A Contrast- and Luminance-driven Multiscale Network Model of Brightness Perception

LUIZ PESSOA,* ENNIO MINGOLLA,*† HEIKO NEUMANN‡

Received 20 May 1994; in revised form 25 October 1994

A neural network model of brightness perception is developed to account for a wide variety of data, including the classical phenomenon of Mach bands, low- and high-contrast missing fundamental, luminance staircases, and non-linear contrast effects associated with sinusoidal waveforms. The model builds upon previous work on filling-in models that produce brightness profiles through the interaction of boundary and feature signals. Boundary computations that are sensitive to luminance steps and to continuous luminance gradients are presented. A new interpretation of feature signals through the explicit representation of contrast-driven and luminance-driven information is provided and directly addresses the issue of brightness “anchoring”. Computer simulations illustrate the model’s competencies.

Brightness  Filling-in  Neural networks  Mach bands  Brightness anchoring  Multiple scales

“The sensation corresponding to a given luminance depends so much upon the other luminances in the array, and upon other conditions in the world, that it can never provide an explanation of the perception of the world. It is a laboratory curiosity of great interest. If we did not know so much about it, the present theory could never have been formulated. But nobody ever really sees a color sensation of this sort in the ordinary environment. Instead, one detects the pigmentation, facing, and illuminatedness of surfaces, all three together; and these invariants of the world can only be specified by invariants among the variable luminances of the optic array. Space perception is not based on sensations of brightness and color. For this reason, we do not need to review all the efforts of psychologists to measure and scale these sensations, and all the controversies to which they have led. Interesting as they may be, they are not in the line of inquiry that ends with perception.”

Gibson (1966, p. 222)

INTRODUCTION

While Gibson’s (1966) warning about the complexities of a full account of our experience of brightness is well-taken, brightness phenomena provide important constraints for understanding the visual system’s mechanisms for encoding spatial luminance variations in a scene. Theories of brightness perception fall into four main classes: (a) contrast-sensitivity-function (CSF) models assert that appearance is determined largely by the CSF of the visual system (e.g. Campbell & Robson, 1968; Cornsweet, 1970); (b) integration models attempt to recover object lightness§ (apparent reflectance) by using operations of differentiation, thresholding and integration (e.g. Land & McCann, 1971; Arend & Goldstein, 1987); (c) “filling-in” models propose that spreading of neural activity within filling-in compartments produces a response profile isomorphic with the percept (Fry, 1948; Walls, 1954; Gerrits, de Hann & Vendrik, 1966; Gerrits & Vendrik, 1970; Davidson & Whiteside, 1971; Hamada, 1984; Cohen & Grossberg, 1984; Grossberg & Todorović, 1988); and (d) symbolic models provide symbolic descriptions (i.e. tokens) of salient local luminance changes in the visual stimulus (Marr, 1982; Watt & Morgan, 1985; Kingdom & Moulden, 1992).

In this paper we develop a neural network model of brightness perception in the tradition of filling-in theories (Cohen & Grossberg, 1984; Gerrits & Vendrik, 1970; Grossberg & Todorović, 1988). The output of the model is an activity profile that is, ideally, isomorphic with a human's brightness distribution in response to the corresponding stimulus, and therefore the model can be said to be isomorphicastic, as discussed by Todorović (1987); see also Ratliff and Sirovich (1978). The neural network developed here uses the boundary contour system/feature contour system (BCS/FCS) proposed by Cohen and Grossberg (1984) and Grossberg and Mingolla (1985a,b) to explain brightness data.
Relative to the implementation used for the simulations of Grossberg and Todorovíc (1988), the present model includes several modifications: (a) on and off channels with separate filling-in domains; (b) multiple spatial scales; (c) a new non-linear circuit for simple cells; (d) boundary computations that engage a recurrent (feedback) competitive circuit; (e) a new treatment for the FCS, including contrast-driven and luminance-driven channels. The combination of the non-linear simple cell with the feedback circuit in the BCS produces qualitatively different boundary signals for shallow (e.g. ramp-like) and sharp (e.g. step-like) luminance transitions. Depending on the input luminance distribution, very localized or spatially extended boundaries may result. The new treatment for the FCS, which now includes a luminance-driven channel (Neumann, 1993a,b, 1994), modifies the role of diffusion mechanisms have on predicting brightness appearance.

The power of the present approach lies in the proper interaction of the above model components in a way that allows the model to handle several brightness phenomena previously considered challenging to, or even inconsistent with, filling-in or diffusion theories of brightness. One example concerns shallow brightness gradients that originate from luminance ramps (which, incidentally, may produce Mach bands) or sinusoidal waves. Perhaps an even greater challenge to filling-in models is a luminance staircase distribution. The “steps” of the staircase presumably block diffusion, and it is not evident how a filling-in model can predict that different steps appear with different brightnesses (since “border contrast” is the same everywhere).

Recently, models of brightness perception have explicitly incorporated multiple scales in their mechanisms. The MIRAGE model of Watt and Morgan (1985) was the first to incorporate such mechanism in computer simulations of brightness perception. The work of Morrone, Burr and colleagues (Morrone & Burr, 1988; Burr & Morrone, 1992) on the local energy model employs multiple spatial scales and has been used to account for brightness stimuli (e.g. Ross, Morrone & Burr, 1989). Kingdom and Moulden (1992) have proposed a new model of brightness perception, called MIDAAS, which draws upon the MIRAGE model. We began the present research intrigued by Kingdom and Moulden’s (1992) approach, seeking to demonstrate that results comparable or superior to those of MIDAAS could be obtained with an (isomorphic) neural network model of brightness perception. In this article we show that functions analogous to those performed by several of MIDAAS’s interpretation rules are obtained as emergent properties of our system of equations.

Kingdom and Moulden (1992, p. 1579) have claimed that “it is not clear how Grossberg’s model, at least in its single filter version, could predict the presence of Mach bands in for example, the trapezoid wave”. Simulations we present in a subsequent section account for human’s perception of a wide variety of stimuli, including Mach bands in trapezoidal and triangular waveforms, as well as stimuli whose brightness contains shallow spatial gradients—e.g. a sinusoidal brightness percept. The set of stimuli used in order to demonstrate the model’s capabilities was chosen so as to include several of the stimuli employed by Kingdom and Moulden (1992) as it constitutes a rather challenging data set and can be used as a “benchmark” in order to evaluate competing brightness models.

**BRIGHTNESS GRADIENTS WITHIN THE BCS/FCS THEORY**

A fundamental idea of the BCS/FCS theory is that *boundaries* are used to generate filling-in compartments where *featural* quality (“brightness” in our case) is diffused, or spread. The final diffused activities correspond to the predicted brightness, whereby boundaries control the process of filling-in by forming gates of variable resistance to diffusion. Note that while it may seem natural to assume that boundary signals only exist in locations corresponding to discontinuities of luminance (i.e. “edges”), Grossberg and Mingolla (1987) showed that spatially dense “boundary webs” can form in regions of continuous luminance gradients. Therefore, the process of diffusion may be totally or partially blocked within extended regions, yielding a percept of spatially gradual changes in brightness. Boundary signals work to contain diffusion; large boundary values do not allow a featural value at a given spatial position to affect a neighboring one. In regions with zero boundary activity, featural quality is free to diffuse, while in regions containing spatially dense boundary signals of sufficient amplitude, little diffusion of featural quality throughout a large area may occur. In other words, in such regions featural quality cannot be spread, and the corresponding predicted brightness will be similar to the profile of featural quality derived by the initial filtering of the scenic input (image) at those spatial positions. In this sense, spatially extended boundaries of sufficient strength can be thought of as *print signals*, i.e. signals that replicate the filtered input to the filling-in stage.

Figure 1 illustrates how dense boundaries can potentially be employed to account for Mach bands in a luminance ramp, while sharp boundary signals are required for a luminance step. The luminance ramp profile (input) is initially filtered (feature). If spatially extended boundaries of sufficient amplitude can be generated (boundary and boundary feedback), the overshoots and undershoots present in the featural signal are to some extent preserved in the final brightness stage, therefore predicting the appearance of Mach bands on the ramp stimulus (brightness). For a luminance step no Mach bands are generated since spatially localized boundary signals allow diffusion to proceed unimpeded—“spreading” the overshoot and undershoot present in the featural signal. Note that for this to occur boundary signals are sharpened (boundary feedback) so as not to incorrectly block diffusion. For the luminance ramp, no such sharpening is triggered. As will be shown, the power of the present approach is such that the same system of equations that “isomorphistically” generates Mach
bands for a luminance ramp does not generate them for a luminance step, for identical system parameters. At this point it may be useful to briefly summarize the three ways in which a diffusion process responsible for filling-in could result in brightness gradients. First, the mechanisms specifying diffusion in the FCS are nonconservative, i.e. they include an activity source (FCS input) as well as a sink (decay of activation). Therefore, the diffusion process is not equivalent to the classical heat equation, in which the overall energy is preserved (i.e. it is conservative), and at equilibrium "activity" is uniform. FCS diffusion is such that at equilibrium values within compartments are not necessarily uniform (see Arrington (1994) for some examples). Second, as pointed out above, boundaries can provide partial blockage of diffusion. In fact, this was the mechanism able to explain the assimilation data of Shapley and Reid (1986); see Grossberg and Todorović's (1988) simulations (their Fig. 13). Third, boundaries can be broad (print signals) and trap modulation present in the FCS [Fig. 1(A)]. Featural signals that are spatially within strong boundary signals cannot spread, while those that are not can. At equilibrium, the feature signals trapped by the boundaries preserve their initial spatial profile. In summary, in the BCS/FCS, filling-in is not just an averaging of activity within "abrupt" compartments (as claimed, e.g. by Paradiso & Nakayama, 1991). Boundaries are not always "all-or-none" but can be graded, as can the brightness variations computed by the filling-in process.

**Boundary computations**

Analysis of several brightness stimuli indicates that stimuli with abrupt luminance transitions (e.g. luminance steps) generally require sharp boundary signals to create spatially abrupt barriers between regions of discretely differing brightness levels, while stimuli containing smooth luminance modulations will require broader boundaries to be able to trap (at least some of) the modulation that is present in the featural quality to create smoothly varying brightness distributions. How then can the portions of the visual system responsible for brightness perception, or a model of it, "decide" (without a homunculus) whether or not to sharpen boundary signals?

What differentiates the situations requiring sharp and extended boundary signals? Consider a system where the input waveform is filtered by both on and off center-surround (circularly-symmetric) operators—allogous to, e.g. retinal ganglion cell receptive fields. The solution originates from the observation that abrupt transitions of luminance produce strong responses in both the on and off channels. In other words, in the region surrounding an "edge", there will be strong on-activity (at the "light" region) and strong off-activity (at the "dark"

---

*Such operators are similar to lateral inhibition models such as the ones employing "mexican-hat" functions (see, e.g. Marr, 1982).
region). This is illustrated in Fig. 2(A), where on and off channel responses to a step are sketched. In sketching these responses it was assumed that the input luminance is filtered by center–surround operators with zero d.c.-level (i.e. the "excitatory" and "inhibitory" areas of the kernels are the same). Figure 2(A) shows that on and off channel activities are spatially adjacent for a step in luminance. On the other hand, waveforms with more gradual variations of luminance lead to a different distribution of on and off responses. Figure 2(B) shows on/off filtering responses for a luminance ramp, illustrating the fact that the latter are spatially separated. In both cases above, strong responses occur only at positions of abrupt change in slope since a center–surround operator approximates a second-order derivative in one dimension.

The above analysis indicates that on and off responses can be used to guide the computation of boundaries: on/off spatial juxtaposition (Gove, 1994; Iverson & Zucker, 1990; Marr & Hildreth, 1980; Gove, Grossberg & Mingolla, 1993; Grossberg, Mingolla & Williamson, 1993; Cruthirds, Gove, Grossberg, Mingolla, Nowak & Williamson, 1992) should eventually lead to sharp boundaries while on/off separation should produce spatially broad and shallow boundary signals. We propose a two-stage process for achieving the computational competencies to generate the proper activity distribution within the BCS: (1) initial boundary responses to stimuli with on/off juxtaposition should be large; and (2) responses that are "large" (relative to neighboring ones) are sharpened while those that are small are not. Moreover, we will demonstrate that a recurrent network employing a single set of parameters

can automatically "switch" its behavior to provide appropriate output in each case (Grossberg, 1973; Grossberg & Marshall, 1989). The next two subsections expand on these points.

Simple and complex cells

In the BCS/FCS, boundaries originate from computations starting with simple and complex cells.* One of the contributions of the present work is to propose a new circuit for simple cells. Figure 3 sketches the circuit of a simple cell with light–dark polarity. The circuit itself has four stages and contains two streams, or channels,

*In many of the early reports on the BCS/FCS, a simple-cell computational stage involving odd-symmetric, polarity specific operators, has been employed. A complex-cell computational stage pools responses from opposite polarity simple-cells and thus is polarity insensitive. These basic properties, which resemble in vitro cell responses, have motivated the "simple" and "complex" names for these stages (see, e.g. Jacobson, Gaska, Chen & Pollen, 1993) for a more detailed physiological model; also Grossberg, 1994.)
stage four. The opponent inhibition associated with the within-channel inhibition provides a mechanism for disinhibition. As a consequence, the output of the circuit will be large only when the inputs from both channels are strong—since disinhibition of activation in each subfield occurs. The self-inhibition/disinhibition mechanisms therefore realize the functionality of a "smooth" AND-gate (Iverson & Zucker, 1990; Neumann & Pessoa, 1994).

Therefore, one of the main properties of the operation of simple cells in the current model is that they behave (partially) as "coincidence" (juxtaposition) detectors. Note that model simple cells do respond when stimulated by either the on or the off channel alone. They behave as "coincidence" detectors only inasmuch as they favor—i.e., non-linearly amplify—co-activations of the two channels. In summary, the simple cell design is such that it meets the first computational requirement outlined in the end of last section, namely that initial boundary-related responses to stimuli with on/off juxtaposition should be large in comparison to the case where a single channel is activating the circuit.

**Non-linear recurrent competitive circuit**

Complex cells responses activate a feedback network that is capable of spatially sharpening strong signals. However, the network should not sharpen weak complex cell responses. The recurrent competitive circuit illustrated in Fig. 4 is able to generate such a behavior. The circuit has two layers, \( F_1 \) and \( F_2 \), with the inputs exciting the cells in \( F_1 \). Within layer \( F_2 \), cells interact via a feedback network.

This type of network has been carefully investigated by Grossberg and Marshall (1989) and been shown to contrast enhance certain types of patterns (see also Grossberg, 1973). A key property of the circuit is that it can sharpen the spatial distribution of signals that are "large" relative to their neighbors. The behavior

![Figure 5](image-url)
representation of contrast-driven and luminance-driven information in the visual system. In Neumann’s proposal (Fig. 6), the visual scene is initially filtered with center-surround on and off operators and then combined in two separate ways: (1) in an opponent (inhibitory) fashion that generates zero d.c. level signals (on/off contrasts); and (2) in a pooled (excitatory) form that produces a “low-pass” version of the input distribution (low-pass luminance). More precisely, the “low-pass luminance” corresponds to a low-pass filtered and non-linearly compressed version of the input luminance distribution, which is formally defined in Appendix A. Luminance-driven cells have not been reported in primates, although they have been found in cats (Barlow & Levick, 1969).

In the present model, a related scheme is used for the initial FCS computational stages. We embed on and off domains of filling-in in the circuit of Fig. 6. As sketched in Fig. 7, the inputs to the on/off diffusion domains are the respective contrast signals. The equilibrated on/off filling-in signals are then used to modify the low-pass “luminance”. In other words, on/off domains are combined with luminance-driven signals in order to produce (single-scale) brightness predictions. Figure 7 also shows that boundary signals generated by the BCS are used to regulate the diffusion process. Such signals are generated by the combination of simple and complex cell circuits and the feedback competitive circuit, as discussed above. In the current implementation the luminance-driven pathway does not undergo filling-in—this was adopted for simplicity. Instead this signal is used as a “baseline”

**FCS computations and the “luminance-driven” pathway**

Recently, Neumann (1993a, 1994) presented an analysis that clarified how the interaction of on/off signals for brightness computation (Grossberg, 1987, 1994; Grossberg & Wyse, 1991) can be thought of as part of a parallel scheme for “contrast-driven” and “luminance-driven” processing. Indeed, it was proposed that the existence of initial retinal on and off channels provides a highly efficient coding and transmission scheme for the
of activity which is then modified by the outputs of the on and off filling-in domains.

The segregation of contrast-driven and luminance-driven signals by the early visual system also addresses the anchoring problem (Gilchrist, Cattaliotti, Bonato & Li, 1992; Gilchrist, 1993): how are relative luminances in the visual field transformed into absolute brightness values? This important question goes back to Hering's (1874/1964) assignment of "middle gray" to the luminance that corresponds to the midpoint of the range of perceptual responses. It is interesting to note that while confronting this problem, Arend (1973, p. 391) proposed that "an additional channel of information from the retina" that is not differential in response, may be responsible for attributing absolute brightness levels.

**MULTIPLE SCALE BRIGHTNESS SYSTEM**

Figure 8 illustrates the four key computational stages of the model: (1) the input stimulus is split into separate contrast-driven and luminance-driven representations; (2) contrast-driven signals are employed to produce

![Diagram of model components. Stages that perform multi-scale computations are specified by several "shifted boxes". Simple and complex cells are indicated by polarity specific and direction of contrast insensitive "cells", respectively. The final output is given by the stage "Multiple Scale Brightness". The names employed are indicated in parentheses in the formal description of the model stages. The numbers in each box are keyed to the equations that specify the model in Appendix A.](image)
set of graphs the activity levels of certain model stages at equilibrium as a function of spatial position. Plots displaying on and off contrast-driven activities use, respectively, solid and dotted lines. Results for on and off filling-in domains are displayed with on activities being positive and off activities negative; the latter are negative for illustration purposes only (since they contribute to “darkness”) and in the model they are positive signals. Multiple scale computations were performed in three scales (“small”, “medium” and “large”) which were sufficient to illustrate the behavior of the model; these were indicated by the numbers 1, 2, and 3 in the graph labels. The same plotting range for the y-axis is used throughout different simulations for the final multi-scale brightness stage.

Trapezoidal wave. The model correctly predicts the appearance of Mach bands on a trapezoidal wave, as shown in Fig. 10. Figure 11 summarizes the key computational stages of the simulation. The left plots show how the final boundary (dotted lines) spatially overlap the feature signals (solid lines), which will, then, not be able to diffuse and become uniform. The equilibrated on (dense line) and off (dotted line) filling-in activities are shown in the middle plot and contribute only to the production of the “light” and “dark” bands—i.e., the trapezoidal modulation is not present in the equilibrated filling-in signal. The right plot shows the activity in the luminance-driven channel. It is this signal that registers the input stimulus modulation. The recombination of the on and off filling-in information with the luminance-driven channel produces the final brightness predictions. Such recombination can be interpreted in terms of the luminance-driven channel providing a “baseline” of activity that is modified by on (excitatory) and off (inhibitory) filling-in activities [see equation (A26)], which produces Mach bands.

The above discussion highlights the several model interactions responsible for its behavior, namely interactions between boundary, feature, and luminance-driven signals. The trapezoidal wave simulation also
demonstrates that filling-in can have a reduced role in the present model when compared to the one proposed by Grossberg and Todorović (1988). In other words, while filling-in is responsible for registering the light and dark bands, it is the luminance-driven channel that provides the information concerning the overall brightness distribution. For other stimuli, filling-in plays a role more similar to the one in the Grossberg and Todorović (1988) proposal, such as in a square wave or in simultaneous brightness contrast (see below). In such cases the contribution of the luminance-driven channel to the final brightness prediction is similar to the filling-in contribution.

The “soft-threshold” behavior of the boundary system also is evident in the trapezoidal wave simulation. Complex cells non-linearly amplify their responses when ON/OFF juxtapositions of initial filtering activities are detected. For a trapezoidal waveform, ON and OFF contrast-driven signals arise only at the inflection points and are associated with OFF responses at the “low knee-point” and with ON responses at the “high knee-point”. Such spatial distribution does not favor strong complex cell responses and does not trigger sharpening of boundary signals. Dense boundary signals result.

For more than a century Mach bands have fascinated vision scientists and an enormous amount of data exists on them (see e.g. Ratliff, 1965). Recently, Morrone, Burr, and colleagues have provided several important measurements related to Mach bands. In one study (Ross et al., 1989) they measured the contrasts needed to see Mach bands on trapezoidal waveforms. Their main finding was that sensitivity rises gently to a peak, then drops sharply as spatial frequency increases. We studied the behavior of the present model when stimulated with trapezoidal waveforms of varying spatial frequencies. Figure 12(A) plots light Mach band strength as a function of spatial frequency (see Appendix C for details). The results show a gentle rise to a peak followed by an abrupt fall. Since the aim of the current investigation is to study many brightness phenomena at the same time in order to best probe underlying mechanisms, we were not concerned with optimizing the parametric fit to any one data set. Note that these fits were produced after all model parameters were fixed once the simulations for Figs 10, 14, 15, 16, 17, 18, 19, and 20 were done—we will refer to the set of simulations shown in these figures as our “initial set.” The results presented in Fig. 12(A) should be compared to the data obtained by Ross et al. (1989) [Fig. 12(B)].

Mach bands are not seen in a luminance step, or in a square wave. As pointed out by several researchers (e.g. Fiorentini, 1972) this fact highlights the inadequacy of models relying primarily on lateral inhibition to explain Mach bands, since the effect should be maximal at a step discontinuity in luminance. In our simulations, the recurrent competitive field sharpens all boundary profiles at step transitions in luminance, a behavior which depends on the non-linear amplification of simple cell responses at step edges that results from ON/OFF signal juxtaposition. The final multi-scale brightness is, consequently, a square wave. Ross et al. (1989) corroborated with simulations the claim that lateral inhibition models of Mach bands are inadequate by comparing results predicted by such model with actual data. In their study contrast thresholds for seeing Mach bands on trapezoidal waveforms that were filtered, or smoothed, to different extents, were measured. The main result was that as the cut-off frequency of the Gaussian filter increased (making the low-pass filtering less severe), sensitivity to Mach bands increases [Fig. 13(B)]. The results of the model when simulated with such stimuli are shown in Fig. 13(A). As in the case of the human data, the model becomes more and more sensitive as the cut-off frequency of the smoothing filter increases.

Triangular wave. Mach bands have been studied not only in trapezoidal like modulations of luminance but in other stimuli. A noteworthy example is the presence of brightness overshoots and undershoots reminiscent of

![Diagram](image_url)

**FIGURE 12.** (A) Strength of light Mach band as a function of spatial frequency as predicted by the model. All runs used to generate the plot employed the fixed set of parameters of the initial set of simulations. (B) Threshold contrast for seeing light Mach bands (triangles) as a function of spatial frequency (c/deg) as obtained by Ross et al. (1989); adapted with permission.
Mach bands in triangular waveforms. Recent evidence on detection thresholds for these patterns have been shown to be very similar to the ones for trapezoidal waves (Ross et al., 1989), lending further support that these perceptions have the same origin as Mach bands. The model correctly predicts the presence of bright and dark bands and computationally their origin is the same as for the trapezoidal wave (Fig. 14).

**High and low contrast missing fundamental.** The missing fundamental (MF) wave first shown by Campbell, Howell and Robson (1971), is generated by having the fundamental harmonic component removed from a square wave. This stimulus is particularly interesting, as its appearance changes qualitatively depending on the amount of contrast. For high contrasts it has a more or less veridical appearance (the “cusps” are seen) while for low enough contrasts, subjects report perceiving a square wave. The classical explanation for this stimulus relies on the fact that the MF differs from a square wave only at the lowest spatial frequency, and that low spatial frequencies are attenuated by the visual system. Therefore, only a stimulus of sufficient contrast can be differentiated from a square wave; for a discussion of some weaknesses of this approach see Todorović (1987) and Kingdom and Moulden (1989).

Our simulations correctly predict the effect of contrast on the percept, as shown in Figs 15 and 16. Figure 15 shows several model stages. In the bottom row the input \( I_f \) is shown to the left of the final multiple scale brightness prediction \( f_e \). The input luminance is initially filtered by ON and OFF channels and the results generate contrast-driven and luminance-driven signals; the second row shows ON/OFF filtering \( y_i^+ \) and \( y_i^- \), ON/OFF contrasts \( c_i^+ \) and \( c_i^- \), and the luminance-driven signal \( l_i \). Rows 3–5 show multi-scale computations. Column 1, complex cells \( x_i^\pm \); column 2, boundaries \( \omega_i^\pm \); column 3, equilibrated ON and OFF filling-in \( e_i^+ \) and \( e_i^- \); column 4, single scale brightness \( u_i^\pm \). The model predicts that a high contrast MF is perceived more or less veridically, since dense boundary signals present in scales 1 and 2 work to contain diffusion, thereby preserving the activity gradient present in the ON/OFF filtering feature. Although the input luminance contains abrupt transitions (“edges”) that produce sharp boundary signals, small complex cell responses between the “edges” are able to survive as non-zero boundary signals. In other words, while the recurrent circuit producing boundaries sharpens the large activities associated with the “edges”, it also allows preservation outside the domain of the feedback (regulated by spatial kernels). Note that the complex cell responses in between the “edge” positions are small since at these positions only ON or OFF (but not both) contrast-driven information is non-zero, and the simple cell circuit is designed so as to amplify responses only when ON/OFF juxtaposition occurs.

The simulation for the high contrast missing fundamental shows the multi-scale behavior of the model. As discussed, in scale 1 dense boundary signals survive which allow the model to capture the large variations in brightness induced by the stimulus (see ON/OFF filling-in I). At the same time, the sharp boundaries generated for scale 3 allow the system to capture the overall modulation of the stimulus (see ON/OFF filling-in 3). The
combination of these two scales, and indeed of all three scales, provides a more accurate prediction of how a high contrast missing fundamental is perceived by the human visual system than is possible by any single scale.

A low contrast MF, on the other hand, is seen as a square wave (Fig. 16). For this stimulus, all boundary signals are sharp and all feature signals are free to diffuse, generating square wave (single-scale) predictions. The final predicted brightness is a square wave.

*Simultaneous brightness contrast.* Grossberg and Todorović (1988) have shown that their implementation of the BCS/FCS theory can account for a multitude of brightness contrast phenomena. The model proposed here is also capable of correctly predicting this type of phenomenon (Fig. 17).

*Sine wave.* Another luminance waveform whose appearance depends on stimulus contrast is the sine wave. When contrast is small, subjects perceive the luminance
variations essentially veridically. On the other hand, if contrast is high enough, subjects perceive the luminance waveform in a highly distorted way; the “positive-going” segments of the modulation being perceived as “flattened.” We agree with Kingdom and Moulden (1992) that such deformation most likely originates from the initial stages of visual processing that are responsible for gain control. Figure 18 shows the outcome of the model for a high contrast sine wave. The initial ON/OFF filtering compresses the input pattern (due to gain control) and is responsible in great part for the final brightness appearance. However, single-scale predictions are the outcome of the interaction between boundary, feature and luminance-driven signals and we can see that boundary signals have a further contribution to the deformation of the brightness percept.

**Luminance staircase.** A luminance staircase illustrates well how the luminance-driven channel contributes to brightness perception. Figure 19 shows some of the model stages for this distribution. By definition, the luminance-driven channel faithfully follows the modulation present in the input and since it provides a “baseline” of activity that is combined with ON and OFF filling-in information to derive brightness, the model correctly predicts that a series of brightness steps is perceived. Note also that brightness is not perfectly

![Figure 16](image1.png)

**FIGURE 16.** Low contrast missing fundamental.

![Figure 17](image2.png)

**FIGURE 17.** Brightness contrast.

![Figure 18](image3.png)

**FIGURE 18.** Nonlinear contrast effects associated with high contrast sine waves. Only computations for scale 1 are shown; other scales have similar trends.
uniform along the luminance plateaus. Smaller regions would produce even more uneven brightnesses, as predicted by the Chevreul illusion. This once more illustrates the nature of nonconservative filling-in, in which final equilibrated activities are not necessarily uniform.

We also investigated the pattern shown in Fig. 20 where a series of luminance “cusps” is present. It is known that while one or two cusps can induce vigorous brightness steps, the effect is greatly diminished for a series of cusps (Coren, 1983; Cohen & Grossberg, 1984). In many cases, only the outer regions produce brightness differences that are perceptually salient, as shown in the simulations. Incidentally, in such displays the cusps themselves are more salient, i.e. they appear more like cusps than like steps. Note that our model is perfectly capable of handling brightness steps induced by a single cusp, as is evident from the brightness associated with the last cusp in our stimulus. The simulations for the low contrast missing fundamental also shows that the model correctly handles such cases; in fact, a low contrast missing fundamental can be seen as an alternative method to generate a luminance cusp. Finally, note that brightness models such as MIDAAS (Kingdom & Moulden, 1992) that propose perfect brightness transitivity would predict that a luminance staircase and a series of cusps, as shown in the two stimuli above, look approximately the same. As the present model predicts, transitivity is not preserved completely and is contingent on specific stimulus parameters (Shapley & Reid, 1985; Reid & Shapley, 1988; Arend & Goldstein, 1987).

Recently, Morrone, Burr and Ross (1994) have discovered an interesting illusion related to a luminance staircase. As indicated in Figure 19, a regular luminance staircase is perceived to be “scallopy,” the well known Chevreul illusion. Such appearance changes when a series of lines (either light or dark) are added to the steps (see Fig. 21 left), as the steps no longer look “scallopy” but of uniform brightness on both sides of the line. Morrone et al. (1994) point out that the introduction of the lines creates an illusory brightness step, i.e. there is a net change in brightness at the line where in the stimulus there is no associated luminance change. Figure 21 shows that our model correctly predicts such illusory steps. In the model, the lines in the stimulus produce sharp boundaries that block diffusion. Since the domains of filling-in created by such boundaries are much smaller than the entire plateau in a regular staircase (less than half the size), diffusion will produce equilibrated activities that are more uniform. Figure 21 also shows the activity in the luminance-driven channel. While such activity underlies the final steps in brightness that are associated with the luminance steps, it cannot, by itself, explain the illusory steps at the line. These originate from feature signals, in a way similar to a simulation of brightness contrast. Note that the region to the left of a line is brighter because the incoming feature signal contains an overshoot from the dark-to-light step, while the region to the right is darker due to the undershoot in the feature signal associated with the light-to-dark step. Such feature signals are used as inputs to on and
off filling-in domains which contribute to the final brightness. Not only is the model's behavior related to that of simultaneous brightness contrast, but we believe that phenomenologically the modified Chevreul illusion is related to such effect. The regions separated by the line appear different in a way that can be predicted by brightness contrast.

Aside from correctly predicting the appearance of the overall stimulus, our model accounts for two key facts revealed by Morrone et al. (1994). First, the thin lines can be either light or dark for the illusion to work. In the model, what is necessary is that these be strong enough to generate sharp boundaries. Second, low contrast lines do not change the Chevreul illusion producing the illusory brightness step at the line, only contrasts that are high enough, which are typically much higher than the ones required to detect the lines in isolation. This can also be understood in the context of our model: as the contrast of the line is increased there is a point where there is a switching in the model's behavior and the production of sharp boundaries. This behavior is reminiscent of the phenomenology of the stimulus: "as the contrast of the bar increased, it reaches a point where there is a qualitative change in the pattern, as the scalloppness gives way to two separate regions on each step, each of distinct brightness" (Morrone et al., 1994, p. 1568). We are also confident that the present model can account for the perceptual modifications certain stimulus manipulations produce, such as high-pass filtering (Morrone et al., 1994). For example, employing a filter with a cut-off frequency of 2.75 c/step dramatically reduces the illusion [see their Fig. 1(E)]. Due to the high-pass filtering the staircase luminance distribution is basically removed and the luminance edges lead to Craik-O'Brien type "cusp"transitions. The present model will then generate a brightness step at the cusp, but these will be basically the same everywhere. The resulting prediction is that the modified Chevreul illusion disappears.

Morrone et al. (1994) also show the behavior of the local energy model when simulated with such stimuli. with excellent data fits. It is interesting to note that while the local energy model produces good results for the modified Chevreul illusion, as Burr and Morrone (1992) state, it does not yet provide a satisfactory explanation of the classical Chevreul illusion. The present model, on the other hand, provides a unified account for both stimuli and a host of others through the interaction of contrast-driven (feature) and luminance-driven channels and the switching behavior of the boundary computations.

**DISCUSSION**

**Brightness perception models**

Next we briefly discuss models that have been recently proposed and that have directly tried to account for some of the brightness percepts we have simulated. *MIRAGE*. The MIRAGE model of Watt and Morgan (1985) was proposed to provide a symbolic description of local luminance changes in visual stimuli. MIRAGE has attempted to explain, among other things, brightness perception, in particular effects such as the Chevreul illusion and Mach bands. Although MIRAGE produces good matches for "edge" and "blur" data, we do not think it provides the correct level of analysis for brightness data. The visual system is confronted with a multitude of stimuli, the majority of which have shallow variations in brightness.* MIRAGE, however, is only concerned with "important" features and while it can properly determine "edges" and "bars" it does not describe the "other" visual properties. For concreteness, consider a trapezoidal wave, which Watt and Morgan (1985) claim the model predicts as having Mach bands. The output of the model states that two bars of opposite polarity are present and these can be associated with Mach bands. However, the bars are joined by a region corresponding to a luminance plateau, therefore predicting that there is no ramp connecting the light and dark bands (Watt & Morgan, 1986; their Fig. 6). This point has been previously made by Fiorentini et al. (1990) (see also Kingdom & Moulden, 1992). Thus, by providing scene descriptions that employ a vocabulary that only includes "edges", "bars", and "plateaus", MIRAGE does not properly describe brightness phenomena.
FIGURE 22. Four of the interpretation rules employed by MIDAAS. Rules specify how filtered responses are interpreted in terms of brightness “predictions” according to properties of the filtered responses (e.g. zero-crossings). Left: input stimuli. Middle: filtering responses. Right: brightness interpretations. Adapted from Kingdom and Moulden (1992) with permission.

FIGURE 23. Interpretation rules and “uneven” contrast. (A) Luminance distribution. (B) How does MIDAAS process this stimulus? Brightness prediction. (C) Application of rule (C) of Figure 22 together with “extrapolation” produces a brightness ramp where one should not exist.

MIDAAS. Kingdom and Moulden (1992) have proposed a model of brightness perception using multiple spatial scales (MIDAAS) and have simulated a large set of brightness stimuli. MIDAAS has five processing stages but we will concentrate on its most important stage which generates symbolic descriptions of brightness changes for multiple spatial scales. The model proposes that after the input is convolved, interpretation rules are used to determine the brightness prediction associated with each spatial scale. Figure 22 displays four of the interpretation rules used by MIDAAS. Fixed interpretation rules as used by MIDAAS will often err and cannot provide a general framework for brightness perception. Consider the situation depicted in Fig. 23(A), where MIDAAS is to process a stimulus similar to the one used for rule (D), but where the middle plateau extends in space such that the central lobe of a triphasic response [Fig. 23(B)] is “cusped” and uneven (for any given spatial scale). This spatial distribution of luminance does not trigger rule (D), but instead rule (C); unlike the situation depicted for rule (C) where the contrasts are the same at both steps, the contrasts now differ. According to MIDAAS, when rule (C) is triggered, “the brightness profile is interpreted as a ‘filled-in’ version of the lobe” (Kingdom & Moulden, 1992, p. 1567). However, the filtered profile in Fig. 23(B) is asymmetric and it is not clear how to treat this case.
properly. Kingdom and Moulden (1992, p. 1573) propose that their algorithm “extrapolates” between the two edges as shown in Fig. 23(C)—the “extrapolation” constitutes yet another (implicit) rule of the model. This solution is clearly inappropriate since no brightness ramp is perceived for this stimulus. Even more serious problems arise when evaluating MIDAAS within the framework of regularization theory (Terzopoulos, 1983; Poggio, Torre & Koch, 1985; Blake & Zisserman, 1987). It can be shown that MIDAAS employs a sequential “scanning” technique—i.e. another rule—so as to connect local brightness steps by brightness plateaus (see Pessoa, Mingolla & Neumann, 1994). Finally, Kingdom and Moulden (1992) state that as defined MIDAAS can only deal with stimuli with one-dimensional symmetry. The arbitrary nature of the interpretation rules is highlighted by investigating how the model would behave with two-dimensional data. Although it can be postulated that a different set of “two-dimensional rules” should be applied if the model is to be expanded to account for such stimuli, we believe that the approach is greatly undermined by necessitating a revision of its set of rules, which constitutes the core of the model. The present work, on the other hand, has a natural extension to a two-dimensional implementation, which is currently underway. Indeed an implementation of a related multi-scale BCS/FCS network for image processing has already been described in Grossberg et al. (1993).

Rule-based models in general, and MIDAAS in particular, have to offer some rational as to why particular rules are employed by them as opposed to other possible ones. The network model developed here offers a mechanistic explanation for several of the rules employed by MIDAAS by the instantiation of a dynamical system that behaves as if it were a rule based system following some of the Kingdom and Moulden (1992) prescriptions. Consider, for example, rules (A) and (B) outlined in Fig. 22. Intuitively, they state that undershoots (or overshoots) in the initial filtering stages of the visual system are only perceived as brightness undershoots (or overshoots) when gradual variations of luminance occur. As shown in the computer simulations above, in our model, the registration of brightness undershoots (or overshoots) is obtained by shallow boundary signals that “print” the input to diffusion by blocking filling-in. At the same time, abrupt discontinuities of luminance produce sharp boundaries that allow for featural diffusion or uniformization. In terms of the model, this “switching” behavior originates from the recurrent circuit’s ability to distinguish which responses from the simple and complex cell stage originate near steps and which originate near ramps of luminance. Therefore, these two symbolic rules employed by MIDAAS serve as descriptions of the complex interactions present in our model, i.e. they are useful at a descriptive level. Indeed, in conducting our study we devoted considerable time to asking precisely how competencies such as present in the MIDAAS rule set could be instantiated in an analog neural network. It should be pointed out, however, that the rules per se lack the flexibility and context-dependency necessary to account for a wider range of stimuli, and thus cannot be taken as the sole explanation of the phenomena at hand. Systems of ordinary differential equations such as the one we used, on the other hand, cannot fail to specify model behavior for all possible input distributions, including those that “fall between” the rules of symbolic approaches.

Local energy model. Although models of early visual processing differ widely, they share the property that incoming inputs are first filtered by even-symmetric receptive-field like operators; a few use odd-symmetric operators (e.g. Canny, 1986) instead. Morrone and Burr (1988) proposed that by combining the output of both types of operators—even and odd-symmetric—it is possible to account for a large body of psychophysical data. Their model employs two sets of matched operators and uses them to obtain a “local energy” measure at every location of the visual scene. Local energy is defined as the square-root of the sum of the squares of the filter responses and is used to indicate the positions of “features”. As with other models, Morrone and Burr (1988) are interested in determining the positions of “lines” and “edges”, and in their theory, these correspond to peaks (local maxima) of local energy. The operations of the model are performed in multiple spatial scales, although these are not critical to its functioning.

The model has been used to account for several brightness phenomena (see Burr & Morrone, 1992) and its performance when compared to the perceptual detection of Mach bands is particularly striking (Ross et al., 1989). The model encounters problems with other stimuli, however. As Ross et al. (1989) have themselves pointed out, the local energy profile is flat for single sinusoids. Therefore, the model predicts that a, say, high contrast sine wave will be “featureless”. Nevertheless, if an initial compressive non-linearity is assumed, the local energy profile will not be flat. The local energy model is primarily interested in determining the localization of visually salient features (“lines” and “edges”) and it is not clear what the output of the model would represent in this case.

The local energy model has problems in order to explain the fact that for many stimuli, contrast determines how they are perceived. As pointed out by Kingdom and Moulden (1992), this behavior is due to the fact that the position of the peaks in local energy that constitute the output of the model and signal important “features” in a scene (“lines” and “edges”), are invariant with input stimulus amplitude. Thus, the model cannot, without modifications, account for why the missing fundamental stimulus is perceived differently as a function of contrast; as a square-wave for low contrast and with the “veridical” cusps for higher contrasts, as shown in our simulations. The same problem is encountered when processing low and high contrast sinusoidal waves—the latter is perceived in a deformed way while the former in a more veridical form.
"Lines" and "edges" in human vision

Morrone and Burr (1988) have shown how the local energy model (see above) correctly accounts for several results on the perception of "lines" and "edges". A rather striking set of demonstrations employed four different stimuli whose amplitude spectrum was that of a square wave, differing in phase only. Subjects indicated the major features of the pattern and described their nature. It is interesting to study the behavior of the present model, which unlike alternative approaches does not code lines and edges as "primitives", with such stimuli. Figure 24 shows the prediction of the model when simulated with a stimulus that elicited the impression of having "edges with superimposed lines, somewhat like a Mach band" (Morrone & Burr, 1988, p. 231). Both the "edge" discontinuity in brightness and the "line" are well captured by the model. Figure 25 shows how the model behaves with a waveform that elicited the appearance of dark and bright slightly blurred lines, separating regions that were fairly homogeneous. Again the predicted brightness closely matches that of observers.

Morrone and Burr (1988) propose that the visual system detects "lines" and "edges" by computing local energy and by comparing the outputs of even and odd symmetric operators. The present approach postulates that the interactions of the non-linear circuits that produce boundary signals—i.e. the non-linear simple cell and the feedback circuit—with feature signals will determine final perceived brightness. "Edges" will typically be associated with the sharpening of boundaries, and "lines" with the printing of filling-in features that can be registered in the final brightness percept. Morrone and Burr (1988) propose that the signal for an "edge" may also be the signal for a brightness change (even if not accompanied by a corresponding luminance change), and suggest that this may account for the Craik–O'Brien effect. In the present approach, brightness changes originate from different strengths of feature signals feeding into separate filling-in regions and combining with luminance-driven signals. Consider, for example, the waveform in Fig. 24. In our simulations, the intermediate spatial scale produces sharp boundaries at every inflection point of the waveform and the corresponding single scale brightness prediction is a square wave (with "edges"). The largest scale produces dense boundaries "around" the inflection points which register the "lines" that are present in the final multi-scale prediction. In all, different spatial scales are sensitive to distinct stimuli characteristics and combine to produce the final model prediction.

Other factors in brightness perception

At least since Wertheimer (see Beck, 1972) it has been known that figure-ground relationships may modify brightness perception. In fact, several ingenious demonstrations have shown that factors such as shape, transparency, and shadows, affect brightness perception. Recent examples include the work of Knill and Kersten (1991) showing that the Craik–O'Brien–Cornsweet effect depends on surface curvature perception; Grunewald, Pessoa, and Ross (1994) have shown that three-dimensional surface interpretation also affects the Craik–O'Brien–Cornsweet effect. Aldelson (1993) has shown the strong influence of spatial organization and transparency on brightness.

While no current implemented computational model of brightness perception can claim to account for all such factors, an important criterion in evaluating the many existing proposals is the extent to which they can be extended to account for such phenomena—or at least are compatible with them. Our neural network approach is embedded in a larger framework that has studied problems such as three-dimensional figure-ground separation, including three-dimensional neon color spreading and transparency (Grossberg, 1994). In fact, a preoccupation with issues such as surface perception has always been present in our approach (Grossberg, 1987; Grossberg & Mingolla, 1987).

The amount of circuitry employed in our brightness simulations is non-trivial. Apart from brightness perception, one may ask what does it do for biological vision in general. As stated, our work is part of a larger framework that is interested in different aspects of visual perception. The equations of model have a natural, indeed "obvious" extension to two dimensions, and at the moment such implementation is being pursued.
CONCLUSIONS

We have presented a neural network model of brightness perception that can account for some challenging data involving slow variations in brightness; as well as sharp transitions. The model can account for the classical phenomenon of Mach bands as well as other stimuli (e.g. brightness contrast). Our proposal was compared to alternative approaches attempting to explain similar sets of brightness data, such as MIRAGE (Watt & Morgan, 1985), MIDAAS (Kingdon & Moulden, 1992), and the local energy model (Burr & Morrone, 1988).

The network model presented in this paper is in the tradition of the BCS/FCS style of filling-in theory as developed by Grossberg and colleagues (Grossberg, 1983; Cohen & Grossberg, 1984; Grossberg & Mingolla, 1985a,b; Grossberg & Todorović, 1988). In the context of brightness simulations, the model is closely related to Grossberg and Todorović's (1988) work and relies on the key concepts of the BCS/FCS theory, including containment of featural diffusion by boundary signals and modulation of boundary signals by a feedback circuit. The main point of departure of the current model is the proposal that two parallel channels convey information about the input: a contrast-driven and a luminance-driven channel.

Grossberg (1987, 1994) proposed the FACADE theory to explain a vast range of visual phenomena, including aspects of three-dimensional perception. The present work dealt directly with brightness perception and, therefore, did not invoke all mechanisms postulated by FACADE theory. For example, although our model includes multiple spatial scales, it is assumed, for simplicity, that they are kept independent of each other and that the final brightness prediction is a linear average of the single-scale predictions. On the other hand, FACADE theory postulates several interscale interactions responsible for figure-ground separation and other aspects of three-dimensional visual perception. It remains to be seen to what extent the present work can be further extended through the insights of FACADE theory to meet the challenge posed in the remark by Gibson quoted in the Introduction of this article.

REFERENCES


Acknowledgements.—Luiz Pessoa was supported in part by CNPq and NUTES/UFJF, Brazil. Ennio Mingolla was supported in part by the Air Force Office of Scientific Research (AFOSR-F49620-92-J-0334), the Northeast Consortium for Engineering Education (NCCEE-A303-21-93), and the Office of Naval Research (ONR N00014-91-J-1100). Heiko Neumann was supported in part by the German BMFT, grant 413-5839-01 IN 101 C/1. The authors thank Stephen Grossberg and three anonymous reviewers for helpful comments on an earlier draft of this article.

**APPENDIX A**

The implementation employed is one-dimensional, i.e. stimuli of interest are actually slices through full two-dimensional stimuli with one-dimensional symmetry. All model parameters (Greek letters) not involved in spatial weighting functions (upper-case letters) have been kept constant across scales: spatial parameters that vary according to scale have an index $S$. Variables that are computed in multiple scales also contain the index $S$ unless noted otherwise. All model variables are in lower case letters. Associated with each computational stage we include a reference to Fig. 9 so as to guide the reader in understanding the computations of the model. Appendix B discusses parameter choices.

The input pattern to the model is a spatial pattern of luminance given by $I(x,y)$.

**Center–surround antagonism and ON and OFF channels** [Fig. 9(a, b)]

The input pattern is processed by both ON and OFF cells in a way similar to retinal ganglion cells. An ON cell at position $i$ obeys a membrane equation of the form (Grossberg & Todorovic, 1988)

$$\frac{dv_i^+}{dt} = -2v_i^+ + (\beta - \gamma) (v_i^- - v_i^+) + E$$

$$\frac{dv_i^-}{dt} = 2v_i^- + (\beta - \gamma) (v_i^+ - v_i^-)$$

$$\frac{dv_i^+}{dt} = -2v_i^+ + (\beta - \gamma) (v_i^- - v_i^+) + E$$

$$\frac{dv_i^-}{dt} = 2v_i^- + (\beta - \gamma) (v_i^+ - v_i^-)$$

This model captures the essential features of the ON-OFF interaction through the inclusion of the center-surround antagonism and the use of two different sets of parameters for ON and OFF cells.
where \( \alpha, \beta \) and \( \gamma \) are constants; \( C_i^* \) is the total excitatory input to \( y_i^* \) and \( E_i^* \) is the total inhibitory input to \( y_i^* \). These terms denote discrete convolutions of the input \( I_i \) with spatial weighting functions, or kernels, as in

\[
C_i^* = \sum_j I_j C_{ij} \quad \text{and} \quad E_i^* = \sum_j I_j E_{ij}
\]

where the weighting functions are defined by normalized Gaussians for the center and surround mechanisms as in

\[
C_{ij} = \frac{1}{\sigma_i \sqrt{2\pi}} \exp \left( -\frac{(j-i)^2}{2\sigma_i^2} \right)
\]

and

\[
E_{ij} = \frac{1}{\sigma_s \sqrt{2\pi}} \exp \left( -\frac{(j-i)^2}{2\sigma_s^2} \right)
\]

where \( \sigma_s > \sigma_i \) ("surround" broader than "center").

Since the system is assumed to reach equilibrium quickly, on responses are solved at equilibrium, \( dy_i^*/dt = 0 \), and rectified so that

\[
y_i^+ = \begin{cases} 
\beta C_i^* - \gamma E_i^* \\ x + C_i^* + E_i^* \end{cases}
\]

off cells also obey membrane equations calculated at equilibrium as in

\[
y_i^- = \begin{cases} 
\beta C_i^* - \gamma E_i^* \\ x + C_i^* + E_i^* \end{cases}
\]

where \( \{x\}^+ = \max(x, 0) \). It was assumed that center and surround mechanisms provide input to both on and off pathways, such that off cells are obtained by having \( C_i^* = E_i^* \) and \( E_i^* = C_i^- \). The non-linear processing of membrane equations (A5) and (A6) produces an imbalance in the amplitudes of \( y_i^+ \) and \( y_i^- \), with the resulting activity in the off channel being larger. We stress that such imbalance does not contribute in any important way to the model’s behavior.

Grossberg and Wyse (1991) and Grossberg et al. (1993) have employed an on/off circuit in which the off channel has a positive baseline activity while the on channel has a zero baseline activity, thereby creating a “tonic” off signal. The present formulation (Grossberg & Toddorović 1988) does not produce “tonic” off signals on and off cells perform “lateral inhibition” in a way similar to models employing ‘mexican-hat’ operators (see, e.g. Marr, 1982). However, the membrane formulation used here has several additional properties such as “reflectance processing” (Grossberg, 1983).

**Segregation of contrast-driven and luminance-driven channels [Fig. 9 (2a, 2b, 3)]**

The representation of contrast-driven and luminance-driven information is obtained via interactions of on and off channels. On and off channels compete, eliminating the d.c. component, leaving contrast information [see Fig. 9a, 2b]:

\[
c_i^+ = \{y_i^+ - y_i^-\}^+ \quad \text{and} \quad c_i^- = \{y_i^- - y_i^+\}^+.
\]

"Baseline," or luminance-driven, activity is obtained by pooling the output of on and off channels, obtaining a low-pass filtered and non-linearly compressed transformation of the input [see Fig. 9a(3)]

\[
l_i = y_i^+ + y_i^-.
\]

See Neumann (1993a, 1994) for further details.

**Simple and complex cell responses [Fig. 9a, 4, 5, 6]**

Simple and complex cells are the first major stages leading to the computation of boundaries. Before feeding into simple cells, on and off contrast signals are blurred by [see Fig. 9b(4)]

\[
h_i^+ = \sum_j c_j^+ K_i^+ \quad \text{and} \quad h_i^- = \sum_j c_j^- K_i^-
\]

where

\[
K_i^+ = \frac{1}{\sigma_i \sqrt{2\pi}} \exp \left( -\frac{(j-i)^2}{2\sigma_i^2} \right)
\]

and \( \sigma_i \) determines the spatial extent of the blurring Gaussian kernels, where \( S = 1, \ldots, k \), and \( k \) is the number of scales. Note that these are the first multi-scale computations in the model; the initial stages of on/off center-surround filtering and segregation into contrast-driven and luminance-driven channels are computed for a single scale.

The model employs simple cells that respond to light-dark and dark-light luminance transitions [Fig. 9(5)], as functionally described in the section Simple and Complex Cells. These are obtained by collecting contrast-driven information from spatially different branches. Consider a simple cell at position \( i \). For a light-dark cell, on information originates from the “left” and off information from the “right!” (with respect to position \( i \)). To simplify the notation below we use the following index convention: \( l = l - \sigma_l \) and \( r = l + \sigma_r \), where \( l \) and \( r \) are the “left” and “right” spatial offsets and \( \sigma_r \) is a scale dependent constant [as used in equation (10)]. The equations for the on channel of a light-dark simple cell are

\[
dy_i^+ = -\xi y_i^+ + \beta_i^+ + \eta_i^+ q_i^+.
\]

and

\[
dy_i^- = -\mu y_i^- + \beta_i^- + v_i^- z_i^-.
\]

where \( \xi, \eta, \mu \) and \( v \) are constants. The spatial scale superscript \( S \) has been dropped to simplify the notation. Equations (A11) and (A12) determine how the second and third stages of the on channel of a light-dark cell behaves (see Fig. 3). The (blurred) on contrast excites it and the (blurred) off contrast inhibits it; inhibition is performed in a multiplicative, or shunting, form. Note that the inhibition of the off channel on \( q_i^+ \) (opponent inhibition) realizes the mechanism of disinhibition referred to before. By inhibiting stage 2 it inhibits the inhibitory effect \( q_i^+ \) has on stage 3 \( z_i^+ \) in equation (A12), thereby producing disinhibition. These processes are assumed to reach equilibrium quickly and are thereby computed as

\[
q_i^+ = \frac{h_i^+}{\xi + \eta_i^+ q_i^+} \quad \text{and} \quad z_i^- = \frac{h_i^-}{\mu + v_i^- z_i^-}.
\]

Similarly for the off channel of a light-dark cell, the equilibrated values are

\[
q_i^+ = \frac{h_i^+}{\xi + \eta_i^+ q_i^+} \quad \text{and} \quad z_i^- = \frac{h_i^-}{\mu + v_i^- z_i^-}.
\]

The final response for a light-dark simple cell is computed as

\[
s_i^+ = z_i^+ + \tau_i^+.
\]

The dark-light response, \( s_i^- \), is obtained in a similar manner.

Complex cell responses are insensitive to direction of contrast and are obtained by summing light-dark and dark-light simple cell responses [Fig. 9(6)]. It is assumed that the complex cell output is a scaled-thresholded version as in

\[
x_i = n s_i^+ + s_i^- - \theta i.
\]

where \( n \) and \( \theta \) are constants.

**Boundaries: feedback competition of complex cell responses [Fig. 9(7)]**

Boundaries are obtained by processing complex cell responses through a recurrent competitive network. The system sharpens strong inputs and leaves small signals largely unmodified, as discussed in the section. Non-linear recurrent competitive circuit. The multi-scale non-linear feedback network was constructed along the lines of Grossberg and Marshall (1989) (see also Grossberg, 1973) and is given by [spatial scale superscript \( S \) omitted]

\[
dw_i = -d w + (z_i - w)(F_i^+ + B_i^+) - (\phi + w)(F_i^- + B_i^-)
\]

where \( \tau, \chi, \phi \) and \( \phi \) are constants. Term \( F_i^- \) is the total excitatory feedforward signal for node \( i \), \( B_i^+ \) is the total feedback signal, \( F_i^- \) is the total inhibitory feedforward signal, and \( B_i^- \) is the total inhibitory feedback signal. All the signals are discrete convolutions:

\[
F_i^- = \sum_j x_j F_j^+
\]
NETWORK MODEL OF BRIGHTNESS PERCEPTION 2221

![Diagram of network model with boundary and feature signals.]

**FIGURE A1.** Implementation of filling-in. Boundary signals can only affect permeability between filling-in positions which receive direct inputs from featural signals. Adapted with permission from Grossberg (1987).

\[ F_i = \sum_j x_j F_j \]  
(A19)

\[ B_i^+ = \omega \sum_j h(w_j) B_j^+ \]  
(A20)

\[ B_i^- = \psi \sum_j h(w_j) B_j^- \]  
(A21)

where \( \omega \) and \( \psi \) are constants. The feedback signal function was chosen “faster-than-linear” above threshold

\[ h(x) = \max(x - \delta, 0)^\theta \]  
(A22)

where \( \delta \) is a constant. The feedback signals \( B_i^+ \) and \( B_i^- \) are chosen non-linear (“faster-than-linear”) so as to generate the noise suppression and contrast enhancement properties desired (see Grossberg, 1973; Grossberg & Marshall, 1989). The spatial kernels used in the above equations \( F_j, F_j^+, F_j^- \), and \( B_j, B_j^+, B_j^- \) were all defined as Gaussians whose spatial parameters depend on scale as in equation (A10); the standard deviations of the Gaussians are, respectively, \( \sigma_{b+}^i, \sigma_{b-}^i, \sigma_{b+}^i, \sigma_{b-}^i, \sigma_{b+}^i, \sigma_{b-}^i \), and \( \sigma_{b+}^i, \sigma_{b-}^i \). In order to obtain the desired sharpening behavior, those space constants are defined such that \( \sigma_{b+}^i \) is very small (“strong” excitatory feedback limited to a single node) and \( \sigma_{b-}^i = \sigma_{b+}^i \) (Grossberg & Marshall, 1989).

**ON and OFF feature filling-in** [Fig. 9(8a, 8b)]

The above stages have specified how, after the initial ON/OFF filtering and the segregation into contrast-driven and luminance-driven channels, the contrast-driven signals are used in order to produce boundary signals. Contrast-driven signals are also used as featural quality and together with the luminance-driven channel, determine (single-scale) brightness. Filling-in performed for both ON and OFF domains (Grossberg, 1987; Grossberg & Wyse, 1991; Arrington, 1993; Gove, 1994; Grossberg et al., 1993).

Diffusive filling-in for the ON domain is implemented as [Fig. 9(8a)]

\[ \frac{d c_i^+}{dt} = -\lambda c_i^+ + \sum_{j,k} (c_j^+ - c_i^+) p_{ik}^+ + c_i^+ \]  
(A23)

and similarly for the OFF domain [Fig. 9(8b)]

\[ \frac{d c_i^-}{dt} = -\lambda c_i^- + \sum_{j,k} (c_j^- - c_i^-) p_{ik}^- + c_i^- \]  
(A24)

where \( \lambda \) is a constant and \( N_i \) specifies the neighborhood of node \( i \) (Cohen & Grossberg, 1984; Grossberg & Todorovic, 1988). Terms \( c_i^+ \) and \( c_i^- \) are the ON and OFF contrasts which are computed at a single scale in equation (A7). Diffusion is limited to nearest neighbors so that \( N_i = \{i - 1, i + 1\} \). The diffusion coefficients, \( p_{ik}^\pm \), regulate the magnitude of cross influence of location \( j \) on location \( i \) and depend on boundary signals as

\[ p_{ik}^\pm = \frac{\rho}{1 + \epsilon w_i^\pm} \]  
(A25)

where \( \rho \) and \( \epsilon \) are constants.

Filling-in is implemented as indicated in Fig. A1. Boundary signals always have their effect “between” diffusion locations; i.e. they do not affect permeability of diffusion within a filling-in position but between positions. In practice this was done by having filling-in spatial locations and boundary spatial locations separated with filling-in positions given by \( k = 2n \) and boundary positions by \( k = 2n + 1 \). See Geman and Geman (1984) for a similar scheme.

**Single scale brightness prediction** [Fig. 9(9)]

The filled in activities in the ON and OFF domains are used in conjunction with the luminance-driven channel to determine the single scale brightness prediction, \( u_i \). The luminance-driven channel provides a baseline of activity that can be modified by the equilibrated ON and

**TABLE A1.** Processing stages of the system

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Equation</th>
<th>Figure 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Center–surround filtering</td>
<td>ON/OFF filtering</td>
<td>A1</td>
<td>1a/1b</td>
</tr>
<tr>
<td>Contrast-driven and luminance-driven channels</td>
<td>ON/OFF contrast-driven</td>
<td>A7</td>
<td>2a/2b</td>
</tr>
<tr>
<td>Simple and complex cells</td>
<td>Luminance-driven</td>
<td>A8</td>
<td>3</td>
</tr>
<tr>
<td>ON/OFF blurred contrast-driven</td>
<td>A9</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>ON/OFF simple cell stage 2</td>
<td>A13 and A14</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>ON/OFF simple cell stage 3</td>
<td>A13 and A14</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Light-dark simple cell</td>
<td>A15</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Dark-light simple cell</td>
<td>A16</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Complex cell response</td>
<td>A17</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Boundary signal (feedback)</td>
<td>A17</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Filling-in</td>
<td>ON/OFF filling-in</td>
<td>A23 and A24</td>
<td>8a/8b</td>
</tr>
<tr>
<td>Diffusion permeability</td>
<td>A25</td>
<td>8a and 8b</td>
<td></td>
</tr>
<tr>
<td>Single-scale brightness prediction</td>
<td>A26</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Multiple scale brightness</td>
<td>A28</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

The first column specifies the variable name, the second column provides a brief functional description, the third column indicates the equation where it is defined, and the fourth column indicates where in Fig. 9 the variable is used. Indices \( i \) and \( j \) refer to spatial positions. Every variable with the superscript \( S \) is computed for multiple scales; in the equations \( S \) was sometimes omitted to improve readability.
off contrasts. The interactions are governed by the following differential equation
\[
\frac{du}{dt} = \dot{u}u + l + \phi\dot{u}^2 - \omega u^2 \tag{A26}
\]
which is computed at equilibrium
\[
u^* = \frac{l + \phi u^2}{\dot{u} + \omega u^2} \tag{A27}
\]
where \(\dot{u}, \phi, \) and \(\omega\) are constants. Equation (A26) says that the single scale brightness is determined by having a luminance-driven signal \(l\) in equation (A8) computed for one scale and the off filling-in activity \(u^2\) in equation (A23) have an excitatory effect on \(u\), and the off filling-in activity \(u^2\) in equation (A24) have an inhibitory effect on \(u\). This computational stage formally defines how contrast-driven and luminance-driven information are recombined in order to predict brightness appearance.

Single scale brightness predictions use on and off filling-in combined with luminance-driven information which does not go through a process of diffusion. We have adopted this scheme for simplicity but should analysis of a larger body of data indicate that the luminance-driven channel also undergoes filling-in, our present scheme may have to be revised. It should be noted, however, that both schemes can account for the results shown in this paper (involving only minor parameter changes).

Multiple scale brightness pooling [Fig. 9(10)]

The final brightness percept is obtained by taking the mean of all spatial scales:
\[
f = \frac{1}{k} \sum_{k} u^k \tag{A28}
\]
where \(k\) is the number of scales being employed, \(f\) is the final output of the model.

Table A1 provides a summary of the variables used. Together with the variable name, a brief description of its functional significance is given, the equation where it is formally defined is provided, and the location of the computational stage employing the variable is indicated with reference to Fig. 9.

APPENDIX B

Although the model formally defined in equations (A1)-(A28) has a large number of parameters, the model is robust in that most of them can be selected from a large array of values. Moreover, many of the parameters simply determine scale-dependent weighting functions, such as the Gaussians used in several of the model stages; in such cases, larger scales are obtained by multiplying the space constants of the smallest scale by constant factors [see Grossberg’s (1987, 1994) discussion of “self-similarity” across scales]. In what follows we discuss the rationale for several of the parameter choices for certain model stages. Other parameters define attributes such as activity ranges and decay of activity; these can be chosen from a vast range of values without affecting the overall behavior of the system.

Simple cell

The choice of parameters for simple cells was driven by wanting trapezoidal waves of varying contrasts and absolute amplitude to generate Mach bands. In other words, both low and high contrast trapezoidal waves should produce Mach bands (of different strength) although the initial filtering responses will be of (very) different magnitudes; these stimuli should therefore produce similar boundary responses. Assume a light-dark simple cell only receives non-zero inputs from, say, the on channel (such as in the higher “knee-point” of a ramp). Equations (A13)-(A15) imply that the final response, \(s\), will be of the form
\[
f(x) = \frac{x}{A + Bx} \tag{B1}
\]

<table>
<thead>
<tr>
<th>TABLE B1. Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
</tr>
<tr>
<td>Center-surround filtering</td>
</tr>
<tr>
<td>(\pi)</td>
</tr>
<tr>
<td>(\beta)</td>
</tr>
<tr>
<td>(\gamma)</td>
</tr>
<tr>
<td>(\sigma)</td>
</tr>
<tr>
<td>(\tau)</td>
</tr>
<tr>
<td>Simple and complex cells</td>
</tr>
<tr>
<td>(\sigma_{i}^{+}, \sigma_{i}^{-})</td>
</tr>
<tr>
<td>(\zeta)</td>
</tr>
<tr>
<td>(\eta)</td>
</tr>
<tr>
<td>(\mu)</td>
</tr>
<tr>
<td>(\nu)</td>
</tr>
<tr>
<td>(\kappa)</td>
</tr>
<tr>
<td>(\theta)</td>
</tr>
<tr>
<td>Boundaries</td>
</tr>
<tr>
<td>(\tau)</td>
</tr>
<tr>
<td>(\chi)</td>
</tr>
<tr>
<td>(\phi)</td>
</tr>
<tr>
<td>(\sigma)</td>
</tr>
<tr>
<td>(\psi)</td>
</tr>
<tr>
<td>(\delta)</td>
</tr>
<tr>
<td>(\sigma_{i}^{+}, \sigma_{i}^{-})</td>
</tr>
<tr>
<td>(\sigma_{i}^{+}, \sigma_{i}^{-})</td>
</tr>
<tr>
<td>(\sigma_{i}^{+}, \sigma_{i}^{-})</td>
</tr>
<tr>
<td>(\sigma_{i}^{+}, \sigma_{i}^{-})</td>
</tr>
<tr>
<td>Filling-in</td>
</tr>
<tr>
<td>(\dot{\lambda})</td>
</tr>
<tr>
<td>(\rho)</td>
</tr>
<tr>
<td>(\epsilon)</td>
</tr>
<tr>
<td>Brightness</td>
</tr>
<tr>
<td>(\zeta)</td>
</tr>
<tr>
<td>(\theta)</td>
</tr>
<tr>
<td>(\epsilon)</td>
</tr>
</tbody>
</table>
From the definition of $q^*$ in equation (A13) and the fact that the off channel is not active (spatial scale superscripts S omitted below)

$$q^* = \frac{b^*}{\zeta}.$$  \hspace{1cm} (B2)

Substituting this value into the definition of $z^*$ in equation (A13) we have

$$z^* = \frac{b^*}{\mu + \nu \zeta}.$$  \hspace{1cm} (B3)

Since the off channel is inactive, $s^*_i = z^*_i$. Therefore, $s^*_i$ is of the form of equation (B1) if $x = b^*_i$, $A = \mu$, and $B = \nu \zeta$. After $\zeta$ was fixed at a value of 1, the remaining parameters were chosen so that $f(x)$ would operate close to its saturation range: $A$ small ($\mu$ small) and $B$ large ($\nu$ large), i.e., increases in $x$ produce small increases in $f(x)$. With such parameters, the model correctly predicts Mach bands as shown in the computer simulations. Moreover, a higher contrast trapezoidal wave will also produce bands in our simulations, although the simple cell inputs (on/off contrasts) may be much larger. That is, complex cell responses will be very similar in magnitude for both stimulus contrast values, and therefore will not be sharpened for either contrast. If only stimulus contrast changes (and not the spatial configuration), eventually the complex cell responses will become larger and start to be contrast enhanced by the recurrent competitive field. This behavior is in accord with data; since in this case only the steepness (slope) of the ramp changes, eventually the Mach bands should disappear.

**Complex cells**

Complex cell activity [equation (A16)] has to exceed a threshold, $\theta$, before having an output. In our simulations, $\theta = 0.2$, and, therefore, for the low contrast missing fundamental, complex cells only have activity associated with “edge” locations. Thus, only sharp boundaries are present and a square wave percept is produced. For all other stimuli, the complex cell threshold does not have a significant effect.

**Recurrent competitive field**

Equations (A17)-(A22) specify how complex cell responses are processed in order to generate boundary responses. Four parameters are important. The strength of feedback is determined by parameters $\phi$ and $\omega$. These were chosen large so as to produce very sharp signals once feedback is “activated”. The behavior of the recurrent competitive field can be described as “winner-takes-all-above-threshold” within local regions—for $t \to \infty$. Therefore, the most critical parameter is the $\delta$-threshold [equation (A22)], which determines the magnitude of the complex cell responses that will “trigger” sharpening. Since the $\delta$-threshold determines only when feedback is triggered, it behaves as a “soft” threshold. For our simulations, $\delta$ was chosen so as to guarantee that a low contrast square wave (or low contrast missing fundamental) would have complex cell responses large enough to trigger feedback, thereby producing sharp boundaries and a square wave appearance.

As mentioned above, the recurrent circuit behaves in the limit as “winner-take-all-above-threshold” within local regions. The spatial extent of these regions is determined by the feedback-off-surround parameter, $\sigma^*_v$, used in equation (A21). In our simulations, $\sigma^*_v$ specifies an area large enough so that very sharp boundary signals ensue when a square wave is processed; this is illustrated when processing the low contrast missing fundamental. Smaller values of $\sigma^*_v$ would leave “surviving” boundary signals around the sharp boundary peak associated with the “edges” of the missing fundamental.

**ON/OFF filling-in**

Parameters $\sigma$ and $\epsilon$ in equation (A25) determine the amount of boundary permeability and are chosen so that “small” boundary signals, such as encountered in the simulations of the trapezoidal wave, can function as “print” signals. In other words, in such cases small boundaries have to be strong enough to effectively “block” diffusion such that the modulation present in the FCS input (to diffusion) is “trapped”—such as the overshoots and undershoots in the trapezoidal wave.

**APPENDIX C**

In order to produce the plots shown in Figs 12 and 13 we employed the same assumptions as those of Ross et al. (1989). Thresholds that take into account probability summation were predicted with an equation for sensitivity given by

$$S = k \left( \sum_{i=1}^{\delta} u_i(x)^\delta \right)^{-\beta} \hspace{1cm} (C1)$$

where $k = 1$ is a proportionality constant, $i$ is a spatial scale index, $x$ is a position index, $\beta$ is a parameter of the Weibull function (Weibull, 1951) related to the slope of the psychometric function with a value of 3.5, $u_i(x)$ is the model response at position $x$ of the $i$th spatial scale, and $\Gamma$ is the spatial interval for which probability summation occurs. In order to simulate thresholds for seeing Mach bands, we only considered the regions of the model’s output that actually comprise the bands themselves, i.e. the undershoots and overshoots of the output waveform [see Ross et al. (1989) for further details]. The parameters used to generate the plots in Figs 12 and 13 were the same as for all other simulations.

Equation (1) and (2) of Ross et al. (1989) were used to generate the input stimuli. The ratio of ramp width to period was $\tau = 0.125$. 

2223