decent accord with known physiological results of the time. It turned out NOT to be in complete accord with the physiology, however, as is discussed below. Hence a distinction is made here between a *classical V1 simple cell* (one that is perfectly described by the classical model) and a *simple cell* (any V1 cell that would be classified as a simple cell by the criteria ordinarily used by physiologists of Hubel and Wiesel's time or today).

The simple multiple-analyzers model shown in the top panel of Fig. 1 was and is a very good account, qualitatively and quantitatively, of the results of psychophysical experiments using near-threshold contrasts. And by 1985 there were hundreds of published papers each typically with many such experiments. It was quite clear by that time, however, that area V1 was only one of 10 or more different areas in the cortex devoted to vision. See sketch in Fig. 1 bottom panel. (Lennie (1998) and Hochberg (1998) give an interesting perspective on the complexity and functionality of a subset of these cortical areas, V1 through V4 and MT.) The success of this simple multiple-analyzers model seemed almost magical therefore. How could a model account for so many experimental results when it represented most areas of visual cortex and the whole rest of the brain by a simple decision rule? One possible explanation of the magic is this: In response to near-threshold patterns, only a small proportion of the analyzers are being stimulated above their baseline. Perhaps this sparseness of information going upstream limits the kinds of processing that the higher levels can do, and limits them to being described by simple decision rules because such rules may be close to optimal given the sparseness. It is as if the near-threshold experiments made all higher levels of visual processing transparent, therefore allowing the properties of the low-level analyzers to be seen.

Even for near-threshold experiments, there were hints of extra non-linear inhibition among analyzers (Graham, 1989). And for supra-threshold psychophysical results (although not in the 1989 book, there were many tens of published papers I knew very well) a satisfactory decision rule would either be very complicated or very vague. Given that V1 is one area of many known visual areas in the brain, and that even V1's physiology was known to be more complicated than the classical model of V1 simple cells, this was not very surprising. But it was unclear how to improve the model and yet keep it tractable and useful.

In the last 25 years a number of processes have been suggested as possible additions to the simple multiple-analyzer model of Fig. 1, additions which have the flavor of intermediate stages of visual processing, of stages for which the physiological substrate might be V1 (or perhaps V2 or V3). These stages might be called the "hidden stages" as they are far from the light image that stimulates the eye and far also from both conscious perception and the control of action. Several of these suggested additions to the simple multiple-analyzers model are the substance of this review. They have been suggested as explanations of pattern vision in general, both for psychophysical and neurophysiological results. Here the discussion is focused on the psychophysical side, but the neurophysiological is too intertwined in the history to be ignored entirely. (A multi-author paper from a mini-symposium in the early 2000s – Carandini et al., 2005 – is one convenient source for more about the physiological side as are other of the articles in this volume.)

I will discuss these additional processes as falling into the five categories listed below, and the rest of the article will be organized by these five categories. The general categories are neither mutually exclusive nor exhaustive. There are specific examples in each category, however, which are distinct from examples in other categories and which seem to present distinct computational advantages and to give different perspectives on desirable functionality. The list below is ordered for ease of exposition as I could find no more systematic order (e.g. chronological) that turned out to be satisfactory or useful.

0.1. List of five categories of additional processes

Addition 1. Higher-order processes (including FRF structures).
Addition 2. Divisive contrast nonlinearities (including contrast normalization).

Addition 3. Subtractive contrast nonlinearities (including contrast comparison).
Addition 4. Non-classical receptive fields (including surround suppression and facilitation, cross-orientation or overlay suppression).
Addition 5. Contour integration.

The sections in the main text below on *Additions 2* and *3* form a natural pair. They both focus on the intensive characteristics of neurons or model units, and, in particular, on the magnitude of response as a function of magnitude of input. The section on *Addition 2* starts with an introduction to both categories, and the section on *Addition 3* ends with an example involving both categories.

The sections on *Additions 4* and *5* return to the question of spatial characteristics of behavior not explainable by the simple multiple-analyzer model. In so doing, they overlap with each other, and with all three of the preceding categories.

0.2. Scope of this review

This review focuses on static two-dimensional visual stimuli (spatial patterns). Motion, color, and depth will only be mentioned occasionally in this article. See other articles in this volume for them. While it has proven useful to separate issues of spatial vision from those of color, motion, and depth vision – as is done here – this separation should be treated as a convenience and not as a good representation of the visual system. Evidence is accumulating that they are in fact intertwined at all stages of processing (see Shapley & Hawken, 2011; also Hochberg, 1998; Lennie, 1998).

Unfortunately I will only be able to present a fraction of what is known about each category of process listed above. For each category, a description will be given of how the processes in the category work, with an attempt to give some intuition as well as some idea of the formalisms with which such processes are incorporated into models. But much else beyond this is known. Very few of the empirical results from psychophysics and physiology are covered. From these results, much could be deduced about how these processes' characteristics vary with different dimensions of pattern vision (spatial frequency, orientation, retinal position, temporal frequency or speed of motion, color, and their interactions). And also omitted are many interesting suggestions about the possible functionality of these processes, and about the possible evolutionary pressures that led to such processes existing. Since most of these topics cannot be covered here in any detail, a large number of studies and investigators are not directly referenced at all. To try to compensate somewhat for this lack, references here have been chosen not only for the work they directly report but also for their references to other studies covering omitted knowledge.

Also outside this review's scope are two topics that are closely related to the material covered. One topic is the question of how to adequately model the known variability (noise) in both physiology and psychophysics.

The second topic is the question of how to adequately model the "missing link" between the entities of interest e.g. (analyzers in the top of Fig. 1) and the observed measurement in the experiment (e.g. the observer's responses in the top of Fig. 1). Since, for psychophysics, this missing link (the decision rule in Fig. 1) is likely to be representing a great deal of processing in the brain – visual and cognitive – the best way to specify it is far from trivial.

The problem of a missing link exists, however, even when the entity of interest is a single neuron's firing and the measurement is the output from an electrode (How does one know that the electrode found a representative sample of cells or that it is not distorting the response of the cells?) And it is a well-known problem using the increasingly popular fMRI measures.

In the bulk of the material reviewed in this article, with its emphasis on modeling and on psychophysics, the aim is to keep the implicit or explicit model of this missing link (the decision stage in psychophysics) simple enough to allow investigation of the preceding stages without being so simple as to be misleading. There is some further material in Appendix A.

In general, I tried to include some cross-references to other articles in this volume but am sure many appropriate cross-references are missing. Morgan's article has particularly strong cross-ties with this one. The general topic of both his article and this one is the same: considering advances in our understanding of spatial vision using a modeling approach and starting from a pure linear-filtering approach. The specific focus of his article is quite different from that of mine. In Morgan's article, the specific focus is the question of the existence of spatially-localized features or primitives (e.g. edges) in human vision: the physiological and psychophysical evidence for them; their appearance in the perceptions of the observer; whether they are or are not the same as features importantly present in the natural patterns. In my article, the specific focus is on candidates for general-purpose calculations done at intermediate levels of processing, and these intermediate calculations need have little to do with spatially-localized features. The relationship of these intermediate calculations' outputs to human perception (or control of action) is assumed to be rather distant. When psychophysical results are considered, the experiments are assumed to belong in a special category allowing the observer's response to be calculated from the intermediate calculations' outputs by a simple decision rule (even though that simple decision rule is representing a great deal of cortical processing). When physiological experiments are considered, the neural substrate is intermediate in the approximate stream from light on the eye to the highest areas, e.g. V1 through V4 and MT.

One line of work that could well belong within either specific focus is described in Nachmias (1999) and the references therein. This line of work considers observers' likely use of spatially-localized features to determine their responses in rather simple pattern masking experiments. These simple masking experiments are a class of experiments often used in the study of intermediate level calculations like the multiple analyzers of Fig. 1 and the second-order processes described next.

1. Addition 1. Higher-order processes (including *FRF* structures)

The first category of additions to the simple multiple-analyzers model (Fig. 1) is called here *higher-order processes*. Prominent among these are *second-order processes*, processes in which linear units with one kind of receptive field – and with outputs that are rectified or otherwise nonlinearly transformed – serve as inputs for units with another kind of receptive field. A very large number of investigators seem independently to have come up with suggestions of this type. One very early suggestion was that of Henning, Hertz, and Broadbent (1975). But these suggestions accelerated and matured during the last 25 years. Much of the work up through the early 1990s can be found in *Higher-order processing in the visual system*, a book published in 1994 from a symposium the previous year. There was also a special issue of *JOSA A* in September 2001 on second-order processes in vision. Reviews or overviews of higher-order processes can also be found in many of the references listed throughout this section.

Suggestions of second-order processes were frequently motivated by the use of special patterns (often called *second-order or higher-order patterns*) that have custom-made and perceptually-salient characteristics. The simple multiple-analyzers model of Fig. 1 cannot explain these perceptions. Examples are shown in Figs. 2 and 3. Most observers see (among other things) relatively

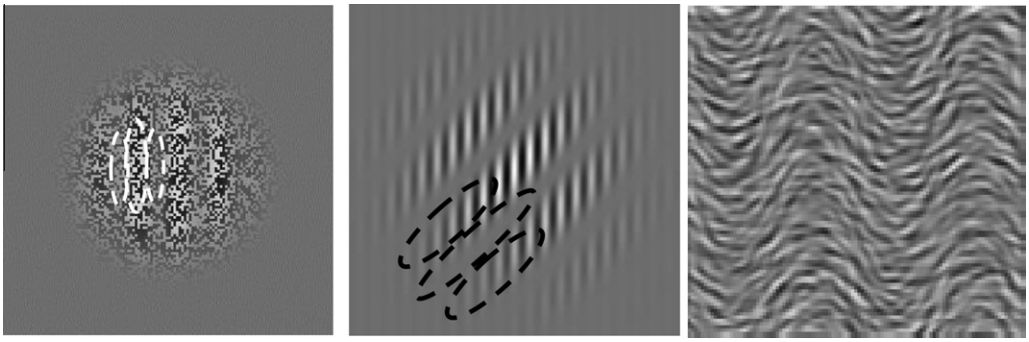

Fig. 2. Three examples of second-order patterns. Adapted from Figs. 1 and 6 of Schofield (2000). Dashed ellipses are superimposed on the patterns in left and middle panels to show receptive fields of linear filters corresponding to perceptually salient aspects of the patterns.

wide stripes of one orientation in each of these patterns. These global stripes are vertical in the left panel of Fig. 2, oblique in the central panel, and vertical in the right panel. In Fig. 3 they are vertical in the left two panels and horizontal in the right. Most human observers can easily see the global stripes in both contrast-modulated and orientation-modulated patterns (as in Fig. 3). Furthermore, most observers can easily discriminate between different orientations (or spatial frequencies) of global stripes in rigorous psychophysical experiments.

That these characteristics could be readily perceived demonstrated weaknesses in the simple multiple-analyzers model of Fig. 1. The information necessary to identify these custom-made characteristics would be lost if the observer were completely characterized by simple linear filters by themselves (followed by only a simple decision rule). To see why, consider the dashed ellipses in Fig. 3. Each set of three ellipses shows a receptive field of spatial frequency and orientation matched to the global stripes; the center ellipse represents the center of the receptive field and the flanking ellipses represent the inhibitory surround. Notice that the response from each receptive field would be approximately zero because each region (either inhibitory or excitatory) is stimulated by equal amounts of light and dark which cancel out. Further, no matter where these receptive fields are located relative to the pattern, they will continue to produce outputs of approximately zero. Thus observers cannot use receptive fields these to perceive the orientation, spatial frequency, or even the existence of global stripes.

The only receptive fields that produce (substantial) non-zero outputs to the patterns in Fig. 3 are those small enough to respond to the details in the individual Gabor patches. (A concrete example showing these small receptive fields and their outputs will appear in the top two rows of Fig. 6, described more completely in subsection 1.1.1.) However, the outputs of these small receptive fields at different spatial positions just mimic the original pattern and show all the local features but do not encode the fact of the global stripes in any more direct way. Thus, a simple decision rule has no way of lining up the non-zero responses from the little receptive fields to figure out that those responses are arranged in wider stripes. More generally, there is no size of receptive field that would allow a simple multiple-analyzer model (Fig. 1) to predict the fact that observers reported perceptions or performance in psychophysical experiments.

1.1. FRF (Filter, Rectify, Filter) processes (structures, channels)

In the pattern-vision literature, the most common form of higher-order process is often known by the name *FRF channel or FRF process* where *F* is for "Filter" and *R* is for "Rectify." We will look here at this structure in some detail and, in particular, at how it finds the global stripes in Figs. 2 and 3.

As shown in Fig. 4, an *FRF* channel consists of three parts: a linear filter characterized by: a relatively small weighting function (a relatively small receptive field); followed by a point-wise nonlinearity of the rectification type; followed by a second linear filter that is characterized by a relatively large spatial weighting function (a relatively large receptive field).

The assumption illustrated Fig. 4 is that each filter is not only a linear system but is also *translation-invariant* (having the same receptive field or weight function at each spatial position). This is a false assumption about the visual system of humans and mammals. In fact, sensitivity is generally highest at the fovea and declines with eccentricity (although this depends somewhat on spatial and temporal characteristics of the pattern). However, the assumption of translation-invariant systems is not very restrictive. One could easily substitute non-translation-invariant systems as is occasionally done. Or, more commonly and usually implicitly, different regions of the visual field are considered separately and each region (e.g. just the fovea by itself) is small enough that translation-invariance is a pretty good description within that region.

The rectification *R* at the in-between stage can take various forms, including half-wave or full-wave *rectification* (of the usual piecewise-linear type). Or, as shown in the diagram of Fig. 4, it can be a function that is not made of linear pieces at all but is a power function or some other simple function.

What is common to all these functions *R* is that they are operate on individual points of the output from the filter, independent of the outputs at other points. They are sometimes referred to as instantaneous nonlinearities (perhaps because much work was originally done with filtering in the time domain) but here the word *point-wise* is used because it applies equally to points in space and points in time.

(For more about linear systems and rectification, see glossary.)

As diagrammed in Fig. 5, a large number of such *FRF* channels are assumed to exist in the visual system. This is analogous to the assumption of a large number of simple linear channels in the simple multiple-analyzers model. The *FRF* channels differ from one another in the size and orientation of the receptive fields at the first and second filters. In Fig. 5, the intermediate rectification function (middle box in each row) is a conventional *full-wave rectification function*.

One of the attractions of continuing to use linear systems in models like these is the accumulated knowledge about how to work with them, knowledge known by the names *linear-systems analysis*, *Fourier analysis*, and other related terms. And the computation of the response of a linear filter by computer was made very fast by the invention of an algorithm (called the Fast Fourier Transform).

Similarly one of the attractions of rectification stages (point-wise nonlinearities) is that they also have proven mathematically tractable.

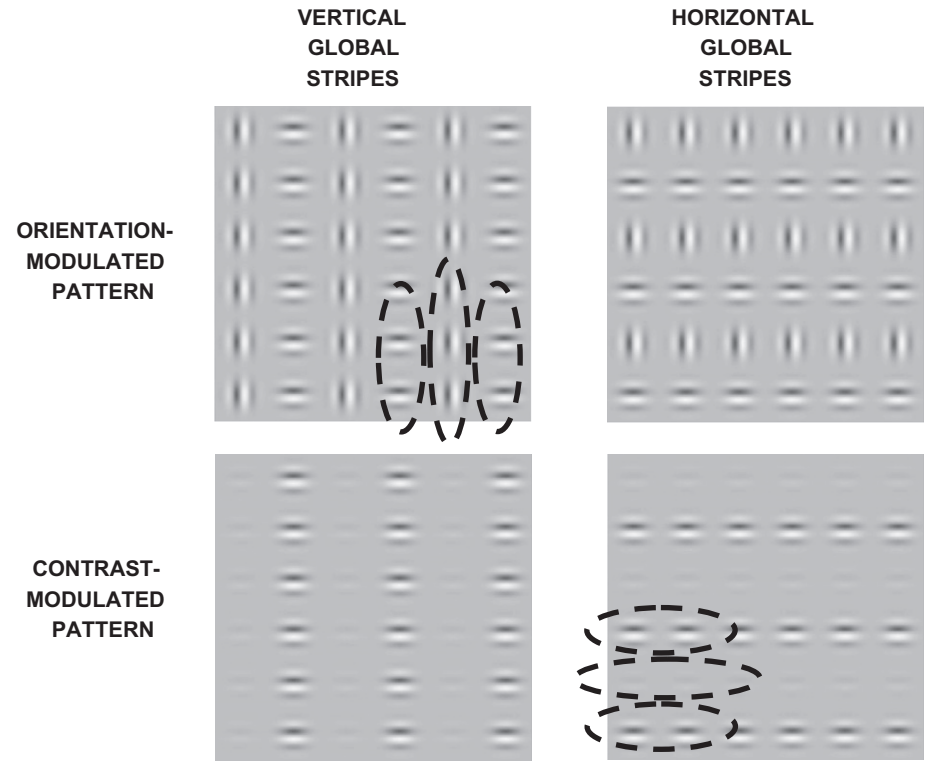

Fig. 3. Four examples of second-order patterns. Each set of three dotted ellipses (upper left and lower right panels) shows a receptive field of spatial frequency and orientation like that of most observers' perceptions of the contrast-defined stripes. The center ellipse represents the receptive-field center and the two flanking ellipses the receptive-field surround. No matter where those receptive fields are located in the four patterns, they will produce outputs of approximately zero, and thus cannot be used by the observer to discriminate between horizontal and vertical global stripes.

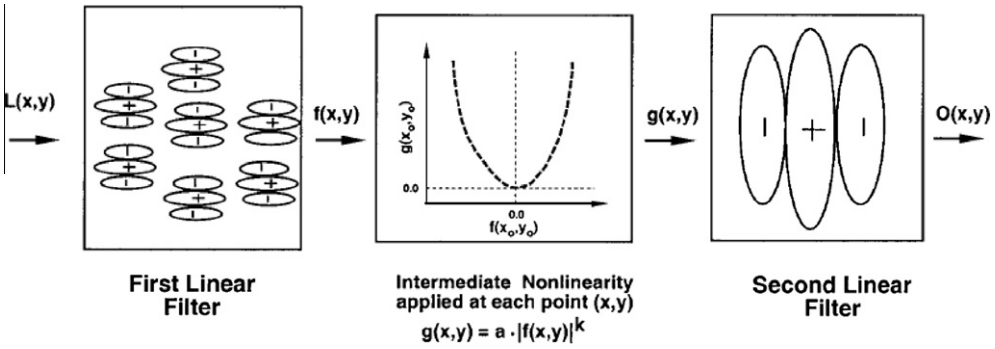

Fig. 4. Wiring diagram of an *FRF* (filter, rectify, filter) *channel*. The symbols on the diagram, e.g. $L(x, y, t)$, have meanings that can be inferred from the diagram, but they can also be ignored with little loss for the purposes of this review. They are explicitly defined in Graham and Sutter (1998).

1.1.1. Illustration of how a "tuned" FRF channel responds to contrast-modulated and orientation-modulated patterns

Fig. 6 illustrates the responses of *FRF* channels to patterns like those in Fig. 3. The middle and right columns show the case of an orientation-modulated pattern. The left column shows the case of a contrast-modulated pattern. The contents of each row are described in the labels on the left and show the results step by step of processing in the *FRF* channel.

Each column shows the output of an *FRF* channel that has spatial characteristics "matched" or "tuned" to the pattern in question in at least spatial frequency. The spatial frequency of the channel's first filter is tuned to the Gabor patches. For each pattern, two channels are shown: a channel having a first-stage filter tuned for orientation and a channel having a first-stage filter with a concentric receptive field (and therefore not tuned for orientation). These two channels are shown separately for the orientation-modulated pattern (in the right two columns) but superimposed for the contrast-modulated pattern (in the left column, to save space and emphasize the point that they both lead to the same predictions). The spatial frequency and orientation of the channel's second filter is tuned to the global stripes.

Notice that the output of the *FRF* channel to a contrast-modulated pattern (bottom row, left column) shows a modulation that reveals the global stripes in the pattern; this is true whether the first-stage is concentric or tuned to the Gabor-patch orientation. (Of course if the first-stage filter had been tuned to an orientation other than that of the Gabor patches, there would have been no output from such a channel.)

The middle column shows that when the first-stage filter has concentric receptive fields (and thus is not selective for orientation) and the pattern is orientation-modulated rather than

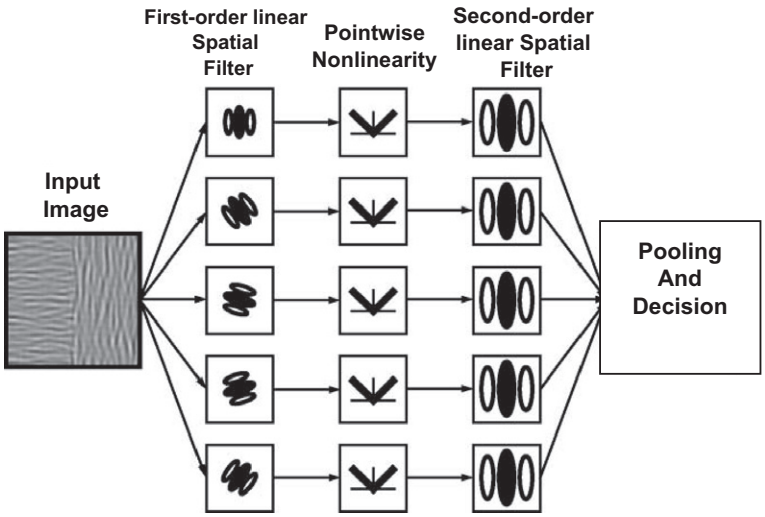

Fig. 5. Multiple *FRF processes* followed by decision rule. From Fig. 1, with slight changes, of Landy and Oruc (2002).

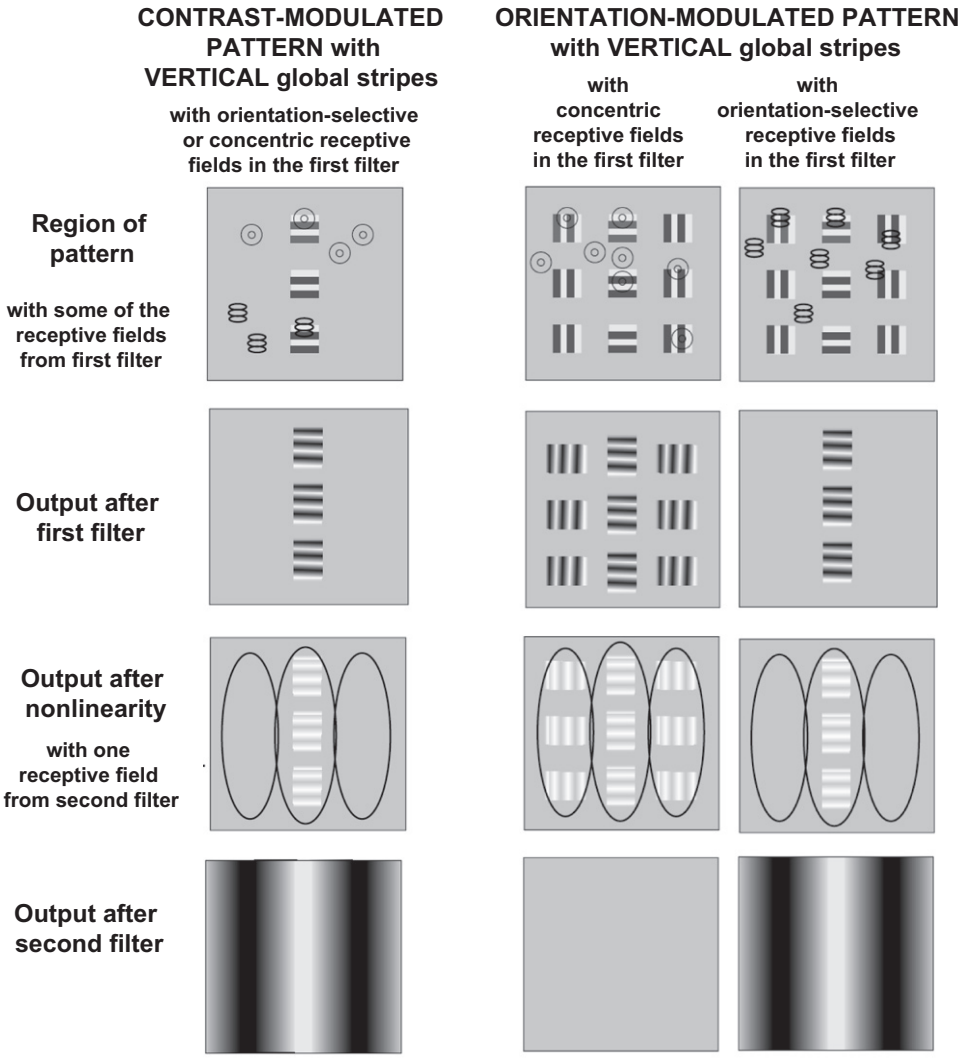

Fig. 6. Responses to contrast-modulated (left) or orientation-modulated (right) patterns by *FRF processes*. The patterns contain square-wave patches, and the responses are semi-schematic. See text for further details.

contrast-modulated, the *FRF* channel produces an output that cannot reveal the global stripes (bottom row, middle column).

The right column shows that when the first-stage filter is itself orientation-selective, the *FRF* channel produces an output

that does reveal the global stripes (bottom row, right column).

As mentioned above, human observers can see the global stripes in both contrast-modulated and orientation-modulated patterns. Therefore, human performance can not be explained using only *FRF* channels having concentric first-filter receptive fields, but could be explained using only *FRF* channels having orientation-selective first filters (and could also be explained using both kinds of *FRF* channels in a model).

Examples of the stage-by-stage outputs of *FRF* channels to various artificial patterns can be found in Graham, Beck, and Sutter (1992), Graham, Sutter, and Venkatesan (1993), Graham and Sutter (1998) and Graham and Wolfson (2004). These are illustrated as in Fig. 6 but with longer verbal descriptions than that here.

Examples of the responses of final outputs of simple linear channels (a single filter) and *FRF* channels to photographs of natural scenes can be found in Johnson, Kingdom, and Baker (2005) and Johnson and Baker (2004).

1.2. Other higher-order processes – second-order and even more general

More elaborate than *FRF* but still in the general category of higher-order are a number of other structures hypothesized for pattern vision. Examples from eight studies are illustrated in Fig. 7 and described briefly in the next several paragraphs. (The reader may need to consult the original references to fully understand these examples.)

Some of these more elaborate schemes share the characteristic of *FRF* processes in that outputs from units with small receptive fields at different spatial positions go into units with bigger receptive fields, and thus are being called *second-order processes* here. Instances like this in Fig. 7 include the orientation-modulation unit of Motoyoshi and Nishida (2004), and orientation-opponent channels of Graham and Wolfson (2004), and spatial frequency contrast units of Arsenault, Wilkinson, and Kingdom (1999). Many structures in this larger category of second-order processes can be seen as veering into the contour-integration processes discussed in the *Addition 5* section. For example, various schemes (called names like collator and collector units, e.g. Levi & Waugh, 1996, shown in Fig. 7) are frequently discussed as ways of detecting contours, and in that sense are doing contour integration.

There are other higher-order processes that differ from all those described so far in an important way. In the examples so far, the second stage collects outputs from first-stage receptive fields that are centered at a variety of spatial positions in the visual field. The second stage of other higher-order processes, however, collects outputs from first-stage receptive fields that are all centered at the *same* spatial position. These receptive fields differ along some dimension other than spatial position. For example, in the MIRAGE model, the outputs of receptive fields at the same position but with different preferred spatial frequencies were half-wave-rectified

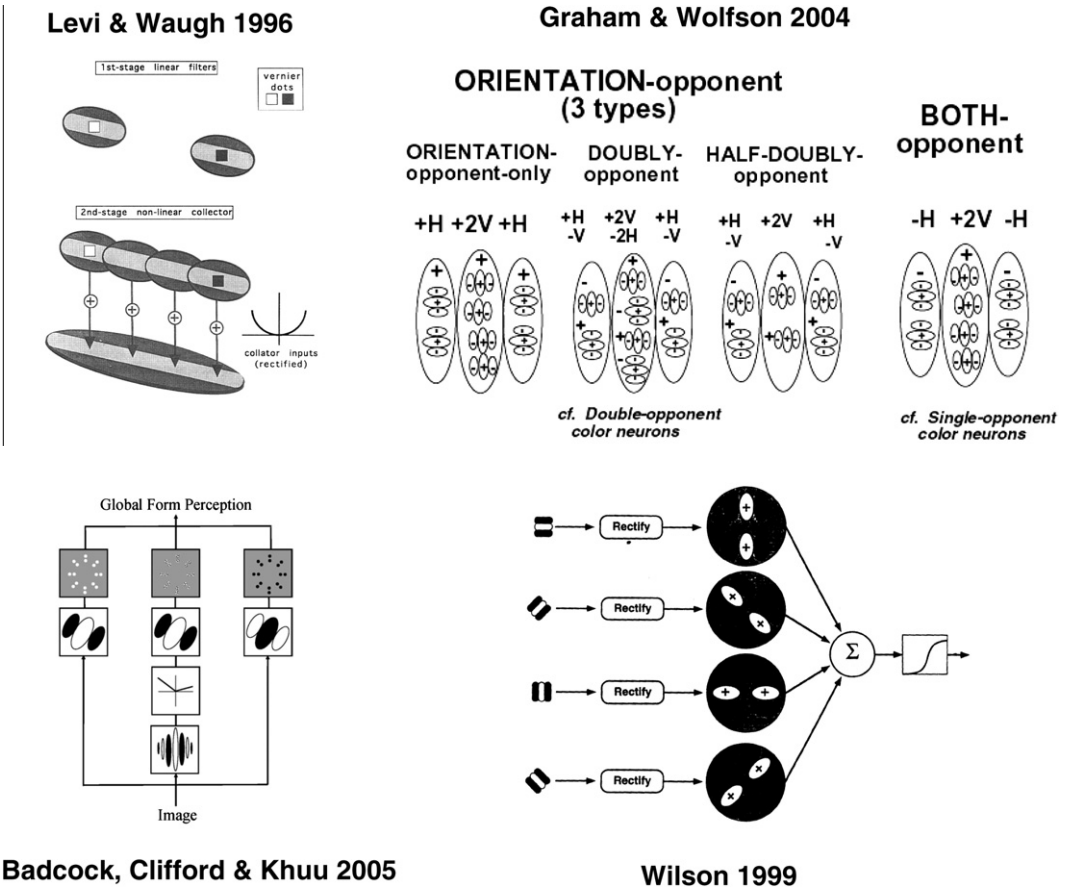

Fig. 7. Sample of higher-order processes that have been suggested. From the publications indicated by labels on the figure, slightly modified in a number of cases for consistency. Most of the diagrams are flow diagrams with receptive-field sketches. However the pair of diagrams from Olzak and Thomas (1999) are plotted in two-dimensional frequency space where the distance from the center point represents spatial frequency and the angle between the line from the center to a point and the horizontal represents orientation. Pooling across spatial frequency (respectively orientation) is shown by the gray area in the diagram on the left (respectively right). Figure numbers in the original publications from which these diagrams here were adapted are (from upper left to lower right in ordinary reading order): Fig. 14 in Levi & Waugh, 1996; Fig. 2 in Graham & Wolfson, 2004: Fig. 5 in Badcock et al., 2005; Fig. 5 in Wilson, 1999; Fig. 6 in Olzak & Thomas, 1999; Fig. 9 in Arsenault et al., 1999; Fig. 8 in Motoyoshi & Nishida, 2004; Fig. 4 in Motoyoshi & Kingdom, 2007.

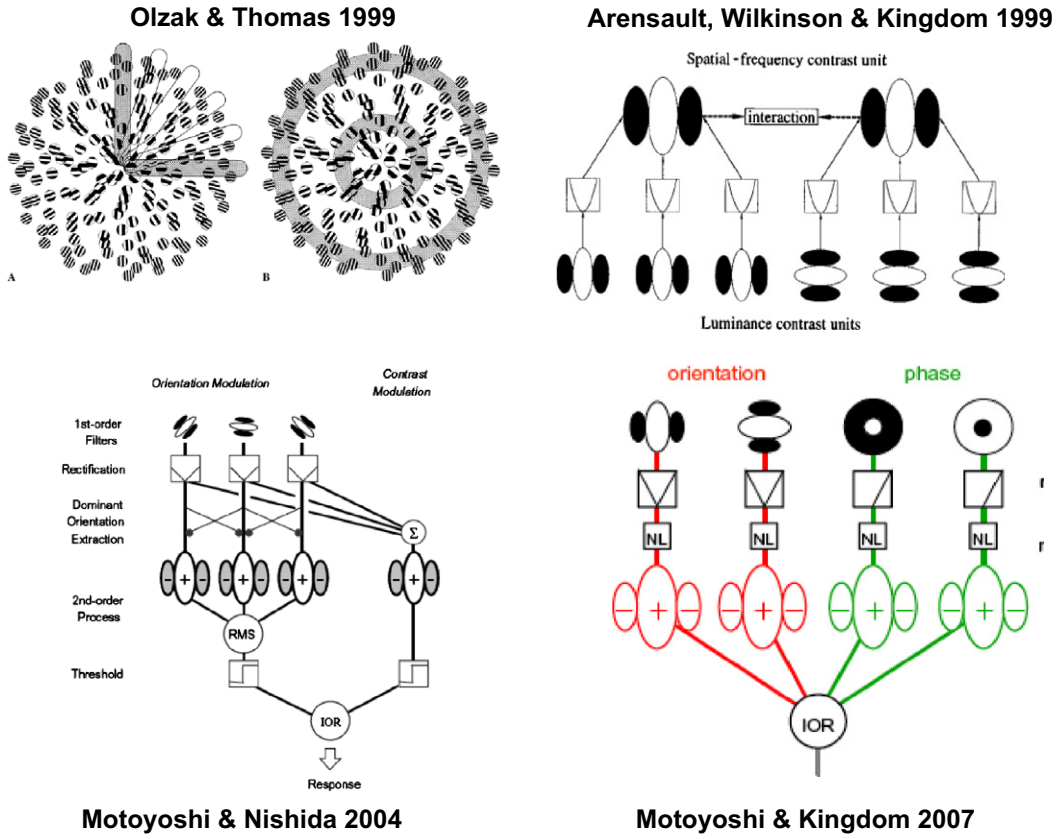

Fig. 7 (*continued*)

and added in order to create an edge-finding mechanism that combined information across scale. (See Morgan's article in this volume.) In another example, receptive fields at the same position but differing in either spatial frequency or orientation (but not both) were used in order to explain performance in different tasks: pooling across preferred spatial frequency was useful for some tasks and orientation for others (Olzak & Thomas, 1999, shown in Fig. 7).

Many other varieties of higher-order channels – wiring diagrams with layers of different receptive fields and rectification type nonlinearities in-between – have been suggested for the explanation of many other perceptual phenomena as shown in several further drawings in Fig. 7 (Badcock, Clifford, & Khuu, 2005; Motoyoshi & Kingdom, 2007; Wilson, 1999).

And a caution: Some of these higher-order processes – if considered not as hard-wired but instead somewhat plastic in response to particular task demands – might have their primary substrate at very high levels of the visual system (even if feedback to the intermediate levels means that the intermediate levels' responses show evidence of these processes).

1.3. A sample of references for further reading

Much is now known about the spatiotemporal properties of higher-order processes in human vision and their dependence on the other parameters of pattern vision, and some references already given here include information of this sort. A small haphazard sample of additional studies (with a bias toward recent and toward authors not previously cited here) includes: Allard and Faubert (2007), Ellemberg, Allen, and Hess (2006), Hess, Baker, May, and Wang (2008), Manahilov, Simpson, and Calvert (2005), Schofield and Georgeson (1999), Vakrou, Whitaker, and McGraw (2007).

Other studies slightly farther afield include: reading 2nd-order letters in Oruc, Landy, and Pelli (2006), 2nd order learning in Dosher and Lu (2005), 2nd-order illusions in Lu and Sperling (1996); high-order phase correlations in Victor and Conte (1996).

In other places in this volume, second-order mechanisms are discussed in motion perception by Burr and Thompson, in stereopsis by Blake and Wilson in attention by Carrasco, and in visual search by Morgan.

There has been a small amount of work on the possible neural substrate of *FRF*, second-order, and higher-order channels. There is some work on cat single neurons (e.g. Baker & Mareschal, 2001), primate single neurons (e.g. El-Shamayleh & Movshon, 2006; El-Shamayleh, 2009) and human fMRI (e.g. Hallum, Landy, & Heeger, in press; Larsson, Landy & Heeger, 2006). The amount of physiological work to date is small, however, and this addition to the simple multiple-analyzers model of pattern vision has been motivated primarily by perceptual observation and psychophysical experiments rather than by physiology.

In the last several decades there has been a growing emphasis on understanding the role of proposed visual mechanisms – and, in particular, of *FRF* or more general higher-order mechanisms – in everyday vision. It remains rather unclear what that role might be for higher-order mechanisms – or why evolution might have produced such mechanisms. For a starting point on this topic, a reader could consult two papers that analyze the output of groups of multiple simple linear and *FRF* channels to natural images (Johnson et al., 2005; Johnson & Baker, 2004). There is also a third paper following on the first two that uses artificial stimuli based on natural images in psychophysical experiments (Johnson, Prins, Kingdom, & Baker, 2007). It concludes that, when local second-order and first-order information are present in the artificial images in ecologically valid way, discrimination performance in the laboratory is improved. These papers provide many other references as well.

	Divide by *a* (Multiply by *1/a*)	Subtract *a* (Add *-a*)
Operates on *R*	***R = f(c)/a*** **Graph: translates on log R axis** *Response gain change*	***R = f(c) - a*** **Graph: translates on linear R axis** *Response subtraction*
Operates on *c*	***R = f(c/a)*** **Graph: translates on log c axis** *Contrast gain change*	***R = f(c - a)*** **Graph: translates on linear c axis** *Contrast subtraction*

Fig. 8. Table of pure divisive and subtractive effects on response ***R*** as a function ***f*** of contrast ***c***. At the top of each cell in the table is a formula for a family of functions that can be generated by varying the value of the parameter ***a***. In the middle of each cell is a fact about the relationship among the individual members of that family when they are graphed. At the bottom of each cell is the name we will use for that family.

2. Addition 2. Divisive contrast nonlinearities (including contrast normalization)

2.1. Introduction to this section and the next: divisive and subtractive contrast nonlinearities

When supra-threshold patterns are used, quantitative results both from human psychophysical experiments and also from physiological experiments on V1 simple cells show clear nonlinearities that are direct violations of the simple multiple-analyzer model of Fig. 1. In particular, nonlinearity is immediately apparent when response magnitude is plotted as a function of stimulus contrast magnitude. In physiological experiments, the response is measured directly from single cells or from groups of cells. In psychophysical experiments, the response is inferred indirectly from psychophysical results in contrast-discrimination, adaptation, masking and other experiments. This section (*Addition 2*) and the next (*Addition 3*) are about two broad classes of contrast nonlinearities (divisive and subtractive) that can describe the non-linear relationship between the contrast magnitude and response magnitude and the systematic changes that occur in this relationship across different conditions, e.g. when the function is measured in the presence or absence of other patterns.

Consider any function ***R = f(c)***. For our purposes here ***R*** is response magnitude and ***c*** is stimulus contrast magnitude. As displayed in Fig. 8, you can form a new function in any of four simple ways, each involving only a single parameter ***a***. These four ways are the four possible combinations of:

- whether ***R*** or ***c*** is manipulated (top vs. bottom of table);
- whether the manipulation is *division* of the variable by ***a*** or *subtraction* of ***a*** from the variable (left vs. right of table).

The table also shows a possible name for each kind of process (in italics at the bottom of each cell). Other terms are used as well in the literature, and the differences among terms can be quite confusing. We will use the terms in the table. They will often refer, however, not to the absolutely pure cases shown in this table but to near relatives.

Empirical and theoretical values of ***R*** as a function of contrast ***c*** have been plotted many ways. Some of the most frequent involve using a log or linear axis for ***R***, and using a log or linear axis for ***c***. All four possible combinations are used from time to time. In the middle of each cell in the table is a particularly useful fact about the graphs for the case in that cell. These four facts can be summarized as saying: if parameter ***a*** is dividing a variable (***R*** or ***c***), it is useful to plot that variable (***R*** or ***c***) on a log axis. Similarly, if the value of the parameter ***a*** is subtracting from a variable, it is useful to plot that variable on a linear axis. The usefulness arises from the fact that, when plotted in the stated way, the family of functions produced by varying the value of ***a*** is a set of curves that are simply translations (shifts) of each other on the relevant axis. Such shifts can be easily and immediately perceived when looking at plots.

2.1.1. A point of terminology

Rather than referring to the categories as divisive and subtractive, they are sometimes called multiplicative and additive. Division by ***a*** is the same as multiplication by **1/*a***. Similarly subtraction of ***a*** is the same as addition of $-\boldsymbol{a}$.

2.1.2. A caveat about divisive vs. subtractive

Making a distinction between divisive and subtractive processes is often quite useful. It should be treated with caution, however. Remember that the log of a quotient equals the log of the numerator minus the log of the denominator. If a process of interest simply takes the quotient of two inputs to produce an output, consider what happens if you now redefine the inputs and outputs of that process to be the logarithms of the original values. The process now subtracts one of the two newly-defined inputs from the other to produce its newly-defined output. Thus what might have been called a divisive process originally looks like a subtractive process now. Any difference between the old and the new case is simply a matter of exactly what the "input" and "output" are. For our purposes here, it makes sense to consistently use contrast, not log contrast, as the input, and thus we avoid most of these problems.

2.1.3. A caveat about whether the variable affected is response or contrast

If ***f*** is a linear function, then the top and bottom rows of Fig. 8 are indistinguishable. For in this case, anything that divides ***c*** can be rewritten as dividing response ***R*** (left column). And the same is true for subtraction (right column). But since we are usually talking about non-linear functions ***f***, this matters little here.

2.2. More about the two divisive cases: response-gain and contrast-gain change

That the two divisive cases (left column of table in Fig. 8) generally make different predictions can be seen in Fig. 9, which shows a family of functions illustrating response-gain change in the left panel and contrast-gain change in the right panel. The axes for this figure are logarithmic, i.e. log response versus log contrast. Each family is fitted as well as possible to empirical results from a representative cortical area V1 neuron (Albrecht & Hamilton, 1982). The different functions in these empirical results (shown as data points) result from varying the spatial frequency of the pattern stimulating the neuron. Notice in the predictions that there are vertical translations in the case of response-gain change and horizontal translations in the case of contrast-gain change predictions. (To see an example plotted in a different way, look ahead to Figs. 12 and 13 for some functions plotted as *linear* response versus log contrast.) The set of empirical results (data points) in Fig. 9 is well described by the response-gain change family of functions and not well described by the contrast-gain change family.

This figure from Albrecht and Hamilton (1982) uses a 3-parameter non-linear function that had been used for some decades for describing responses of neurons and for many other reasons in many fields of science. It was called the *hyperbolic ratio*

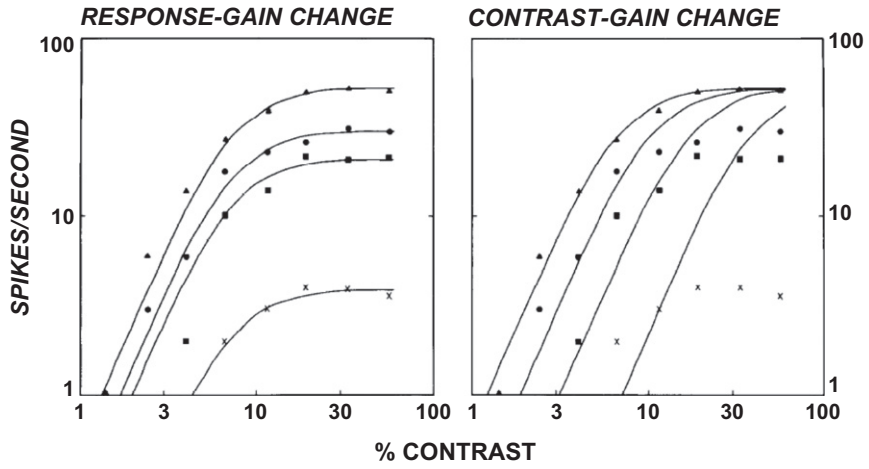

Fig. 9. Response (data points) versus contrast for a typical V1 simple cell for four different patterns. The solid lines show a fitted family of functions of the response-gain-change type (left panel) or contrast-gain-change type (right panel). Both the horizontal axes and the vertical axis are logarithmic. The horizontal axis gives the contrast of the stimulus. There are four sets of data points (indicated by different symbols) showing the responses of the cell to four different spatial frequencies. From Fig. 9 in Albrecht and Hamilton (1982) relabeled to match the terminology here. Used with permission, Am Physiol Soc.

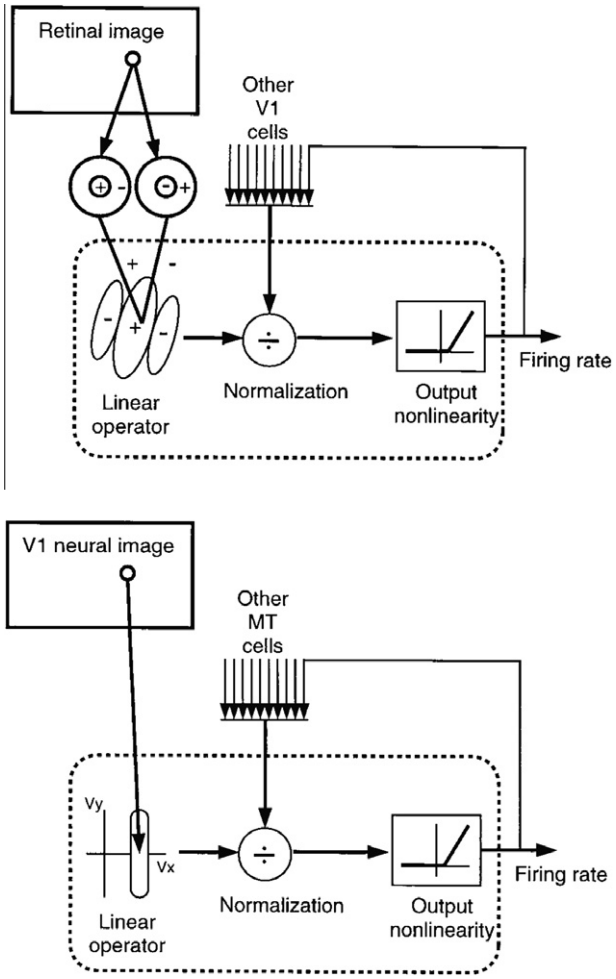

Fig. 10. Models of normalization as feedback circuit for V1 cells (top) and MT cells (bottom). From Fig. 1 in Heeger, Simoncelli, and Movshon (1996).

(or H ratio) function in this study. Other names used for it – or very close relatives to it – include *Michaelis–Menten* function and *Naka–Rushton* function.

$$R(c) = R_{max} \cdot \frac{c^n}{c^n + c_{50}^n} \tag{1a}$$

It has three parameters. R_{max} is the maximum (or saturation) response as c goes to infinity. The parameter c_{50} is the value of contrast that produces a response exactly equal to half the maximum (or saturation) response R_{max}. It is often called the half-saturation contrast. The parameter n affects the steepness of the curves.

As you can verify by examination of Eq. (1a) and the equations in the table of Fig. 8, varying the maximum response R_{max} in Eq. (1a) is exactly the response-gain-change type (the upper-left cell of the table of Fig. 8). Therefore, this variation produces a family of functions that are exactly vertical translations of one another on the log response axis (Fig. 9 left).

Varying the half-saturation contrast c_{50} is exactly equivalent to the contrast-gain-change case (lower-left cell of table of Fig. 8). To see this more easily, first divide both the numerator and the denominator by c^n producing the expression:

$$R(c) = R_{max} \cdot \frac{1}{1 + (c_{50}^n / c^n)} \tag{1b}$$

So c_{50} is playing the role of $\boldsymbol{a}$ in the contrast-gain change cell of the table (Fig. 8). As expected, therefore, varying c_{50} produces functions that are horizontal translations of one another on the log contrast axis (Fig. 9 right).

Terminological notes: The value of the half-saturation contrast c_{50} can also be referred to more generally as a *contrast threshold*, where the associated threshold criterion is the half-maximum response. The reciprocal of c_{50} (that is, the value of $1/c_{50}$) is called the *contrast gain* or *contrast sensitivity*. The smaller the value of threshold is, the greater the value of gain or sensitivity.

One might refer to the value of R_{max} as the value of the *response gain*, but that seems to be less common usage. However, situations in which the value of R_{max} changes are frequently referred to as response-gain changes.

2.3. Contrast normalization model – a divisive nonlinearity

The function in Eqs. (1a) and (1b) was in the pattern-vision literature originally as a simple description of response versus contrast (as in Fig. 9) with little commitment to the process that might have led to these equations. This was true earlier than 25 years ago. (Before being used for describing response as a

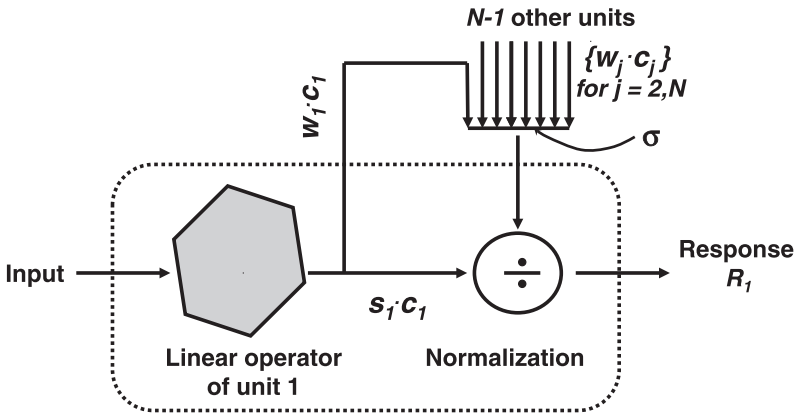

Fig. 11. Model of normalization as feedforward circuit. Symbols defined in text.

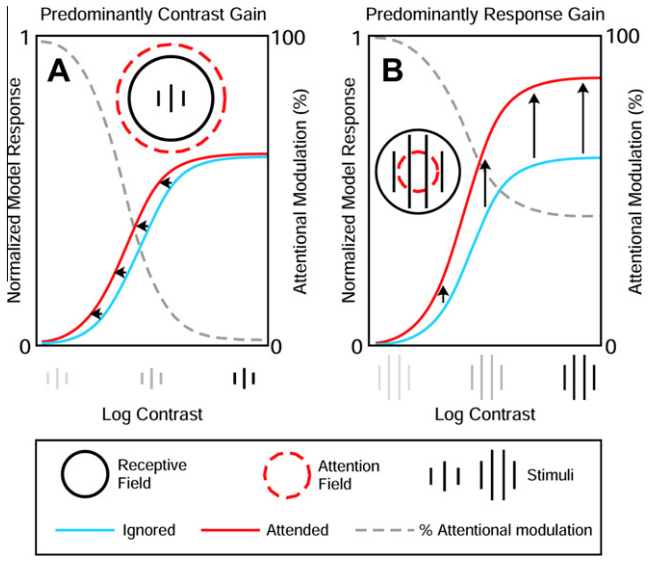

Fig. 12. Each panel shows a pair of contrast-response functions for a stimulated neuron, as predicted by a normalization model of attention. One is the function for attending to a stimulus within the neuron's receptive field (black or red, "Attended") and the other when attending to a stimulus in the opposite hemifield (light gray or blue line, "ignored"). The contrast of the test stimulus is plotted on the horizontal axis. The normalization model of attention predicts predominantly contrast-gain change in the left panel and predominantly contrast-response gain in the right panel. Which it predicts depends on the stimulus size and the size of the attention field (sketched in the inset of each panel). These two cases are reversed left-to-right compared to those in Fig. 9. Note also that here the vertical axis is in linear units while it was logarithmic in Fig. 9. In both figures, the horizontal axis is logarithmic. From Fig. 2 in Reynolds and Heeger (2009).

function of stimulus contrast, it was used for response as a function of stimulus luminance.)

Also in the pattern-vision literature earlier than 25 years ago were suggestions of various kinds of inhibitory processes that were non-linear (as distinguished from the linear subtraction acting in the classical V1-simple-cell model). Many of these suggestions were qualitative but there were some models of inhibition that could be applied, e.g., the well-known inhibition in the eye of the horseshoe crab (the limulus) that can be modeled by the Hartline–Ratliff equations for *recurrent (feedback)* inhibition. And there were a number of computer instantiations of multiple interacting neurons as well. At that earlier time, however, these models seemed to be cumbersome and leading to little intuition. So they did not seem to be a tractable and useful way of adding this suggested inhibition to models of psychophysics. (Of course, at that time, computers were much slower and had much smaller memory capacity.)

Then about 20 years ago an influential class of ideas about non-linear inhibition surfaced from two different (although probably not totally independent) sources. One source was physiological research (e.g. Bonds, 1989; Heeger, 1991, 1992) and the other was psychophysical research (e.g. Foley, 1994; see also Legge & Foley, 1980). The name contrast normalization is often used for this class of process and we will use it here. In a normalization network, the response of each neuron is divided by (is normalized by, has its contrast gain set by) the total output from a pool of neurons. Such a process may prevent overload on higher levels and overcome the limitations of a restricted dynamic range while simultaneously preserving selectivity along dimensions like orientation and spatial frequency. Perhaps it helps encode images more efficiently. (Many references to the early history and to other investigators working on these ideas can be found in, e.g. Graham & Sutter, 2000; Reynolds & Heeger, 2009.)

One striking aspect of this class of ideas is its tractability in models. The suggested equations from both sources (e.g. Heeger, 1991; Foley, 1994) look much like Eq. (1a) with the added feature that the denominator is elaborated to include contributions from multiple neurons or units that are inhibiting the *signal* neuron. (We use *signal* unit or neuron here to mean the unit or neuron from which the response ***R*** is being measured or inferred in these equations.) The equations from the two sources are not precisely identical (except in certain sub-cases) but indeed the differences seem minor relative to the similarities and we will look in some detail at a particular example below.

Fig. 10 shows simple diagrams of normalization networks in two different places in the visual cortex – V1 and MT – for which Heeger and his colleagues propose a normalization model as a good account of many experimental results. In these diagrams the normalization is a feedback effect onto the signal neuron from a pool of neurons (potentially including the signal neuron).

A feedforward normalization network is shown in Fig. 11. As it turns out, the same equations can be used for either the feedforward or feedback case if one is not concerned with dynamics responses near transients. Discussion and further references can be found in Reynolds and Heeger (2009).

The symbols that label connections in Fig. 11 are used in the equations below. They are:

R_j = Response from unit j,
c_j = contrast stimulating unit j,
s_j = sensitivity along excitatory path for unit j,
w_j = weight on unit j's input to normalization network,
σ = parameter characterizing the amount of normalization.

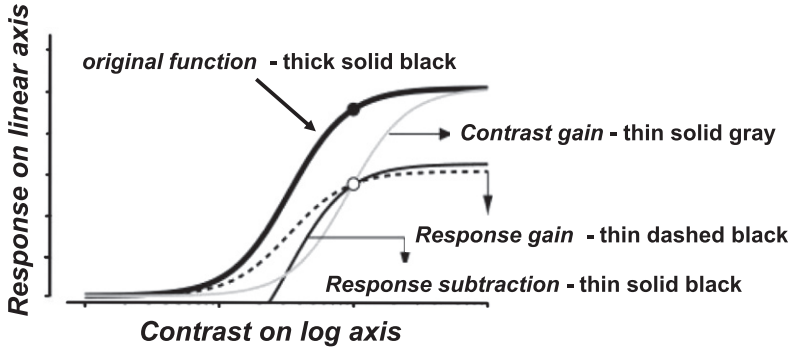

Fig. 13. Examples of three types of contrast nonlinearities (*Contrast gain, Response gain,* and *Response subtraction*) plotted as linear response versus logarithmic contrast. The original authors use the term *Subtraction* for what is called *Response subtraction* here. (With slightly expanded labeling from Fig. 5A of Cavanaugh, Bair & Movshon (2002). Used with permission, Am Physiol Soc.

The smaller the value of σ the larger the effect of normalization. (See further description in the beginning of Appendix B.) All the above symbols have values greater than or equal to zero except that σ is strictly greater than zero (so that the denominator can never be zero).

A typical equation for contrast normalization then assumes the form shown in Eq. (3). In Eq. (3) and the following equations, the *signal* neuron or unit – the neuron or unit of interest – will be numbered as the first unit and given the subscript 1.

$$R_1 = \frac{s_1 \cdot c_1}{H(\sigma, w_1 \cdot c_1, w_2 \cdot c_2, \ldots, w_N \cdot c_N)} \tag{2}$$

where the denominator is a function H of many variables. R_1 is also a function of many variables (the ones in the denominator as well as in the numerator). The arguments of R_1 are not explicitly written out in Eq. (3), however in order to make it easier to read. The arguments are written out in Appendix B.

The input from each unit to the normalization network is assumed to be $w_j \cdot c_j$ (the weight factor times the contrast stimulating that unit). These inputs can in general include one from the signal unit itself, as indicated in Eq. (3) and sketched in Figs. 10 (as feedback) and 11 (as feedforward). The function H is often considered to be a *power-summation (Minkowski distance, Quick pooling – see Appendix A for more about such measures)* with an exponent of 2 or higher, or something very similar. But for simplicity in exposition – and transparency of the basic concepts – in Eqs. (2) and (3) and Appendix B we will use an exponent of 1.

$$R_1 = \frac{s_1 \cdot c_1}{\sigma + w_1 \cdot c_1 + \sum_{j=2}^{j=N} (w_j \cdot c_j)} \tag{3}$$

Note that the denominator in Eq. (3) is the sum of several terms: the parameter σ that characterizes the normalization; the contribution of the signal unit to its own inhibition $w_1 \cdot c_1$; and the summed contribution of other units to inhibition of the signal unit.

In the Appendix B, Eq. (3) is further reduced to represent a model in which there are only two units: the signal unit and one other unit.

Normalization models incorporating equations like Eqs. (2) or (3) can make predictions of several kinds. Four of these will be listed below without attempting here to derive these equations. Some derivations can be found in Appendix B here and also in the other references given.

(i) Situations showing predominantly *contrast-gain change*.

See Case 1 in Appendix B here for a two-unit model example. This could be applied to an experiment in which the contrast of a test pattern varied while the contrast of a mask (or adapt) pattern was held constant. It predicts pure contrast-gain change.

See Heeger, 1992 (Fig. 2 and accompanying text) for an example of pure contrast-gain-change predictions from normalization that explain the effects of contrast adaptation on the contrast response function of V1 cells.

A more complicated situation in which the normalization model can predict predominantly contrast-gain change is illustrated in the left panel of Fig. 12. This is from a study of attention. For these predictions the normalization model as we have looked at it was augmented with assumptions about attention. (See Reynolds and Heeger (2009), Fig. 2A and associated text on p. 271. See also Carrasco (2009))

(ii) Situations showing *mixtures of response-gain and contrast-gain changes*.

See Case 2 in Appendix B here for a two-unit model example. This could describe experiments in which the contrasts of a mask pattern and a test pattern both vary together. It predicts substantial changes of both contrast gain and response gain. Heeger (1992) discusses cases like this.

(iii) Situations showing predominantly *response-gain change*.

See Case 3 in Appendix B here for a two-unit model example. This example could, for example, describe experiments in which the response to a pattern was measured as a function of its contrast (for any one curve) but the spatial frequency or orientation or other similar value varied between curves. (This is the kind of experiment already shown in Fig. 9.) The predictions in Case 3 Appendix B are for pure response-gain change.

Also Heeger (1992, Fig. 3 and accompanying text) provides a prediction of pure response-gain change for this kind of experiment from a more general normalization model.

Normalization preserves selectivity. If it is predominantly the response gain that changes when spatial frequency or orientation is varied between curves, then the unit's spatial frequency and orientation-selectivity will remain quite narrow even at the highest contrasts. This is experimentally found for many cortical neurons (e.g. the results shown in Fig. 9 here and also shown compared to normalization predictions in Heeger, 1992, where Fig. 3 again shows response vs. contrast and Fig. 4 shows response vs. spatial frequency). The ability of the normalization model to predict the preservation of selectivity across the contrast range for, e.g., spatial frequency and orientation was one of the major motivations for considering contrast normalization seriously. It is often listed as one of the possible reasons evolution might have favored the development of this kind of process.

A situation involving attention that shows response-gain change, perhaps mixed with a tiny bit of contrast-gain change, is shown in the right panel of Fig. 12 (Reynolds & Heeger, 2009, Fig. 2B and accompanying text on p. 172).

(iv) The normalization model can predict Weber-law behavior.

For psychophysical contrast-discrimination experiments (where an observer's ability to discriminate between a pattern at one contrast and the same pattern at a different contrast is measured and typically the data presented as contrast thresholds), the normalization model can predict the Weber-law or near-Weber-law behavior that is typically observed.

The normalization model can also predict Weber-law behavior in somewhat more complicated pattern-discrimination experiments that measure above-threshold behavior as well as near-threshold behavior (Graham & Sutter, 2000; Wolfson & Graham, 2009).

2.3.1. About a missing exponent to incorporate an accelerating pointwise nonlinearity

An exponent found in many models containing contrast normalization was omitted in the equations here for the sake for clarity. This is an exponent on the contrast values – e.g. all the c_j values in Eqs. (2) and (3) might be raised to an exponent. When this exponent is greater than 1, it can represent an accelerating early nonlinearity that operates on the stimulus contrast before the unit's input to the normalization network. Models including this exponent on contrast can account for the so-called pedestal or dipper effect found in psychophysical contrast-discrimination results. And they can predict other related effects found in the more complicated pattern-discrimination experiments. (See, for example, Foley, 1994; Graham & Sutter, 1998. This exponent explicitly appears in this article in the model at the end of *Addition 3* as k_m in the equations of Fig. 18.)

2.3.2. About a missing decision rule

For ease of exposition, predictions for the response of a single unit were discussed here. This is reasonable when the unit is a neuron, and that single neuron's responses can be measured directly. However, in psychophysics, the observer's responses are necessarily a result of many different units'/neurons' responses. Simply looking at the most sensitive unit – as we implicitly did above – is not in general satisfactory for psychophysics. More sophisticated decision rules are required. Discussion of this stage is largely omitted from this article, but there is some in Appendix A.

2.4. A sample of references for further reading

Many other studies by many other investigators deal with divisive contrast nonlinearities in general, and with the model of contrast normalization in particular. Only a very few of these have been explicitly referenced so far. What follows here is a sample of a few more references, rather widely scattered over the possible set that could be listed here, to allow a reader to pursue an interest in this topic: Bex, Mareschal, and Dakin (2007), Bonds (1993), Carandini (2004), Carandini, Heeger, and Movshon (1997), Goris, Wichmann, and Henning (2009), Itti, Koch, and Braun (2000), Meese and Holmes (2002), Olzak and Thomas (2003), Ringach (2010), Schwartz and Simoncelli (2001), Victor, Conte, and Purpura (1997), Watson and Solomon (1997). Many other studies can be found in the reference lists of the ones explicitly mentioned.

In this volume, the use of contrast-gain controls (e.g. contrast normalization) is discussed in models of brightness by Kingdom and in models of binocular vision by Blake and Wilson.

3. Addition 3. Subtractive contrast nonlinearities (including contrast comparison)

The table in Fig. 8 shows two subtractive contrast nonlinearities as well as the two divisive nonlinearities described in the last section (*Addition 2*). These two cells for subtractive nonlinearities are described further in this section (*Addition 3*). Response subtraction is covered very briefly first and then contrast subtraction at greater length. This section ends with an example of a model incorporating both a divisive and a subtractive nonlinearity.

3.1. An example of response subtraction

Fig. 13 shows predictions from three of the four contrast nonlinearities listed in the table of Fig. 8. (The missing fourth one is contrast subtraction – the lower right cell in Fig. 8. It is discussed in the next subsection.) The horizontal axis in Fig. 13 is *logarithmic* contrast. Thus the relationship between the original function (thick solid black line) and a function with a different amount of contrast gain (thin solid gray line) is a horizontal translation. The vertical axis is response in linear units. Thus the relationship between the original function (thick solid black line) and a function with a different amount of response subtraction (thin solid black line) is a vertical translation. But on this linear axis, the relationship between the original function (thick solid black line) and the function with a different amount of response gain (dashed black line) is not a simple vertical translation; rather, the two curves diverge and get further and further apart from one another as contrast gets higher.

Predictions like those in Fig. 13 were explicitly compared to response vs. contrast functions from single V1 neurons in a study of surround suppression (Cavanaugh, Bair & Movshon, 2002). See next section of this review *(Addition 4)* for more about surround suppression.

3.2. An example of contrast subtraction: Contrast comparison

An example of a subtractive nonlinearity that operates on contrast (the lower right cell in Fig. 8) is the recently suggested process named *contrast comparison* (Graham & Wolfson, 2007; Wolfson & Graham, 2007). In explaining visual perception, it is usually assumed that the action of contrast is monotonic with 0% contrast (unchanging gray) being least effective and contrast values greater than 0% producing greater effects. However, some recent psychophysical evidence suggests that, at least for some pathways between visual image and perception, something highly non-monotonic with contrast occurs, something like the suggested contrast-comparison process, which goes like this: A comparison level is computed at each spatial position in the pattern; this comparison level equals the recent weighted average contrast at that position (averaged over a short period, a fraction of a second). The comparison level at a position is subtracted from the current value of contrast at that position, and thus the comparison level plays the role of the constant $\boldsymbol{a}$ in Fig. 8, making this process an example of contrast subtraction. To form the final output of the process, the magnitude of the difference between current and recent average contrast is retained, but information about the sign of the difference (increase vs. decrease in contrast) is lost or degraded.

Fig. 14 shows the input–output function for the proposed contrast-comparison process: the input–output function at a particular spatial position translates horizontally on an axis linear with contrast, when the recent history of contrast at that position changes as shown by the arrows.

The input–output function is drawn in Fig. 14 as a *full-wave rectification* centered at the current comparison level, that is, the output is the absolute value of the difference between the current contrast and the comparison level. This full-wave rectification function would predict complete loss of information about the sign of the contrast change. A complete loss turns out to be too dramatic for any of our observers so far. Thus the actual function used in the model will be less extreme (as will be discussed below in reference to the bottom panel Fig. 17).

3.2.1. An experiment showing the straddle effect

The evidence for this proposed contrast-comparison process comes from an effect we initially found rather surprising (called

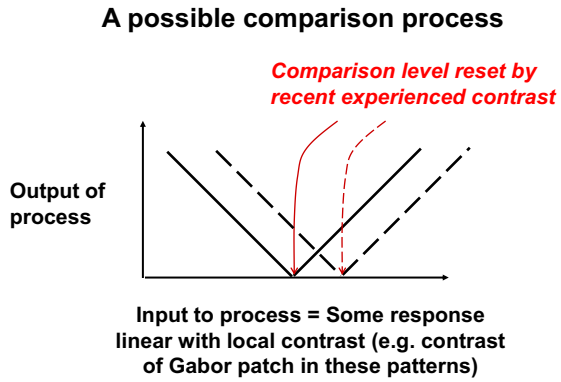

Fig. 14. A proposed *contrast-comparison process*. This is an example of a *contrast subtraction* process. Modified from Graham and Wolfson (2007), Wolfson and Graham (2007).

the *straddle effect* here) of the contrast of one pattern on the perception of an immediately-following test pattern. In particular, when a test pattern is composed of two contrasts straddling the contrast of the pattern that immediately preceded it – that is, when one test contrast is above the previous contrast and the other test contrast is below – the test pattern is very difficult to perceive correctly.

Fig. 15 shows the characteristics of one experimental task producing the straddle effect. An *Adapt pattern* – composed of four identical Gabor patches – is presented for a short period of time. The Adapt pattern is immediately followed by a brief-duration *Test pattern*. The task of the observer is to say whether the global stripes are horizontal or vertical. See figure and its legend for more details. The adapt stimulus as shown in Fig. 15 is only on for a fairly short time (1 s). Thus, while we are use the language of adaptation experiments here, the experiment might also be called a study of masking. It might also reasonably be called simply a study of the temporal processing of visual contrast.

Some results from an experiment like that in Fig. 15 are shown as the data points in Fig. 16. These results are from an experiment where five different *adapt contrasts* and many average *test contrasts* were used; the difference between the two test contrasts was always 10%. In the left half, all the results from one observer (from all five adapt contrasts) are plotted on a horizontal axis that gives difference between adapt and average test contrast. In the right half, they are plotted in separate panels by adapt contrast (for each of the two observers). The horizontal axis is the average test contrast and a diamond on the horizontal axis marks the value equaling the adapt contrast. (The solid lines are predictions discussed below.)

The *straddle effect* is the sharp dip in performance, which produces a notch in the curve, when the average test contrast is very near the adapt contrast so that the two test contrasts straddle the adapt contrast. For the results in Figs. 16 the adapt duration was 1 s, but much shorter adapt durations, e.g. 50 ms, will still produce the straddle effect (Foley, 2010; Graham & Wolfson, in press).

Notice that the lowest point in the notch (the minimal performance in the straddle effect) does not always get as low as chance (50% in this experiment). See, for example, the panels for 25% adapt contrast. This above-chance performance is an example of partial retention of information about the sign of a contrast change, of the visual system's partial although far from perfect ability to distinguish increases and decreases of equal size.

Also notice that, for average test contrasts just above or just below the contrast range showing the straddle effect (above or below the average test contrasts in the notch), the observer's performance is excellent. However, as the average test contrast gets very far below or very far above the adapt contrast, performance declines again. In other words, adaptation generally improves performance for test contrasts that are close to the adapt contrast relative to those further away but there is a dramatic exception for the straddle effect. Or to put it in other words: If you view each curve in the right side of Fig. 16 as the silhouette of a mountain that happens to have a canyon at its very peak, then you could say that that mountain's peak moved from the left side (near 0%) to the right side (near 100%) as the adapt contrast moved from the left side to the right side.

Improving the observer's performance for contrast values near the recent adapt contrast (moving the mountain's peak in the figure to follow the value of the adapt contrast) is often suggested as a valuable function of adaptation. Perhaps the bad performance shown by the straddle effect (the canyon at the very peak, that reflects the inability of the system to keep track of contrast increases

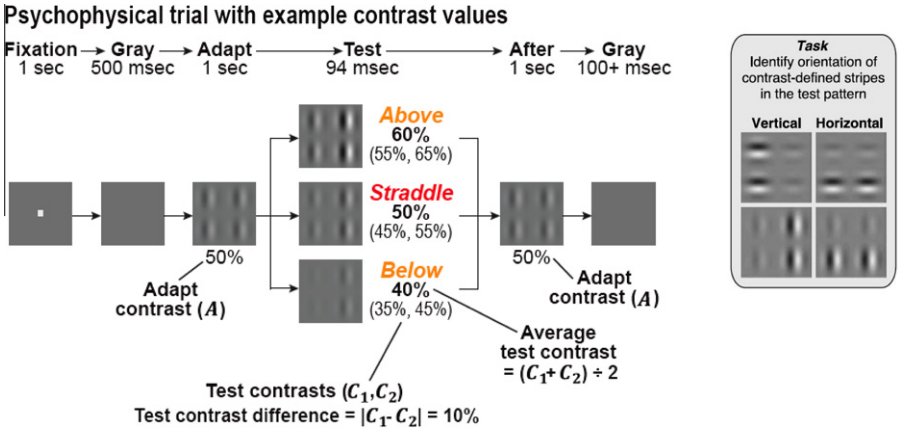

Fig. 15. A diagram of a single trial from an experiment that demonstrates *the straddle effect* and thus suggests the existence of *contrast comparison*. Although only one test pattern is presented on any single trial, three possible ones are shown here to illustrate the 3 cases where the two test contrasts are *above* the adapt contrast, or *straddle* the adapt contrast, or are *below* the adapt contrast. The test patches always are identical except for contrast to those in the Adapt pattern and are in the same spatial position as those in the Adapt pattern. The observer's task is to identify the orientation of the contrast-defined stripes (the global orientation) as horizontal or vertical. The orientation of the patches (local orientation) and the orientation of the global stripes (global orientation) can be either horizontal or vertical, and they vary independently from trial to trial. See Wolfson and Graham (2009) for more details of this and related experiments and their results. In the task shown here, the Test pattern is then followed by another presentation of the Adapt pattern. However, replacing that second Adapt presentation with gray does not erase the straddle effect in the typical situation. (See Supplementary Fig. 1 in Wolfson & Graham, 2009.) This experiment can be called an adaptation experiment (as implied by the terms used in this figure), or a masking experiment, or simply an experiment to explore the temporal properties of visual contrast processing.

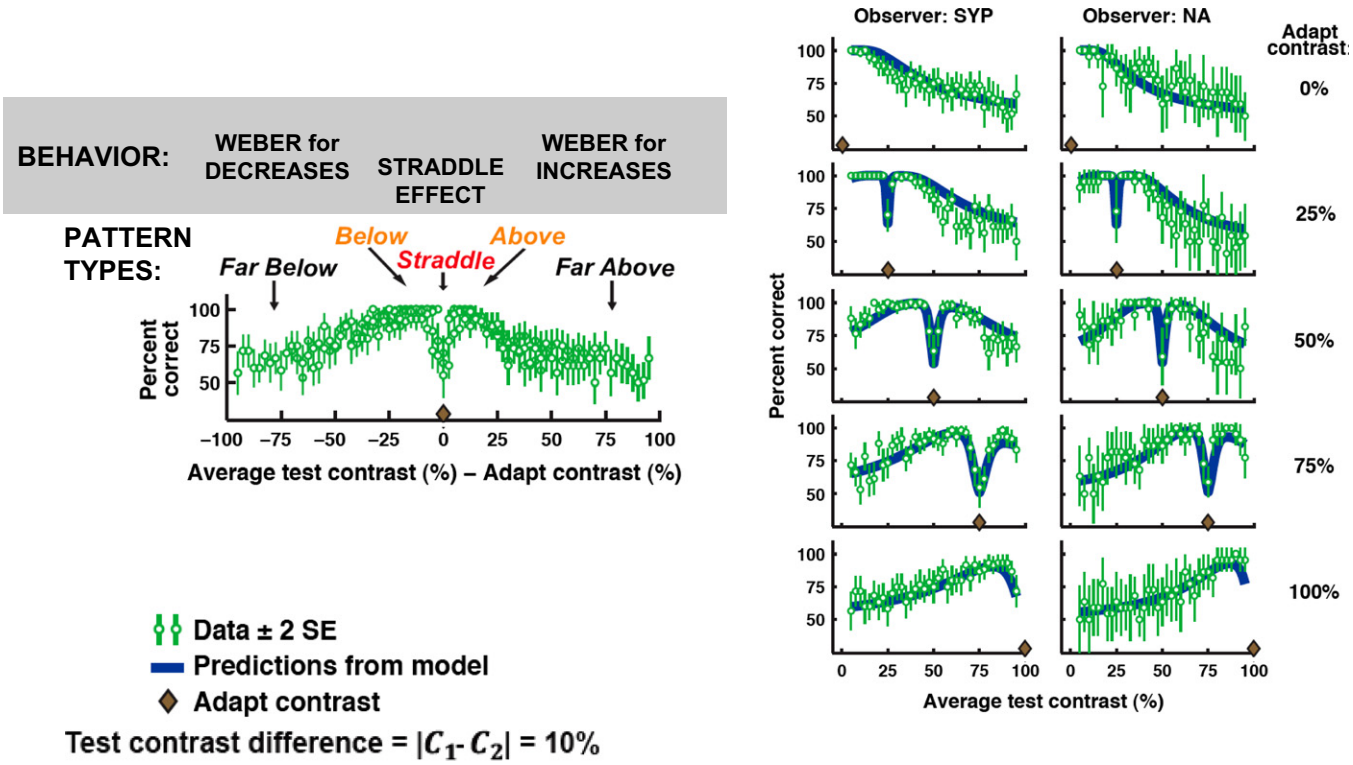

Fig. 16. The data points in the left part of this figure show one observer's results from the experiment in Fig. 15, for five different adapt contrasts, superimposed so that the adapt contrast is at the center of the horizontal axis. The right part of this figure shows the results for the five different adapt contrasts separately for two observers. Error bars show ±1 standard error calculated across sessions. The solid lines are the predictions from the model of Figs. 17 and 18. Experimental results from Graham & Wolfson, 2009; Predictions from Graham et al., 2009.

versus decreases) is an undesirable but relatively minor side effect of the overall useful value of this adaptation. Or, to speak in terms of process, perhaps the bad performance shown by the straddle effect is due to some design constraint; perhaps evolving a system to produce the best (fastest, most efficient) contrast-change detection is very difficult unless there is some accompanying loss of information about the direction of the change. See further discussion in Graham and Wolfson (2007, in press).

3.2.2. Weber-law behavior along with the straddle effect

The decline at the tails of the function is an instance of *Weber-law behavior*, since, in the tails, performance depends on the ratio of two intensities and depends in such a way that performance declines as that ratio becomes closer to 1. In this case the ratios are on an intensity dimension that is not something as simple as luminance or contrast; instead the dimension is the unsigned difference between current contrast and recent time-averaged contrast. For this experiment, it is the unsigned difference between a contrast in the Test pattern and the contrast in the Adapt pattern that preceded it. See further discussion in Wolfson and Graham (2009).

As was mentioned in the previous section (*Addition2)* contrast-normalization models can predict Weber-law behavior.

Shortening the adapt duration much below 1 s dramatically changes the behavior at the tails of the function, particularly the left tail, an outcome quite different from the lack of change of the notch (Graham & Wolfson, in press).

3.3. Model incorporating a divisive nonlinearity and a subtractive nonlinearity

The abstract model shown in the diagram of Fig. 17 and embodied in the equations of Fig. 18 contains both a divisive nonlinearity (contrast normalization) and a subtractive nonlinearity (contrast comparison). The diagram in Fig. 17 only shows in detail one channel, but the model contains many channels sensitive to different ranges of spatial frequency and orientation. The responses of all the channels are assumed to form the input to a decision stage (as in the "Decision rule" oval of Fig. 1 or the "Pooling and Decision" square of Fig. 5) although that decision stage is not explicitly shown in Fig. 17.

Contrast normalization (the divisive nonlinearity) appears in Fig. 17 as a *normalization pool* (upper right oval) that exerts influence on the signal channel. This normalization pool can include responses from many channels.

Contrast comparison (the subtractive nonlinearity) also appears in Fig. 17. Its input–output function (as in Fig. 14) is shown as Part 2 of the abstract model, and there is also a *comparison pool* (upper left oval) the output from which is integrated to determine the comparison level. The comparison level in any one channel might depend not only on its own stimulation but also on the contrast stimulating other channels. The comparison pool in the figure allows for this possibility, but there is no evidence for it yet.

Parts 1 and 3 are abstract in the model of Fig. 17. One concrete model for all 3 parts of the channel is given in Graham & Wolfson, 2007. That concrete version of a channel is like an *FRF* process except that instead of a simple rectification-type function *R* there is the machinery of the contrast comparison.

To explain the fact in experimental results that there is only partial loss of information about sign of a change (e.g. the notch in many panels of Fig. 16 does not dip all the way down to 50%), any number of modifications of the model in the top half of Fig. 17 would undoubtedly work. The modification used here is to assume pairs of channels where the two members of a pair are identical except the input–output functions of the contrast-comparison process are shown in the bottom of Fig. 17. The functions are intermediate between a half-wave type and a full-wave

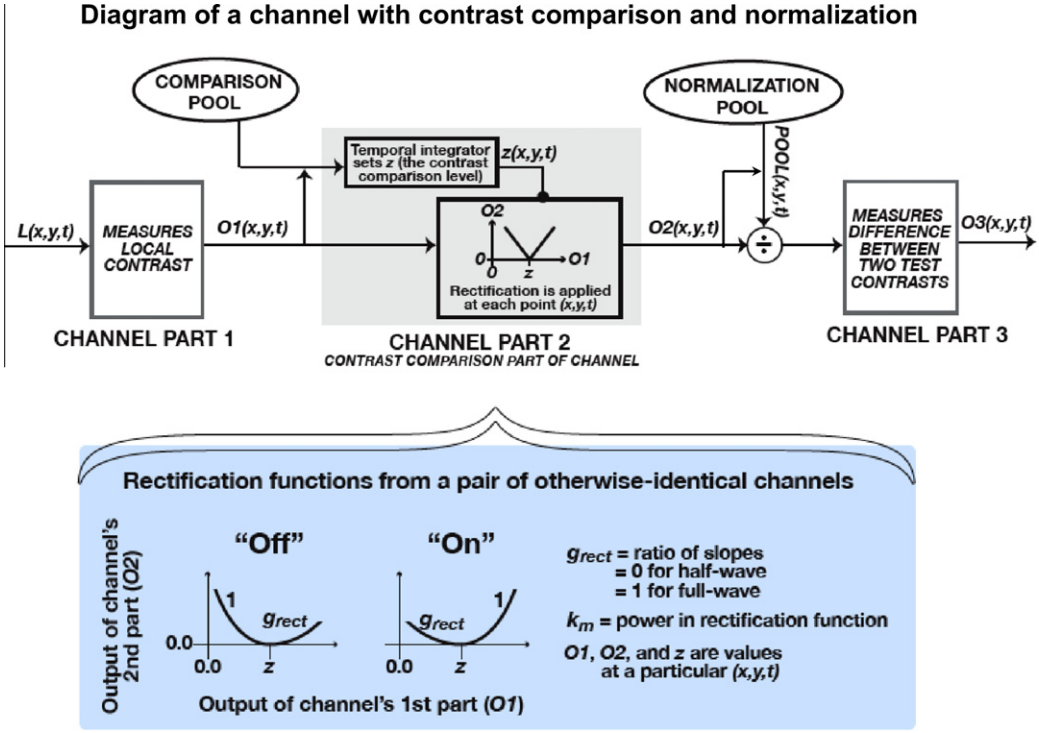

Fig. 17. *Top:* Diagram from a model containing both contrast normalization and contrast comparison. $L(x, y, t)$ is the luminance at each position in space and time. $O1(x, y, t)$, $O2(x, y, t)$, and $O3(x, y, t)$ are the outputs of the three parts of the channel. The output of the comparison pool (upper left) sets the contrast-comparison level z in Part 2 of the channel. The output of the normalization pool (upper right) divides the output from Part 2 of the channel. Parts 1 and 3 of the channel could both be linear filters as in an ordinary *FRF* process. *Bottom*: Diagram of the pair of input-output functions at the contrast-comparison stages in a pair of otherwise identical channels. The "On" channel responds to increases in contrast more than to decreases. The "Off" channel does the opposite.

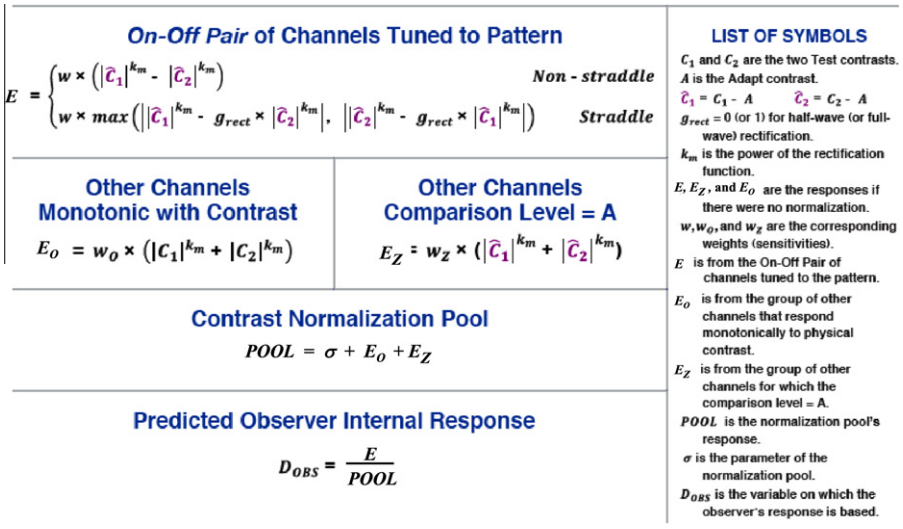

Fig. 18. Equations derived from the model shown in Fig. 17. Fitted predictions from these equations are the solid lines in Fig. 16.

rectification, and one member of the pair is more sensitive to increases in contrast (the "On" channel) and the other more sensitive to decreases in contrast (the "Off" channel). Further, to fit other details of the experimental results, each half of each input–output functions needs to be slightly accelerating upward (as shown) rather than being a straight line.

The Weber-law behavior in the experimental results (Fig. 16) for average test contrasts far from the adapt contrast (in either direction) can be easily explained by adding contrast normalization to the model as in the previous section (*Addition 2*). It is necessary to assume that channels other than the channel at issue contribute to the normalization (e.g. Graham, Beck, & Sutter, 1992). Or, to put it another way, the *normalization pool* in Fig. 17 must contain other channels than the channel itself to predict the experimental results.

One of the advantages of relatively abstract models with parts which can themselves be represented by relatively simple equations is that often they can then be used to generate simple

equations that describe the joint action of many channels (units) and thus of an observer. The abstract model of Fig. 17 containing contrast comparison and normalization is an example of this advantage. As it turns out, it can generate the simple set of equations shown in Fig. 18. These equations were used to compute the predictions (solid lines) in Fig. 16.

The predictions are an excellent fit not only to the overall shapes of the curves but also to differences among the curves for different adapt contrasts. Individual differences between observers are predicted as well. Some more information about the derivations of these equations and their fit to experimental results is given in Appendix C.

An aside about static vs. dynamic modeling. The equations above are all static equations: they do not contain any explicit modeling of time that would allow predictions for arbitrary time-courses of visual stimulation. Instead the dynamic properties of the visual system are only represented implicitly by the differing locations of the comparison level depending on recent history and by the approximation of spatiotemporal filters with purely spatial filters.

3.4. Several further references

We know of several other studies in the literature that differ in important ways from the experiment and results in Figs. 15 and 16 but that are nonetheless suggestive of a subtractive comparison from moment to moment (Heeger & Ress, 2004) or something like a contrast-comparison process including both contrast subtraction and rectification (Kachinsky, Smith, & Pokorny, 2003; Zenger-Landolt & Koch, 2001). These are discussed further in Wolfson and Graham (2009).

What the physiological substrate might be for the contrast-comparison process is unclear at this time. Of possible relevance is a human fMRI study demonstrating that in V4 (but not earlier areas) the response to contrast decreases is of the same sign as the response to contrast increases (Gardner et al., 2005). The duration of patterns in the fMRI study are very different from those in the psychophysical straddle-effect studies, however, so caution should be maintained about any analogue.

4. Addition 4. Non-classical receptive fields (including surround suppression and facilitation, cross-orientation or overlay inhibition)

The preceding two sections (*Additions 2 and 3*) concentrated on the intensive characteristics of neurons and units (the magnitude of response as a function of input). We return here and in the section after this (*Addition 5*) to the question of spatial characteristics.

The receptive field of a neuron was defined earlier as all positions on the retina at which stimulation will directly evoke a response. For the *classical V1 simple cell* the receptive field could be modeled as a linear filter (an adding and subtracting device) followed by a half-wave rectification (to account for the fact that the rate of spikes could never go below zero). This earlier model is referred to as the *classical V1 receptive field* to avoid possible confusion with a more general notion of receptive field we are about to discuss – that of the classical receptive field with an additional *non-classical surround* enlarging it and/or *non-classical processes* spatially juxtaposed with it.

The classical receptive field of a V1 simple cell is very small relative to the distances over which visual perception has to function. Indeed, the classical receptive field is typically composed of only a few inhibitory and excitatory sub-sections (like those sketched in the figures up to this point). Or, in other words, the classical receptive field has a width of only 1 to 3 periods of the cell's preferred spatial frequency.

Anything that is not predicted by the earlier model could be referred to as a non-classical response, but the typical non-classical response talked about at present is a response to one pattern that is affected by the simultaneous presentation of a second, adjacent pattern where the second pattern, alone, does not produce a response. These non-classical responses of V1 simple cells can occur over a substantially larger area than the classical receptive field. This is one reason that non-classical receptive fields are now frequently invoked in explanations for perceptual phenomena.

Non-classical receptive fields of V1 cells were first described more than 25 years ago. (Carandini, 2004 credits Levick, Cleland, & Dubin, 1972 with first using the name "suppressive field" for some of these non-classical characteristics.) But their exploration has been much more extensive in the last couple of decades of physiological research (e.g. Jones, Grieve, Wang, & Sillito, 2001; Levitt & Lund, 1997; and see bibliographies of the model-containing papers described below for many more). Further, these non-classical receptive fields have been invoked to account for a number of psychophysical phenomena as well (see references below and their bibliographies).

The top panel of Fig. 19 sketches typical stimuli used to explore – both physiologically and psychophysically – non-classical effects coming from the surround outside the classical receptive field. The test pattern (middle column) is smallish and centered on a classical receptive field. The mask pattern (left column, top panel in figure) is offset from the test pattern in such a way that it does NOT stimulate that classical receptive field. The physiological or psychophysical response is measured both when the test pattern is presented by itself (middle column) and when it is presented together with the mask pattern (right column).

The assumption that the test pattern is centered on a classical receptive field is supported straightforwardly when studying single neurons. In psychophysical experiments the observer's response always reflects responses from many units, so the assumption of centering is supported by the following argument. One unit (or a small group of very similar units) is assumed to be the unit most sensitive to the test pattern and therefore the

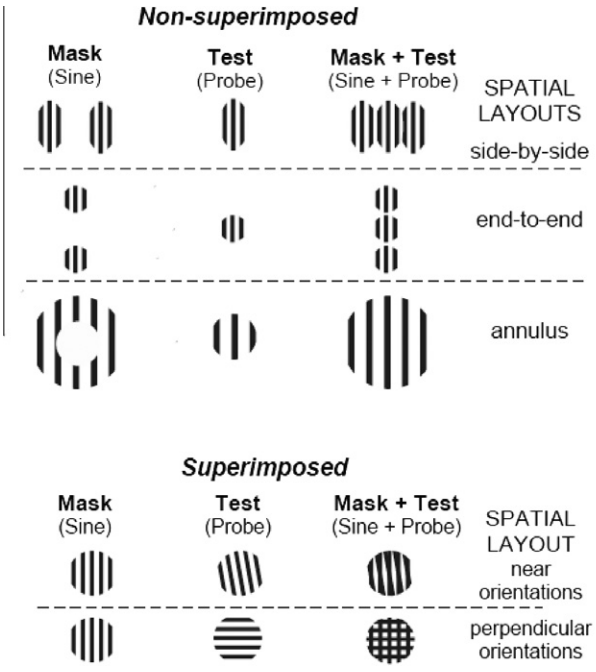

Fig. 19. Arrangements of stimuli used to study non-classical receptive field phenomena.

primary determiner of the observer's response. That unit (or small group) has a receptive field at the position of the test pattern.

The typical effect of a surround stimulus is to suppress the response to the center stimulus, and the suppression is strongest when the spatial frequency and orientation of surround and center stimuli are similar. Hence *surround suppression* is a frequently-used term. Two early studies showing such suppression in psychophysical experiments (with masks flanking a test of the same orientation as in top row of Fig. 19) were Rogowitz (1983) and Ejima and Miura (1984).

In addition to the possibility of non-classical suppressive (or facilitatory) effects from outside the classical receptive field, there is also a possibility that the same non-classical effects extend *inside* the classical receptive field. And perhaps there are different non-classical processes that exist inside the classical receptive field but not outside. Evidence regarding this possibility of a juxtaposed or overlaid process can be investigated by using mask patterns of approximately the same size and location as the test pattern (as in the bottom panel of Fig. 19).

The exact features of non-classical inhibitory and facilitatory effects are now known to depend not only on the stimulus characteristics represented in Fig. 19 but on many other parameters of pattern vision (retinal location, exact distance between the mask and test, spatial frequency content of each of mask and test, temporal frequency content, color content, dichoptic vs. binocular vs. monoptic, etc.). And this is true in both V1 physiology (see, e.g., Carandini, 2004; Cavanaugh et al., 2002; Webb, Dhruv, Solomon, Tailby, & Lennie, 2005) and as inferred from human psychophysics (e.g. see Meese, 2004; Petrov, Carandini, & McKee, 2005; Snowden & Hammett, 1998; Yu, Klein, & Levi, 2003). As in previous sections, however, none of these dependencies will be described here in detail. Many appropriate references can be found in the bibliographies of the references given here.

4.1. How are these non-classical effects explained and modeled?

The next figure (Fig. 20) illustrates two of the possible non-classical V1 effects that have been documented in the physiological and psychophysical literature. The two floating squares in a panel represent the same area in the visual field. The classical receptive field of the neuron being studied is shown in the bottom floating square of each panel, and it has the classical arrangement of an excitatory center and inhibitory side flanks. The classical receptive fields of the neurons exerting the suppression are shown in the top floating square of each panel. *Cross-orientation overlay suppression* from superimposed test and mask stimuli of perpendicular orientations is shown in panel A, and *surround suppression* from non-superimposed test and mask stimuli of the same orientation is shown in panel B. Although both panels here show inhibitory (suppressive) effects, facilitatory effects have also been suggested.

An aside about terminology: Many terms are used with slightly different meanings in the vision literature. The same term can sometimes mean an effect in empirical results, sometimes a process to explain such an effect, sometimes the experimental procedure that produces the effect, and sometimes still other concepts. *Light adaptation* is one common example of such term. And *surround suppression* is one (but not the only) in this review. I have tried to be clear at each point I use such a term. I have undoubtedly failed.

Existing attempts to explain these suppressive and facilitatory effects can be roughly characterized in terms of 2 characteristics that we discuss in turn.

4.1.1. Lateral (e.g. intracortical) connections or feedback connections

One of the major characteristics distinguishing the explanations of these non-classical effects is whether they assume the effects are primarily due to lateral connections (e.g. intracortical excitation or inhibition) or primarily due to feedback connections.

This distinction is illustrated in the contrast between Fig. 21 (an explanation based on lateral connections) and Fig. 22 (an explanation based on feedback from higher areas). See legends for more details.

These two figures were from a study of single V1 neurons, but the distinction can easily be invoked for psychophysics as well by considering generally top-down effects of various sorts.

4.1.2. Simple equations or large-scale numerical computer simulations

Some explanations are simply verbal accounts of possible mechanisms. These verbal explanations are often very valuable in the initial studies of any scientific question. The study of vision is far enough along, however, that developing more rigorous explanations – models – is often possible. Models themselves can be of very different flavors as illustrated here with brief discussions of two published models of non-classical receptive field effects. Both models address whether suppression is divisive or subtractive, and both address whether suppression arises from intracortical, lateral connections or feedback from higher visual areas. The models differ in their complexity: one model examines the steady state (of either behavior or of neuronal firing) using relatively simple computational stages expressed as simple equations; the other is elaborate and reductionist, incorporating anatomical and biophysical facts about many neurons requiring large-scale numerical computer simulations.

The example of the simple-equations abstract-model approach is from Cavanaugh et al., 2002. (This is the paper from which Fig. 13 was taken. A later paper from the same lab produced Figs. 21

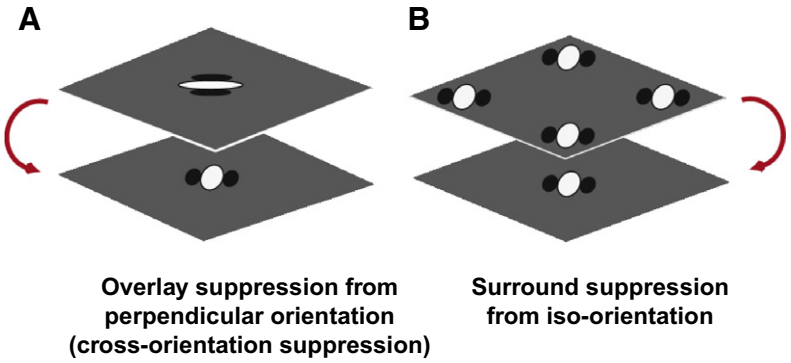

Fig. 20. Interpretation of suppression phenomena in V1 in terms of intracortical inhibition. A: Cross-orientation suppression might be explained by inhibition between V1 neurons with overlapping receptive fields and different preferred orientations. B. Surround suppression might be explained by inhibition from V1 neurons with displaced receptive fields and similar preferred orientations. (With slightly expanded labeling from Fig. 8, Carandini, 2004, with permission MIT Press.).

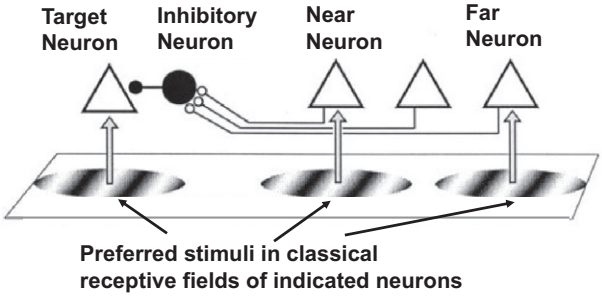

Fig. 21. Possible circuit for iso-orientation surround suppression in V1. The target neuron (left triangle) receives inhibition from a nearby neuron (large filled circle) that is driven by excitatory neurons displaced laterally in cortex (three triangles at right) that have orientation tuning similar to the target cell. Preferred stimuli confined to the classical receptive fields for three excitatory neurons are indicated by the circular patches of sinusoidal grating shown on the tilted plane that represents a two-dimensional visual field. Block Arrows extending vertically from the stimuli (grating sketches) to the neurons (triangles) indicate localized feedforward inputs to the excitatory neurons. Figure and legend modified from Fig. 1 of Bair, Cavanaugh, and Movshon (2003).

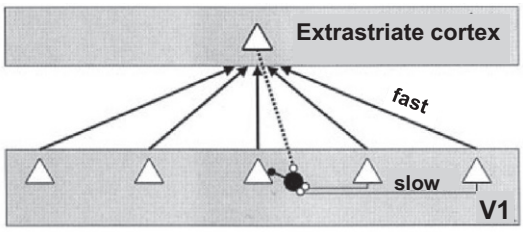

Fig. 22. Cells (triangles) in V1 (bottom gray box) project via fast axons (arrows) to a higher visual area (top gray box) where neurons have larger classical receptive fields created by convergent input. If cells from the higher area project back (dotted line) directly or indirectly to local inhibitory neurons in V1 (black circle) then suppression from far regions of the surround could arrive on target cells in V1 with little additional delay compared with suppression from the near surround. Figure and legend modified from Fig. 8 of Bair et al. (2003).

and 22 as well.) This model extends previous work on divisive suppression (e.g. Carandini et al., 1997; Chen, Kasamatsu, Polat, & Norcia, 2001). This allows them to write simple equations like those shown elsewhere in this article. These equations depend on very few parameters and the input is simply represented often as just two contrast values – that of the test pattern and that of the mask pattern. They then compare these equations to the results. For their results, surround suppression was best described as a divisive change in response gain (Cavanaugh et al., 2002).

A dramatic contrast to this simple-equations model can be found in a very large-scale neural model (McLaughlin, Shapley, Shelley, & Wielaard, 2000; Wielaard & Sajda, 2006). This model depends both on analytic work and on extensive computation and explicitly computes the activity of many neurons. This is a reductionist model that tries very carefully to represent in each model neuron as much detail as possible based on anatomical and physiological studies at a somewhat sub-neuronal level. For the V1 model, it incorporates the known lateral connections in detail. It also models the dynamics not just the steady state. The output from each of many neurons is a spike train that is realistic in its probabilistic as well as deterministic properties. There is nothing simple or transparent about this large-scale model of V1 cells, however. Understanding why it produces any particular prediction becomes an investigation in itself. I find this a drawback. One thing that is desirable from a theory is that it provides a way for human thought to encompass the phenomena and understand their implications. On the other hand, a complicated reductionist model is impressive when it predicts phenomena that were not built into it. And this model successfully predicts non-classical receptive-field phenomena including surround suppression. The successful prediction by the model depends on the lateral connections in V1, and it does NOT depend on any feedback from later stages. Lateral connections in a dynamic model like this have indirect effects, so that effects can extend over distances substantially longer than one connection length if sufficient time has elapsed.

4.2. At the boundary between this section and the next

This section (Section 4, on non-classical receptive fields) and the next (Section 5, on contour integration) overlap in a number of ways, but perhaps in no way more than in the studies using the stimulus configuration shown in the top two rows of Fig. 19 where the test pattern is a Gabor patch and the mask is two Gabor patches, one on either side of the test. This configuration has been well studied, particularly in psychophysics. A great deal of work of this sort was done by or inspired by Polat and Sagi and their colleagues (e.g. Chen et al., 2001; Polat & Sagi, 1993, 1994; Sterkin, Yehezkel, Bonneh, Norcia, & Polat, 2009). These studies are frequently interpreted as giving information about the properties of cortical lateral connections (as in the model Fig. 21, and also the large-scale neural model just discussed). And thus these studies belong in this section on non-classical receptive-field mechanisms.

On the other hand, these studies also belong in the next section (on *Addition 5*, see especially the Section 5.1.3 on *Contour integration explained by association fields*). Particularly when the configuration in the middle row of Fig. 19 is used (when patches are of same orientation and are in a line), the results tend to show facilitation. This facilitation is then frequently discussed as part of the process that allows us to perceive contours, e.g., to see a contour as continuous even if parts of it are occluded.

4.3. A sample of further references

Color-selective influences from outside the classical receptive field, e.g. surround suppression, have been seen in single neuron recordings in V1 and higher levels. See Foster (2011) and Shapley and Hawken (2011).

Empirical and theoretical relationships between surround suppression and second-order mechanisms are explored in an fMRI study of human cortex (Hallum et al., in press).

Surround suppression may be a general computational principle at many levels of the visual system. For example, it has been reported for singe-neuron recordings in the lateral intraparietal area (LIP), an area that is often described as forming a priority map directing attention and eye movements (Falkner, Krishna, & Goldberg, 2010). Also see Carrasco (2009).

A recent large psychophysical study measuring the effects of flankers on several psychophysical tasks in human observers with normal vision and human observers with amblyopia – as well as references to a wide variety of previous literature – is reported in Levi and Carney (2011).

5. Addition 5. Contour integration

I suspect the intuition that contours are important for perception has existed in humans for tens of thousands of years. It is difficult to believe, for example, that the outline animal figures in the stone-age cave paintings (e.g. search the web for Peche Merle, France, to see some done 15,000–25,000 years ago) were done by someone without such an intuition. We do not yet understand how the visual system discovers and processes the contours. Over the last 25 years, however, there has been substantial development of possible ideas.

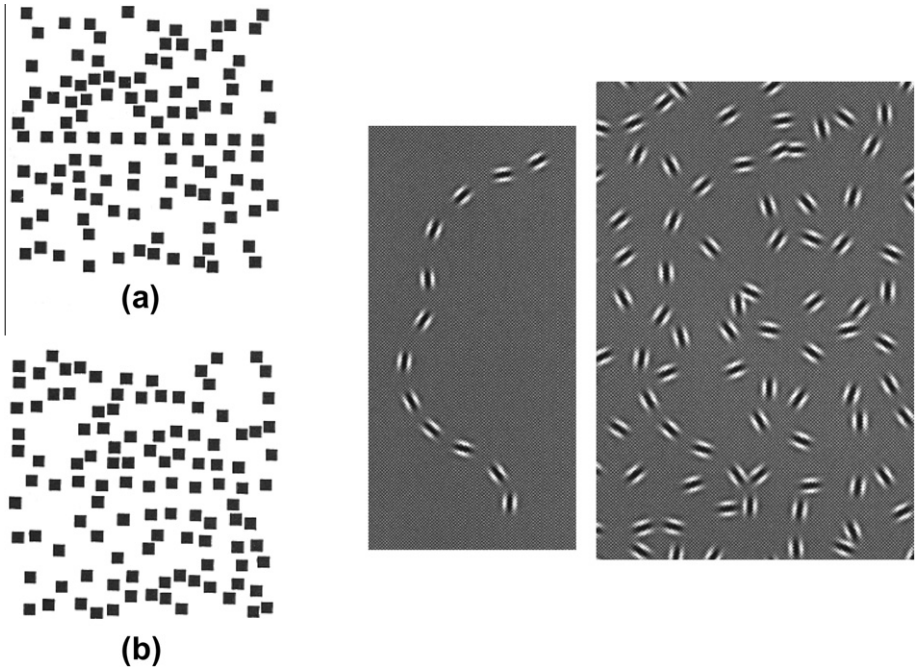

Fig. 23. *Left part.* There is a horizontal line of aligned squares in the top panel, and a horizontal line of misaligned squares in the bottom panel. Example of pattern used in Beck et al. (1989). *Right part.* The curve of Gabor patches in the left image is repeated in the right image but other distracter Gabor patches are added. Based on Fig. 3 of Field et al. (1993).

Fig. 23 shows two of the visual patterns that have been used in the experimental study of contour integration. Each pattern is composed of many elements, and one subset of the elements lies along a path (a contour) while the others are randomly scattered around that path. In the left panel the elements are solid black squares (Beck, Rosenfeld, & Ivry, 1989). In the right side of the figure the elements are Gabor patches (Field, Hayes, & Hess, 1993). The characteristics of the elements are often varied – e.g. in orientation or spacing – in an attempt to find out what influences the ease with which the human observer can see the contour.

Much of the research on contour integration was influenced by the Gestalt ideas about perceptual grouping, and how such ideas might explain our perceptual ability to see what pieces of a contour belong together even when intervening regions of a contour are obscured. The research was also influenced by work in computer science and, in particular, by attempts to develop computer algorithms that could make different points along the length of a curved edge cohere and compare the outputs of the algorithms to perceptions. References to these earlier influences can be found in the bibliographies of Beck, Rosenfeld, and Ivry (1989), Field et al. (1993) and Prins, Kingdom, and Hayes (2007).

5.1. Possible processes for contour integration

Ideas for explaining contour integration vary among themselves. And they overlap with all of the preceding ideas. Some overlap with higher-order processes (in *Addition 1*); others incorporate well-specified contrast nonlinearities (and thus overlap with *Additions 2 and 3*); and still others invoke interactions across spatial positions resulting from lateral connections as in non-classical receptive fields (and thus overlap with *Addition 4*).

5.1.1. Contour integration as higher-order processing

Fig. 24 illustrates two suggestions made for contour perception. Both models include two stages and are thus analogous to the second-order processes of *Addition 1*. But the suggestions in Fig. 24 have a slightly different flavor than typical examples of higher-order processes, at least in my reading. Underlying these two suggestions in Fig. 24 there seems to be an idea of actively finding a contour rather than just a notion of describing a fixed entity that could be a general-purpose processing tool. (For example, is there a fixed entity for every degree of curvature? Or is it more actively formed when there is a curve in the field?) I may well be wrong in my reading of particular examples. But even if I am right, the general point holds that higher-order mechanisms like those of *Addition 1* may be a possible piece in solving the puzzle of how we see contours.

Lets look briefly at the two examples of Fig. 24 more closely. Both these suggestions suppose that pieces of contour are first sensed by simple linear units, that is, by units described by the classical V1 cell model. And then the pieces are put together nonlinearly by a second stage that has a larger spatial reach. In the top diagram of a border-detecting mechanism, the non-linear process operates between two different phases of local unit at each position and then along the length of a number of local units (from Shapley & Gordon, 1985). The bottom diagram is a schematic used to argue that the information from the various orientations of simple linear units along a curve is sufficient to tell later stages about the curve (figure from Prins et al., 2007, based on ideas of Wilson (1985) and Wilson and Richards (1989)).

5.1.2. Contour integration explained by association fields

Fig. 25 shows a diagram of a suggested process, an *association field.* (This comes from the Field et al., 1993 study that used the Gabor-patch contours in the right part of Fig. 23.) These investigators hypothesized that an association field integrates information across neighboring first-stage linear units tuned to similar (but not necessarily) identical orientation. The top panel in Fig. 25 shows the constraints specifying which Gabor patches in the field at different positions can be associated. The bottom panel shows the specific relationships allowed between the orientations of associated patches at different relative positions. The solid lines indicate orientation–position relationships that lead to "association" and the dotted lines indicate relationships that do not lead

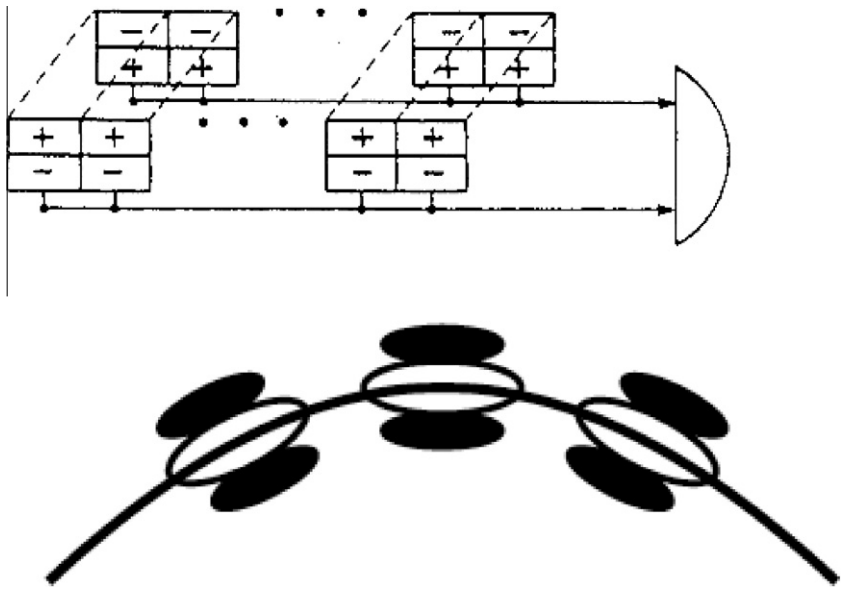

Fig. 24. *Top panel.* Diagram of a border-detecting mechanism. Outputs from local contrast-sensitive units sensitive to one polarity are summed in a non-linear manner with those from local units of opposite polarity. (Each local unit is shown as a collection of four boxes, two marked + and two marked −. The non-linear summation may be absolute value, square, etc.) There is also non-linear summation along the length of the border detector's receptive field, and thus a border detector's receptive field is longer than that of individual local detectors. Diagram from Fig. 4 of Shapley & Gordon (1985). *Bottom panel.* A diagram from Fig. 9 of Prins et al. (2007) of a curvature processing model from Wilson (1985) and Wilson and Richards (1989).

to "association". While not explicitly represented in this figure, the authors expect the association field to show dependence on parameters other than orientation and position, e.g. spatial frequency. This diagram represents the connection pattern for one unit (neuron) and every unit has this same connection pattern suitably translated and rotated.

The authors speculate about possible physiological mechanisms and also about the relationships of this idea to existing computational models of curve extraction. In so doing, they make further suggestions about the possible nature of the association-field process. One suggestion is that long-range lateral connections in primary visual cortex (known to exist anatomically and physiologically) could be considered as a possible basis for the associations. As the authors point out, however, it is not clear whether the anatomical/physiological connections in V1 have the right properties to be the substrate for the hypothesized association field based on psychophysical results. More abstractly, their suggestion can be interpreted in terms of possible lateral connections among entities that are not necessarily in V1.

The example model we look at next is one that specifies many details of the lateral connections, and is used to calculate detailed predictions. This example is a model published in 1998 that took the increasing body of physiological and anatomical data on V1 available at that time, and incorporated this information into an elaborated model that included temporal dynamics as well as spatial position (Zhaoping, 1998).

Predictions were calculated to see how well the model could account for perceptual phenomena (e.g. contour integration) where the features of the neurons in the model that were based on available physiological data were kept constant, but other features were allowed to change. These fits of predictions to perceptual results led to two kinds of conclusions, one of which is of more concern to us here – namely, a conclusion about the extent to which processes of the kind found in V1 could serve as an explanation for perception. (The other kind of conclusion – beyond our scope here – is making predictions of physiological results not yet found by using constraints on the model produced by fitting it to perceptual phenomena.)

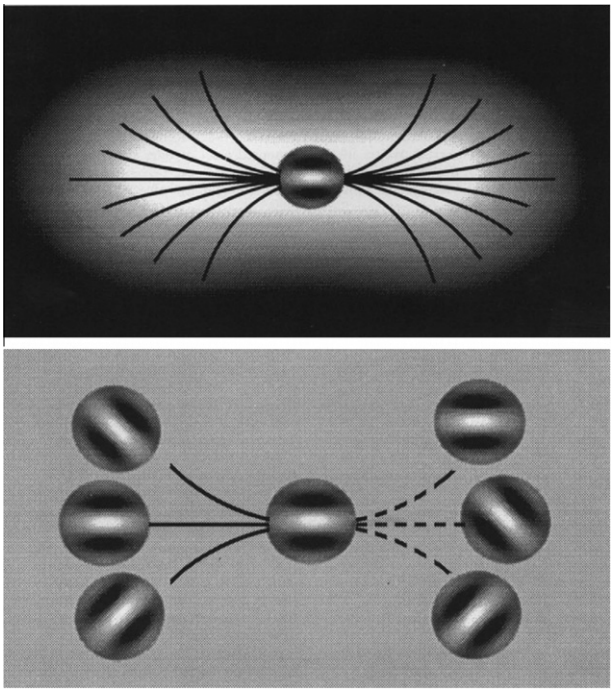

Fig. 25. Diagram of an *association field*. The top panel represents the rules by which the elements in the association field are associated and segregated from the background. The bottom panel represents the specific rules of alignment. The solid lines indicate orientation–position relationships that do lead to "association", and the dotted lines indicate relationships that do not lead to "association". These relationships are those suggested by the experimental results. Based on Fig. 16 of Field et al. (1993).

This model incorporates initial processing by linear units with orientation-selective receptive fields, and then has further processing by a network of recurrently-connected excitatory and inhibitory neurons. These lateral recurrent interactions modify initial activity patterns. The modifications selectively enhance activity that forms smooth contours in the image.

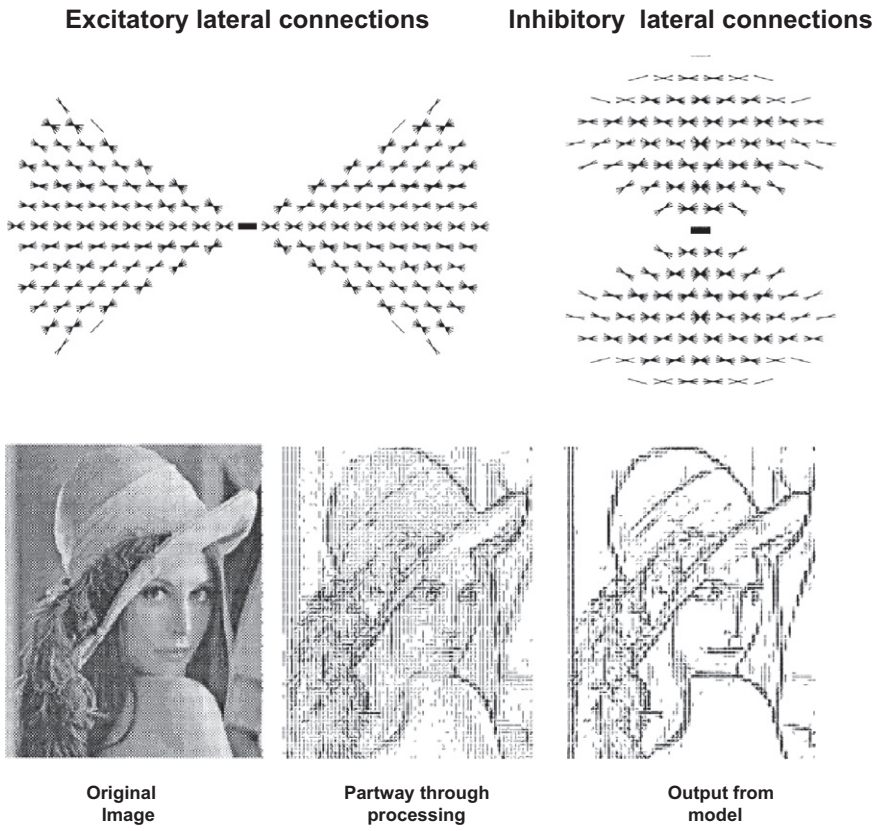

Fig. 26. *Top part:* The diagrams at the left show the excitatory (top left) and inhibitory (top right) lateral connections from one unit in the model (the unit at the center of each diagram) to all other units. The orientation and spatial location of a symbol represents the orientation selectivity and spatial location of the unit receiving the excitation or inhibition from the center unit. The boldness of the symbol represents the strength of the connection between that unit and the center unit. *Bottom part:* Some steps in the processing of an image by the model. Illustrations are adapted from Zhaoping (1998) Fig. 4 and Fig. 5.

The lateral connections are diagrammed in the top of Fig. 26 with the excitatory connections on the top left and the inhibitory connections on the top right. The orientation of each little symbol in these diagrams reflects the orientation preference of the unit at that location and the boldness of each little symbol reflects the strength of the connection. These diagrams represent the connection pattern for one unit only, but every unit has this same connection pattern suitably translated and rotated. The excitatory connections in Fig. 26 top left clearly resemble that of the association field in Fig. 25, and the inhibitory connections have the same arrangement but rotated 90°.

The predictions of this model are illustrated in the bottom of Fig. 26 by three images: The left one shows the original photograph. The middle one shows the response of the model at an intermediate stage in the processing. (White is the background showing places of zero response and gray to black indicates the magnitude of non-zero responses.) The right image shows the model's final output. In the model's final output most of the long extended contours were discovered and enhanced even when originally of very low contrast. For example, the weak contours at the chin and the top of the hat (and many noisy edges) are originally sub-threshold but many of them are well preserved at the model's output as a result of the lateral connections. (Finding such contours may be important for visual object formation even if the contours themselves are not salient in perception.) Further, the model's predictions would presumably be even better if the model were modified to contain more than a single spatial scale. For example, if the model included a range of units of different spatial-frequency preference it would correct some of the fine details like those around the eyes in the photos and perhaps help rescue the long weak contour at the brim of the hat above the eyes. (Introducing multiple spatial scales can still be expensive in terms of computing time and was more so at the time this model was published.)

The success of the model in carrying out predictions is valuable at least as a partial validation of the qualitative kind of idea presented by Field et al. (1993). More generally, it is evidence (beyond one's intuition) that lateral connections can indeed produce contour integration. This model is, however, very complicated, containing many parameters that were not even mentioned here. This complexity has drawbacks. For example, one would need to examine rather closely the effects of varying the many parameters to be sure that the work was really being done by the association-field like pattern of connections and not by fine-tuning the quantitative predictions of other features of the model.

5.1.3. Contour integration explained as dynamic cooperative activity

Many recent investigators have been attracted by the notion that the perception of contours and boundaries must require requiring dynamic developing cooperative activity among hosts of neurons. Two of the models we have already mentioned (Wielaard & Sajda, 2006; Zhaoping, 1998) incorporate some such features. There are other models or verbally expressed ideas that seem beyond the general framework of this review. They are substantially more complicated and have not been as rigorously compared to either physiological or psychophysical empirical results as the bulk of ideas mentioned here. On the other hand, to not mention such models at all seems unreasonable when talking about current ideas about contour perception, since such models may prove valuable, and so here are two examples: Thielscher and Neumann (2007) present a model inspired by the extensive work of Grossberg and Mingolla and

colleagues; Legendy (2009) attempts an abstract understanding of how brain circuits might compute what needs to be computed in order to do shape processing in visual perception.

5.2. A sample of further references

Section 4.2 (called *At the boundary between this section and the next*) gives several references that are relevant here for *Addition 5* as well as previously for *Addition 4*.

A recent review (Roelfsma, 2006) presents and discusses contour integration from the perspective of perceptual grouping, and references a wide variety of behavioral and physiological results as well as a model of them somewhat different from those presented here.

For a study of synchronous activity in visual cortex for collinear and cocircular contours see Samonds, Zhou, Bernard and Bonds (2006).

For a study on the development of contour integration (in macaques) see Kiorpes and Bassin (2003).

The perception of shape in smooth and jagged contours and its implications for understanding the filtering (including *FRF* filtering) of contours is studied by Prins et al. (2007).

An example of investigation into the spatiotemporal properties of contour integration can be found in the following articles: Dakin and Hess (1998, 1999) published results indicating the contour integration mechanism was very limited with respect to spatial frequency tuning, but Persike, Olzak, and Meinhardt (2009) recently published evidence that as long as the frequencies were within 2 or so octaves apart, there was only a very tiny impairment of performance, and no impairment at all at separations of 1.25 octaves.

Recent explorations of contour integration – and the role of occlusion in it – using illusory contours can be found in Maertens and Shapley (2008).

The problem of finding the contours in a scene even when they are partially occluded (as in Fig. 26) easily generalizes to the problem of finding the surfaces in the scene even when they are partially occluded. See Caputo (1996) and Wolfson and Landy (1999) for exploration of this problem with an psychophysical experimental paradigm that implicates interaction across quite long distances in human perception.

6. Summary

We now have more ideas than we did 25 years ago about the hidden stages of the visual system – the intermediate stages far from the image on the retina and far from the visual system's output in perception or action – and, in particular we have more ideas about the intermediate stages' processing of visual patterns.

This review briefly presented processes that have been suggested and studied in the last couple of decades, processes that are additions to the simple multiple-analyzers model based on linear filters (on the classical model of V1 simple cells). They were presented in five categories:

1. Higher-order processes (including *FRF* structures).
2. Divisive contrast nonlinearities (including contrast normalization).
3. Subtractive contrast nonlinearities (including contrast comparison).
4. Non-classical receptive fields (including surround suppression and facilitation, cross-orientation or overlay suppression).
5. Contour integration.

These hidden stages can be difficult to study. They are far from the input (which we can control) and far from the output (which is relatively easy to study through direct report or responses in psychophysical experiments, or through measurement of actions dependent on visual information). But these hidden stages form a large part of the visual system. And I think it will probably be worth the effort to learn more about them. I suspect that understanding these stages is going to be necessary for anything approaching a satisfactory understanding our own visual systems. And I suspect also that understanding them will be very important in making progress on important practical applications. It seems unlikely that we can make software or hardware see or perceive as humans do when, in fact, we are missing huge amounts of knowledge about how existing human (or similar) visual systems see and perceive. In the search for more understanding of these hidden stages, the five categories of processes listed above seem promising.

7. Glossary of terms

The glossary contains terms of several kinds. At one extreme are technical words with a well-established meaning that many readers may know. They are not defined in the text at all but are defined here for the benefit of readers who do not know them.

At the other extreme are terms that do not have well-established meanings at all. They are still vague or fuzzy because they are an attempt to label concepts that are still developing. These are defined in the text briefly and repeated in the glossary both to help a reader encountering the term in another place in the text and (sometimes) to add other material that seemed better here than in any one place in the text.

All these terms are in italics at their first presentation in the main text. (Note, however, there are many other words in the main text that are italicized because they were names of other sections in the review, or for momentary emphasis in a discussion, or for other conventional reasons, but that did not seem useful to repeat in the glossary.)

7.1. Analyzer

The word "analyzer" as used here is a general concept that includes any entity sensitive to a range of values along some dimension of interest. Dimensions along which analyzers have been postulated in pattern vision include orientation, spatial frequency (crudely size), spatial position, direction of motion and many others (reviewed in Graham (1989)).

7.2. Association field

See Section 5.1.2 and Fig. 25.

7.3. Buffy adaptation

An informal name originally used by Graham and Wolfson (2007) for a result found in psychophysical experiments, the result that is called the *straddle effect* here. The same term was also used for the process that is called the *contrast-comparison process* here. (The name came from *Buffy the Vampire-Slayer* for reasons described a bit more in Graham & Wolfson, 2007.)

7.4. Channel

This word in general has been used to mean a set of units where the set is homogeneous in some sense. For example, if you let all units with peak sensitivity to a particular orientation (e.g. oblique right, regardless of their spatial position, spatial frequency, etc.) be a channel, this set of units could be called an orientation-sensitive channel having peak sensitivity to oblique right.

In one very common use, the set of units comprising a channel would have receptive fields that were distributed densely across the retina but were otherwise identical (e.g. they had the same shape excitatory and inhibitory sections so they were sensitive to the same range of spatial frequencies and orientations). In this usage, a first-order channel is identical to a filter as defined here (a linear, translation-invariant system) and has been called by a number of other names as well, including a Fourier mechanism and a simple channel.

An *FRF structure* as defined here has identical receptive fields at all places in the visual field and thus is an example of a second-order channel.

7.5. Classical V1 simple cell (classical V1 receptive field)

This term *classical V1 simple cell* is to be distinguished from *simple cell*, which refers to a broader category that includes all cells that would be called simple cells by the common criteria of physiologists. The term *classical V1 simple cell* refers to an entity that behaves like the original Hubel & Wiesel descriptions of simple cells in the cortex. Namely, it refers to a linear system (an adding and subtracting device) followed by a half-wave *rectification*. See further discussion in *Introduction* in main text. The receptive field of a classical V1 simple cell has elongated excitatory and inhibitory regions. See more discussion at beginning of text Section 4.0 called *Addition 4: Non-classical receptive fields*.

A caution in reading past literature: The phenomena we are calling NON-classical V1 cell phenomena include only effects that can NOT be modeled by a classical V1 cell model (a linear system followed by a half-wave rectification). Occasionally in the past literature, however, any non-linear effect at all (even if it could be accounted for by the half-wave rectification) was talked about as if it required additions to the classical V1 simple-cell model. Neglecting the implications of the assumed (if sometimes implicitly) rectification seems to have happened particularly when the rectification occurred not at the exact balance point between inhibition and excitation but was biased one way or the other.

7.6. Comparison level, comparison pool, comparison process

See Figs. 14 and 17 and surrounding text.

7.7. Complex channel

The term used by Graham, Beck, Sutter and colleagues for what is here called a *second-order process.*

7.8. Contrast

The word contrast can be technically defined in a number of different ways for different applications. In this article it is used to describe a property of visual patterns. Its default definition here is the difference between the peak and trough (maximum and minimum) luminance in the pattern divided by the mean luminance in the pattern. Some studies described here have defined it in slightly different ways, but for purposes of understanding this chapter, the default definition just given should be sufficient.

7.9. Contrast-gain change

See table in Fig. 8 and accompanying text in Section 2.1.

7.10. Contrast comparison

See table in Fig. 8 and accompanying text in Section 2.1. Also called *Buffy adaptation*.

7.11. Contrast normalization

See normalization.

7.12. Contrast response function

A common name for a function giving the response of some entity (e.g. neuron, abstract component of a model) as a function of contrast of the visual pattern to which that entity is responding. See Figs. 9, 12 and 13 for examples.

7.13. Contrast subtraction

See table in Fig. 8 and surrounding text.

7.14. Decision rule

A simple rule in a model that allows the observer's response in a psychophysical experiment to be calculated from the outputs of the analyzers (units, neurons, any entities of interest).

7.15. Divisive (multiplicative) contrast nonlinearity

See table in Fig. 8 and accompanying text in Section 2.1.

7.16. Feedback and feedforward connections

See higher and lower cortical areas (stages or levels of visual processing, etc.). Also see recurrent connections.

7.17. Filter (linear filter, simple linear filter)

A filter here will be used to mean a linear system that has an input and an output that are both functions of (x, y) or sometimes (x, y, t). The same concept will sometimes be referred to as a linear filter (or, for even more emphasis, as a simple linear filter). A filter will generally be considered to be translation-invariant although that is not any real limitation for discussion at the level of this review. (See further comments in the description in the main text of Fig. 4).

A filter is an appropriate model for a large group of V1 classical simple cells centered at many spatial positions densely covering the visual field where the output of the filter at each location (x, y) is just the output before rectification of the V1 classical simple cell centered at that position (x, y). If the receptive fields of all the simple cells are identical in spatial characteristics (in the sizes and orientations of the excitatory and inhibitory regions) this is a translation-invariant linear filter.

Caution: The word *filter* is sometimes used in this literature to mean the mathematical abstraction corresponding to a single classical V1 cell, to a single unit in the terminology here, instead of to a whole collection of units.

7.18. First-order processes

First-order processes are, at their core, linear systems; but it is important to remember that, when the term is used psychophysics, a first-order process must necessarily accompanied by an assumption relating the output of the process to the response of the observer (as in the multiple-analyzers model of Fig. 1).

7.19. Fourier, Fourier analysis, Fourier transforms

A set of mathematical theorems and methods that depend on theorems by the mathematician Fourier and are particularly well adapted to analysis of linear systems or of systems made up of

linear sub-parts. Roughly, the important theorem says that there is a way (and furthermore the way is always unique) in which any function can be decomposed into (i.e. considered to be the sum of) sinusoidal (harmonic) functions. For functions of two dimensions of space (as in visual patterns), the components are sinusoidal gratings of varying orientation and spatial frequency. The Fourier transform can be plotted two-dimensionally (as in the Olzak and Thomas diagram of Fig. 7). The length of the line from the center in the diagram to a given point is proportional to the spatial frequency of the component sinusoid represented by that point. The angle that line makes with the positive horizontal axis gives the orientation of that component sinusoidal grating. Review 2 of Graham (1989) is an introductory description of the dimensions and use of Fourier analysis in spatiotemporal pattern vision. One textbook that I have found useful over the years is *Fourier transform and its applications* (Bracewell, 1965, 1978, 1999).

7.20. FRF structures (processes, channels) also called LNL or sandwich systems

The *F*'s in *FRF* stand for "filter", and the *R* for "rectification". An *FRF* process is one in which a linear filter that is characterized by relatively small receptive fields has outputs which are rectified and become the inputs to a second linear filter, where this second filter is characterized by relatively large receptive fields.

It is frequently assumed that each filter in an *FRF* process is not only linear but also translation-invariant (having the same receptive field or weight function at each spatial position).

Notice that there must be a rectification or similar nonlinearity in-between two stages of linear filtering, or else the two stages of linear filters are exactly equivalent to a single linear filtering (that has a receptive field that is a simple combination of the two filter's receptive fields).

FRF structures have also been known as *LNL* (*LNL* for linear–non-linear–linear) or *sandwich* systems in general engineering literature where their mathematical properties have been extensively studied (e.g. Korenberg & Hunter, 1986).

In the visual literature people have also used the words second-order channels, complex channels, and Non-Fourier mechanisms sometimes to be synonymous with *FRF* structures but sometimes (as in this article) to mean something more general which includes but is not limited to *FRF* structures.

7.21. Gabor patch

Gabor patches are a very commonly used visual pattern of the last 25 years. They look like small patches of fuzzy bars. If you measure the luminance at each point on a line perpendicular to the perceived bars, the function giving luminance at each position will be a Gabor function: that is, a sinusoidal function multiplied by a Gaussian function. (The bell-shaped function which characterizes the so-called normal probability distribution is a Gaussian function). See Figs. 3, 15 and 23 for several examples (although these will have been distorted by the reproduction processes).

7.22. Global control of contrast gain

See Case 3 in Appendix B.

7.23. Higher and lower cortical areas (or stages, or levels of visual processing, etc.)

It is convenient to talk in terms of higher and lower cortical areas within visual cortex, where: "lower" means roughly the areas closest to or the fewest steps away from the light stimulus (eye, LGN, V1 are the first few large steps); and "higher" means furthest from (or more steps away from, etc.) the light stimulus (e.g. areas in the parietal and temporal lobe). It is convenient but should not be taken to mean that there is no information flow from the so-called higher levels back to the lower levels. Visual information processing is definitely not a one-directional sequence of process. The information flow from the light toward the higher levels (respectively, in the other direction) is also referred to as *upstream* (respectively, *downstream*), or *feedforward* (respectively, *feedback*), or *bottom-up* (respectively, *top-down*).

7.24. Higher-order patterns

Higher-order patterns are patterns in which perceptually-salient characteristics are not easily accountable for by models like that in Fig. 1. Since "perceptually salient" and "simple decisions rules" are somewhat fuzzy concepts, this is not a totally rigorous definition.

A more rigorous definition that corresponds in many cases is as follows: A pattern is called higher-order if it is one for which knowledge of the power spectrum of the Fourier transform is NOT sufficient basis to explain the observer's performance in some perceptual task. Knowledge of the phase spectrum is necessary.

Slightly less rigorously: a pattern is higher-order if there is negligible energy in the Fourier transform at the spatial frequency and orientations corresponding to the salient perceptual characteristics.

See second-order patterns, second-order processes, higher-order processes.

7.25. Higher-order processes

In this article, higher-order processes are processes which are able to do perceptual tasks that first-order processes (linear units – the simple multiple-analyzers model of Fig. 1) cannot do. Slightly differently, they could be defined as processes which can perform pattern discriminations that depend on the phase spectrum as well as the power spectrum of the patterns. See also *second-order processes*.

7.26. Hyberpolic-ratio function (Michaelis–Menten, Naka–Rushton)

See Fig. 9 and accompanying text in Section 2.2.

7.27. Linear filter

A word often used, as here, to refer to a translation-invariant linear system where the input and output are the same number of dimensions. For example, a filter might have as an input a visual pattern of two spatial dimensions and as an output the neural response of some stage of the visual system corresponding to each spatial position in the pattern.

7.28. Linear unit (simple unit)

A linear unit is a unit that can be modeled as a linear system followed by a half-wave rectification. (For the meaning of *unit* in this article see its entry in this glossary.) The classical V1 simple-cell model is a model of a linear unit.

7.29. Linear system

A linear system is an adding and subtracting device. More formally, the most substantial part of the definition of a linear system requires such a system to show the *superposition property*. A system has the superposition property if and only if the following holds: when there is a compound input that equals the SUM of

two component inputs, then the system's response to the compound input must be the SUM of its two responses to a component alone.

What SUM means in the previous sentence can vary with the kind of system, but it has to have the same abstract properties as summing in ordinary arithmetic. The kind of summation that is relevant in this article is the point-wise summation of functions. That is, the function R is the sum of the functions R_1 and R_2 if and only if $R(x) = R_1(x) + R_2(x)$ at every point x, (where + is the ordinary arithmetic summation).

Technically, for a system to be linear, it needs a bit more than the superposition property. It needs something that can be called the *scalar multiplication property*: If (and only if) an input equals k times another input then the response to the first input must be k times the response to the second.

For practical purposes the scalar multiplication property gives little more information than the superposition property. It is a property necessary to deal with irrational numbers. But considering the superposition and scalar-multiplication properties separately often helps a human thinker notice the important similarities and differences between different non-linear systems.

Many parts of the nervous system can be approximately described overall as linear systems, at least for some range of stimuli. However, this overall linear behavior occurs in spite of the fact that the component sub-parts each act nonlinearly; there is some sort of compensation of one for another in such a way that the overall effect is linear. The fact evolution produced this linearity in the visual system in spite of the complexity of doing so suggests that the linearity itself may have desirable effects in visual perception. See the examples and discussion in Shapley (2009).

7.30. Linear-systems analysis

The term linear-systems analysis tends to refer to *Fourier analysis*. More generally it refers to all mathematical and computational ways of dealing with linear systems.

7.31. Michaelis–Menten equation (hyberbolic-ratio, Naka–Rushton)

See Fig. 9 and accompanying text in Section 2.2.

7.32. Minkowski distance (power-summation, Quick pooling)

See discussion in Appendix A. It is defined in Eq. (A.1).

7.33. Naka–Rushton equation (hyberbolic-ratio, Michaelis–Menten)

See Fig. 9 and accompanying text in Section 2.2.

7.34. Near-threshold contrast

A pattern is said to be at near-threshold contrast if its contrast is so low that it is imperfectly discriminable from a steady blank field of the same space-average luminance.

7.35. Normalization

As used here refers to a class of quite abstract models introduced by Heeger (1991) and Foley (1994) and others. No particular physiological mechanism is implied by our use of the term here. Or to say it another way, any mechanism that produces an output obeying these equations is included in our meaning here. In the general literature, however, some people use it to mean a particular physiological mechanism. For further information see Section 2.3.

7.36. Normalization pool (network)

Set of neurons or units that exert divisive suppression via a normalization network on the unit under discussion (the *signal* unit in terminology here). Whether or the normalization pool should be considered to include the signal unit itself seems to be inconsistent from one usage to another. And there is the related question of whether or not the signal unit does exert normalization on itself. I try to be explicit on these matters whenever normalization pool is used here.

7.37. Other units

Units other than the *signal unit.*

7.38. Pattern, pattern vision, spatial vision

Pattern here means a visual stimulus thought of as only depending on two spatial dimensions, or sometimes two spatial dimensions and time. The third spatial dimension – depth – and also color – are ignored, generally by being held constant throughout the comparisons made experimentally or theoretically. As used here the word *pattern* does NOT imply that the visual stimulus contains any repetition or intended design (unlike the use of the word in everyday English).

Pattern vision refers to any discussion of vision limited primarily to two dimensions and neglecting color and motion.

This same concept is sometimes called *spatial vision* in the visual literature, and I use it here sometimes. This usage may be unfortunate, however, since – at least in my experience – persons outside this technical literature almost always think the term *spatial vision* applies to the perception of three-dimensional space.

7.39. Probability summation

The term probability summation can refer to any advantage that accrues to a decision-maker when there are multiple uncorrelated (although not necessarily completely uncorrelated) sources of information. One common model of probability summation is given at the end of Appendix A.

7.40. Point-wise nonlinearities

A nonlinearity that acts on individual points (e.g. points in a pattern or points in the output from the filter) independent of what is happening at other points. The word can refer to points in space or points in time. The word *instantaneous* is sometimes used instead of point-wise, perhaps because many of the early applications were in the time domain, e.g. AM and FM radios. See also *rectification*.

7.41. Power-summation

See *Minkowski distance* and associated discussion in Appendix A.

7.42. Quick pooling

See *Minkowski distance* and associated discussion in Appendix A.

7.43. Receptive field, weighting function, impulse-response function

This word will be used both for neurons and for their more abstract counterparts *units*. The receptive field of a neuron or unit is the set of positions on the retina at which visual stimulation produces a response from the neuron (unit). In the language of linear systems, one would be likely to refer to a weighting function or an

impulse-response function. (The concepts of receptive field, weighting function, and impulse-response function are closely related but not identical. More precise definitions can be found, e.g., on pp. 68–72 of Graham, 1989.)

7.44. Rectification, half-wave rectification, full-wave rectification and their properties

The most frequent use of the word *rectification* is as a description of a piece-wise linear function (a function made up of pieces of straight line) where the value of the function is always zero or greater.

The most usual use of the term *full-wave rectification* is to name a function where y is proportional to the absolute value of *x*. See the function shown in each panel in the central column of Fig. 5 and the functions shown in Fig. 14.

The usual *half-wave rectification* is identical to a full-wave rectification for half of the *x* values (either for all positive values of *x* or for all negatives values of *x*) but equals zero on the other half.

Sometimes the word rectification generalizes to include cases where, rather than being a piecewise-linear function, the function can be any non-linear function with the following condition: the output of the function is always greater than or equal to zero. The term full-wave rectification can generalized, for example, to power functions where the power is an even number. An example is shown in Fig. 4 middle panel.

The name rectification can also generalize to cases where the center between the two halves of the function is not at zero but at some other value (as in Fig. 14). The center point of a positive half-wave rectification is sometimes referred to as a threshold (since for inputs below the center point, the output of the function is always zero).

When used in models of visual processing, rectification functions are applied at each point in space and time – that is each spatial position in a spatial image and/or at each moment in time. The output of the rectification stage at any one point is always entirely independent of what is going on at any other point in space and time, a property that is often called *point-wise* or *instantaneous* and sometimes called *static*.

(In this article we talk about rectifications functions as being functions of one real-number variable into another real-number variable. However, the word sometimes is applied to a function of many variables onto many variables. It retains the characteristic of operating point-wise in those situations.)

Why have rectification functions been so frequently used in models of vision? There are at least two important reasons. (i) One is a practical reason. The point-wise nature of rectifications has turned out to make them computationally tractable (easy to deal with) when working in the context of linear systems (as in the *FRF* process). (ii) The second reason comes from physiology. When the output of process under discussion is neural spikes, the number of spikes can never be less than zero although the entity driving the number of spikes (e.g. a slow potential) may well go far below the value at which it first produces a non-zero response. Therefore using a half-wave rectification at the terminal stage of the model of such a process is often convenient. Indeed it was so obvious to physiologists that there must be something like a half-wave rectification at the final stage of a model of V1 cells, that in many discussions of those V1 cells are just referred to as linear systems, omitting entirely even a mention of the fact the number of spikes can never be negative.

See Morgan (2011) for some further discussion of the role of rectifications, particularly half-wave rectifications, in modeling spatial vision.

7.45. Recurrent vs. non-recurrent connections

Recurrent connections are connections that run in both directions between two units. Thus, after unit A has affected unit B, then unit B can affect unit A, etc.

7.46. Response-gain change

See table in Fig. 8 and accompanying text in Section 2.1.

7.47. Response subtraction

See table in Fig. 8 and accompanying text in Section 2.1.

7.48. Second-order patterns

Second-order patterns are a subset of *higher-order patterns*. In one definition, second-order patterns are higher-order patterns in which the salient perceptual aspects (in experiments, the aspects to which an observer must respond) are computable by *FRF processes*.

A different but related definition and a useful comparison to the definition in the statistical literature is given by Johnson et al. (2005), p. 2051:

> "An abundance of psychophysical and single unit neurophysiology supports the existence of visual cortex neurons sensitive to changes in dimensions other than luminance or color—for example, variations in contrast, element size, or orientation—here collectively termed "texture variations." Such texture variations, and the visual mechanisms that are believed to detect them, have typically been designated as second-order in the biological vision literature. Note that in a context of image statistics, such stimuli might be considered fourth-order, because the detection of differences between textures requires the comparison of at least four image points. However, following common usage we will refer to stimuli that vary in luminance or chromaticity (and that minimally require the comparison of two image points) as first-order and stimuli that vary in texture (and that require the comparison of four image points), as second-order. Analyses of unfiltered single points in the image (statistically speaking, first-order) will be referred to as pixel analyses."

7.49. Second-order processes (channels)

In this article 2nd-order processes will mean a general category of processes (a subset of *higher-order processes*) in which *units* with little receptive fields (with outputs that are rectified or otherwise nonlinearly transformed) serve as inputs for units with bigger receptive fields. (If the second-order process is translation-invariant, it can be called a second-order channel). This is a more general category in this article than *FRF processes*.

The word "second-order" has been used in a number of subtly and not-so-subtly different ways in the visual psychophysical literature. Some discussion of many related terminological difficulties can be found in Landy & Graham, 2003. As discussed further in the entry for *second-order patterns*, the use of "second-order" in statistics and image processing is definitely different from that here although related.

Related terms here in the glossary are: FRF processes, higher-order patterns, higher-order processes, second-order patterns.

7.50. Signal unit or neuron

Signal is an adjective used here to indicate the neuron or unit from which a response is being discussed. For example, the unit

producing the R's on the left side of Eqs. (1)–(3). The signal unit might be literally a neuron being recorded from. Or it might be a neuron (or hypothetical entity in a model) for which the response is being inferred indirectly.

7.51. simple cell

A *simple cell* here is a term used to mean any V1 cell that would be classified as a simple cell by the criteria ordinarily used by physiologists of Hubel and Wiesel's time or today. It is to be distinguished from a *classical V1 simple cell*.

7.52. simple multiple-analyzers model

The name used here for the model of Fig. 1 top panel. It is a set of linear units – of classical V1 simple cells – followed by a simple decision rule. See further description in text near Fig. 1. In the pattern-vision literature it is referred to by many different multiple-word phrases, frequently with slightly different connotations relevant to the context.

7.53. Spatial vision

See entry 7.38.

7.54. Straddle effect

See Fig. 16 and accompanying text in Section 3.2.1. See *Buffy adaptation*.

7.55. Subtractive (additive) contrast nonlinearity

See table in Fig. 8 and accompanying text Section 2.1.

7.56. Translation-invariant linear system

A linear system is translation-invariant if and only if: when its input is shifted by a given amount, then the output is shifted by that same amount. Or, to put it more informally for the case of interest here, the filter is translation-invariant if and only the receptive fields (weighting functions) at all spatial positions are identical (in, e.g., spatial frequency and orientation preference) except that each is centered at a different spatial position. Also see *filter*.

7.57. Tuned FRF channel,

A tuned *FRF* channel for a given spatial pattern is one in which the first filter has receptive fields matched to the local spatial frequency and orientation in the pattern, and the second filter has receptive fields matched to the global characteristics in the pattern.

7.58. Unit

A unit is an abstract entity that can be used to explain behavior and that is analogous to a single neuron. It is something that produces an output that is a function of time but not of space. Even time can be (and is often) ignored by, for example, using the maximal instantaneous response, or by integrating over some time period. Words used in other contexts for what is called "unit" here include: neural unit, mechanism, and detector.

7.59. Upstream and downstream

See higher and lower cortical areas (stages or levels of visual processing, etc.).

7.60. Weber's law, Weber-law behavior, Weber behavior

The most common form of Weber's law to psychophysicists is the following: the threshold ΔI for perceiving a change on some intensity dimension I is proportional to the background level I (for a large range of background intensities I). Or, equivalently, the ratio $\Delta I/I$ at threshold is constant over a large range of non-zero background intensities.

More generally, and as used here, Weber behavior is behavior which is constant (not just at threshold but above threshold too) if and only if the ratio of the intensities involved is constant. Or, equivalently, Weber behavior is behavior where the determining factor is the ratio of the intensities, not their actual values.

For the Weber behavior in results like those shown in Fig. 16, the intensity dimension I is the absolute value of the difference between the current contrast and comparison level, or, more concretely, the absolute value of the difference between the test contrast and the adapt contrast.

7.61. Weighting function

See entry 7.43.

Acknowledgment

I thank the many people who contributed advice and knowledge and comments during the writing of this article. Particular thanks belong to Sabina Wolfson for help at all stages of the work, and to the two referees and to Luke Hallum for very helpful reviews of an earlier version of the article.

Appendix A. Non-linear pooling, decision stage, variance (noise)

A.1. A convenient family of functions for non-linear pooling: Power-summation, Minkowski distance, Quick Pooling

There is a very convenient family of functions that can approximate the action of non-linear pooling occurring at a number of places in psychophysical and physiological models. One important such place in the psychophysical models is at the decision stage (e.g. the decision rule in Fig. 1 or the decision-and-pooling stage in Fig. 5). For this stage the effect of multiple analyzers' (units', neurons') outputs on the observer's decision must be described. Another example of this convenient family's use is to describe non-linear pooling among units in a normalization network as discussed in Section 2.0 on *Addition 2*.

This family is known by various names in related literature, including *power-summation functions, Minkowski distances,* and *Quick Pooling*. The beginning of this family's use in this field occurred longer ago than 25 years. To the best of my knowledge, its usefulness was brought to the attention of investigators of pattern vision by Mostafavi and Sakrison (1976) and Quick (1974).

This family of functions is defined in the equation below where each member of the family is specified by a particular value of the parameter k. This parameter k is sometimes called the *order* of the function or the *power* or *exponent* of the function.

Let X and Y be two points in an n-dimensional space (which might for example represent the responses of n different units or channels to two different patterns), that is, let:

$$X = (x_1, x_2, \ldots, x_n) \quad \text{and} \quad Y = (y_1, y_2, \ldots, y_n)$$

Then the Minkowski distance of order k (with exponent k) between x and y is:

$$d(X,Y) = \left\{ \sum_{i=1}^{n} |x_i - y_i|^k \right\}^{1/k} \quad \text{(A.1)}$$

Members of this family provide a good approximation to many commonly used assumptions, e.g.

- the maximum-rule or winners-take-all (k = infinity),
- the city-block or Manhattan metric (k = 1),
- Euclidian distance or ordinary shortest-distance between two points (k = 2),
- probability summation among independent entities.

A.1.1. About probability summation

As Quick (1974) pointed out, the pooling rule in Eq. (A.1) can be derived from the following assumptions: (i) the multiple entities are all probabilistically independent, (ii) the overall probability that as a group they detect the stimulus is the probability that one or more of them detects that stimulus, and (iii) the probability that any individual entity detect the stimulus is well described by a convenient form known as the Weibull function. Assumptions (i) and (ii) are known as high-threshold signal-detection theory. They also are the most common model for the empirical phenomenon known as probability summation: an advantage accrued when there is more than one probabilistically-independent chance of getting an answer correct. In practice, k = 3 or 4 is a good approximation to many probability-summation situations in human psychophysics. Further it is quite a good approximation to more sophisticated detection theories. Graham (1989) Chapter 4 covers this material.

A.2. Tractability of this family of functions

Importantly, the particular construction of this family as a sum of powers (with the sum then raised to the reciprocal of that power) turns out to be tractable in many situations encountered in modeling.

One very convenient property it has is the following: When using a function in this family to nonlinearly pool over a large number of entities, you can frequently reduce the problem to pool over a small number of "composite" entities. Each of these composites is formed by pooling over a subset of the original large set, and this is usually a natural and meaningful subset. (See a use of this on pp. 415–416 in Robson and Graham (1981); another use can be found on p. 191 of Graham (1989)).

A.3. Use of this family

How do you use a family of functions like this if you do not have a pre-conceived idea about what form non-linear pooling should take? One approach is to calculate predictions using a variety of values of k (e.g. 1, 2, 3 or 4, and infinity) and to see whether the ability of the model to fit the data is robust against these changes or is responsive to them. What can happen?

(i) In many situations the predictions turn out to depend very little if at all on this exponent k (to be robust against changes in it). In other words, the same data can be fit with any exponent, and the resulting values for other parameters (e.g. spatial-frequency bandwidth) are not much affected by a choice of exponent. In these situations you have learned that the other parameters estimated from the results are stable and trustworthy. But you have also learned that the results at issue are not sufficient to decide among different exponents in the non-linear pooling rule. For much of the psychophysical work on the hidden stages of visual processing – the kinds of things discussed here – this is the result the experimenters would be happy to find.

(ii) In some situations it turns out that NO exponent assumed for the non-linear pooling will fit the results. In this case, the whole structure of the model may be wrong (although there are possible non-linear pooling rules that would not be covered by this family).

(iii) In a third class of situations – the most interesting kind if one particularly cares about the properties of the pooling under question – one can find that the results can only be fit by a limited range of exponents in the family and therefore you have learned something about the decision stage per se (although only very conditional knowledge about the rest of the model).

(iv) In a fourth class of situations – which is perhaps the most frustrating – one finds that almost any exponent will produce predictions that fit the results, but the exact values of the parameters differ wildly among the exponents in such a way one cannot even characterize the way they differ.

(Further discussion and examples of use of this family can be found in, e.g. Graham, 1989, 1991; Graham et al., 1992; Graham & Wolfson, 2004).

A.4. Variance (noise): Deterministic vs. probabilistic models

Variability in the responses of neurons and variability the responses of observers in psychophysical experiments is ignored in this review. For many purposes this is not a fatal omission. In probabilistic models, it is frequently only the ratio of the mean to the variance that is of substantial significance and that is a deterministic quantity. A related way of approximating a probabilistic model by a deterministic model was mentioned above (a probabilistic high-threshold decision model approximated by a deterministic Minkowski-distance function).

Appendix B. Three example cases using a two-unit model of contrast normalization

Eq. (3) from the main text is repeated here. See definitions of symbols in the main text. For clarity of argument here, the left hand side is augmented explicitly here to show the dependence of the response on all the various variables.

$$R_1(\sigma, s_1, w_1 \ldots w_N, c_1, \ldots c_N) = \frac{s_1 \cdot c_1}{\sigma + w_1 \cdot c_1 + \sum_{j=2}^{j=N} (w_j \cdot c_j)} \quad \textit{repeat of (3)}$$

The parameter σ characterizes the amount of normalization by setting the extent to which R_1 depends linearly or nonlinearly on the value of of c_1. More specifically, the larger the value of σ, the less the effect of normalization in the network, i.e., the more the network acts just like a linear system. To see this, consider the following:

If σ is very very large relative to the other terms in the denominator, the value of the denominator is equal to σ (approximately equal but the approximation is very very good). Then the value of R_1 is proportional to c_1 with a constant of proportionality equal to s_1/σ. Thus the system is equivalent to a linear system (approximately equivalent but the approximation is very very good).

If σ becomes smaller and smaller relative to the other terms in the denominator, then R_1 depends more and more nonlinearly on c_1, In other words, there is a greater and greater effect of the other units via the normalization pool.

For ease of algebraic manipulation and transparency in the results we are going to consider here a model of only two units. One unit is the signal unit (numbered as the 1st unit by convention here). The second unit will represent all the other units that contribute to the normalization of the signal unit. Although this may

seem like a ridiculous over-simplification, it can be done without loss of generality for a number of situations. (As mentioned in Part A of this Appendix, a single unit can replace a set of units in power-summation expressions, without loss of generality, whenever the contrasts stimulating the different units in the set are all changing proportionally to one another.)

Reducing Eq. (3) to just two units gives:

$$R_1(\sigma, s_1, w_1, w_2, c_1, c_2) = \frac{s_1 \cdot c_1}{\sigma + w_1 \cdot c_1 + w_2 \cdot c_2} \tag{B.1}$$

This two-unit model is used in several example cases below to illustrate how the normalization model can lead to contrast-gain change or to response-gain change or to both. In each of these cases the model is asked to predict a family of functions like those shown in Fig. 9, where each curve plots R_1 vs. c_1 (response of the signal unit as a function of contrast affecting the signal unit) and some aspect of the situation changes from curve to curve (while every other aspect stays the same).

B.1. Case 1. Pure contrast-gain change

In Case 1, the values of all quantities in Eq. (B.1) except contrasts stay fixed throughout (are the same for all points on a curve and for all curves in the family). The value of c_2 (the contrast affecting the other unit) stays fixed for each curve, but it changes from curve to curve. Along each curve, the value of c_1 is varied (and is what is plotted on the horizontal axis in Fig. 9).

An experiment to which Case 1 might reasonably apply would be a smallish test pattern (which affects the signal unit but not the other unit) with a largish mask or adapt pattern (that affects the other unit but not the signal unit) where the procedure is as follows: The mask-pattern contrast (which is c_2 in the equation) is held constant while the test-pattern contrast (c_1 in the equation) is varied to measure the signal unit's response for one curve. (The measurement of the signal unit's response could be either direct measurement of a neurophysiological response or inferred measurement of an internal psychophysical response.) The mask pattern's contrast c_2 is different for different curves. Such an experiment might be done to explore surround suppression, for example. Let us compute the predictions from Eq. (B1).

Remember that the value of c_2 is fixed on any one curve and the values of σ and w_2 are fixed throughout a family. Thus the full quantity $\sigma + w_2 \cdot c_2$ stays fixed on any one curve. This quantity forms a large part of the denominator in Eq. (B.1). It will be useful to substitute the name *alpha* (spelled out to keep it distinct) for that fixed quantity to make the derivation more readable. Let

$$alpha = \sigma + w_2 \cdot c_2 \tag{B.2}$$

Then Eq. (B.1) can be rewritten for each curve as follows. All variables that are constant throughout a family of curves are dropped entirely from the argument list for R_1. And the value of c_2, which is fixed for any one curve and varies from curve to curve, is put in the argument of R_1 to the right of a *vertical* bar to indicate this status. (Caution: This vertical bar is different from the diagonal bar that indicates division in many places in these equations.)

$$R_1(c_1|c_2) = \frac{s_1 \cdot c_1}{w_1 \cdot c_1 + alpha} = \frac{s_1}{w_1 + (alpha/c_1)} \tag{B.3}$$

What happens to $R_1(c_1|c_2)$ as c_1 goes to infinity? The term $(alpha/c_1)$ in the denominator of the rightmost form of the Eq. (B.3) goes to zero as c_1 goes to infinity (remember that *alpha* is fixed). Thus, the value of $R_1(c_1)$ gets bigger and bigger until it asymptotes at s_1/w_1. This asymptotic maximum value will be labeled R_1^{max}. That is,

$$R_1^{max} = s_1/w_1 \tag{B.4}$$

Note that the maximum response R_1^{max} is not affected by the values of σ, w_2, or c_2. And, in particular, changing the fixed value of c_2 from curve to curve does not change the value of R_1^{max}. In other words, we could say that there is no response-gain change in this case (Case 1).

To know whether changing the fixed value of c_2 changes what we could call the contrast gain (of the signal unit's response R_1 as a function of contrast c_1), we need to be able to characterize the function's placement on the contrast axis. This characterization can be done by calculating contrast threshold.

We will define the contrast threshold (denoted by symbol c_1^{th}) to be the value of c_1 that produces a criterion level of response where the criterion level is denoted by *crit*. It is straightforward algebraic derivation from Eq. (B.3) above to show that for any value of *crit*,

$$c_1^{th} = \frac{alpha \cdot crit}{s_1 - w_1 \cdot crit} \tag{B.5}$$

In this expression, contrast threshold depends on *alpha*, which itself depends on the fixed value of c_2, which varies from curve to curve. And thus, for this case (Case 1), the contrast threshold does change from curve to curve.

Let's consider a more specific case. For *crit* equal to half the maximum response, let's replace c_1^{th} by the symbol c_1^{50}. Then it can be shown straightforwardly that:

$$c_1^{50} = \frac{\sigma + w_2 \cdot c_2}{w_1} \tag{B.6}$$

Thus, in Case 1, as can be seen in Eqs. (B.5) and (B.6), changing the value of c_2 from curve to curve means that contrast threshold changes from curve to curve. Combining both the statements about R_1^{max} and c_1^{th} gives the following conclusion: Case 1 is a situation in which there is no response-gain change, there is only a contrast-gain change.

B.2. Case 2. Mixed response-gain change and contrast-gain change

To describe this case it will be convenient to let the symbol *cratio* mean the ratio of the contrast stimulating the other unit to the contrast stimulating the signal unit, that is:

$$cratio = c_2/c_1 \tag{B.7}$$

In Case 2, the values of all parameters except contrasts will be held fixed throughout the case. Also the value of $cratio = c_2/c_1$ stays fixed for each curve, but it changes from curve to curve.

Another way to describe Case 2 is to say that both c_2 and c_1 are changing for any individual curve, but they are changing proportionally.

An experiment to which Case 2 might reasonably apply would be a smallish test pattern with a largish mask or adapt pattern (as in Case 1) but with a different procedure. In case 2, the mask-pattern contrast is NOT held fixed while measuring the signal unit's response as a function of test-pattern contrast. Instead it varies in direct proportion to the test-pattern contrast. Such an experiment might be done to further explore surround suppression, for example.

For Case 2, Eq. (B.1) can be rewritten for each curve as follows. All variables that are constant throughout a family are dropped entirely from the argument list for R_1. The value of *cratio*, which is fixed for any one curve and varies from curve to curve, is put to the right of a vertical bar to indicate this status. The variable c_2 is not in the argument list as it is completely determined by c_1 and *cratio*, both of which are arguments already. This rewriting of Eq. (B.1) produces the following function for each curve in a family from Case 2:

$$R_1(c_1|cratio) = \frac{s_1 \cdot c_1}{\sigma + w_1 \cdot c_1 + w_2 \cdot cratio \cdot c_1} = \frac{s_1 \cdot c_1}{\sigma + [w_1 + w_2 \cdot cratio] \cdot c_1} \quad (B.8)$$

It will be convenient to introduce another temporary symbol *gamma*, where

$$gamma = (w_1 + w_2 \cdot cratio) \quad (B.9)$$

Notice that in this Case 2, *gamma* stays constant on any curve measuring R_1 as a function of c_1, since w_1, w_2, and *cratio* all stay constant. Expression (B.8) can then be rewritten as:

$$R_1(c_1|cratio) = \frac{s_1 \cdot c_1}{\sigma + gamma \cdot c_1} = \frac{s_1}{(\sigma/c_1) + gamma} \quad (B.10)$$

Letting c_1 go to infinity produces the maximum possible value on the curve:

$$R_1^{max} = \frac{s_1}{gamma} = \frac{s_1}{w_1 + w_2 \cdot cratio} \quad (B.11)$$

Thus, for Case 2, the maximum response R_1^{max} on a curve is strongly influenced by value of the ratio between the contrasts (*cratio*) which is held constant along each curve but changes from curve to curve. In Case 2, therefore, response gain is changing from curve to curve.

The contrast threshold for Case 2 also changes from curve to curve. We will skip the expression for c_1^{th} for this case, giving only the one for the criterion of half-maximum response, namely:

$$c_1^{50} = \frac{\sigma}{gamma} = \frac{\sigma}{w_1 + w_2 \cdot cratio} \quad (B.12)$$

Thus Case 2 shows both response-gain change and contrast-gain change.

B.3. Case 3. Pure response-gain change

In Case 3, the value of σ is held fixed for all points on all curves in the family. Furthermore c_2 equals c_1 for all points on all curves in the family. The excitatory sensitivities of the signal unit (s_1, s_2) as well as the weights of all the inputs into the normalization network (w_1, w_2) remain fixed along any single curve; however, these vary from curve to curve.

An experiment to which Case 3 might reasonably apply would be one in which a sinusoidal grating was used as test stimulus and both its spatial frequency and contrast were varied as follows: The grating's spatial frequency is held fixed for the measurements of the signal unit's response at different grating contrasts to produce one curve. But the spatial frequency is different for each curve. (This is the experiment done on a typical V1 neuron in Fig. 9.) Again dropping as arguments the variables that stay fixed throughout a family and segregating those that are constant on a curve but vary among curves to the right of a vertical bar gives:

$$R_1(c_1|s_1, s_2, w_1, w_2) = \frac{S_1 \cdot c_1}{\sigma + w_1 \cdot c_1 + w_2 \cdot c_2} \quad (B.13)$$

If nothing further is assumed about the excitatory sensitivities and the normalization weights, a wide variety of families of functions could be predicted. However, Case 3 makes one other and crucial assumption. It assumes that the excitatory sensitivities and the normalization weights vary in a coordinated way such that the denominator of the normalization equation remains approximately equal. That is, it assumes that the total amount of inhibition exerted by the normalization network remains approximately constant when spatial frequency (or orientation) is changed. This seems a reasonable assumption given what we have known for many years about this aspect of pattern vision. When spatial frequency (or orientation) is changed, then exactly which V1 cell responds most to it changes, and which V1 cell responds to it second-most changes, and so on. But there is always some V1 cell responding very well, and some others responding somewhat less, and some others responding still less, and so on. Thus, if the whole set of neurons sensitive to all spatial frequencies or orientations (the set with receptive fields in more or less the same position) is contributing to inhibition of any one neuron, this assumed constancy of the total amount of inhibition is precisely what would be seen. This kind of situation is often called a *global control of contrast gain*.

There would be a number of ways of instantiating this assumption of a constant total amount of inhibition in the two-unit model here. We will use a very simple way (specialized to fit with using an exponent of 1 in our equations):

$$s_1 + s_2 = 1; \quad w_1 = q \cdot s_1 \quad \text{and} \quad w_2 = q \cdot s_2 \quad (B.14)$$

where q is a parameter that is assumed to be held fixed for a family of curves. It would be very easy to generalize the constraint in Eq. (B.14) to a larger number of units.

In words this assumption says: As the spatial frequency changes, the total excitatory sensitivity of the two units stays the same. The normalization weight from a unit changes proportionally to its excitatory sensitivity. This latter statement means that when a neuron or unit fires less, it produces less inhibition.Now simplifying Eq. (B.13) using the equivalences in assumption (B.14) and then simplifying still further gives:

$$R_1(c_1|s_1) = \frac{s_1 \cdot c_1}{\sigma + q \cdot s_1 \cdot c_1 + q \cdot s_2 \cdot c_2} = \frac{s_1 \cdot c_1}{\sigma + q \cdot c_1} = \frac{s_1}{(\sigma/c_1) + q} \quad (B.15)$$

It is straightforward to show that, on each curve $R_1\,(c_1|s_1)$ the maximum response and contrast threshold will be:

$$R_1^{max} = \frac{s_1}{q} \quad (B.16)$$

$$c_1^{50} = \frac{\sigma}{q} \quad (B.17)$$

Remember that, for this Case 3, the values of q and σ are held fixed throughout a family of curves, but the value of s_1 is different for different curves. Hence, for this case, the maximum response is varying from curve to curve in a family, but the contrast threshold is not. Thus, Case 3 is a pure case of response-gain change with no contrast gain. A pure response-gain change is consistent with the empirical results in Fig. 9.

Appendix C. More about the example containing both contrast normalization and contrast comparison

C.1. Some specific points about the equations in Fig. 18

The particular equations used in Fig. 18 depended on grouping the channels into three sets and approximating each set by a single composite channel. The responses of these three sets then enter into the equations as numbers (E, E_O, E_Z) giving what the responses would be if there were *no* normalization in the system. The equation for D_{obs} computes what happens when normalization is added into the system. Although the equations and the symbol definitions have already been given in Fig. 18, we repeat some of them here in order to discuss them a bit further. Let's start with the final equation. It gives the variable that determines the observer's response D_{obs} in the form familiar from the contrast-normalization examples; the numerator is the excitatory response from the channel that responds to the pattern on which the observer's response depends, and the denominator is the action of the normalization network.

$$D_{obs} = \frac{E}{(\sigma + E_O + E_Z)} \quad (C.1)$$

The numerator E is from the *tuned channel*, the channel for which the first filter has receptive fields matched to the local spatial frequency and orientation of the visual pattern, and for which second filter is matched to the global characteristics of the pattern. This is the channel that can "do the task", or more precisely, it is the channel that provides information allowing the observer to do better than chance on the task.

σ is the usual parameter characterizing the normalization network. Both E_Z and E_O in the denominator are from sets of other channels; these other channels cannot themselves do the task, but they contribute to the normalization pool for the tuned channel. The difference between them is as follows.

The channels producing E_Z contain a contrast-comparison process, which adapts rapidly enough that the comparison level at the time of the test pattern equals the adapt contrast A. Hence the output E_Z is *not* monotonic with test contrast.

The channels producing E_O are monotonic with physical contrast. This monotonicity could be produced in at least two ways: These channels might be monotonic with physical contrast because their comparison level changes so slowly it is still set at 0% contrast at the time of the test pattern. Alternately, these channels might simply not have a contrast-comparison process at all but just be conventional *FRF* processes.

Let us look a bit further at the expressions for E, E_Z and E_O.
A is the contrast of the adapt pattern.
C_1 and C_2 (upper case letters C) are the two different contrasts in the test pattern.

[*An aside to prevent possible confusion.* Note that the meanings of C_1 and C_2 (upper case) on the equations of this figure are the contrasts of two different Gabor patches. This is NOT always identical to the meaning of c_1 and c_2 (lower case) when those terms appeared previously [main text Eqs. (2) and (3) and many equations in Appendix B]. The lower case values c_1 and c_2 referred to the contrasts affecting two different units, not (necessarily) the contrasts of two different Gabor patches.]

Not surprisingly, the expression for E from the tuned channel depends on the two test-contrasts, C_1 and C_2, and on the adapt contrast A. (The adapt contrast A does not directly appear in the equations of Fig. 18, because the term C_j with a hat over it is used instead, and that term equals the difference between C_j and A.)

The expression for E also depends on two important parameters k_m and g_{rect} that can be seen in the diagram at the bottom of Fig. 17 as well as being defined in the list in Fig. 18. The parameter k_m is the exponent in the piecewise power function that is assumed for the input–output function at the comparison stage. The parameter g_{rect} multiplies the steeper side of this input–output function to produce the shallower side. It is zero for a function of the half-wave rectification type (the shallow half is a horizontal line at zero) and 1 for a function of the full-wave type (an even-symmetric function).

What about E_Z and E_O? These are from channels that *cannot* do the task as they do not have the proper receptive fields to see both the local information and the global. And these are channels that have no sensitivity to the difference between the two test contrasts so there is nothing in the expressions for them like a difference between C_1 and C_2. These channels are sensitive instead to the total amount of contrast in both patches; the relevant contrast for E_Z is the test contrast minus the adapt contrast; and the relevant contrast for E_O is the test contrast itself. The parameter k_m appears here again for the same reason as for E. But g_{rect} is irrelevant.

C.2. Why the derivation of the equation in Fig. 18 was reasonably simple

First, the fact that the contrast normalization and contrast comparison processes are themselves expressible as simple equations helps make the resulting equations simple.

Secondly, there was a simplification that was more specific to the kind of test patterns we used. As we had shown by extensive numerical filtering previously (e.g. Graham et al., 1992), the responses of linear filters to patterns like those in Fig. 15 can be very well approximated by simple combinations of the responses to the two contrasts of Gabor patches. This approximation is made in the equations of Fig. 18 for the responses of all these channels. Thus we did not have to actually do any filtering or indeed even consider the full luminance profiles as a function of space or time. One could work simply with C_1 and C_2. [*A further comment about this approximation procedure.* Although we have only extensively documented this possibility for patterns we have used, I suspect that there are many other patterns routinely used in experiments for which such approximations could be concocted if the filtering became burdensome. And searching for such approximations seems often to produce insights into what is really driving the results in the cases at issue.]

Thirdly, we use the very convenient family of Minkowski-distance functions (see Appendix A) to approximate the action of non-linear pooling both in the normalization pool, and for the effect of multiple channels' outputs on the observer's decision.

Then the derivation of the expressions for E, E_O, E_Z and of the expression for the normalization pool and the decision variable D_{obs} is parallel to derivations done earlier for tasks not involving adaptation and therefore not showing any effects of contrast comparison (e.g. Graham & Sutter, 1998, 2000; Graham et al., 1992). The only place which is more complicated here than in those earlier studies is the derivation of the expression for the tuned channel in the numerator (since here it is necessary to keep track of what both the On and the Off members of the pair of channels are doing).

To produce the equation shown here, we set exponents equal to 1 wherever a Minkowski-distance function was used to represent non-linear pooling across multiple channels. And those exponents that equaled one were not written into these equations at all to reduce visual clutter. (This had also been done with the two-unit model in Appendix B). As it happens, the predictions are not particularly sensitive to those exponents anyway as long as the numerator and denominator are kept in balance.

Fourthly, we make no attempt to explicitly include noise. (See a related comment in Appendix A) But we need to predict a probabilistic quantity: the observer's percent correct. Hence, as part of the fitting process we use a three-parameter S-shape function to predict the observer's percent correct as a function of the calculated value of D_{obs} from the equations above.

C.3. About fitting the predictions to the results

The fits were done by eye and with trial-and-error sampling of parameters (Graham, Wolfson, Pan, Wauble, & Kwok, 2009). To do the fits in Fig. 16, nine parameters were varied for each observer: the parameter characterizing the normalization network σ; the three parameters in the s-shaped function converting D_{obs} to percent correct; the three weights w, w_o, w_z on channel sensitivities and normalization strengths; and the two parameters k_m and g_{rect} that characterize the input–output function of the contrast-comparison process.

Nine parameters may seem a large number, but these nine were used to fit a very large number of different data points. Thus the goodness of fit in Fig. 16 is very impressive.

Of course not all nine of these parameters are independent of one another. Thus for most of the parameters no conclusions can be drawn from what values did or did not produce predictions that fit the results.

There were two interesting ways, however, in which the fit to the results put serious constraints on the parameters.

(1) It was necessary that the value of k_m be somewhat greater than 1.0 and the predictions here are for a value of 2. This is necessary here in order to account for the fact that the notch of the so-called *straddle effect* contains test patterns with contrasts which do not really straddle the adapt contrast. (See more on p. 10 of Wolfson & Graham, 2009).

(2) There is also a strong constraint on the ratio of w_o to w_z, that is, on the relative strength of input to the normalization network from channels that are monotonic with contrast compared to the input from channels with a shifting contrast-comparison level. This constraint is provided by two rather subtle effects in the results: the widening of the notch seen in the results as adapt contrast gets higher and the left–right asymmetry of individual curves. If one attempts to fit these results with a denominator that contains ONLY channels which themselves show the comparison process (that is, with $w_z > 0$ and $w_o = 0$) these two effects disappear. On the other hand, if there is too much input from the channels that do not show the comparison-level adaptation (that is if w_o gets too large) the functions rapidly get wildly asymmetric and do not show a orderly shift with adapt contrast at all. It is the ratio of w_o to w_z that matters over much of the range. And it is this parameter that provides much of the leverage over individual differences in the several observers we have looked at so far (p. 15–16 in Wolfson & Graham, 2009).

References

Albrecht, D. G., & Hamilton, D. B. (1982). Striate cortex of monkey and cat: Contrast response function. *Journal of Neurophysiology, 48*, 217–237.

Allard, R., & Faubert, J. (2007). Double dissociation between first- and second-order processing. *Vision Research, 47*, 1129–1141.

Arsenault, A. S., Wilkinson, F., & Kingdom, F. A. A. (1999). Modulation frequency and orientation tuning of second-order texture mechanisms. *Journal of the Optical Society of America A, 16*(3), 427–435.

Badcock, D. R., Clifford, W. G., & Khuu, S. K. (2005). Interactions between luminance and contrast signals in global form detection. *Vision Research, 45*, 881–889.

Bair, W., Cavanaugh, J. R., & Movshon, J. A. (2003). Time course and time–distance relationships for surround suppression in Macaque V1 neurons. *Journal of Neuroscience, 23*, 7690–7701.

Baker, C. L., Jr., & Mareschal, I. (2001). Processing of second-order stimuli in the visual cortex. *Progress in Brain Research, 134*, 171–191.

Beck, J., Rosenfeld, A., & Ivry, R. (1989). Line segregation. *Spatial Vision, 4*, 75–101.

Bex, P. J., Mareschal, I., & Dakin, S. C. (2007). Contrast gain control in natural scenes. *Journal of Vision, 7*(11), 12.

Blake, R., & Wilson, H. (2011). Binocular vision. *Vision Research, 51*(7), 754–770.

Bonds, A. B. (1989). Role of inhibition in the specification of orientation selectivity of cells in the cat striate cortex. *Visual Neuroscience, 2*, 41–55.

Bonds, A. B. (1993). The encoding of cortical contrast gain control. In R. M. Shapley & D. M. Lam (Eds.), *Contrast sensitivity* (pp. 215–230). Cambridge: MIT Press.

Bracewell, R. (1965, 1978, 1999). *The Fourier transform and its applications.* New York: McGraw-Hill.

Burr, D., & Thompson, P. (2011). Motion psychophysics: 1985-2010. *Vision Research, 51*(13), 1431–1456.

Caputo, G. (1996). The role of the background: Texture segregation and figure-ground segmentation. *Vision Research, 36*, 2815–2826.

Carandini, M. (2004). Receptive fields and suppressive fields in the early visual system. In M. S. Gazzaniga (Ed.). *The cognitive neurosciences* (Vol. III, pp. 313–326). MIT Press. 3rd ed..

Carandini, M., Demb, J. B., Mante, V., Tolhurst, D. J., Dan, Y., Olshausen, B. A., et al. (2005). Mini-symposium: Do we know what the early visual system does? *Journal of Neuroscience, 25*(46), 10577–10597.

Carandini, M., Heeger, D. J., & Movshon, J. A. (1997). Linearity and normalization in simple cells of the macaque primary visual cortex. *Journal of Neuroscience, 17*, 8621–8644.

Carrasco, M. (2009). Visual attention: Neurophysiology, Pscychophysics and cognitive neuroscience. *Vision Research, 49*(10), 1033–1036.

Cavanaugh, J. R., Bair, W., & Movshon, J. A. (2002). Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. *Journal of Neurophysiology, 88*, 2530–2546.

Chen, C. C., Kasamatsu, T., Polat, U., & Norcia, A. M. (2001). Contrast response characteristics of long-range lateral interactions in cat striate cortex. *Neuroreport, 12*, 655–661.

Dakin, S. C., & Hess, R. F. (1998). Spatial-frequency tuning of visual contour integration. *Journal of the Optical Society of America A, 15*, 1486–1499.

Dakin, S. C., & Hess, R. F. (1999). Contour integration and scale combination processes in visual edge detection. *Spatial Vision, 12*, 309–327.

Dosher, B. A., & Lu, Z. L. (2005). Level and mechanisms of perceptual learning: Learning first-order luminance and second-order texture objects. *Vision Research, 46*, 1996–2007.

Ejima, Y., & Miura, K. (1984). Change in detection threshold caused by peripheral gratings: Dependence on contrast and separation. *Vision Research, 24*, 367–372.

Ellemberg, D., Allen, H. A., & Hess, R. F. (2006). Second-order spatial frequency and orientation channels in human vision. *Vision Research, 46*, 2798–2803.

El-Shamayleh, Y. (2009). *Neuronal and behavioral responses to visual form*. Ph.D. dissertation. New York University.

El-Shamayleh, Y., & Movshon, J. A. (2006). *Responses to texture-defined forms in macaque V2 neurons*. Program No. 604.9. Neuroscience meeting planner. Atlanta, GA. Society for Neuroscience.

Falkner, A. L., Krishna, B. S., & Goldberg, M. E. (2010). Surround suppression sharpens the priority map in the lateral intraparietal area. *Journal of Neuroscience, 30*, 12787–12797.

Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local 'association field'. *Vision Research, 33*, 173–193.

Foley, J. M. (1994). Human luminance pattern mechanisms: Masking experiments require a new model. *Journal of the Optical Society of America A, 11*, 1710–1719.

Foley, J.M. (2010). The Role of Temporal Transients in Forward and Backward Masking. VSS spring 2010. Journal of Vision, 10, 7, article 1390.

Foster, D. H. (2011). Color constancy. *Vision Research, 51*(7), 674–700.

Gardner, J. L., Sun, P., Waggoner, R. A., Ueno, K., Tanaka, K., & Cheng, K. (2005). Contrast adaptation and representation in human early visual cortex. *Neuron, 47*, 607–620.

Goris, R. L. T., Wichmann, F. A., & Henning, G. B. (2009). A neurophysiologically plausible population code model for human contrast discrimination. *Journal of Vision, 9*(7), 15.

Graham, N. (1989). *Visual pattern analyzers*. New York: Oxford University Press.

Graham, N. (1991). Complex channels, early local nonlinearities, and normalization in texture segregation. In M. L. Landy & J. A. Movshon (Eds.), *Computational models of visual processing*. MIT Press.

Graham, N. (1992). Breaking the visual stimulus into parts. *Current Directions in Psychological Science, 1*, 55–61.

Graham, N., & Wolfson, S. S. (in press). Two visual contrast processes: One new, One Old. In C. Chubb, B. Dosher, Z. Lu, & R. Schiffrin, (Eds.), *Vision, memory, and attention*. American Psychological Association.

Graham, N., Beck, J., & Sutter, A. (1992). Nonlinear processes in spatial-frequency channel models of perceived texture segregation: Effects of sign and amount of contrast. *Vision Research, 32*, 719–743.

Graham, N., & Sutter, A. (1998). Spatial summation in simple (Fourier) and complex (non-Fourier) channels in texture segregation. *Vision Research, 38*, 231–257.

Graham, N., & Sutter, A. (2000). Normalization: Contrast-gain control in simple (Fourier) and complex (non-Fourier) pathways of pattern vision. *Vision Research, 40*, 2737–2761.

Graham, N., Sutter, A., & Venkatesan, C. (1993). Spatial-frequency- and orientation-selectivity of simple and complex channels in region segregation. *Vision Research, 33*(14), 1893–1911.

Graham, N., & Wolfson, S. S. (2004). Is there opponent-orientation coding in the second-order channels of pattern vision? *Vision Research, 44*(27), 3145–3175.

Graham, N., & Wolfson, S. S. (2007). Exploring contrast-controlled adaptation processes in human vision (with help from Buffy the Vampire Slayer). In M. Jenkins & L. Harris (Eds.), *Computational vision in neural and machine system* (pp. 9–47). Cambridge University Press.

Graham, N., Wolfson, S. S., Pan, S., Wauble, G., & Kwok, I. (2009). Modeling the interaction of two rapid adaptation processes: Contrast comparison and contrast normalization. Program No. 651.74. 2009 Neuroscience Meeting Planner. Chicago, IL. Society for Neuroscience 2009.

Hallum, L. E., Landy, M. S., & Heeger, D. J. (in press). Human primary visual cortex (V1) is selective for second-order spatial frequency. Journal of Neurophysiology.

Heeger, D. J. (1991). Computational model of cat striate physiology. In M. S. Landy & J. A. Movshon (Eds.), *Computational models of visual processing*. Cambridge, MA: MIT Press.

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience, 9*, 181–198.

Heeger, D. J., & Ress, D. (2004). Neuronal correlates of visual attention and perception. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 339–350). MIT Press.

Heeger, D. J., Simoncelli, E. P., & Movshon, J. A. (1996). Computational models of cortical visual processing. *Proceedings of the National Academy of Sciences of the United States of America, 93*, 623–627. Copyright National Academy of Sciences U.S.A.

Henning, G. B., Hertz, B. G., & Broadbent, D. E. (1975). Some experiments bearing on the hypothesis that the visual system analyzes spatial patterns in independent bands of spatial frequency. *Vision Research, 15*, 887–899.

Hess, R. F., Baker, D. H., May, K. A., & Wang, J. (2008). On the decline of 1st and 2nd order sensitivity with eccentricity. *Journal of Vision, 8*(19), 1–12.

Higher-order processing in the visual system (1994). G. R. Bock (Organizer) & J. A. Goode (Eds.), *Ciba foundation symposium 184. Symposium held October 1993 based on a proposal by Michael J. Morgan*.

Hochberg, J. (1998). Gestalt theory and its legacy. Organization in eye and brain. In attention, mental representation. In J. Hochberg (Ed.), *Perception and cognition at century's end* (pp. 276–278). New York: Academic Press.

Itti, L., Koch, C., & Braun, J. (2000). Revisiting spatial vision: Toward a unifying model. *Journal of the Optical Society of America A, 17*, 1899–1917.

Johnson, A. P., & Baker, C. L. Jr., (2004). First- and second-order information in natural images: A filter-based approach to image statistics. *Journal of the Optical Society of America A, 21*(6), 913–925.

Johnson, A. P., Kingdom, F. A. A., & Baker, C. L. Jr., (2005). Spatiochromatic statistics of natural scenes: First-and-second-order information and their correlational structure. *Journal of the Optical Society of American A, 22*, 2050–2059.

Johnson, A. P., Prins, N., Kingdom, F. A. A., & Baker, C. L. Jr., (2007). Ecologically valid combinations of first- and second-order surface markings facilitate texture discrimination. *Vision Research, 14*, 2250–2281.

Jones, H. E., Grieve, K. L., Wang, W., & Sillito, A. M. (2001). Surround suppression in primate V1. *Journal of Neurophysiology, 86*, 2011–2028.

Kachinsky, E. S., Smith, V. C., & Pokorny, J. (2003). Discrimination and identification of luminance contrast stimuli. *Journal of Vision, 3*(10), 599–609.

Kingdom, F. A. A. (2011). Lightness, brightness, and transparency: A quarter century of new ideas, captivating demonstrations, and unrelenting controversy. *Vision Research, 51*(7), 652–673.

Kiorpes, L., & Bassin, S. A. (2003). Development of contour integration in macaque monkeys. *Visual Neuroscience, 20*, 567–575.

Korenberg, M. J., & Hunter, I. W. (1986). The identification of nonlinear biological systems: LNL cascade models. *Biological Cybernetics, 55*, 125–134.

Landy, M. S., & Graham, N. (2003). Visual perception of texture. In L. M. Chalupa & J. S. Werner (Eds.). *The visual neurosciences* (Vol. 2, pp. 1106–1118). MIT Press.

Landy, M. S., & Oruc, I. (2002). Properties of second-order spatial frequency channels. *Vision Research, 42*, 2311–2329.

Larsson, J., Landy, M. S., & Heeger, D. J. (2006). Orientation-selective adaptation to first- and second-order patterns in human visual cortex. *Journal of Neurophysiology, 95*, 862–881.

Legendy, C. (2009). *Circuits in the brain. A model of shape processing in the primary visual cortex*. NYC: Springer-Verlag.

Legge, G. E., & Foley, J. M. (1980). Contrast masking in human vision. *Journal of the Optical Society of America, 70*, 1458–1471.

Lennie, P. (1998). Single units and visual cortical organization. *Perception, 27*, 889–935.

Levi, D. M., & Carney, T. (2011). The effect of three flankers on three tasks in central, peripheral, and amblyopic vision. *Journal of Vision, 11*(1), 1–23.

Levi, D. M., & Waugh, S. J. (1996). Position acuity with opposite-contrast polarity features: Evidence for a nonlinear collector mechanism for position acuity? *Vision Research, 36*, 1907–1918.

Levick, W. R., Cleland, B. G., & Dubin, M. W. (1972). Lateral geniculate neurons of cat: Retinal inputs and physiology. *Investigative Ophthalmology, 11*, 302–311.

Levitt, J. B., & Lund, J. B. (1997). Contrast dependence of contextual effects in primate visual cortex. *Nature, 387*, 73–76.

Lu, Z. L., & Sperling, G. (1996). Second-order illusions: Mach Bands, Chevreul, and Craik-O'Brien-Cornsweet. *Vision Research, 36*(4), 559–572.

Maertens, M., & Shapley, R. (2008). Local determinants of contour interpolation. *Journal of Vision, 8*(7), 3.

Manahilov, V., Simpson, W. A., & Calvert, J. (2005). Why is second-order vision less efficient than first-order vision? *Vision Research, 45*, 2759–2772.

McLaughlin, D., Shapley, R., Shelley, M., & Wielaard, J. (2000). A neuronal network model of macaque primary visual cortex (V1): Orientation selectivity and dynamics in the Input layer 4Ca. *Proceedings of the National Academy of Sciences of the United States of America, 97*, 8087–8092.

Meese, T. S. (2004). Area summation and masking. *Journal of Vision, 4*, 930–943.

Meese, T. S., & Holmes, D. J. (2002). Adaptation and gain pool summation: Alternative models and masking data. *Vision Research, 42*, 1113–1125.

Morgan, M. (2011). Features and the 'primal sketch'. *Vision Research, 51*(7), 738–753.

Mostafavi, H., & Sakrison, D. J. (1976). Structure and properties of a single channel in the human visual system. *Vision Research, 23*, 997–1004.

Motoyoshi, I., & Kingdom, F. A. A. (2007). Differential roles of contrast polarity reveal two streams of second-order processing. *Vision Research, 47*, 2047–2054.

Motoyoshi, I., & Nishida, S. (2004). Cross-orientation summation in texture segregation. *Vision Research, 44*, 2567–2576.

Movshon, J. A. (1990). Visual processing of moving images. In H. Barlow, C. Blakemore, & M. Weston-Smith (Eds.), *Images and understanding*. Cambridge, England: Cambridge University Press.

Nachmias, J. (1999). How is a grating detected on a narrowband noise masker? *Vision Research, 39*, 1133–1142.

Olzak, L. A., & Thomas, J. P. (1999). Neural recoding in human pattern vision: Model and mechanisms. *Vision Research, 39*, 231–256.

Olzak, L. A., & Thomas, J. P. (2003). Dual nonlinearities regulate contrast sensitivity in pattern discrimination tasks. *Vision Research, 13*, 1433–1442.

Oruc, I., Landy, M. S., & Pelli, D. G. (2006). Noise masking reveals channels for second-order letters. *Vision Research, 46*, 1493–1506.

Persike, M., Olzak, L. A., & Meinhardt, G. (2009). Contour integration across spatial frequency. *Journal of Experimental Psychology: Human Perception and Performance, 35*, 1629–1648.

Petrov, Y., Carandini, M., & McKee, S. P. (2005). Two distinct mechanisms of suppression in human vision. *Journal of Neuroscience, 25*, 8704–8707.

Polat, U., & Sagi, D. (1993). Lateral interactions between spatial channels. Suppression and facilitation revealed by lateral masking experiments. *Vision Research, 7*, 993–999.

Polat, U., & Sagi, D. (1994). The architecture of perceptual spatial interactions. *Vision Research, 34*, 73–78.

Prins, N., Kingdom, F. A. A., & Hayes, A. (2007). Detection low shape-frequencies in smooth and jagged contours. *Vision Research, 47*, 2390–2402.

Quick, R. F. (1974). A vector-magnitude model of contrast detection. *Kybernetik, 16*, 65–67.

Reynolds, J. H., & Heeger, D. J. (2009). The normalization model of attention. *Neuron, 61*, 168–185.

Ringach, D. (2010). Population coding under normalization. *Vision Research, 50*, 2223–2232.

Robson, J. G., & Graham, N. (1981). Probability summation and regional variation in contrast sensitivity across the visual field. *Vision Research, 21*, 409–418.

Roelfsma, P. R. (2006). Cortical algorithms for perceptual grouping. *Annual Review of Neuroscience, 29*, 203–227.

Rogowitz, B. (1983). Spatial/temporal interactions: Backward and forward metacontrast masking with sine-wave gratings. *Vision Research, 23*, 1057–1073.

Samonds, J. M., Zhou, Z., Bernard, M. R., & Bonds, A. B. (2006). Synchronous activity in cat visual cortex encodes collinear and cocircular contours. *Journal of Neurophysiology, 95*, 2016–2602.

Schofield, A. J. (2000). What does second-order vision see in an image? *Perception, 29*, 1071–1086.

Schofield, A. J., & Georgeson, M. A. (1999). Sensitivity to modulations of luminance and contrast in visual white noise: Separate mechanisms with similar behaviour. *Vision Research, 39*, 2697–2716.

Schwartz, O., & Simoncelli, E. P. (2001). Natural signal statistics and sensory gain control. *Nature Neuroscience, 4*, 819–825.

Shapley, R. (2009). Linear and nonlinear systems analysis of the visual system: Why does it seem so linear? A review dedicated to the memory of Henk Spekreijse. *Vision Research, 49*, 907–921.

Shapley, R., & Hawken, M. (2011). Color in the cortex. *Vision Research, 51*(7), 701–717.

Shapley, R., & Gordon, J. (1985). Nonlinearity in the perception of form. *Perception & Psychophysics, 37*, 84–88. Fig. 4 from this publication used in Fig. 24 here with kind permission from Springer Science+Business Media B.V.

Snowden, R. J., & Hammett, S. T. (1998). The effects of surround contrast on contrast thresholds, perceived contrast and contrast discrimination. *Vision Research, 38*, 1935–1945.

Sterkin, A., Yehezkel, O., Bonneh, Y. S., Norcia, A., & Polat, U. (2009). Backward masking suppresses collinear facilitation in the visual cortex. *Vision Research, 49*, 1784–1794.

Thielscher, A., & Neumann, H. (2007). A computational model to link psychophysics and cortical cell activation in human texture processing. *Journal of Computational Neuroscience, 22*, 255–282.

Vakrou, C., Whitaker, D., & McGraw, P. V. (2007). Extrafoveal viewing reveals the nature of second-order human vision. *Journal of Vision, 7*(14), 1–15.

Victor, J. D., & Conte, M. M. (1996). The role of high-order phase correlations in texture processing. *Vision Research, 36*(11), 1615–1631.

Victor, J. D., Conte, M. M., & Purpura, K. P. (1997). Dynamic shifts of the contrast-response function. *Visual Neuroscience, 14*, 577–587.

Watson, A. B., & Solomon, J. A. (1997). Model of visual contrast gain control and pattern masking. *Journal of the Optical Society of America A, 14*, 2379–2391.

Webb, B. S., Dhruv, N. T., Solomon, S. G., Tailby, C., & Lennie, P. (2005). Early and late mechanisms of surround suppression in striate cortex of macaque. *The Journal of Neuroscience, 25*(50), 11666–11675.

Wielaard, J., & Sajda, P. (2006). Extraclassical phenomena and short-range connectivity in V1. *Cerebral Cortex, 16*, 1531–1545.

Wilson, H. R. (1985). Discrimination of contour curvature: Data and theory. *Journal of the Optical Society of America A, 2*, 1191–1199.

Wilson, H. R. (1999). Non-fourier cortical processes in texture, form, and motion perception. In P. S. Ulinski & E. G. Jones (Eds.). *Cerebral cortex* (Vol. 13, pp. 999). New York: Academic/Plenum Publishers. Fig. 5 from this publication used in Fig. 7 here with kind permission from Springer Science+Business Media B.V.

Wilson, H. R., & Richards, W. A. (1989). Mechanisms of curvature discrimination. *Journal of the Optical Society of America A, 6*, 106–115.

Wolfson, S. S., & Graham, N. (2007). An unusual kind of contrast adaptation: Shifting a contrast-comparison level. *Journal of Vision, 7*(8), 12.

Wolfson, S. S., & Graham, N. (2009). Two contrast adaptation processes: Contrast normalization and shifting, rectifying, contrast comparison. *Journal of Vision, 9*(4), 30.

Wolfson, S. S., & Landy, M. S. (1999). Long range interactions between oriented texture elements. *Vision Research, 39*, 933–945.

Yu, C., Klein, S. A., & Levi, D. M. (2003). Cross- and Iso-oriented surrounds modulate the contrast response function: The effect of surround contrast. *Journal of Vision, 3*, 527–540.

Zenger-Landolt, B., & Koch, C. (2001). Flanker effects in peripheral contrast discrimination. Psychophysics and modeling. *Vision Research, 41*, 3663–3675.

Zhaoping, L. (1998). A neural model of contour integration in the primary visual cortex. *Neural Computation, 10*, 903–940.